Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Google timeline accuracy assessment and error prediction

Andrea Macarulla Rodriguez[a], Christian Tiberius[b], Roel van Bree[a] and Zeno Geradts[a]

[a]Netherlands Forensic Institute, Den Haag, The Netherlands; [b]Delft University of Technology, Delft, The Netherlands

ABSTRACT

Google Location Timeline, once activated, allows to track devices and save their locations. This feature might be useful in the future as available data for evidence in investigations. For that, the court would be interested in the reliability of this data. The position is presented in the form of a pair of coordinates and a radius, hence the estimated area for tracked device is enclosed by a circle. This research focuses on the assessment of the accuracy of the locations given by Google Location History Timeline, which variables affect this accuracy and the initial steps to develop a linear multivariate model that can potentially predict the actual error with respect to the true location considering environmental variables. The determination of the potential influential variables (configuration of mobile device connectivity, speed of movement and environment) was set through a series of experiments in which the true position of the device was recorded with a reference Global Positioning System (GPS) device with a superior order of accuracy. The accuracy was assessed measuring the distance between the Google provided position and the de facto one, later referred to as Google error. If this Google error distance is less than the radius provided, we define it as a hit. The configuration that has the largest hit rate is when the mobile device has GPS available, with a 52% success. Then the use of 3G and 2G connection go with 38% and 33% respectively. The Wi-Fi connection only has a hit rate of 7%. Regarding the means of transport, when the connection is 2G or 3G, the worst results are in Still with a hit rate of 9% and the best in Car with 57%. Regarding the prediction model, the distances and angles from the position of the device to the three nearest cell towers, and the categorical (non-numerical) variables of Environment and means of transport were taking as input variables in this initial study. To evaluate the usability of a model, a Model hit is defined when the actual observation is within the 95% confidence interval provided by the model. Out of the models developed, the one that shows the best results was the one that predicted the accuracy when the used network is 2G, with 76% of Model hits. The second model with best performance had only a 23% success (with the mobile network set to 3G).

Introduction

Google tracks the user's device location through Google Maps, which also works on the iPhone and the web. It's possible to see the Timeline from the user's settings in the Google Maps app on Android. It even shows if the user walked, drove or was in a plane [1].

To access the Timeline, it is necessary to turn on Location History. It can be enabled or disabled in Google Settings on phones running Android 2.3 or higher [2]. When an Android phone is first set up, Google will likely ask to turn Location History on (it's not turned on by default).

Google Location History Timeline has the potential of acting as evidence or assistance in investigations in the future. It can help to open possibilities to track mobile devices of suspects, gathering information about their whereabouts.

In order to be able to harness this information in court or for justice purposes, an assessment of the accuracy has to be performed. Google registers that the mobile device of the suspect was at a certain time in a certain position, and it estimates its own error of X meters. This paper focuses on quantifying this error and the initial steps to develop a linear least squares multivariate model that can potentially predict the actual error with respect to the true location, considering the environmental variables. This work was based on the Thesis of the same name (http://resolver.tudelft.nl/uuid:d8653d95-b2ec-48c5-9e28-0e69758d9053).

Related work and background

Google withholds information about its algorithms on how the location estimation is computed, and which variables and parameters influence on it. For that, in this paper we have studied what variables have the potential to affect Google accuracy, first

quantifying the accuracy, and secondly developing a model that could predict it, when the information of *ground truth* is missing.

Mobile location network

A mobile location network uses a signal from mobile provider. The technology of localization is based on measuring power levels and antenna patterns and uses the concept that a powered mobile phone always communicates in a wireless manner with the base station or cell tower with the best quality signal, which in most cases is the most nearby base station. Knowledge of the location of this base station implies the cell phone is nearby [3,4].

Advanced systems determine the sector in which the mobile phone is located and roughly estimate also the distance to the base station. Further determination can be done by employing signals from adjacent antenna towers [5,31].

Basic positioning methods

Dead reckoning

Dead reckoning is the process of calculating one's current position by using a previously determined position, or fix, and advancing that position based upon known or estimated speeds over elapsed time and course [5,6].

It is useful because this is the simplest way finding the approximate position, although it is the least accurate method [7]. This method could be used when the phone is not receiving signal from Global Navigation Satellite System (GNSS), such as Global Positioning System (GPS) or the Russian Global Navigation Satellite System (GLONASS); or can't connect to a cell tower or base station. It may use the last position known and estimate the new position based on direction and velocity.

Proximity sensing: signal signature

The mobile position is derived from base-station coordinates. It is usually determined by tracking signal signatures or cell identity (Cell ID) of neighbouring base stations [5,8].

Every base station has its own signal pattern, which is usually embedded into its pilot and some synchronization channels. It normally comprises: signal signature estimation, neighbour list update and mobile location analysis.

On the other hand, in an indoors situation, and using Wi-Fi, this method may be used too. Some Wi-Fi have location services capabilities. Wi-Fi positioning takes advantage of the rapid growth in the early 21st century of wireless access points in urban areas [9].

A technique called *fingerprinting* relies on a calibration survey which consists on the recording of the signal strength from several access points in range and storing this information in a database along with the known coordinates of the client device (as an offline phase). One advantage of this approach [10], is that no special hardware is required on the user mobile station (MS) side. A big disadvantage is that trees or buildings may change the fingerprint that corresponds to each location, requiring an update to the fingerprint database, and that the calibration survey has to be done beforehand in the zone of the study.

Google has a fleet of "War-Cars" that systematically photograph streets and gather 3D images of cities and towns around the world, Google's Street View cars are fitted with antennas that scan local Wi-Fi networks and use the data for its location services [11]. This has been quite controversial but can explain the accuracies obtained when Wi-Fi is activated but no connection is established.

Trilateration

Trilateration is the process of determining absolute or relative locations of points by measurement of distances, using the geometry of circles, spheres or triangles [12].

Normally the position of the device is determined using trilateration with time of arrival (ToA) [13,14].

Time difference of arrival (TDOA)

This method follows the same principle as ToA, but this time the measurement is difference in the arrival times between two stations. In this way the location to look is some point of a branch of a hyperbola. Repeating the process with a third tower, another hyperbola is obtained. The intersection of both branches gives the location of the point [12,15].

With this method, the receiver's clock does not have to be synchronized with the network time, because it is the difference in time that is measured.

Angle of arrival (AoA)

This method is based on the measurement of the angle of arrival (AoA) of the signal. Two (or more) oriented bases with directional antennas are necessary. These antennas are capable of measuring the signal arrival angle from the device, and subsequently communicate the information to it. With a simple calculation, the device can determine its own position [12].

Multiple receivers on a base station would calculate the AoA of the cell phone's signal, and this

information would be combined to determine the phone's location on the earth [16].

Generally this measurement is made by measuring the difference in received phase at each element in the antenna array. The delay of arrival at each element is measured directly and converted to an AoA measurement [17].

Location by GNSS

GNSS is a constellation of satellites that transmits signals used for positioning in any part of the globe, whether on land, sea or air. A navigation system based on artificial satellites can provide users with information about the position and time (four dimensions) [18].

For this project, even though there are other GNSS (such as GLONASS or Beidou [19]) we focused on the GPS, given that the experiments were carried out in the Netherlands (Europe) and android version 4.4 states that it uses either mobile networks, Wi-Fi or GPS to locate the device. GPS is a space-based radio-navigation system owned by the United States government and operated by the United States Air Force. It is a GNSS that provides geolocation and time information to a GPS receiver anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites [18].

GPS satellites continuously transmit their current time and position. A GPS receiver tracks multiple satellites and solves equations to determine the position of the receiver and its deviation from true time.

Smartphone devices can also determine their position via Assisted GPS, also known as A-GPS or AGPS. It enhances the performance of standard GPS in devices connected to the cellular network.

A-GPS uses proximity to cellular towers to calculate position when GPS signals are degraded.

The A-GPS servers download the orbital information from the satellite and store it in the database. An A-GPS capable device can connect to these servers and download this information using mobile network radio bearers such as Global System for Mobile communications (GSM), Code Division Multiple Access (CDMA), Wideband Code Division Multiple Access (WCDMA), Long-Term Evolution (LTE) or even using other wireless radio bearers such as Wi-Fi [20].

Received signal strength indication

Received signal strength indicator (RSSI) is a measurement of the power present in a received radio signal [21]. The RSS values are measured in dBm and have typical negative values ranging between 0 dBm (excellent signal) and –110 dBm (extremely poor signal) [9].

The distance is estimated in relation to the strength of the received signal. Estimating the distance to three nearby towers and using trilateration, the position is obtained [12].

Observing the received power, and knowing the emitting power and the loss factor, the distance could be determined comparing the measurements to a reference.

IP address location

This method will detect the device's location based on nearest Public IP Address on the device connected to. They can be either a computer, router or the ISP provider. Depends on the IP information available, but in many cases where the IP is hidden behind Internet Service Provider Network Address Translation (NAT), the accuracy is at the level of city, region or even country [22].

There are several public databases, like Google (https://developers.google.com/maps/documentation/geolocation/intro) or WiGLE (https://wigle.net/), which store SSID of wireless networks, linked to their locations. Google can use WiGLE, database or its own to locate mobile devices based on the Wi-Fi networks nearby.

Multi-linear regression model

Theory

The simplest idea of linear regression summarizes the relationship between a quantitative predictor variable (x) and a quantitative response variable (y) with a straight line [23]. This model is used in the paper to establish a relationship between the environmental variables (mobile network and distance to the towers) and the accuracy and error expected from Google, in order to estimate what these values will be when *ground truth* is no longer present. This model can be extended to handle:

- Several explanatory variables.
- Categorical independent variables and interactions between independent variables.
- Nonlinear relationships.

It is important to note that regression models with observational data can only describe outcomes of processes, but they cannot explain them.

The structure of the data is established by the following equation [24] (process that generated the observations):

$$y_i = \alpha + A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n + e_i$$
$$\text{for } i = 1, 2, ..., m. \quad (1)$$

Where m is the number of observations and n the number of predictor variables or regressors represented by A_{ij}.

A_{ij} is the value of the regressor j in the experiment i.

The parameters (x_i) represent partial effects. Each slope is the effect of the corresponding regressor holding all other predictor variables in the model constant. These parameters define a hyperplane. e_i is the perturbation error term and α the intercept. e_i is the distance from the observations y to the hyperplane. The objective of the least squares method is to find the hyperplane that better adjusts to the observations.

Figure 1 shows a graphic interpretation of the adjusted hyperplane.This gure represents the linear adjustment of z variable for 5 observations. The x, y variables are the regressors. The distances d from observations to the plane represent the residuals e and the projections of the observations on the plane are the predicted values.

With this model we can not only do approximations with the original variables, but we also can extend it to quadratic, cubic, etc. terms or even cross products. In some cases, this will allow for a better fit of the model.

$$\underset{m \times 1}{y} = \underset{m \times (n+1)}{A} \times \underset{(n+1) \times 1}{x} + \underset{m \times 1}{e}. \qquad (2)$$

A is called the model matrix, because it contains all the values of the explanatory variables for each observation in the data.

Every coefficient $x_i (i \geq 1)$ measures the marginal effect that over the response variable y when a predictor variable A_i is incremented leaving the rest of the variables A_j constant, with $j = i$.

How to generate a model for multilinear regression

To develop the multi-linear model, a road map of five steps was followed:

Check the data. Acquire relevant dataset, adequate volume.

Select variables. Choose the variables that have the most direct relationships with the chosen response variable. The aim when selecting variables is to collect the maximum amount of information possible from a minimum number of variables [25].

Test model. The test for significance of regression is performed studying the variance of the errors. The total sum of squares is partitioned into a sum of squares due to regression and a sum of squares due to error.

Correct model problems. The linear model must meet a set of assumptions. If any of these assumptions is violated (i.e. if there are nonlinear relationships between dependent and independent variables or the errors exhibit correlation, heteroscedasticity or non-normality), then the forecasts, confidence intervals and scientific insights yielded by a regression model may be (at best) inefficient or (at worst) seriously biased or misleading [26].

Validate model. The model will be validated using the cross validation k-fold method. Cross-validation is a model assessment technique used to evaluate a machine learning algorithm's performance in making predictions on new datasets that it has not been trained on [27]. This is done by partitioning a dataset and using a subset to train the algorithm and the remaining data for testing. Because cross-validation does not use all of the data to build a model, it is a commonly used method to prevent overfitting during training.

Experiments & terminology

Terminology

In order to assess the accuracy of Google Location Timeline, a series of terms are coined to fill this purpose:

Provided accuracy. The provided accuracy is the radius Google Timeline application gives when registering a position. It is expressed in meters, and it is represented as the radius of the circle around the provided location.

Google error. Google error is the distance between the location provided by Google Timeline and the actual position at that moment. It is calculated based on the position provided by a reference GPS device that will be considered as *ground truth*.

Hit or miss. A location provided by Google with a certain accuracy, it is considered to be a hit if the distance between the actual position and the location is less than the accuracy provided by Google. That is to say, that the actual location falls inside the circle whose centre and radius are the Google location, and the provided Google accuracy. Otherwise, we'll say it is a miss (Figure 1).

All the relevant data concerning the experiment has to be registered for later analysis in the logbook. This data include:

Date and time. To avoid mistakes, Coordinated Universal Time (UTC) is used for everything. Google files and GPS registers use this convention, so only local files in the phones use local time, that has to be translated to UTC before processing. At the end of the experiment the time has to be written too.

Phone 1/2 configuration. This means the kind of signal each device will use. For example, phone 1 with 2G and phone 2 with 3G.

Environment. Rural or urban.

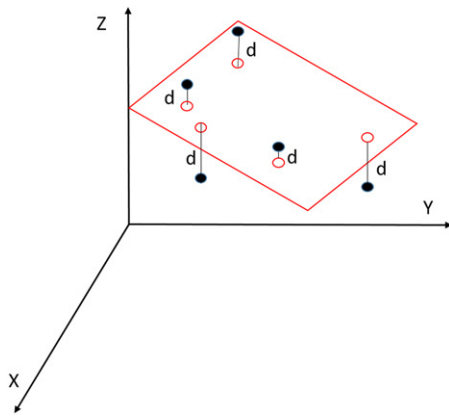*Weather.* Clear, cloudy, rainy.

Traffic. Light, normal, busy.

Figure 1. Graphic interpretation of the adjusted hyperplane. This figure represents the linear adjustment of z variable for 5 observations. The x, y variables are the regressors. The distances "d" from observations to the plane represent the residuals e and the projections of the observations on the plane are the predicted values. Picture based on [29].

Means of transportation. These can be divided into:

- Still. These experiments were at the Netherlands Forensic Institute (NFI), and at home, in Delft.
- Walking. Most of the experiments were between home and the tram stop.
- Bike riding. Most of these experiments were between Delft and NFI.
- Tram travel. Same endpoints as bike, but in public transport.
- Car travel in a rural area. A circuit in the zone of Gouda was run several times with different phone configurations.
- Car travel in an urban area. A circuit in The Hague was run several times with different phone configurations.

In a logbook, important data is noted down to afterwards, record it in an excel sheet. The information registered is:

Date. Current date the experiment takes place.

Start time. The moment the experiment starts.

Mobile configuration. Each mobile is configured to use a unique signal source (2G, 3G, Wi-Fi and GPS).

Environment. As the mobile phones use radio signals to communicate and calculate position, the number of cell-towers and obstacles are important parameters to be considered. So, two environments were defined as *rural* and *urban*.

Weather. This circumstance was noted down for each experiment, to check its dependency with the studied variables. Three values were used: *clear*, *cloudy* and *rainy*.

Traffic. Also the saturation of traffic was registered to check its influence in the study. Three values were used: *light*, *normal* and *busy*.

Experiments

Google Location History is an application that registers the location of the users' smartphones. The first hypothesis we assume is that Google location performance depends of the mobile phone configuration. In order to study the data collection, two phones Huawei G6-U10 with Vodafone SIMCARDs are arranged. Once the Google accounts are registered and logged in the application and activating location history option, Google automatically starts collecting data. These data can be retrieved at any moment from the Google Maps website.

The *ground truth* are the locations registered by an independent and reliable device (handheld GPS). For the experiments not based on the phone GPS capabilities (2G, 3G and Wi-Fi connection), the device used was a GARMIN model GPS Garmin GPSmap 76Cx (designed in USA). This model records the location and time in its SD memory card, and can be downloaded later to a computer.

For the experiments based on the mobile phone GPS capabilities, the *ground truth* has to be a more accurate and precise device. Then the device used was an uBLOX model EVK-M8 (designed in Switzerland). This device is not able to store the data in its own memory. It is connected to a computer, so the experiments with this receiver were run on a car.

To collect the information from the devices above described, a laptop with Microsoft Windows is used. This computer is used to:

- Connect to Google's web page to download the file with Google Timeline. This is done after the experiment has finished.
- Copy the locations registered by GPS Garmin. This is done offline too.
- Extract information stored in the mobile phones, after the experiment has taken place.
- Calculate and store the location processing by uBLOX GPS device output, during the experiment.

Depending of the experiment, different means of transport are used. For still experiments no equipment is needed. Apart from this, bike (personal), tram (public transport) and car (NFI van), are used.

For all the experiments, both smartphones were kept in the pocket of the researcher. Only when the position was Still the smartphones were on the table. Thus, smartphone position with respect to the person's body is not a variable considered for the model.

Experiment execution

For every experiment the procedure is as described in next steps:

Time synchronization. Switch on mobile phones and GPS device. Check they are synchronized in time.

Phone Configuration. Change each mobile phone settings to desired configuration.

Experiment conditions. Note down conditions on logbook.

Logcat registration. Start terminal emulators in phones and register logs.

End experiment. Note end time in logbook.

Data processing. Retrieve and gather all data and process.

With the terminal emulator, two log processes are started in each mobile phone. The result of each log process is called logcat.

Data processing

After the experiments all available data are collected from:

- GPS GARMIN device, in GPX format.
- GPS uBLOX device, in NMEA format.
- Google Timeline, from the web in JSON format.
- Smartphone logcats (Standard and Radio logcat), in TXT format.
- Logbook from Excel, in XLSX format.

All the data is processed with Matlab in gathered in a table.

Experiment summary

In Table 1, a summary of invested time in experiments is shown.

Google accuracy and error assessment

Once all available data is collected from different sources and put together, some preliminary studies have been done to have a global impression of the characteristics of this information. In this section these results are shown numerically and graphically.

The data Google provides is a location, and a measure of accuracy. Location is expressed as latitude and longitude. Accuracy is the radius expressed in meters of the circle around the given location, where the mobile device can be. The experiments consist in registering this information, and at the same time the actual location provided by a *ground truth* device. With this information *Google error* can be derived as the distance between the location provided by Google and the actual location. A *hit* is defined when Google provides an accuracy radius larger than the actual error (the mobile device is inside the circle), and a *miss* when the error is larger than the accuracy radius. Figure 2 defines this concept.

A simple and quantitative view of hits and misses is shown in Figure 3. This figure shows that GPS has the highest hit ratio (52% of hits), followed by 3G, then 2G and the latest is Wi-Fi with only 7% of hits. This low hit ratio in Wi-Fi is because Google is too optimistic when calculating locations using Wi-Fi networks.

Having a look at Figures 2 and 3 we could guess that Google, before assigning an accuracy radius to a calculated location, wonders which confidence interval can apply. If it takes a $\pm\sigma$ for its confidence interval, it would be natural (Figure 4) that it had a 68.27% of hits. Nevertheless, the hit rate seems to be lower in all the cases with respect to the expect 1 σ rule.

In Figure 4 it is shown that for the normal distribution, the values less than one standard deviation away from the mean account for 68.27% of the set; while two standard deviations from the mean account for 95.45%; and three standard deviations account for 99.73%.

Accuracy and error based on phone configuration

To evaluate globally these two variables some histograms were obtained. *Accuracy* is the radius provided by Google to define the circle where the mobile device can be, and *Error* is the actual distance between the provided location and the actual location. In Figure 5, the accuracy and error of 2G measurements are shown. But to have a better comprehension of these distributions, the cumulative distribution is used instead as shown in Figure 6.

Table 1. Experiments duration summary. Durations shown in this table correspond to phone switched on and connected to corresponding network (2G, 3G, Wi-Fi and GPS). For these registered time intervals, logcats were retrieved from the phones and Google was able to register locations (hh:mm).

	Rural	Urban					
Time in experiments	Car	Bike	Car	Still	Tram	Walking	Total
2G	0:28	26:39	4:06	565:41	15:32	1:26	613:52
3G	0:56	32:16	4:06	1 230:16	42:36	4:26	1 314:36
Wi-Fi	0:58	36:51	2:25	1 337:17	52:01	4:02	1 433:34
GPS	0:55	9:44	5:10	261:57	18:55	0:10	296:51
Total	3:17	105:30	15:47	3 395:11	129:04	10:04	3 658:53

Figure 2. Description of what is considered a Google hit/miss. Google gives a radius of accuracy in meters where it is possible to find the device at a given time. Measuring the distance to the *ground truth* point (location provided by GPS device) we determine if the device was truly inside the circle. If it is, we call it a hit, otherwise it is a miss.
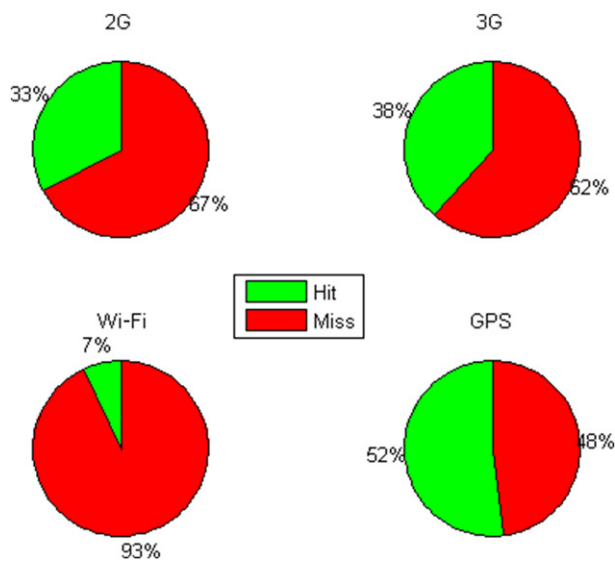

Figure 3. Google hits classified by 2G/3G/Wi-Fi and GPS. Both 2G and 3G show similar results. GPS goes on a first position and Wi-Fi shows the worst result.
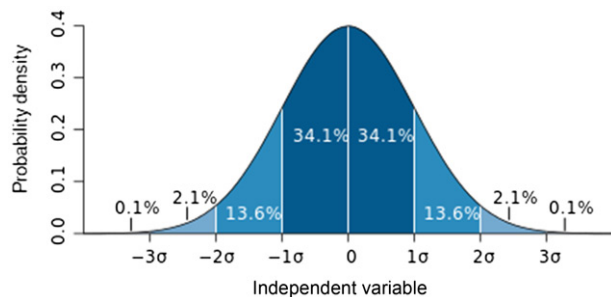

Figure 4. For the normal distribution, the values less than one standard deviation away from the mean account for 68.27% of the set; while two standard deviations from the mean account for 95.45%; and three standard deviations account for 99.73%.

These histograms represent the Google accuracy and error when using 2G network. It can be observed that error has a wider distribution than accuracy. Google location is too optimistic providing small values of accuracy (high precision) when the location has in fact larger error.

In Figure 6A, the cumulative distribution function for Google accuracy is shown. It can be observed that the signal which approaches faster to unity is Wi-Fi. This means that all the accuracies provided by Google when using this signal fall below a lower threshold, that is to say, all the accuracies are small, which means that Google is too optimistic. Next one is GPS, it gives 93% of its accuracies below 80 m and the rest of values are uniformly distributes up to 1 700 m. 2G graph is a stepped one. 52% of its accuracies are below 170 m, then a new increase is about 800 m and then it grow irregularly up to 3 000 m. The last signal, 3G has a uniform increase in accuracy, and is below 2G until 1 800 m. When Figure 6B is studied, one can see that the cumulative distributions or Google error are more uniform than accuracy. 2G and 3G cross each other when error equals 1 800 m, and the order of functions, from best to worse is the same as accuracy: GPS, Wi-Fi, 2G and 3G.

Observing Figure 7, it can be noted that Google is optimistic when providing and accuracy between 300 and 1 250 m when using 2G network because the number of observations in this range is larger than the number of observations with actual error in the same range. In the case of 3G, the results are similar. Google is optimistic between 200 and 1 500 m.

With GPS and Wi-Fi, Google is always optimistic giving any value for accuracy.

Accuracy and error based on environment

In Figure 8, the cumulative distribution functions (CDFs) have been calculated for both environments. In accuracy CDF for rural environment , there is no graph for Wi-Fi because there are no epochs in the
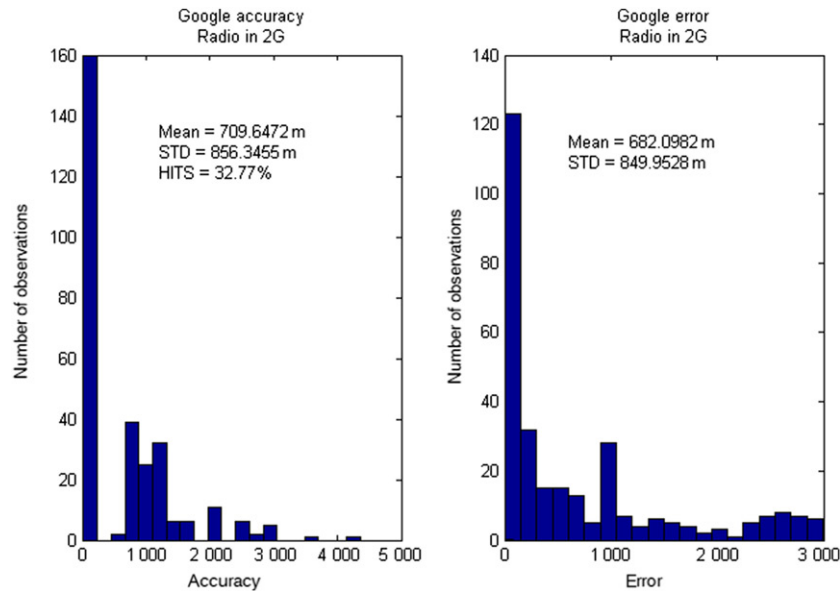
Figure 5. 2G Histograms. These histograms represent the Google accuracy and error when using 2G network. It can be observed that error has a wider distribution than accuracy. Google location is too optimistic providing small values of accuracy (high precision) when the location has in fact larger error.
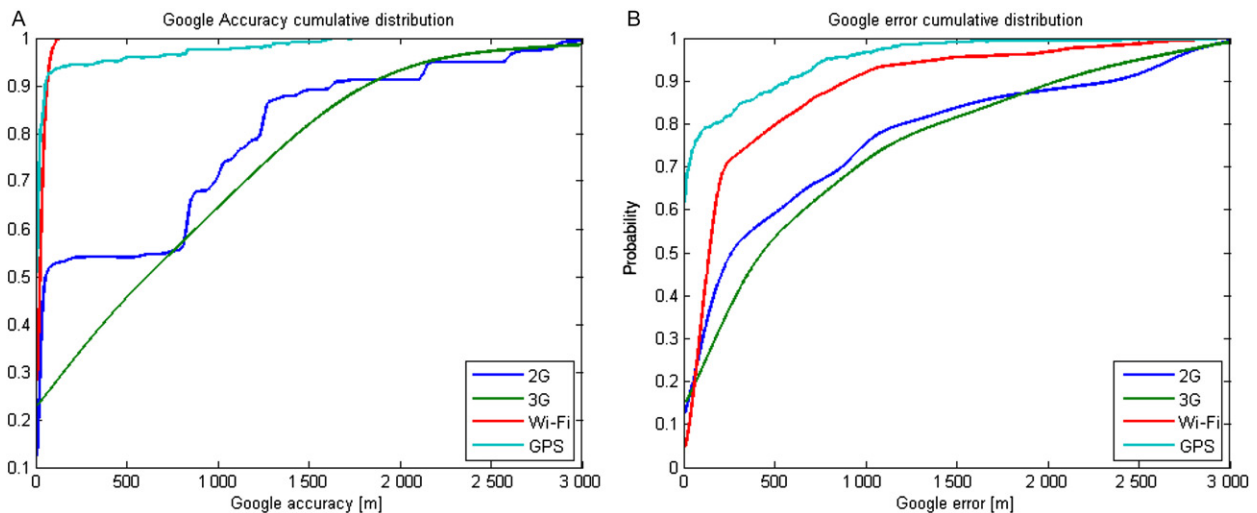

Figure 6. Cumulative distribution functions for accuracy and error. (A) Accuracy cumulative distribution function. (B) Error cumulative distribution function.

experiments. In error CDF for rural environment, 2G and 3G have a similar behaviour and GPS present a soft slope, compared to accuracy. It is strange that GPS doesn't have lower errors. This can be caused due to the elapsed time from the moment the phone is switched on to the time it starts to compute accurate locations, making some blunders using other location methods. The results for urban environment are similar to those explained for Figure 6A, where both environments were not separated. The most important thing to remark is that both, 2G and 3G have very similar behavior. In rural environment it can be checked that the functions are very smooth. This is because there are not many measurements (experiments) and the results are homogeneously distributed. An important aspect is that there is no graph for Wi-Fi in rural

environment. In rural environment, 2G and 3G have similar error distribution, but the accuracy provided by Google is more uniform for 2G than for 3G. Figure 8B shows that in rural environment 2G and 3G have a similar error behaviour and GPS present a soft slope, compared to accuracy. It is strange that GPS doesn't have lower errors. This can be caused due to the elapsed time from the moment the phone is switched on to the time it starts to compute accurate locations, making some blunders using other location methods.

Accuracy based on action

The second division taken into account for studying the data is the speed and kind of movement. This information is registered under the name of *Action*
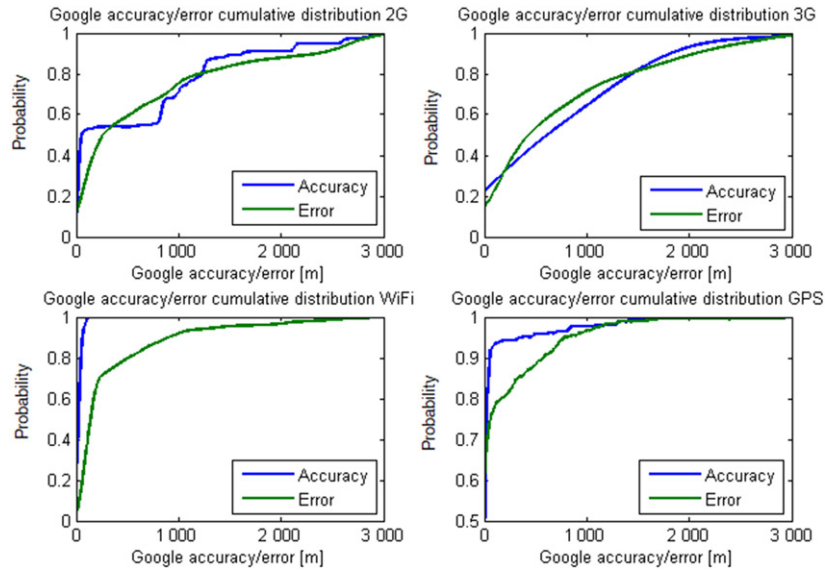
Figure 7. Comparative of Cumulative distributions for accuracy and error for each source of signal.
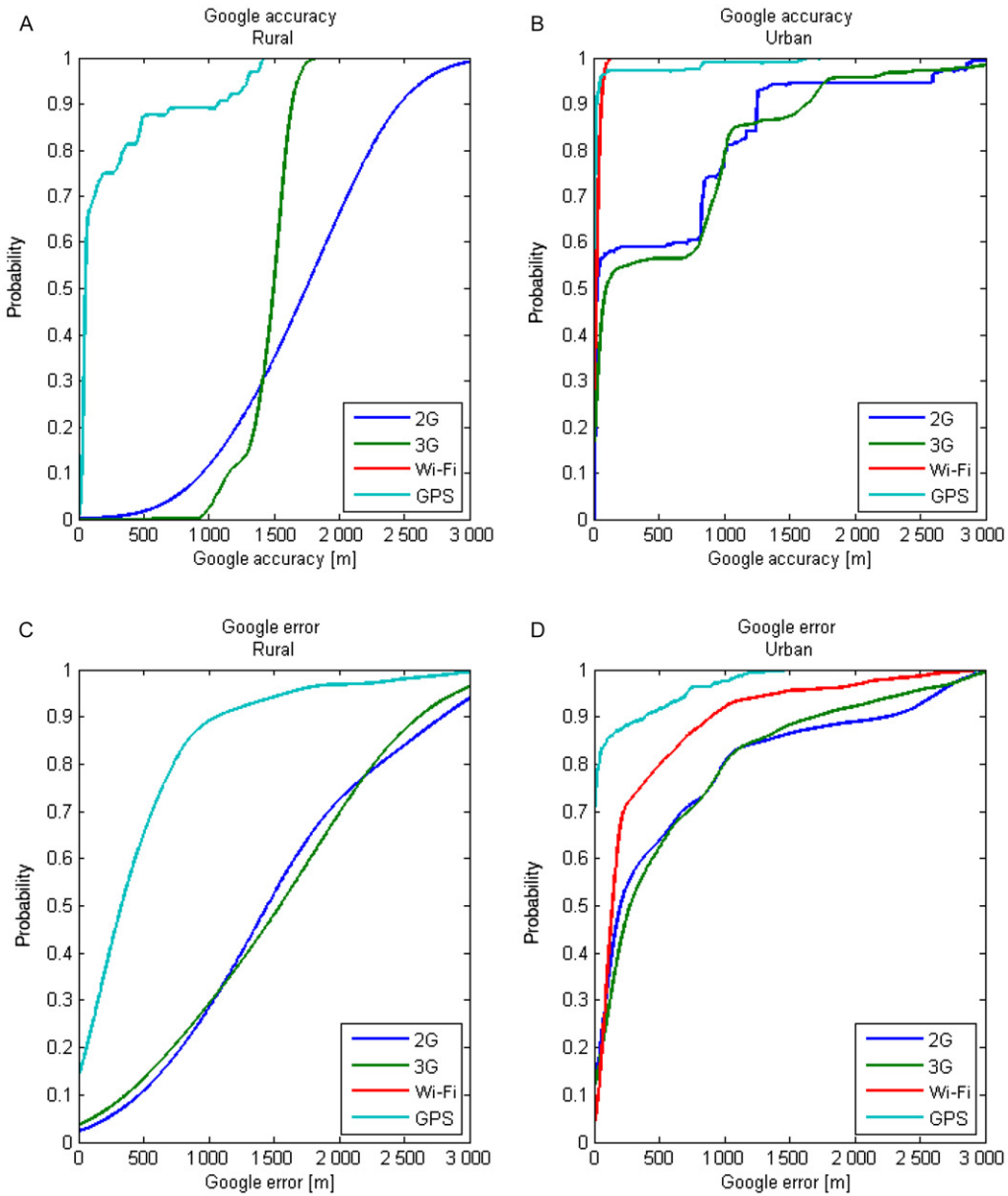

Figure 8. Cumulative distribution function (CDF) for accuracy and error vs. environment. (A & B) Accuracy CDF for different Environment. (C & D) Error CDF for different Environment.

Table 2. Accuracy provided by Google expressed in meters.

Accuracy		2G		3G		Wi-Fi		GPS	
		Median	RMS	Median	RMS	Mian	RMS	Median	RMS
Environment	Rural	1 627.0	1 772.6	1 513.0	1 467.8	–	–	50.0	455.8
	Urban	23.0	922.4	82.0	932.9	20.0	31.3	9.0	198.0
Action	Bike	845.0	842.9	899.0	1 057.2	20.0	31.9	44.0	44.0
	Car	964.0	1 197.4	1 399.0	1 308.9	–	–	9.0	246.8
	Still	20.0	853.0	20.0	344.4	20.0	30.8	20.0	244.9
	Tram	1 169.0	1 255.7	1 000.0	1 276.7	23.0	34.7	13.0	13.0
	Walking	2 857.0	2 857.0	135.0	135.0	11.0	21.8	–	–

RMS: root mean square; -: no data.

and is represented by the means of transport (Figure 9).

In *Still*, Wi-Fi accuracies have small radii, because the measures are taken indoor and the phone connects only to networks in the building, that have all very close locations. 3G gives small accuracies, below 150 m 92% of the measurements. This is very different from 2G that 40% of measurements are quite small and the rest are distributed in increasing distances up and above 3 000 m. In *Tram* the accuracy provided by 2G and 3G are very similar, looking more precise 2G that reaches 98% of measures below 1 400 m. Wi-Fi network gives very low accuracies in every sample, that can be interpreted that Wi-Fi networks are available at every moment in its urban trajectory.

Numerical results

In order to show previous results in a numerical form, some tables have been built.

Tables 2 and 3 give an idea of accuracies reported by Google and Google error values, median and root mean square, dividing the data with two criteria: Source of signal and environment, source of signal and means of transport. The Table 2 is divided into four columns which correspond to data acquired with 2G signal, 3G signal, Wi-Fi signal and GPS. The classification is done with two criteria. First two rows are the Environment division and the other five are the Action. For each division two statistics are shown: median and root mean square (RMS). Urban data is better than Rural, and Tram has the worst results compared to the rest of means of transport, when the signal employed is 2G or 3G. Wi-Fi and GPS has little variations (m). The Table 3 is divided into four columns

which correspond to data acquired with 2G signal, 3G signal, Wi-Fi signal and GPS. The classification is done with two criteria. First two rows are the Environment division and the other five are the Action. For each division two statistics are shown: RMS.

Prediction model results

Next part in the study is to define and evaluate linear regression models for this data. In the Table 3 are shown the six models developed.

Six models were developed (Table 4). Three of them to study Google accuracy and the other three to study the Google error.

Prediction models discussion

A set of 36 random samples has been selected to study the model's performance. With these data all the predictions and measurements are registered in Tables 5 and 6.

For each sample these tables show the limits (Low and High) of the 95% condence interval the linear model provides. The Meas. column gives the real measured value and its background colour is green if the measure falls inside the limits and red if it falls outside.

Data for Google accuracy

The predictions for these random samples, and the real measurements are shown in Tables 3 and 5.

It can be observed that the accuracy estimations when circulating by bicycle have acceptable ranges with 2G. The model is wrong in three out of the three cases. The case 3 is show in Figure 10. This

Table 3. Error measured in meters on Google location.

Google error		2G		3G		Wi-Fi		GPS	
		Median	RMS	Median	RMS	Median	RMS	Median	RMS
Environment	Rural	1 425.2	1 712.0	1 712.6	1 678.8	–	–	300.2	713.6
	Urban	164.6	1 012.0	206.8	938.0	141.7	611.0	4.76	294.7
Action	Bike	1 931.9	1 978.9	416.5	1 321.5	83.3	550.2	26.7	26.7
	Car	654.2	1 207.8	973.1	1 360.0	–	–	4.8	350.3
	Still	118.0	782.2	169.1	280.8	156.1	542.3	395.1	529.5
	Tram	878.1	1 481.1	904.7	1 273.6	96.6	915.1	3.1	3.1
	Walking	2 656.5	2 656.5	21.4	21.4	12.2	28.0	–	–

RMS: root mean square; -: no data.

Table 4. Six linear regression models.

Models	Google accuracy	Google error
2G	model 1	model 4
3G	model 2	model 5
Wi-Fi	model 3	model 6

case is really strange, because Google gave a very good accuracy (44 m) and the value was right (real location at 26.7 m shown in Table 6).

Figure 10 shows the third point for Bike, with 2G connection in Tables 5 and 6. For this point accuracy is 42 m and Google error 26.7 m. The green point represents the real location, which is inside the circle, which represents the position provided by Google and its accuracy. For this point our models predicted an accuracy between 931 and 1 076 m, and an error less than 1 154 m. The confidence interval for accuracy is reasonable and for error is very high. The measures were much better than expected.

On the other hand, when circulating by car the model 2G is right only when the range had negative lower bound (first sample, 63 m of accuracy, when predicted was 271 m or less).

The second sample when traveling by car gives very good accuracy (59 m), better than the predicted by the model (between 318 and 610 m) and the error agrees the accuracy (46 m). The model for

error has a very wide margin (less than 772 m). The position and accuracy of this sample is shown in Figure 11. For this point accuracy is 59 m, and Google error is 47 m (Tables 5 and 6). For this point our models predicted an accuracy between 319 and 610 m, and an error less than 771 m.

The confidence interval for accuracy is reasonable, but the model didn't do a good prediction and for error is very high. The measures were much better than expected.

When the transportation is tram the model always gives high values of accuracy with reasonable margins. Both models (accuracy and error) agree with the measured values in two out of the three cases.

The exception, third case, is shown in Figure 12. In this case, both models gave predicted values higher than real ones: Google accuracy between 989 and 1 211 m, and an Google error between 562 and 2 562 m. The predicted error was greater than (between 562 and 2 562 m) and the real value was very accurate (only 29 m of distance to the real position). The accuracy was 205 m, very small compared to the rest of the Tram samples (1 000 m and 1 247 m).

The confidence interval for accuracy is reasonable, but the model didn't do a good prediction. For Google error, the model gave a very wide interval,

Table 5. Results on accuracy in meters. The predictions are calculated on 36 random samples.

Google accuracy	2G			3G			Wi-Fi		
	Low	High	Mean	Low	High	Mean	Low	High	Mean
Bike	610.2	743.2	845.0	817.8	1 183.5	405.0	30.1	38.2	72.0
	610.2	743.2	845.0	809.3	1 052.8	1016.0	18.6	30.7	32.0
	931.5	1 076.5	44.0	556.3	746.2	21.0	26.7	34.4	36.0
Car	0.0	271.5	63.0	1 112.2	1 662.2	2 0.0	30.9	50.1	45.0
	318.8	610.2	59.0	0.0	357.8	1 571.0	30.3	49.4	20.0
	621.0	883.8	1 347.0	0.0	302.8	1 571.0	22.7	53.1	20.0
						5 0.0			
Still	1 199.2	1 300.4	1 254.0	523.8	629.6	100.0	27.6	32.5	50.0
	730.1	923.7	829.0	485.6	666.8	131.0	27.6	32.5	21.0
	569.6	810.6	171.0	0.0	107.4	131.0	37.7	40.7	83.0
			171.0			28.0			
Tram	1 125.4	1 295.1	1 247.0	788.3	1 013.4	1 000.0	10.7	21.2	24.0
	1 051.6	1 219.0	1 00 0.0	672.1	1 028.1	292.0	46.3	56.0	50.0
	988.6	1 210.5	205.0	718.4	1 070.8	721.0	19.9	29.7	19.0

Table 6. Results on error on selected 36 points sample in meters. The predictions are calculated on the same 36 samples as Table 4.

| Google error	2G			3G			Wi-Fi		
	Low	High	Mean	Low	High	Mean	Low	High	Mean
Bike	0.0	1 578.9	665.3	46.2	1 544.4	300.9	220.4	929.1	45.4
	0.0	1 578.9	640.2	0.0	1 039.6	184.0	132.5	493.0	14.5
	0.0	1 154.1	26.7	659.0	1 287.4	21.1	0.0	351.8	14.9
Car	0.0	825.3	45.2	784.5	2 030.1	11.7	29.8	688.4	39.8
	0.0	771.6	46.6	154.4	569.0	1 054.1	64.5	704.9	17.5
	306.1	1 164.9	799.9	0.0	696.9	15.1	56.7	619.8	26.1
Still	0.0	1 817.2	591.5	0.0	911.4	73.1	0.0	207.4	12.9
	640.8	1 931.6	909.9	0.0	3 747.9	108.9	0.0	207.4	16.5
	0.0	1 020.1	24.5	0.0	2 476.1	20.4	136.9	349.7	39.2
Tram	367.4	1 122.7	539.7	711.7	1 175.5	908.4	322.6	586.6	31.0
	654.4	1 485.5	922.0	0.0	875.3	130.8	269.8	537.1	36.3
	562.1	2 561.9	29.0	122.4	974.0	538.1	90.3	571.4	17.7
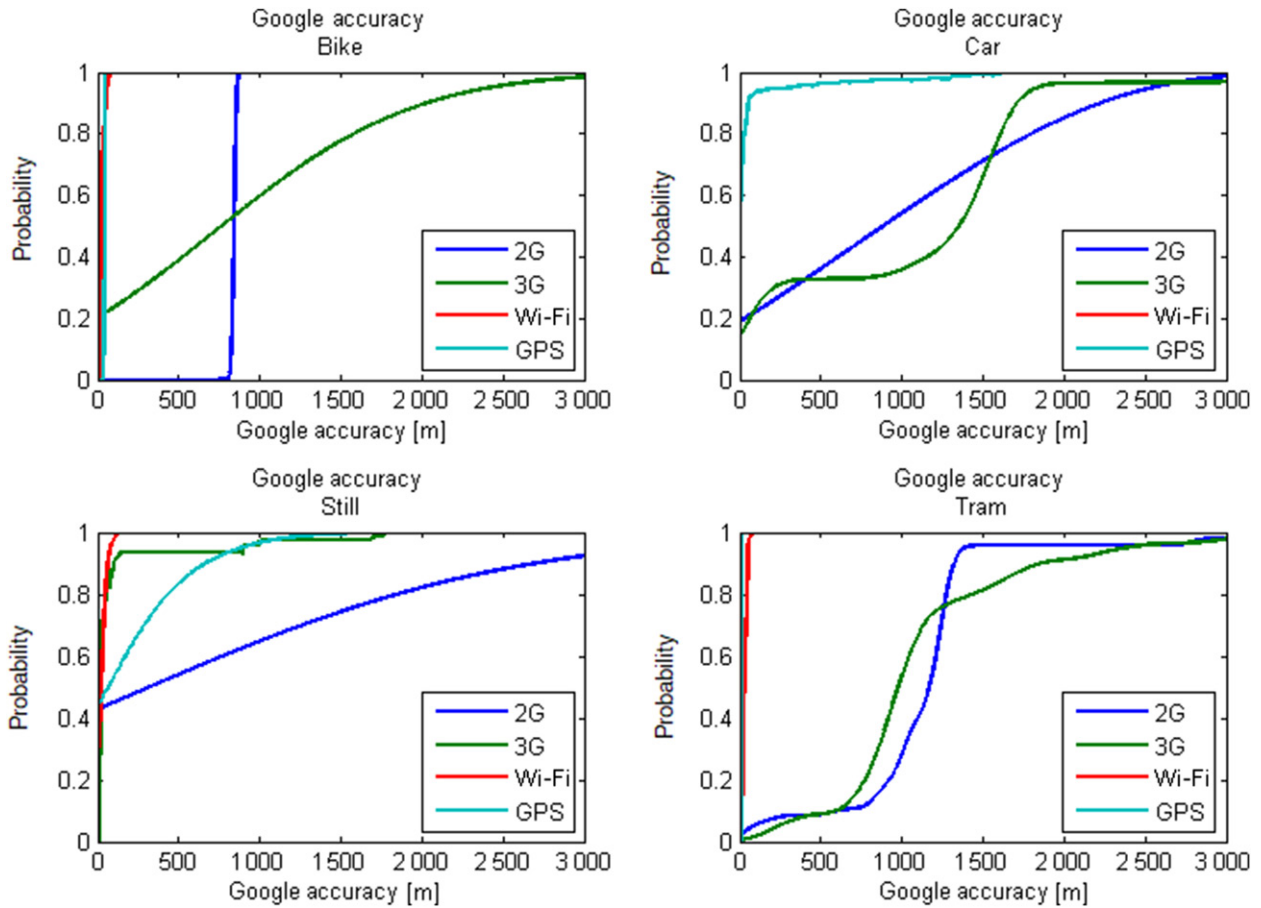
Figure 9. Cumulative Distribution Function for Google accuracy and *Action*. In *Bike*, Wi-Fi network give very small values for accuracy and 2G has a sudden increase at 800 m that means that most of the accuracies have this value. In *Car* GPS give very good values of accuracy.
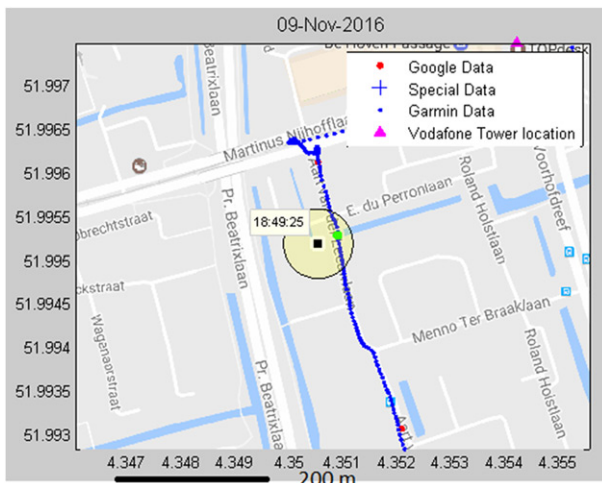

Figure 10. Location of sample 3 on Bike, for model 2G.
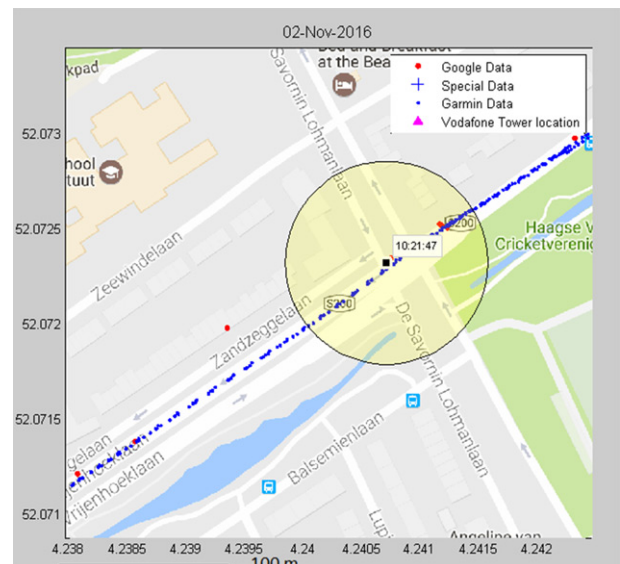

Figure 11. Location of sample 2 on Car, for model 2G.

and the real measure was much lower. The measures were much better than expected, like the other cases, and our models were not able to predict them.

When Google accuracy measures are done with 3G, the model gives reasonable margins except in extreme cases with very high accuracies (more than 1 500 m by car) or very low (28 m when Still or 21 m in Bike).

In these experiments the inferior margin that the model gives is negative and it is used zero instead.

The model is right in the majority of the cases, or it gives accuracy values above the average.

The model which estimates Google accuracy from the Wi-Fi data gives some acceptable margins (none of them negative) and they are close to the registered value given by Google. This value is not correlated with the regressors used, but there is so little variance

that any constant model close to the variable response y can seem valid even when it is not.

Data for Google error

For the positioning error committed by Google some models have been elaborated in order to estimate it. Observing the values obtained with the same sample as the former section it can be observed that many confidence intervals have negative limits, especially when 2G data is taken.

With 2G data, the only limits that maintain themselves positive in all three samples are when traveling by tram. The margins are quite wide and it is right only in two out of the three cases.

In the rest of transportation means, the model is right, but the confidence interval is too wide. In 3G model, the margins of the confidence intervals are

Figure 12. Location of sample 3 on Tram, for model 2G. The green dot inside the circle, indicates the real position at that moment.

also wide, especially in Still. In the tram the model is right in all three cases. The confidence intervals look. The confidence intervals look more uniform in bike than in car. Positioning errors in Google present great dispersion.

In order to evaluate each one of the six developed models in a global way, all the data entries (regressors and observations) have been classified attending to the corresponding model, and for each entry the predicted value has been calculated, as well as the 95% confidence interval.

Thereafter, each observation is looked whether it falls inside its confidence interval or not. If it is inside it is considered as a *Correct Prediction*. Table 7 shows for each model, the number of observations, the Correct Prediction rate, prediction mean and standard deviation, and the confidence interval width (mean and standard deviation). Models which have many Correct Prediction with wide confidence intervals are accurate but not precise. Table 8 shows Median Values provided by Google for Google Accuracy and Google Error expressed in meters and Hit percentage. This table is divided into three columns which correspond to data acquired with 2G signal, 3G signal and WiFi signal. The classication is done with 2 criteria. First two rows are the Environment division and the other four are the Action. For each division three statistics are shown: Google Accuracy and Google Error medians and Hit rate as a percentage.
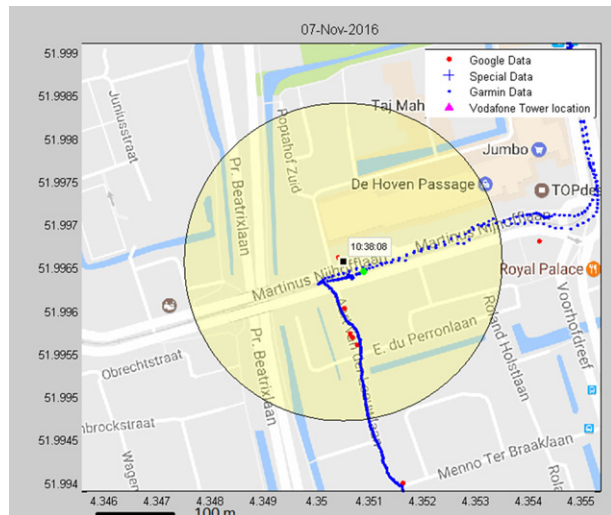
Discussion

In this study we gathered Google Location Timeline data under different circumstances, analyze their veracity, and search for a model in order to predict position accuracy and error. To collect the data

Table 7. Predicted values and 95% confidence intervals.

Linear model		Number of observations	Correct prediction (%)	Predicted value (m) Mean	Std	95% Confidence interval width (m) Mean	Std
Accuracy	2G	297	76	1 111	351	374	167
	3G	779	23	772	388	231	154
	Wi-Fi	2 850	13	31	8.2	5.3	6.1
Error	2G	297	18	460	1 056	2 666	2 088
	3G	842	11	956	1 640	3 404	2 707
	Wi-Fi	3 115	6	297	211	453	622

Table 8. Accuracy, error and hit rate. Median Values provided by Google for Google accuracy and Google Error expressed in meters and hit percentage.

Item		2G Accuracy (m)	Error (m)	Hits (%)	3G Accuracy (m)	Error (m)	Hits (%)	Wi-Fi Accuracy (m)	Error (m)	Hits (%)
Environment	Rural	1 627	1 430	68	1 513	1 710	48.5	–	–	–
	Urban	23	164	24	82	206	33.3	20	142	7.9
Action	Bike	845	1 930	20	899	415	33.3	20.0	8 328.8	28.8
	Car	964	655	59.1	1 399	975	56.9	–	–	–
	Still	20.0	120	9.9	20	170	9.5	20	156	1.6
	Tram	1 169	880	48.9	1 000	905	48.2	23	9 721.5	21.5

several experiments were performed under different conditions on weather, traffic, transportation and environment. To process the data, a multi-linear regression model was designed and applied to the classified data collection. To check the trustworthiness of model we applied it in new data and tested if the prediction adjusts to reality.

Google Timeline provides a dataset with location records. Each register contains among other data, position (latitude and longitude) and a radius (accuracy). In this paper we worked with position and accuracy.

Regarding accuracy stated by Google corresponds to actual accuracy, a summary table (Table 8) is shown below with the experiment measures explained in Section "How to generate a model for multilinear regression" and result tables from Section "Experiment summary".

From this table, we draw the following conclusions:

First, we observe Google's behaviour regarding Environment.

We can see that for both 2G and 3G the hit rate in rural is more realistic (68%–48%) than for the urban environment (24%–33%).

However, looking at the order of magnitude, we can see the interval for both, error and accuracy for the device is narrower in urban (around 20–80 m) than for the rural case, where the accuracy and error are in the order of magnitude of 1500 m. This make measures and positions taken in urban environment more useful to pinpoint where the device was at the given time.

Regarding Wi-Fi connection, we can see in rural environment there is no Wi-Fi connection data and in Urban environment Google is too optimistic, giving smaller radii than the actual errors. This makes Google location by Wi-Fi not reliable.

Secondly, we regard the means of transport:

In 2G and 3G the best results are given by *Car* and they are better in 2G than in 3G. However, with *Still*, the accuracies are small, and although hit rate is low, errors are also small (about 100 m), so Google is not optimistic on its own predictions under these circumstances.

Nonetheless, using Wi-Fi the worst results are at *Still*. This may be because Google likely uses the fingerprinting method with visible networks at any given time. Within a building only the networks of the building itself are visible, and all coincide in position. So Google gives a fixed position when it is inside the building, and the size of the building determines the error. When the device is out of the buildings in the open air, the accuracies that Google gives are very optimistic, and therefore it has low hit rate. However, the errors are not large (less than 100 m).

Examining these results, Google is not totally reliable. Viewing the percentages of hits, none exceeds

70% except when using the GPS that is usually deactivated in mobile phones. But observing the numerical values, although Google does not succeed, the errors are of the same order of magnitude as the accuracies. So even if the phone is not in the circle that Google provides in its Timeline, normally the error is not that big to deny that at least it has been in the whereabouts.

At the same time we discovered that *Weather* and *Traffic* do not affect the performance.

The accuracy and error of Google are very dependent on the four possible phone configurations (2G, 3G, Wi-Fi and GPS). When the connection is 2G, we can suppose that Google determines the device location with RSS.

When 3G is activated, we can only access to the information of the signal strength of the neighbouring towers, but not to their Cell IDs. We can only have access to the complete information about the tower the phone is connected to, which is not enough to perform position computations on our own. However, given that the order of magnitude of both accuracy and error is similar to the one found in 2G, we could also assume that Google has somehow access to the complete information on the neighbouring towers and perform similar computations as 2G configuration.

When using Wi-Fi networks it probably uses fingerprinting methods. While Google vehicles take the Street view information (War-Cars), they take at the same time a fingerprint of available Wi-Fi networks and strengths. This information is compared to the one registered by a mobile device and location is based on best matches. It is possible that these fingerprinting methods are applied by Google with telephony networks too.

It was detected in our experiments that Wi-Fi configuration gives better results outdoors than indoors (Still). This is because when the phone is outside it has several Wi-Fi networks in sight and fingerprinting can be used [28].

It is obvious that when GPS signal is active in the mobile device, Google uses it and with the best results.

In this research we developed several simple linear models to estimate and predict Google provided accuracy and Google error, based on the experiments executed and data recorded by Google and our own location devices. In the case there is a new phone with evidence in it, we would like to know the error really committed by Google, and this way we would be able to do an estimation of the real location of the mobile using the data stored and provided by Google. In this new case there is obviously no *ground truth* data available, we would calculate the location and error with a statistical method.

So, if a phone in a new case has to be investigated as evidence, and we want to apply the linear models

developed in this research, only the Google Timeline *json* file and the Cell Tower database are necessary.

Another source of information could be obtained from the Telecom Company. The cell towers the phone has been connected to, are registered by the Telecom Company and can be used for investigation purposes [8,30]. This is the Call Detailed Record (CDR).

It is evaluated whether the actual measurement of the test data is within the predicted 95% confidence interval. With the hits percentage, the predicted values and the amplitudes of the confidence intervals, Table 6 is elaborated.

With these results we can assure that the best model is Google accuracy for 2G, with a hit rate of 76%. This model presents a large accuracy mean (1 111 m) and a confidence margin of 374 m. The model for Google accuracy for 3G has a success rate of 23%. The predicted Accuracies are smaller and the confidence intervals too.

The other models have few successes and yield only low confidence.

These results indicate that the developed models for Wi-Fi are not adequate for predicting accuracy and error.

However this is a prototype model to put the idea of prediction into practice. With a better tuning and parameter election it could develop into models which provide better predictions and narrower confidence intervals.

Conclusion

Based on the performed experiments, Google locations and their accuracies should not be used in a definite way to determine the location of a mobile device, however, although Google does not succeed, the errors are of the same order of magnitude as the accuracies. So even if the phone is not in the circle that Google provides in its Timeline, normally the error is not that big to deny that at least it has been in the whereabouts.

The linear models developed in this work were improved adding interactions to achieve better predictions and narrower confidence intervals. Even that, the results in this initial study were not satisfactory yet. Further research in the parameters involved and a major collection of data is required.

The linear model is the first step to begin a Big Data Analysis system, and it will surely need much more input than the gathered in this research.

As a final note, the authors remark that this research has been carried out in Europe, thus only considering Google Location Timeline. There exists another map site called site called ditu.google.com, which serves for China location, but the study of it was out of the scope of this research.

References

[1] Heath A. Google can show you everywhere you've been on a map that's surprisingly detailed. Tech Insider; 2016.
[2] Googlemaps. Map types. 2016. Available from: https://developers.google.com/maps/documentation/javascript/maptypes.
[3] STA. Dead reckoning (dr)—ds & ym. 2012. Available from: https://pzsc.org.uk/shorebased/deadreckoning/.
[4] Edens A. Cell phone investigations: search warrants, cell sites and evidence recovery. US: Police Publishing; 2014.
[5] Wang S, Min J, Yi BK. Location based services for mobiles: technologies and standards. Poster session presented at: IEEE ICC: Beijing; 2008.
[6] Hu L, Evans D. Localization for mobile sensor networks. In: Philadelphia PA, editor. Proceedings of the 10th annual international conference on mobile computing and networking. New York (NY); 2004. 45–47.
[7] DEA. Cell phone tracking. 2012. Available from: http://www.largevents.eu/wp/wp-content/uploads/2012/10/Cell_Phone_Tracking.pdf.
[8] Hoy J. Forensic radio survey techniques for cell site analysis. New York (NY): Wiley; 2014. Available from: https://books.google.nl/books?id=ZJLVBQAAQBAJ.
[9] Kotaru M, Joshi K, Bharadia D, et al. Spotfi: decimeter level localization using WiFi. Stanford University; 2015.
[10] Youssef M, Agrawala A. The horus location determination system. In: Wireless networks. New York (NY): Springer;1995. p. 357–374. [Rev Ed. 2009].
[11] Kiss J. Google admits collecting Wi-Fi data through street view scars. 2010. Available from: https://www.theguardian.com/technology/2010/may/15/google-admits-storing-private-data.
[12] Gentile C, Alsindi N, Raulefs R. et al. Geolocation techniques: principles and applications. New York (NY): Springer; 2013.
[13] Hereman W, Murphy WS et al. Determination of a position in three dimensions using trilateration and approximate distances. Decision Sciences; 1995.

[14] Khalaf-Allah M. Time of arrival (toa)-based direct location method. In: 16th International radar symposium (IRS); 2015. p. 812–815.

[15] Munoz D, Lara FB, Vargas C, et al. Position location techniques and applications. Cambridge (MA): Academic Press; 2009.

[16] Hartzell S, Burchett L, Martin R et al. Geolocation of fast-moving objects from satellite-based angle-of-arrival measurements. IEEE J Sel Top Appl Earth Obs Remote Sens. 2015;8:3396–3403.

[17] Fried DL. Differential angle of arrival: theory, evaluation, and measurement feasibility. Radio Sci. 1975;10:71–76.

[18] Misra P, Enge P. Global positioning system: signals, measurements and performance. 2nd ed. New Delhi (India): Ganga-Jamuna Press; 2010.

[19] Ren X, Zhang X, Xie W, et al. Global ionospheric modelling using multi-GNSS: BEIDOU, GALILEO, GLONASS and GPS; 2016. (Scientific Reports).

[20] GPS. Innovation: assisted GPS: a low-infrastructure approach. 2002. Available from: http://gpsworld.com/innovation-assisted-gps-a-low-infrastructure-approach/.

[21] Sauter M. From GSM to LTE: an introduction to mobile networks and mobile broadband. New York (NY): John Wiley & Sons; 2010.

[22] My C. My current location. 2015. Available from: https://mycurrentlocation.net/.

[23] Dekking FM, Kraaikamp C, Lopuhaa HP, Meester LE. A modern introduction to probability and statistics. London (UK): Springer; 2005. Available from: http://www.ebook.de/de/product/3054516/f_m_dekking_c_kraaikamp_h_p_lopuhaa_l_e_meester_a_modern_introduction_to_probability_and_statistics.html.

[24] Tiberius C. Primer on mathematical geodesy. Faculty of Civil Engineering and Geosciences Delft University of Technology; 2014. (Rep. No. CTB3310/CTB3425).

[25] Montgomery DC, Runger GC. Applied statistics and probability for engineers. New York (NY): John Wiley & Sons, Inc; 2014.

[26] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica. 1980;48:817–838.

[27] Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc. 1984;79:575–583.

[28] Chen L, Li B, Zhao K, et al. An improved algorithm to generate a Wi-Fi fingerprint database for indoor positioning. Sensors (Basel). 2013;13: 11085–11096.

[29] Harbaugh JW. A computer method for four-variable trend analysis illustrated by a study of oil-gravity: variations in southeastern Kansas. Kans Geol Surv Bull;1964;171:58.

[30] Zygmunt JH, Samir RD, Ravi J, editors. Proceedings of the 10th annual international conference on Mobile computing and networking; 2004 Sep 26–Oct 1; Philadelphia, PA, USA. New York (NY): ACM; 2004.

[31] Police Publishing. Cell phone investigations series. 2018. Available from: https://books.google.nl/books?id=vTDSrQEACAAJ.