Forensic Science International: Digital Investigation 44 (2023) 301512

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

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

DFRWS 2023 EU - Selected papers of the Tenth Annual DFRWS Europe Conference

A likelihood ratio approach for the evaluation of single point device locations

Hannes Spichiger

Hochschule Luzern Informatik, HSLU I, Suurstoffi 1, Rotkreuz, CH-6343, Switzerland

ARTICLE INFO

Article history:

Keywords:
Localisation evidence
Evaluation
Likelihood ratio

ABSTRACT

There is growing awareness that localisation traces recovered from a mobile device should not be presented as fact. Services providing locations to smartphones are influenced by a multitude of potential error sources which may cause differing observations from where the device was at. This work provides an approach to evaluate a single localisation recovered from mobile devices and present a differentiated result to decision makers. An open source implementation is provided, allowing practitioners to evaluate based on reference data they created for their own case. The effectiveness and forensic value of this method are demonstrated using a simulated scenario illustrating the use of the approach. This study demonstrates the method using locations associated to images, and is generalised to any technology providing location to any type of mobile device.

1. Introduction

Uncertainty and errors linked to location-related evidence has been a subject of growing interest these past few years. A multitude of studies have shown interest in the accuracy of different location services in real world situations (Rodriguez et al., 2018; Merry and Bettinger, 2019; Ryser and Jacquet-Chiffelle, 2021). This interest results from a backdrop of increasing awareness that these results may not simply be presented as statements of fact, but that their reliability must be discussed (Casey et al., 2020). In particular, judges have thrown location-related information out of court due to a failure by the prosecution to properly address these issues (Poser, 2017; Swiss Federal Court, 2019) and standards may require a proper assessment of these issues (FSR, 2021).

Properly handling error and uncertainty related to evidence is of crucial importance, as it has major criminal justice implications. First, there is the obvious risk of sending innocent people to prison. When results that may be affected by error are presented as absolute certainty, justified doubt about the guilt of a person may be swept under the table if an expert manages to convince a judge or jury that his evidence is infallible.

Second, overselling the value of forensic evidence brings about the risk of guilty people going free. The reason for this is threefold: On the one side, if a wrong party is convicted, the true perpetrator generally goes free. On the other side, if the expert work is not done to the highest standards and presented in a realistic manner to the court, this may open up ways for a defendant to vacate convictions and exclude important evidence from trials. Being transparent about uncertainty has been criticised by practitioners for opening up to attacks about the reliability of this approach. This may be true in some judicial systems, however, it is also a very risky strategy as it only works up to the moment that attorneys discover that the value of the evidence has been exaggerated. Some practitioners are afraid that evidence *may* be invalidated if they admit to the limitations of said evidence. This outcome is absolute certainty if they do not find ways to address those limitations when the courts become aware of them. Finally, being unaware of errors may lead to suspects being disregarded during the investigation, if traces suggesting the truth of an alibi are not properly evaluated.

Third, properly addressing uncertainties increases the available evidence. Striving for absolute certainty in forensic evidence leads to practitioners, investigators and prosecutors to only pursue very strong evidence. In turn, judges and juries are looking for the proverbial smoking gun and are very reluctant to convict in the absence of the latter. As a consequence, heaps of weaker potential evidence are disregarded despite their probative value. Quantifying uncertainty linked to these traces has the potential to render them useable evidence. Consider for example the conclusions “Version A and version B may both explain this evidence.”, suggesting that the evidence has no value to distinguish between A and B, and “It is 1’000 times more likely to observe the evidence if A is true, than if B

E-mail address: hannes.spichiger@hslu.ch.

https://doi.org/10.1016/j.fsidi.2023.301512

is true.", which is strong evidence in favour of A. Both these conclusions may describe the same observations and may both be correct assessments of a piece of evidence. However, the evidence having been evaluated in light of A and B, and the uncertainties having been properly quantified, the second conclusion could reasonably be presented as evidence in court, whereas the first would likely lead to the evidence being disregarded.

All these risks can be mitigated by a structured approach for the evaluation of forensic evidence in general and location evidence in special. Basing themselves on established practices for physical evidence, authors have proposed the use of a likelihood ratio approach to resolve this issue (Casey et al., 2020) - an approach that has drawn criticism from some, mostly for the lack of empirical data available to found probabilities upon (Horsman, 2020). As has been discussed by Biedermann and Kotsoglou (2020), empirical evidence is not a requirement to follow an LR approach. Probabilities may be assigned on expert experience alone and the approach is still consistent. No one, however, will dispute, that it is better to base probabilities used in such an approach on empirical measurements. This work presents an approach allowing to draw probabilities for the Bayesian evaluation of single point locations from reference measurements. As such, it refutes the premise of (Horsman, 2020) that it is not possible to use probabilities to evaluate digital evidence using a probabilistic approach. The here presented solution is based on an approach presented in the authors doctorate thesis (Spichiger, 2022) and adapted here for general use.

The remainder of this work is structured as follows: After the introduction in Section 1, existing work is discussed in Section 2. The proposed approach, its implementations and limitations are presented in Section 3. Section 4 presents the application of the approach on real world data before a conclusion is reached in Section 5.

2. Existing work

The existing work section is organised in two parts, the first looking at localisation traces, the second focused on explaining the current state of the art regarding likelihood ratios.

2.1. Localisation traces

A **Localisation** is a location-related digital trace that results from a process attempting to determine the geographical location of the mobile device (Spichiger, 2022).

Localisations generally take the form of two dimensional coordinates, mostly longitude and latitude, indicating where on the planet a device was at a given moment in time. Localisations are the result of measurements and they may have a precision associated with them.

On mobile devices, localisations can be recovered both from native services and from third party apps alike. Both get their localisation from built-in location services that provide locations to all services running on a device upon request. These localisations are generally based on so called *assisted GPS* technology, A-GPS for short, using some sort of fingerprinting approach. A-GPS bases itself on classical GPS, locating a device based on visible GPS satellites. If at least four satellites are visible, it is theoretically possible to locate a device, with accuracy improving the more satellites are visible. A-GPS aims to integrate other signals visible to the device, such as cell towers, WiFi or Bluetooth beacons. If the location of the cell towers, WiFi access points and Bluetooth beacons are known, this information can be used to improve the precision of the localisation. Fingerprinting technology does this by creating an enormous lookup table containing the WiFi networks and Cell Towers visible at any location on the world (Cedergren, 2005). How the technology works in detail, for Android and iOS services is however unknown. Whilst A-GPS does indeed improve the precision in some situations and allows for localisation if no GPS is available, it is also suspected that it is the cause of increased error in some directions observed in real world studies such as Merry and Bettinger (2019) and Ryser and Jacquet-Chiffelle (2021).

Some research has been conducted to assess the accuracy of location services. Rodriguez et al. (2018) studied the accuracy of locations reported in Google timeline history, they found that, only 52% of locations reported by Google are within the indicated error range when GPS was enabled. The median error of observations varied heavily, from 3.1m when in a tram to 395m when standing still. Whilst their results show that the service often fails to correctly locate the device, it is rarely far off, with errors being in the same ordere of magnitude as the accuracy indicated by the service (Rodriguez et al., 2018). A study of temporal influences on device localisation has been conducted by Merry and Bettinger (2019). For eight measurement periods over two years, an iPhone 6 was measuring its location at 6 precise locations around an university building at the University of Georgia in Atlanta, USA. The measurement periods were chosen in a way to have varying times of the day, both during periods where the surrounding trees were carrying leaves and not. They observed errors ranging from 0.05m up to 99.7m. None of the parameters they controlled for had a clear influence on the observed error. They do however report a strong directional dependency of the error, with some directions being observed far more frequently at two of the six locations and the effect increasing in importance when WiFi is turned on (Merry and Bettinger, 2019). A similar observation was presented by Ryser and Jacquet-Chiffelle (2021) when studying impacts on precision localisation. Their full conclusion have not been published yet.

2.2. Likelihood ratios

Throughout Forensic Science, the use of Likelihood Ratios, or LR for short, has been widely accepted as a means to evaluate forensic results. Approaches, solutions and data sets exist for wide ranges of problems. The concept is based on the so called odds form of Bayes' Theorem for conditional probabilities:

$$\underbrace{\frac{Pr(P_1|E;I)}{Pr(P_2|E;I)}}_{Posterior\ Odds} = \underbrace{\frac{Pr(E|P_1;I)}{Pr(E|P_2;I)}}_{LR} \times \underbrace{\frac{Pr(P_1|I)}{Pr(P_2|I)}}_{Prior\ Odds} \quad (1)$$

This theorem mathematically describes the probabilities of two propositions,[1] potential versions of an event, P_1 and P_2, in light of presented evidence E and all available background information I. Read from right to left, it describes the process of a decision maker updating its believe when presented with evidence. At the far right are the *prior odds*, the probabilities before any evidence is presented, in light of the available information. Next comes what is called the *Likelihood Ratio* or LR. It is a quantification of the weight the evidence has in light of the two propositions. Finally, at the left, there are the *posterior odds*. They are the odds of the two propositions after the presented evidence has been presented to the court. It is generally accepted that in most cases, the prior odds are of the courts responsibility, as they address matters that are outside the domain of expertise of the expert and may also pertain to matters unknown to the expert. Consequently, the expert cannot express an opinion on the posterior odds either, as the knowledge of the prior odds would be a requirement. It is generally considered,

[1] Some publications use the term *hypothesis* or *claim* instead.

that only the LR is of the domain of the expert and the result that should be presented to the court. The LR is focused on the evidence: its probabilities describe how likely it is to observe them, if in turn it is assumed that each proposition is categorically known to be true. This probability is not to be understood as a probability in a frequentist sense, as the limit of the fractions when an experiment is repeated infinitely. Instead, it is a subjective measure of incertitude, assigned by the expert. Ideally, the expert assigns this probability based on empirical data, either from available databases or experiments conducted by himself. Absent such data, the expert can assign a value based on their own experience in the field (Aitken and Taroni, 2004).

For digital traces, few approaches exist describing how to assign these probabilities in detail. Casey et al. (2020) described on a general level how to evaluate location-related evidence using a Bayesian approach. Research exists on LR approaches for whether two location-data series were generated by the same device (Galbraith et al., 2020) or whether two devices were carried by the same person based on cell tower data (Bosma et al., 2020). A case example where location data was evaluated using a Bayesian approach in a real world case was presented by Bassi and Scoundrianos (2022) at the EAFS-meeting in Stockholm.

3. Evaluative approach

In this work, the aim is to evaluate a single localisation recovered from a mobile device. This localisation is the evidence E and takes the form (long, lat). This evidence is evaluated in light of a pair of concurring propositions that take the following form:

P_1	The device was at Location X at time t.
P_2	The device was at Location Y at time t.

3.1. Concept

Empirical research has shown that the error on localisations, is not uniform in all directions. To address this, the evaluation of the localisation is split up in two steps. First, the direction is evaluated, then, the distance of the observation to the proposition (d) is evaluated given the direction. The direction is measured as the angle from north (ϕ). This is visualised in Fig. 1a.

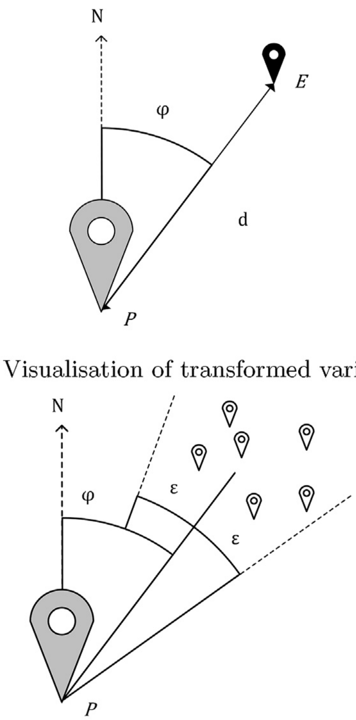
The information residing in E is now transformed into d and ϕ and replaced accordingly in the formula of the LR. As these values are in reference to the locations in the propositions, they are not identical for both the numerator and the denominator. The likelihood ratio becomes:

$$LR = \frac{Pr(E|P_1)}{Pr(E|P_2)} = \frac{Pr(d_1;\phi_1|P_1)}{Pr(d_2;\phi_2|P_2)} \quad (2)$$

It is quite challenging to obtain probability values for two variables at the same time. As d and ϕ are not independent from each other, the LR is transformed using the chain rule to only contain single variable probabilities:

$$LR = \frac{Pr(d_1|\phi_1;P_1)Pr(\phi_1|P_1)}{Pr(d_2|\phi_2;P_2)Pr(\phi_2|P_2)} \quad (3)$$

Four probabilities must be assigned. This is done based on empirical data. A series of reference data is collected at each



(a) Visualisation of transformed variables.

(b) Visualisation of measurements.

Fig. 1. Visualisation of concepts related to the separation of angular and distantial probabilities.

location presented in the propositions. It is here proposed to assign the angular probabilities based on the fraction of measured points that can be found in a wedge of $[\phi + \varepsilon; \phi - \varepsilon]$ surrounding the observed evidence and the distance probabilities based on a density function fitted to the observed distances for points within this wedge. This is shown in Fig. 1b.

The approach is generalized and can be used to evaluate uncertainty in any type of device location trace, regardless of the technology used. For example, an LR can be assigned to a single point device location generated by a vehicle navigation system or a drone, as long as the limitations discussed in Section 3.4 are respected.

3.2. Collection of reference data

To support the probabilities required to obtain an LR, a series of reference measurements is made at the locations indicated in both propositions. It is recommended, that the parties confirm the location chosen based on a aerial image or map and the coordinates that will be used to represent this location. To generate data, a device of the same make and model is used to generate localisations.

At both locations, localisations are generated on the reference device. Ideally, this is done using the same application or service that initially created the trace. For practicability, the expert may choose to generate data in a different manner. A motivation for such deviations may be that the application that generated the trace cannot be provoked to do so, or it can only be done in long time intervals. For example, data collection may be done by capturing images with associated locations or by using an app that requests locations from the device and stores them in a file. On Android devices, it is recommended to automate the data gathering process using an automation app. To the knowledge of the author, no comparable apps exist on iOS. By using multiple devices in parallel, if available, more data can be generated in the same lapse of time, in addition to there being a lower risk of the analysis being impacted from manufacturing defects from a specific device. From experience, device battery is quickly drained when they are used with the intensity required to recover data efficiently. It is therefore recommended to have a power supply available when conducting the measurements. Independently of how the measurement is done, the following aspects are important:

- The devices location and network settings need to be identical to the one of the device of interest.
- The application measuring the location must recover the location using the same service as the application or service that generated the evidentiary location.
- The application measuring the reference locations must have the same location permissions as the application or the service that generated the evidentiary data.
- Based on the measuring protocol, it needs to be possible to determine which data points were generated at which of the proposed locations.

Once the data is created, it has to be recovered from the device and brought into a structured form that can be understood by the used implementation.

It is currently unknown how many measurements should be taken at any given location to obtain a stable picture of the behaviour at this specific location. Future research should aim to address this issue. Existing research has shown that the variability seems to change from location to location (Ryser and Jacquet-Chiffelle, 2021). As variability influences the minimal number required to model a stable distribution, it has to be expected that there may not be a universal value to be used at any place.

3.3. Implementation

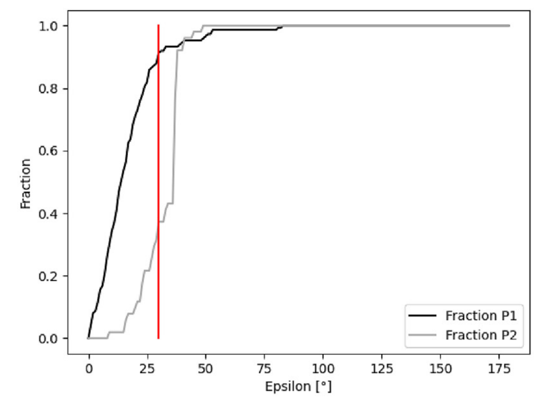
A python script was created to conduct this analysis. The code is available on GitHub (https://github.com/HSpichig/GPS_Evaluation).

The script takes as an input the coordinates of the observed evidence and the two proposed locations, as well as two excel lists containing the coordinates of the locations. For the current implementation, the script expects the list to be in the format of a Cellebrite “device locations”-export, as this is the tool that was used by the author. Future additions may provide alternative input formats. All coordinates are expected to be in the form (lat, long) and based on a WGS 84 geode.[2] A scatter plot of all points is generated by the tool. This plot does not actually project the coordinates. Instead it just assumes latitude and longitude values as x and y coordinates, which makes it impossible to compare distances. It is however a useful tool to visualise the overlap of the two reference populations, which allows to give an idea on the quality of the analysis.

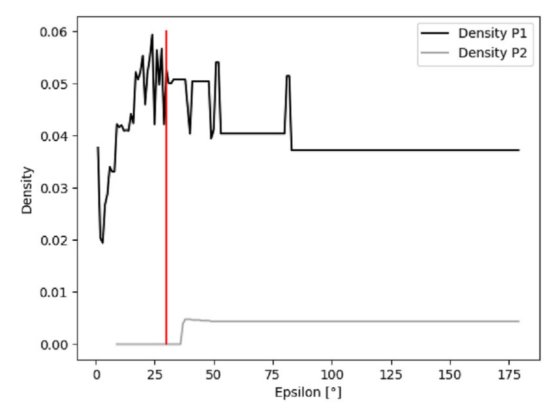
For both locations, the tool then transforms the coordinates into a list of distances and azimuth in relation to their respective ground truth. The distance and azimuth to both locations is calculated for the evidence as well.

For each location, the fraction of points found within $\varepsilon = 30^\circ$ of the observation, resulting in an overall wedge size of 60° centered on the azimuth of the evidence is calculated and outputted. This is assumed as the values for $Pr(\phi_1|P_1)$ and $Pr(\phi_2|P_2)$ respectively. A filtered list containing only the values within this wedge is used for further analysis.

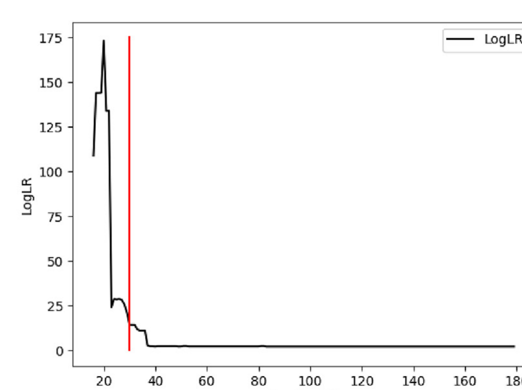
There is no evident way to chose the value of ε. Fig. 2 shows an analysis of the impact of ε on the parameters of the here presented evaluation. If the angle is chosen too wide, the analysis of the angle does not provide any meaningful insight on the direction in which the location was observed. This can be seen in Fig. 2a for values over



(a) Impact of ε on the fraction of measures within the wedge.



(b) Impact of ε on the density obtained from the distance distribution.



(c) Impact of ε on the overall LR.

Fig. 2. Graphs showing the impact of the wedge-size ε on the probabilities and the overall LR.

[2] This is the standard model used for GPS(NGA, 2022). It can generally be assumed that this model is followed by the used tool unless otherwise indicated.

35°. Alternatively, too small values of ε will cause the model to become unstable as there will be too few values available to model a distribution for the distance values. This is visible in Fig. 2b for values under 10° in the P_2-density. The P_1 density seems to mostly stabilise after 20°. This gives an interval of around 20° to around 35° to use. Looking at the overall LR, the LR in this range is of 10^{10} to 10^{30}, in a zone where the observation has an actual impact without skyrocketing.[3] An ε-value of 30°, resulting in an overall wedge of 60°, is chosen, as this is a round value within the previously discussed range.

Using a fitter function, a distribution is fitted to the distance values observed within the wedge. The current approach fits a t-distribution to the data, as this distribution presented good fits with all data sets this approach has been tested on. It is however well possible that this is not the best fit for the distribution and further implementations will likely include a series of distributions to select from. For both propositions, the resulting distribution is then evaluated at the distance value observed for the evidence. The obtained values are outputted and assumed as the values $Pr(d_1|P_1)$ and $Pr(d_2|P_2)$. Based on these values, an LR is calculated and outputted using the formula presented in section 3.1.

3.4. Limitations

Care must be taken when selecting locations from the evidentiary data. Indeed, not all data that takes the form of coordinates are localisations. If for example the coordinates of a cell tower are recovered, they cannot be evaluated in the same manner, as they do not describe the location of the device. Similar reservations should be made towards harvested locations on iOS devices, as their actual meaning is currently unknown (Whiffin, 2021). Also, logically, an opinion on the device location at a given moment in time can only be made when the recovered localisation was indeed created in the time period of interest and by the analysed device. Locations sent to the device or locations associated with an incorrect timestamp can therefore not provide any pertinent information about the devices location and are not to be evaluated with the here presented approach.[4]

The present approach is applicable to one single point of evidence only. In theory, if multiple points of localisation traces were found to be independent from each other, an LR could be calculated for all pieces of evidence independently and then multiplied. However, there is currently insufficient research into the workings of localisation services to justify such an approach.

When presenting the results of this approach, it is essential to state that the results are valid only for the location of the device and do not necessary represent the location of the person of interest. Expressing an opinion on the location of the person would consist an additional inference, linking the device to the person at this particular moment in time. Exploring this inference in detail is out of the scope of this work, but has been done in (Spichiger, 2022).

An issue likely to be encountered, as also seen in the scenario presented in the next section, is very few data points in proximity of the evidence for one of the propositions at hand. This is likely the proposition that is less apt at describing the observed evidence and a large quantity of reference measurements would be required to observe sufficient data to properly model a density distribution in this situation. Alternatively, it may be possible to model a density function, but the observed value is situated at the tail end of the distribution. In such situations, it may not be justified to blindly assign the value obtained by the tool as the respective probability. In these situations, it may be necessary for the expert to assign lower bound probabilities based on his personal experience and the number of reference data he gathered. One possible indication of the lower bound probabilities order of magnitude could be obtained by adding the evidence to the reference data and re-running the simulation. This could be seen as some sort of “worst case scenario”. Basically, this approach simulates what the probability of the evidence would be, if an additional measure was made and a perfect match was obtained. This value should not be assigned as a probability directly, but could help inform the order of magnitude justified with the available number of measurements.

Finally, there is a series of limitations related to the gathering of data, as we currently don't have conclusive research on the factors influencing measurements or the number of measurements that should be gathered at a specific location to obtain stable results.

4. Case example

To illustrate the use of the here presented approach and implementation, data from a simulated scenario is used to calculate an LR. The data was generated and used as an example in the authors doctorate thesis (Spichiger, 2022).

4.1. Parameters of the scenario

An iPhone 6s (A 1688) was used to generate all data. Both locations are on the campus of the University of Lausanne, one in the cafeteria area, the other in the laboratories of the school of criminal justice. The coordinates were obtained through the University of Lausanne interactive campus map (University of Lausanne) and are shown in Table 1.These locations are about 200m separated from each other.

In the scenario, it is assumed that an image containing coordinates in the EXIF-data associated to it was recovered from the device of interest during the period the crime was committed.

The propositions are as follows:

P_1	The device was at Location X at time t.
P_2	The device was at Location Y at time t.

4.2. Generation of reference data

Using an iPhone 6s (A 1688) running under iOS 14.4.4, pictures with location were taken at both locations using the default camera app. The location services, WiFi and Network connections were activated and the device was connected to the *eduroam*-network available on campus. The application had authorisations for precise locations and directly added locations to the images EXIF data. In total 699 pictures were taken.

A file-system device extraction was conducted recovering all

Table 1
Coordinates of the positions considered in each proposition and the observed evidence.

	Longitude	Latitude
P_1: Location X	6.573832039	46.521592273
P_2: Location Y	6.575116326	46.521954786
E: Observation	6.573944444	46.521330555

[3] As will be discussed later on, such values would not be defensible as LR.

[4] For localisations with a wrong timestamp, an evaluation as here described can be conducted if it is possible to confidently reconstruct the true moment of the creation of the trace.

pictures taken in the reference periods using Cellebrite UFED 7.53.0.24 and opened in Physical Analyser 7.54.1.7. An Excel export of the locations associated with pictures was made and the excel was split up in two files containing the measurements at each location. Some filtering was applied to the data: If consecutive pictures had the exact same coordinates associated with them, it was assumed that they result from the same measurement and are therefore not separate observations. All such duplicates were filtered out as they would not have been a reliable representation of the aspect that was intended to be measured. After filtering, 51 pictures from location X (P_1) and 149 pictures from location Y (P_2) remain. The lists containing these locations were provided as an input of the python script.

4.3. Results and discussion

Both locations, their respective reference data and the evidentiary location are plotted in a two dimensional graph. This graph is shown in Fig. 3. It is important to consider that for this graph, latitude and longitude are naively interpreted as X and Y coordinates without projecting on a map. Consequently, the distances in the visualisation do not correspond to the distances in the real world. Nevertheless, this graph provides an overview of the behaviour of both locations. Already, it can be seen that the evidence is found within the cloud of the P_1 reference data. It is therefore to be expected that the final value will be in favour of P_1.

As indicated in Section 3, the evaluation is conducted in two steps, first based on the angle, second based on the distance. For this scenario, the distance of the evidence to both proposed locations, as well as the azimuth of the evidence seen from each location are indicated in Table 2.

The fraction of observations found within a 60° wedge centered on the evidence is used to inform the probability of the angle given each proposition ($Pr(\phi_1|P_1)$ resp. $Pr(\phi_2|P_2)$). Under proposition P_1, this corresponds to 136 out of 149 reference measurements or a fraction of 0.913. Under P_2 this corresponds to 19 out of 51 reference measurements or a fraction of 0.373. These 137 respectively 19 observations within the wedge are the basis of the evaluation of the distance. Over the distances observed among the reference data within the wedge, a distribution is fitted and used to inform the probability of the distance given the angle and the proposition ($Pr(d_1|\phi_1; P_1)$ resp. $Pr(d_2|\phi_2; P_2)$). The histogram of the values within the wedge and the fitted distributions can be seen in Fig. 4. Again, it is visible that the value of E is within the bulk of the values observed

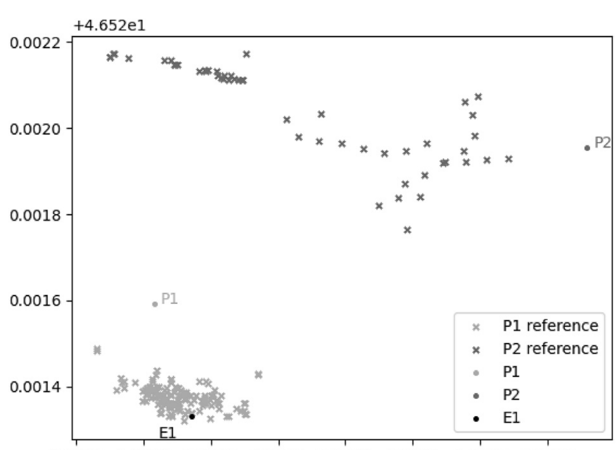


Fig. 3. Plot of the coordinates of the locations (P_1/P_2), the evidence E and the reference data for both locations.

Table 2
Distance and angle observed for each position.

Location	d [m]	ϕ [radians]
P_1	32.3	5.1235
P_2	146.8	3.6310

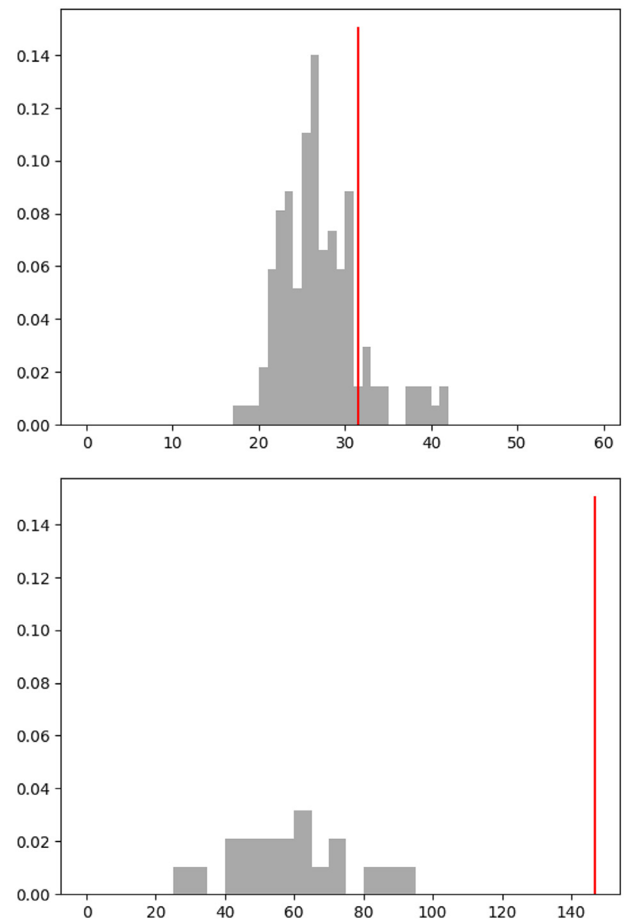


Fig. 4. Distribution of distances within the wedge for P_1 (left, n = 136) and P_2 (right, n = 19). The value observed for E is indicated in red.

under proposition P_1, whilst the under P_2 the value lies outside of the distribution. Doing that, a value of 0.053 was obtained under P_1 and a value of 7×10^{-8} under P_2. Assuming these values as probabilities for the LR, the script provided an LR of 1'769'124. This LR-value is rather high and should not just be used blindly. In particular, the value is dominated by $Pr(\phi_2|P_2)$ which is very low. This value results from the tail end of a distribution based on just 19 samples. As such, it is likely not very stable.

Indeed, it could be argued, that a position close to the evidence would have been observed if more measurements would have been made. In a stable model, the addition of one perfect match should not have major impact on the obtained probability. To assess the viability of this claim, the analysis is conducted again after adding the evidence to the reference data. The result of this analysis indeed shows a much higher value than previously observed, now situated at 2×10^{-4}. This value can be characterised as the worst case scenario: One additional measurement was made, and a perfect match was observed. This situation is rather unlikely, but adapting this value can be considered as a conservative estimate of the probability in question. It is hard to argue, that an even higher value should be assigned. It is considered here, that this second value of 2×10^{-4} is much more justifiable given the sparse number of observations and it is assigned as the probability of $Pr(d_2|\phi_2; P_2)$. If an expert would want to justify assigning a lower value, more measurements should be conducted at that specific location.

All assigned values can be found in Table 3. These values are added to the LR-Formula 3 presented in Section 3.1.

$$LR = \frac{0.053 \cdot 0.913}{2 \times 10^{-4} \cdot 0.373} = 648.646 \quad (4)$$

This obtained value indicates that the observation of the trace is about 650 times more likely if the devices was at location X at time $t(P_1)$ than if it was at location $Y(P_2)$. This is value aligns with the ground truth of this case, namely that the device was at location X when this location data was captured.

This obtained LR will be provided by the expert in his conclusion. It is frequent practice to use a verbal equivalent, called a verbal scale, to either represent or accompany the value (Marquis et al., 2016). For example, the expert may use the UK FSR standards verbal scale as shown in Table 4 and conclude that "in the opinion of the expert, the observations are much more probable if the device was at location X rather than if the device was at location Y." (FSR, 2021).

5. Conclusion

Historically, location-based evidence from mobile devices has been treated without a formal and robust approach to evaluating uncertainty. To mitigate the risk of incorrect conclusions in criminal cases, there is an increasing expectation from all parties that forensic practitioners will use a logical approach to dealing with uncertainty. In some places, this expectation is reflected in department guidelines or binding policy, such as the UK FSR standard for evaluation (FSR, 2021).

This paper presents such a solution for location-based evidence generated by mobile devices. In a simulated scenario it was shown that it is possible to assign probabilities based on observed values in a scientific manner. As such, it refutes a sometimes made claim that this is not possible for digital evidence.

This work provides practitioners with the necessary basis to properly evaluate and express uncertainty in their work. This more formalised approach is an improvement over existing informal approaches to forming conclusions based on digital evidence. Furthermore, this approach provides decision makers with a clearer understanding of the strengths and limitations of digital evidence that they use to reach a decision.

The here presented approach has many limitations. It is currently only applicable to a single data point, requires specific locations to be indicated in both propositions and the approach has not been validated against a large data set. Further research should also focus on properly calibrating the probabilities obtained and should study the parameters of the approach, such as the available distributions, minimal quantity requirements for the reference data, and the size of the wedge for distance-evaluation. Nevertheless, even with this limited scope, this approach is an urgently needed first step towards a varied approach for the evaluation of localisation evidence.

Table 3
Assigned probabilities.

Proposition	$Pr(d_n|\phi_n; P_n)$	$Pr(\phi_n|P_n)$
P_1: Location X	0.052	0.912
P_2: Location Y	2×10^{-2}	0.372

Table 4
Verbal scale according to the UK reporting standard (FSR, 2021). The standard bases itself on the presumption that an LR significantly above 1'000 is unlikely to be attainable for a single piece of evidence.

Range of LR	Verbal Equivalent
1–3	*In my opinion the observations are no more probable if* $[P_1]$ *rather than* $[P_2]$ *were true. Therefore, the observations do not assist in addressing which of the two propositions is true.*
4–10	*In my opinion the observations are slightly more probable if* $[P_1]$ *rather than* $[P_2]$ *were true.*
10–100	*In my opinion the observations are more probable if* $[P_1]$ *rather than* $[P_2]$ *were true.*
100 - 1'000	*In my opinion the observations are much more probable if* $[P_1]$ *rather than* $[P_2]$ *were true.*

Acknowledgments

The author would like to extend his thanks to Prof. Eoghan Casey and Jasmin Wyss for their comments on this paper.

Appendix A. Code

The code presented in this paper is available on GitHub (https://github.com/HSpichig/GPS_Evaluation) and will be updated if new developments become available.

References

Aitken, C., Taroni, F., 2004. Statistics and the Evaluation of Evidence for Forensic Scientists, second ed. John Wiley & Sons, Ltd. URL: https://onlinelibrary.wiley.com/doi/10.1002/0470011238 https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470011238.
Bassi, L., Scoundrianos, A., 2022. Bayesian evaluation of digital location evidence: case report of a homicide investigation. In: EAFS Conference 2022 (Stockholm).
Biedermann, A., Kotsoglou, K.N., 2020. Digital evidence exceptionalism? A review and discussion of conceptual hurdles in digital evidence transformation. Forensic Sci. Int.: Synergy 2, 262–274. https://doi.org/10.1016/j.fsisyn.2020.08.004.
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.
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. URL: https://www.sciencedirect.com/science/article/pii/S2666281720300238. doi:10.1016/j.fsidi.2020.300928.
Cedergren, J., 2005. Assisted GPS for Location Based Services. Master's thesis. Blekinge Institute of Technology.
FSR, 2021. Development of evaluative opinions. Technical report FSR-C-118. UK forensic science regulator. Birmingham. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_Interpretation_Appendix_Issue_1__002_.pdf.
Galbraith, C., Smyth, P., Stern, H., 2020. Statistical methods for the forensic analysis of geolocated event data. Forensic Sci. Int.: Digit. Invest. 33, 301009. https://doi.org/10.1016/j.fsidi.2020.301009.
Horsman, G., 2020. Digital evidence certainty descriptors (DECDs). Forensic Sci. Int.: Digit. Invest. 32, 200896. https://doi.org/10.1016/j.fsidi.2019.200896.
Marquis, R., Biedermann, A., Cadola, L., Champod, C., Gueissaz, L., Massonnet, G., Mazzella, W.D., Taroni, F., Hicks, T., 2016. Discussion on how to implement a verbal scale in a forensic laboratory: benefits, pitfalls and suggestions to avoid misunderstandings. Sci. Justice 56, 364–370. https://doi.org/10.1016/j.scijus.2016.05.009. URL: https://www.sciencedirect.com/science/article/pii/S1355030616300338.
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.
NGA, 2022. World geodetic system 1984 (WGS 84). URL: https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs84.
Poser, N., 2017. Judge blocks Google evidence from Troy murder trial. URL: https://www.hawkanalytics.com/judge-blocks-google-evidence-from-troy-murder-trial/.

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

Ryser, E., Jacquet-Chiffelle, D.O., 2021. Accuracy of geolocation metadata on pictures taken using a mobile phone. URL: https://dfrws.org/presentation/accuracy-of-geolocation-metadata-on-pictures-taken-using-a-mobile-phone/.

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

Swiss Federal Court, 2019. ATF 6B_1074/2018. URL: https://www.bger.ch/ext/eurospider/live/de/php/aza/http/index.php?highlight_docid=aza%3A%2F%2F24-01-2019-6B_1074-2018&lang=de&type=show_document&zoom=YES&.

University of Lausanne, Planète UNIL. URL: https://planete.unil.ch/.

Whiffin, I., 2021. Harvested Locations. URL: https://www.doubleblak.com/blogPosts.php?id=16.