SaPPART Short Term Scientific Mission (STSM)
Scientific Report

“POSITION DATA EVALUATION”

STSM INFORMATION

Reference Number: COST-STSM-TU1302- 29140

Grantee: Dr. Marko Ševrović

Department for transport Planning, Faculty of Transport and Traffic Sciences, University of Zagreb

Zagreb, Croatia

Host: Professor Washington Y. Ochieng and Dr. Shaojun Feng

Imperial college London, Faculty of Engineering, Department of Civil and Environmental Engineering

London, United Kingdom

Period: 07.09.2014 – 18.19.2015

1. PURPOSE OF THE STSM

During STSM a theoretical model of identifying the confidence interval for vehicle position estimation on the expected range of vehicle motion, as a methodological precondition for PVT performance indicators for different applications will be evaluated. Previously collected data set on empirical observations of actual satellite pseudo-distances and/or positions determined in real conditions will be used for error modelling process as well as for statistical analysis of positioning error occurrence probability.

This STSM is linked to the STSM performed by mr. Haris Perakis (COST-STSM-TU1302-21113) and investigates SaPPART key objectives of GNSS receiver evaluation and standardization towards road transport applications. This STSM will further contribute to WP3 of the Action. The results of this STSM can be used by WG1 to identify generic operational scenarios and the development of generic sensitivity analysis method for Deliverable 3 (COST TU1302, MoU, p12).

2. Research problem and work description

During vehicle positioning process, various sources of errors which can include space weather and ionospheric disturbances, ephemeris errors, atomic clock errors and problems with technical equipment can temporary lower the positional accuracy of GNSS system. Additional positioning problems arise in cases where observed vehicle moves through densely populated city centre areas where numerous high objects are present along the road (urban canyons), as well as in cases where vehicle passes under bridges or through tunnels. Deviations which occur during satellite positioning process belong to the group of random errors. Random errors generally occur as a result of larger number of independent causes which influence the overall deviation size at different rates during each individual measurement. Due to that fact, random errors often vary by size and sign, which causes data dissipation and significantly lowers confidence of obtained results. During vehicle positioning process, in cases of more consecutive measurements of vehicle positions and in approximately same external influences, the same GNSS receiver will give results which will have mutually different longitude and latitude error values which will dissipate around certain mean value.

In this STSM, work has been don on testing the possibility of approximating distance error relative frequencies (empirical distance error occurrence probabilities) with various statistical probability distributions during vehicle positioning process inside and outside London city centre area. The main research goal was to determine optimal probability distribution which can be used for precise approximation of random distance error values which occur in the vehicle positioning process in different types of environment.

Representative statistical sample necessary for determination of optimal distance error probability distribution includes datasets of measured and actual vehicle location points (longitude and latitude of current measured and actual vehicle coordinates) on characteristic parts of the observed road network recorded by U-Blox NEO-7P precise point positioning GNSS module and Leica professional Viva GS15 RTK GNSS receiver, respectively. Both datasets were collected in two field surveys conducted at London urban area (outside and inside city centre) by test vehicle equipped with U-Blox NEO-7P GNSS module, professional Viva GS15 Real Time Kinematic GNSS receiver and laptop for storage of measured values.

Relevant characteristics of determined distance errors are described by using the method of descriptive statistic on collected statistical sample. To determine optimal probability distribution, empirical frequencies of distance errors were compared with theoretical frequencies of various statistical probability distributions in EasyFit – Distribution Fitting Software. Three separate statistical tests (Kolmogorov-Smirnov test, Anderson-Darling test and $\chi 02$- test) were performed in order to determine optimal statistical distribution suitable for description of measured GPS positioning error empirical frequencies.

Determined optimal probability distribution can be used in systems that are more vulnerable in terms of positional accuracy such as the Road User Charging systems which require increased precision in determining the location of GNNS receiver.

2.1 Experiment Design and Available Data Analysis

Input empirical data necessary for determination of optimal distance error probability distribution was obtained by using the following equipment in two main datasets:

- Dataset 1: Set of measured vehicle location points (longitude and latitude of current vehicle coordinates) on characteristic parts of the observed road network recorded by U-Blox NEO-7P precise point positioning GNSS module;
- Dataset 2: Set of actual vehicle location points (longitude and latitude of actual vehicle coordinates) on characteristic parts of the observed road network recorded by Leica professional Viva GS15 RTK GNSS receiver.

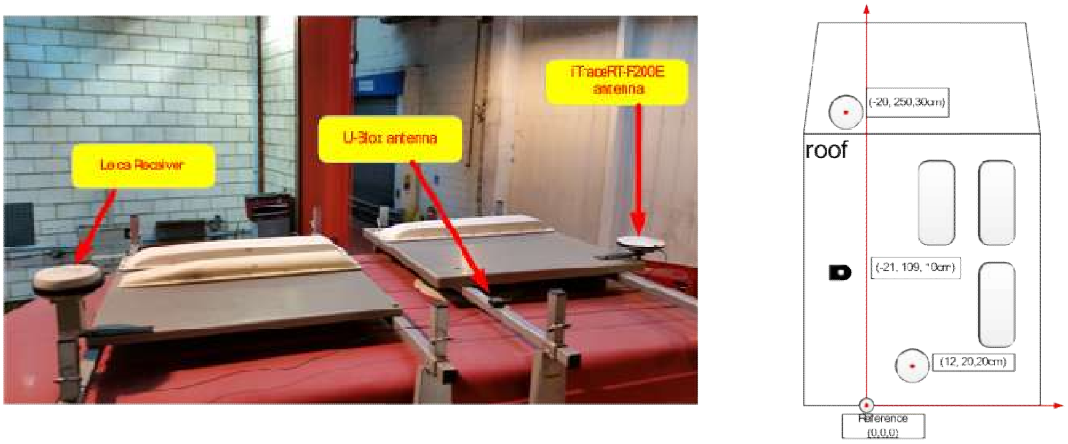
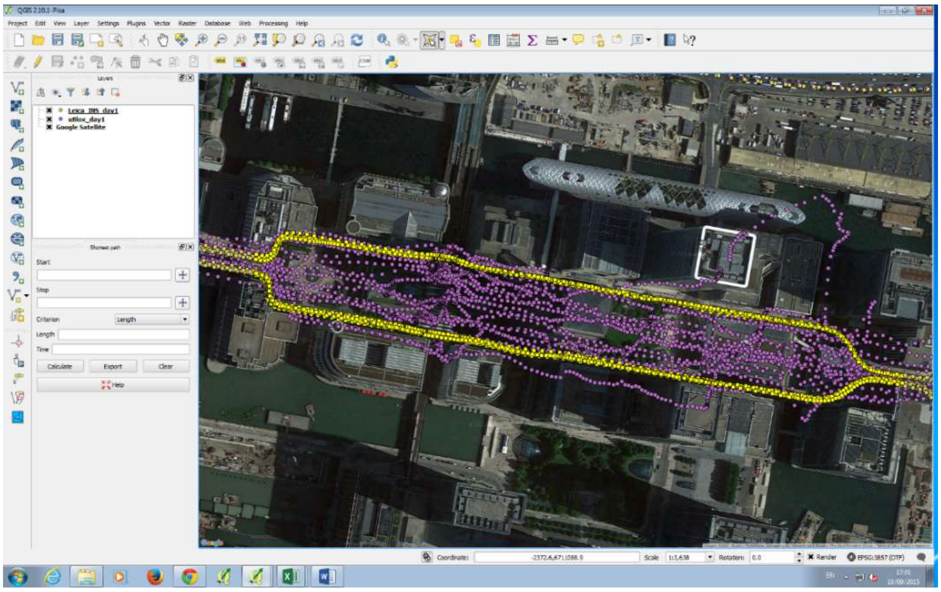
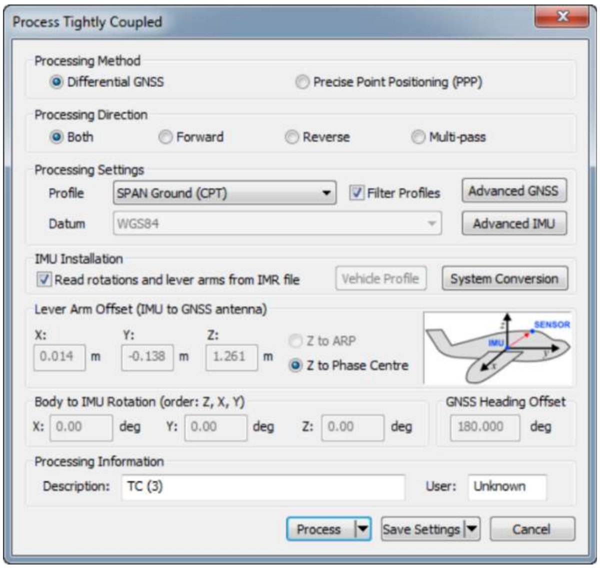


Figure 1-1 Satellite navigation antennas layout of the roof



Both datasets were collected in two field surveys conducted at London urban area (outside

Figure 1-2 Showing uBlox data (purple) vs. ground truth (yellow) in London financial district and inside city centre) by test vehicle equipped with U-Blox NEO-7P GNSS module, professional Viva GS15 Real Time Kinematic GNSS receiver and laptop for storage of measured values. Collected statistical sample includes data on vehicle location points recorded by field measurements conducted in real conditions. Additional data which include information on current date, time, Horizontal and Position Dilution of Precision values were also recorded. The main research goal was to determine optimal probability distribution which can be used for precise approximation of random distance error values which occur in the vehicle positioning process. Due to the fact that Real Time Kinematic GNSS receiver has very high positioning accuracy (distance errors lie in the range from only 1 to 2 centimeters), Viva GS15 proffesional RTK receiver was used determine actual vehicle positions on the observed road network. Very high positioning accuracy is achieved by dual-frequency measurements with real-time corrections, so all expected errors from RTK receiver were negligible and therefore excluded from further analysis .


Process Tightly Coupled

Processing Method
● Differential GNSS ○ Precise Point Positioning (PPP)

Processing Direction
● Both ○ Forward ○ Reverse ○ Multi-pass

Processing Settings
Profile: SPAN Ground (CPT)
☑ Filter Profiles
Datum: WGS84

IMU Installation
☑ Read rotations and lever arms from IMR file

Lever Arm Offset (IMU to GNSS antenna)
X: 0.014 m Y: -0.138 m Z: 1.261 m
○ Z to ARP
● Z to Phase Centre

Body to IMU Rotation (order: Z, X, Y)
X: 0.00 deg Y: 0.00 deg Z: 0.00 deg

GNSS Heading Offset: 180.000 deg

Processing Information
Description: TC (3)
User: Unknown

Figure 1-3 NovaTel Waypoints v8.60. Inertial Explorer was used to post processing of GNSS/INS data in order to get the ground truth

After data were collected, data points from the GNSS receivers were downloaded and stored on a computer and then converted into the appropriate format suitable for statistical analysis (comma-separated values – csv file format). Distance errors were than calculated based on comparison of vehicle location points determined by U-Blox NEO-7P GNSS module and professional Viva GS15 Real Time Kinematic GNSS receiver. Due to relatively large size of statistical sample which was collected on field surveys, detailed statistical analysis was conducted on two stratified random samples (representative sample of 1000 and 500 random vehicle location points).

Vector and raster analysis of collected geospatial data was performed in QGIS - open source GIS analysis application.

In the next steps of data processing, after determined vehicle location points were exported in .csv format, data filtering and data grouping methods were performed. Grouped vehicle location points were than used to calculate values of relevant statistical parameters by methods of descriptive statistics. Descriptive statistical analysis included calculation of absolute and relative frequencies, full and positional mean values as well as dispersion and shape measures for grouped distance errors. Determined empirical probability distributions of absolute and relative distance errors were than compared with various theoretical probability distributions in order to find statistical probability distribution suitable for description of different distance errors values occurrence probability in individual statistical classes.

The first field survey was conducted in London urban area on October 13, 2014. in the period from 09:42:04 to 15:09:39 hours. During that time period, a total of 19656 vehicle location points (longitude and latitude of current vehicle coordinates) were recorded. Second survey inside the London city centre was carried out at the same day in time period from 10:22:15 to 14:44:52 hours during which a total of 3360 vehicle location points were recorded. The geographic coordinates of actual vehicle positions were recorded by Leica professional receiver, while the raw GPS data were recorded by U-Blox GPS receiver (model NEO-7P). On both surveys, individual vehicle location points were recorded at each second of measurement periods. The U-Blox receiver raw data were recorded on laptops, while the Leica receiver data were recorded on its embedded memory card.

For raw GPS data collection, U-Blox NEO-7P precise point positioning GNSS module was used. The NEO-7P module combines the high performance of the u-blox 7 GNSS engine with precise point positioning (PPP) technology for GPS. B-blox precise point positioning algorithm, in combination with SBAS system, provides exceptional precision in clear-sky applications without the need for a reference station. The NEO-7P also supports Differential GPS (DGPS) operation as an alternative to SBAS and PPP technology, using RTCM correction messages from a local reference station or aiding network. Ionospheric corrections received from regional SBAS satellites (WAAS, EGNOS, MSAS) enable the highest stand-alone positioning accuracy from the PPP algorithm. Precise point positioning algorithm delivers its full benefits after the first few minutes of operation with an unobstructed sky view. The NEO-7P also features a front-end SAW RF filter for increased jamming immunity. This is reinforced by sophisticated RF-architecture and interference suppression, ensuring maximum performance even in hostile signal environments. [REF]

The “truth” of the vehicle trajectory is determined by Leica professional Viva GS15 RTK GNSS receiver. This receiver is a dual-frequency, multi-channel instrument, which can lock on to available GPS and GLONASS satellites and receive two signals at different frequencies from each of these satellites if transmitted. The latter feature reduces inaccuracies that result from atmospheric degradation of the satellite signal. As a stand-alone instrument it is capable of giving position and height to an accuracy of about two metres and five metres respectively.

3. EVALUATION AND ANALYSIS OF THE AVAILABLE POSITIONING DATA

3.1 Description of the Methodology

The methodology of selecting optimal statistical probability distribution for approximation of random longitude, latitude and distance error values which occur during vehicle positioning process may be divided in four main parts, i.e. it consists of the following steps:

- Selecting location (suburban, urban area) and time of survey;
- Conducting surveys (collecting input data);
- Conducting statistical analysis of relevant statistical parameters;
- Performing statistical tests to determine appropriate probability distribution.

Values of relevant statistical parameters were determined based on collected data on vehicle location points on the characteristic segments on the observed road network. Statistical analysis of collected data includes the following steps:

Determination of optimal number of statistical classes and grouping of data with defined population size N, permissible deviation level (error value) E, level of confidence and confidence interval;

Descriptive statistical analysis: Calculation of relevant statistical indicators (absolute and relative frequencies, cumulative sequence, full mean values: arithmetic, harmonic and geometric mean, positional mean values: mode, median and percentile values, dispersion measures: variation range, variance, standard deviation and variation coefficient and shape measures: asymmetry and kurtosis coefficients);

Positioning error probability analysis: Statistical hypothesis testing, correlation and regression analysis.

3.2 Descriptive Statistical Analysis

Before performing field surveys, it was necessary to determine optimal size of representative statistical sample. Statistical sample size mainly depends on required accuracy of results, preferred level of confidence during evaluation of relevant parameters for whole population and on standard deviation value of measured parameters. Based on calculated values of arithmetic mean and standard deviation for selected statistical sample it is then possible, with preferred level of confidence, to estimate boundary values within which arithmetic mean and standard deviation of the entire population are present.

In the first phase of the statistical analysis, measured positioning error values were grouped into classes whose width and number were determined pursuant to Sturges rule. After grouping input data into statistical classes, absolute and relative frequencies of measured error values were determined. Other relevant descriptive statistical indicators were also calculated for each data set.

During vehicle positioning process, certain deviations appear between actual vehicle position (real value of observed parameter) and measured vehicle position (measured value of observed parameter). Deviations between current vehicle position determined by GNSS receiver and actual vehicle position occurs due to various errors in satellite positioning system which include appearance of ionospheric and tropospheric delays during signal propagation through Earth's atmosphere, occurrence of multipath effect in urban areas, adverse characteristics of satellite positioning devices, possible additional interferences caused by artificial sources and due to various other imperfections of measurement equipment and GNSS receiver. These deviations are referred to as absolute measurement errors. Size of absolute error (deviation) which occurs during vehicle positioning process is expressed by following expression:

$$E_A = \mathrm{T_i} - \mathrm{T_R} = \sqrt{(x_i - x_R)^2 + (y_i - y_R)^2}$$

where:

E_A − Absolute error (deviation) size during vehicle positioning process;

$\mathrm{T_i}$ − Measured value of observed parameter (vehicle position);

$\mathrm{T_R}$ − Actual value of observed parameter (vehicle position);

x_i, y_i − Measured coordinates of vehicle position (longitude and latitude);

x_R, y_R − Coordinates of actual vehicle position.

For determination of the measured vehicle positions accuracy, in addition to absolute error size, relative error which represents ratio between absolute error value and actual position of the observed vehicle is also calculated. Relative error (deviation) size can be calculated according to following equation:

$$E_R = \frac{\mathrm{T_i} - \mathrm{T_R}}{\mathrm{T_R}} \cdot 100\ [\%]$$

where:

E_R − Relative error (deviation) size during vehicle positioning process;

$\mathrm{T_R}$ − Actual value of observed parameter (vehicle position);

x_R, y_R − Coordinates of actual vehicle position.

Deviations which occur during satellite positioning process belong to the group of random errors. Random errors generally occur as a result of larger number of independent causes which influence the overall deviation size at different rates during each individual measurement. Due to that fact, random errors often vary by size and sign, which causes data dissipation and significantly lowers confidence of obtained results. During vehicle positioning process, in cases of more consecutive measurements of vehicle positions and in approximately same external influences, the same GNSS receiver will give results which will have mutually different longitude and latitude error values which will dissipate around certain mean value.

In order to assess precision of preformed field measurements, it is necessary to evaluate mean squared error for each individual measurement that is the standard deviation value. Standard deviation of individual measured absolute errors (deviations) around determined arithmetic mean value can be calculated by following expression:

$$\mathrm{s} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(E_{Ai} - \overline{E_A})^2} \qquad (1)$$

where:

s – Standard deviation of measured absolute error deviations from arithmetic mean;

$\overline{E_A}$ – Arithmetic mean of measured absolute error (deviation) values during vehicle positioning process;

E_{Ai} – Individual measured absolute error (deviation) values during vehicle positioning process;

n – Number of measurements.

In addition to the mean square error of individual measurements, it is also necessary to calculate error of determined arithmetic mean value. In cases where larger number of vehicle position measurements are performed on different parts of observed road network, different values of deviations are obtained in individual measurements, which results in differences in values of determined arithmetic mean values. Determined arithmetic mean values are, as well as individual deviation values, dissipated around certain mean value. Mean squared error of arithmetic mean values determined at more individual measurements can be calculated according to the following expression:

$$\bar{\mathrm{s}} = \frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(E_{Ai} - \overline{E_A})^2}}{\sqrt{n}} \qquad (2)$$

where:

$\bar{\mathrm{s}}$ – Standard deviation of determined arithmetic mean value;

s – Standard deviation of measured absolute error deviations from arithmetic mean;

n – Number of measurments.

During two separate field surveys, absolute error values between measured and actual vehicle position were determined by GNSS receiver which was installed in the vehicle. Large number of individual measurements were performed (measured current vehicle positions) at consecutive 1 second time intervals. Since performed measurements include large number of individual deviation values, measured values need to be grouped in appropriate number of statistical classes with specified width in order to simplify calculation of arithmetic mean and standard deviation values. Data grouping is performed by dividing error scattering area into several equal parts. Arithmetic mean and standard deviation of grouped data can then be determined by following expressions:

$$\overline{E_{AG}} = E_{A0} + \frac{\mathrm{i}\sum fd}{\sum f} \qquad (3)$$

$$s_G = \mathrm{i}\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} \qquad (4)$$

where:

$\overline{E_{AG}}$ – Arithmetic mean of grouped absolute error (deviation) values during vehicle positioning process;

s_G – Standard deviation of grouped measured deviations from arithmetic mean

i – Statistical class width;

f – absolute frequency of observed parameter,

d – Number of statistical classes between arbitrarily selected null class and observed class.

Besides aforementioned parameters, statistical analysis also included calculation of Longitude and Latitude Error in measured vehicle position (Ex, Ey), Positioning Distance Error (DE), Longitude and Latitude mean and standard deviation, distance root mean squared values (DRMS, 2DMRS) and Circular Error Probability value (CEP).

Longitude and Latitude Error in measured vehicle position can be determined by the following equations:

$$E_x = x_i - x_T; \qquad E_y = y_i - y_T;$$

Where:

Ex, Ey - Longitude and Latitude Error in measured vehicle position;

xi , yi - measured Longitude and Latitude coordinates of vehicle;

xT, yT - the actual geographic coordinates of vehicle determined by Real Time Kinematic GPS receiver.

Positioning Distance Error can be determined using the following expression:

$$D_E = \sqrt{E_x^2 + E_y^2} = \sqrt{(x_i - x_T)^2 + (y_i - y_T)^2}$$

3.3 Descriptive Statistical Analysis Results

Relevant characteristics of observed distance errors are described by using the method of descriptive statistic on collected statistical sample which includes measured GPS positioning error values inside and outside London city area. Due to the fact that collected statistical sample is relatively large in size (it includes total of 19656 vehicle location points), it was necessary to create two additional smaller subsets of data in order to increase statistical analysis results confidence level. This was achieved by random selection of 1000 and 500 vehicle location points from the collected statistical sample, respectively. Determined distance errors inside London city centre area were analysed seperatly in third data subset which includes a total of 3360 vehicle location points. Therefore, relevant descriptive statistical indicators were calculated for four following datasets:

- Full dataset, which includes all recorded vehicle location points inside and outside London city centre area (a total of 19656 vehicle location points).
- First data subset, which includes a total of 1000 radnomly selected vehicle location points from the initial stastistical sample.
- Second data subset, which includes a total of 500 radnomly selected vehicle location points from the initial stastistical sample.
- Third data subset, which includes only vehicle location points recorded inside London city centre area (a total of 3360 vehicle location points).

Extreme distance error values (distance errors larger than 25 meters) were filtered out and excluded from initial statistical sample prior to performing descriptive statistical analysis.

Results of performed descriptive statistical analysis on full dataset are showed on Figures 3-1 to 3-4. Figure 3-1 shows determined distance error absolute values (positioning with with U-Blox NEO-7P GNSS receiver), while Figure 3-2 shows histogram of determined distance errors relative frequencies. Resulting relative cumulative and percentile functions of determined distance errors for full dataset are shown on Figures 3-3 and 3-4.

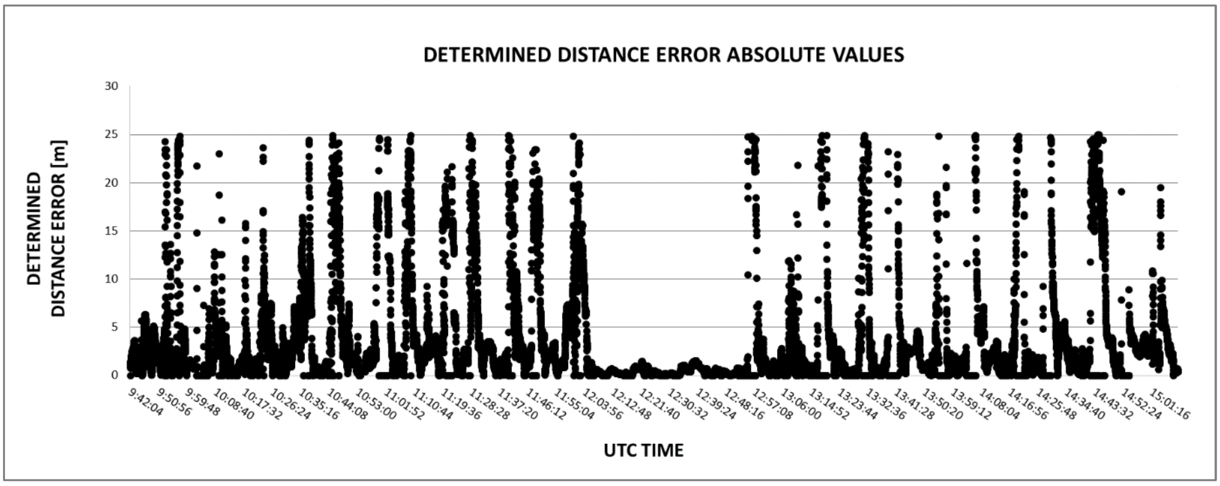


Figure 3-1 Determined distance error absolute values (full dataset).

Based on conducted statistical analysis on full dataset, it is determined that absolute GPS positioning error values lie in the range from 0.00018 to 24.9967 meters with weighted arithmetic mean value of 3.21 meters. Median value is somewhat lower from arithmetic mean value and it amounts to 1.42 meters. Average deviation from arithmetic mean value is 4.77 meters. Variation coefficient shows 148.60 percent of dispersion of measuered deviations from arithmetic mean value. In addition to aforementioned parameters, percentiles of GPS absolute error values, variation range and shape measures were also determined.

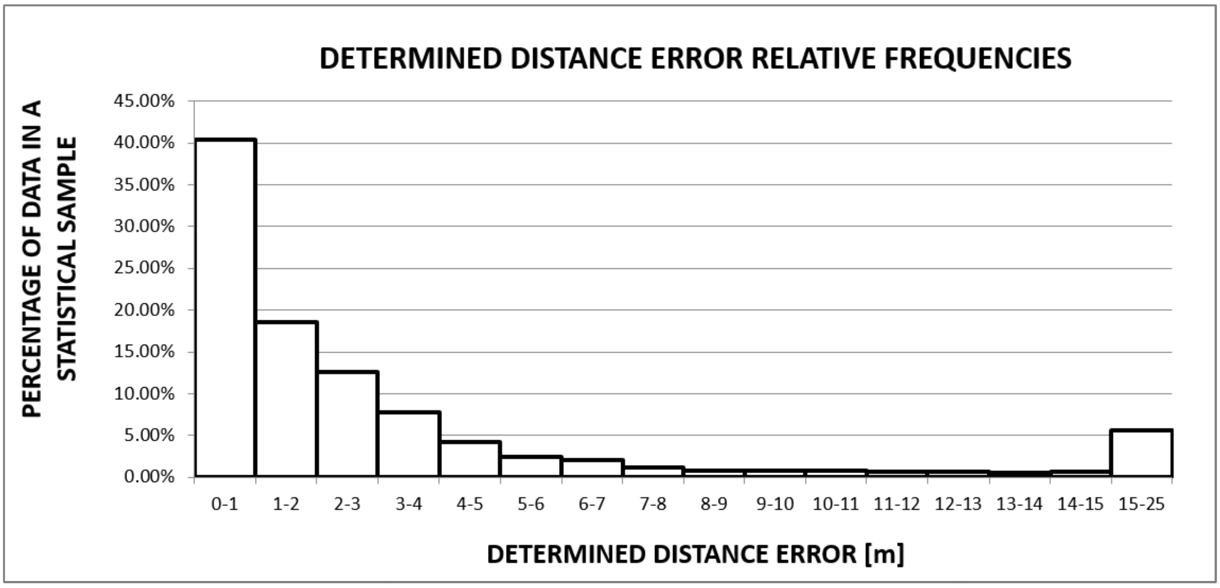


Figure 3-2 Histogram of determined distance errors relative frequencies (full dataset).

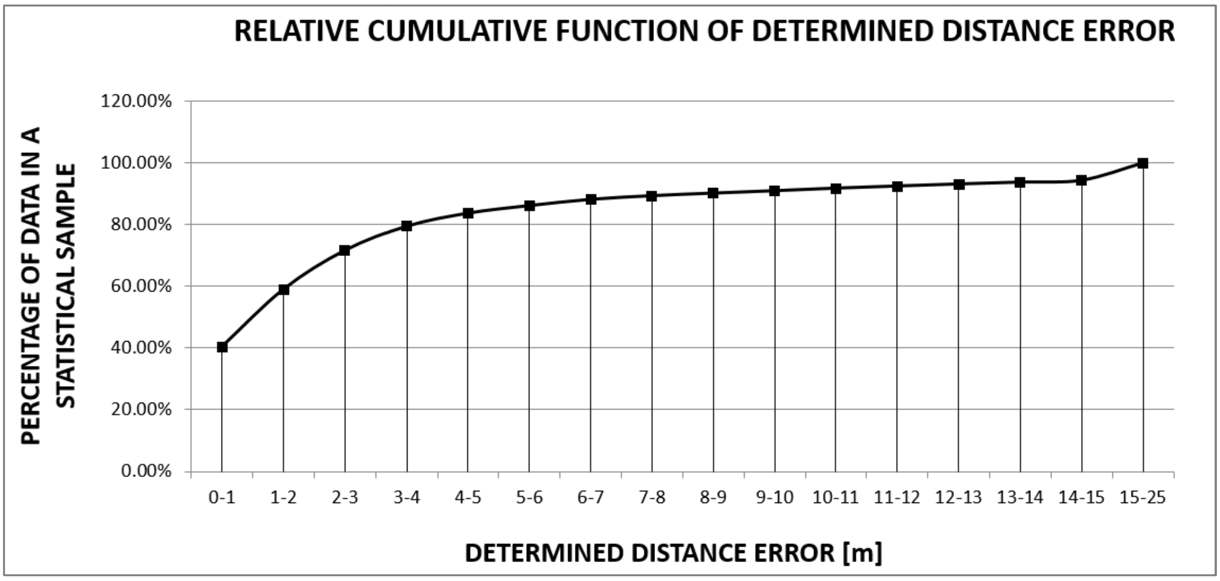


Figure 3-3 Relative cumulative function of determined distance errors (full dataset).

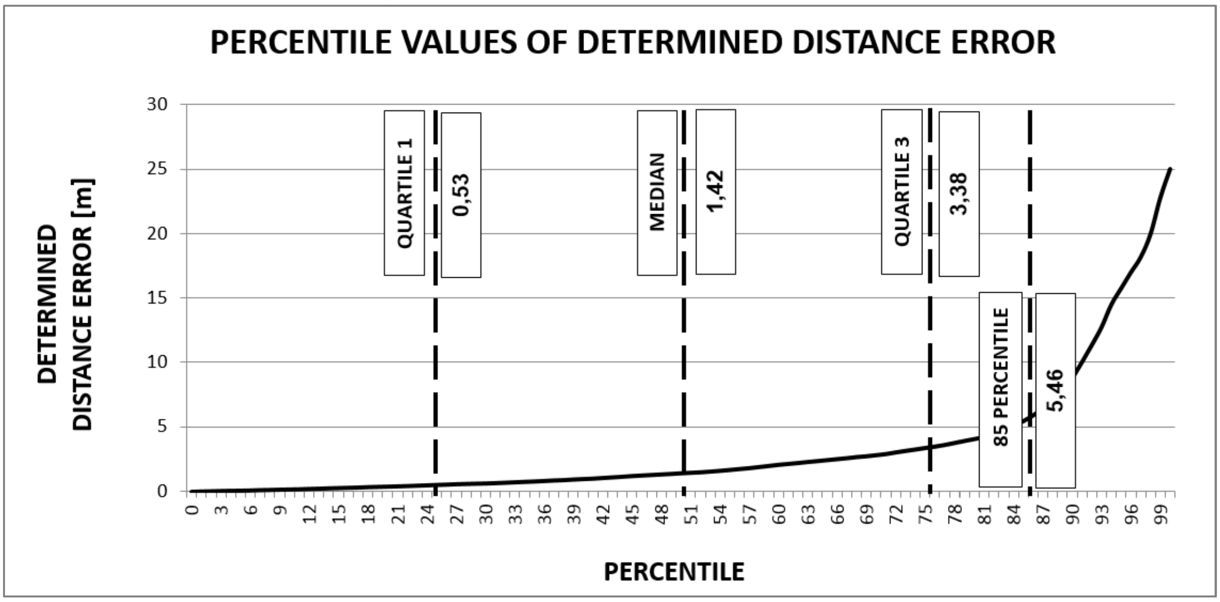


Figure 3-4 Percentile function of determined distance errors (full dataset).

Results of performed descriptive statistical analysis on second data subset (statistical sample of 500 random points) are showed on Figures 3-5 to 3-7. Figure 3-5 shows histogram of determined distance errors relative frequencies, while resulting relative cumulative and percentile functions of determined distance errors for second data subset are shown on Figures 3-6 and 3-7 Based on conducted statistical analysis on second data subset, it is determined that absolute GPS positioning error values lie in the range from 0.00034 to 24.78 meters with weighted arithmetic mean value of 3.17 meters. Median value is somewhat lower from arithmetic mean value and it amounts to 1.58 meters. Average deviation from arithmetic mean value is 4.50 meters. Variation coefficient shows 142.20 percent of dispersion of measuered deviations from arithmetic mean value.

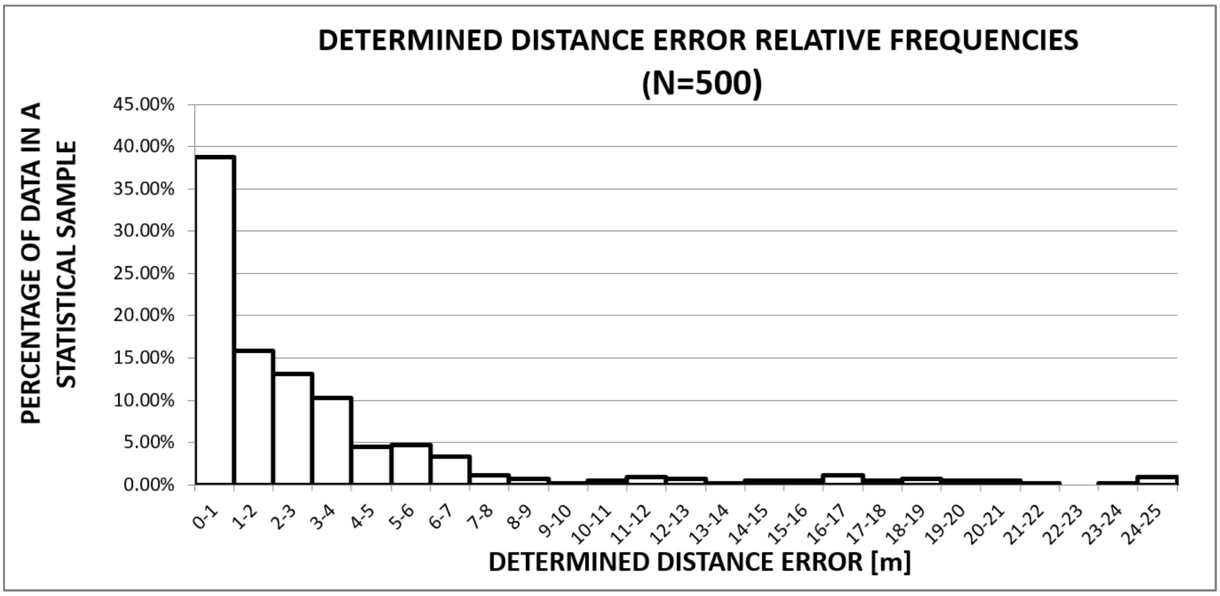


Figure 3-5 Histogram of determined distance errors relative frequencies (statistical sample of 500 random points).

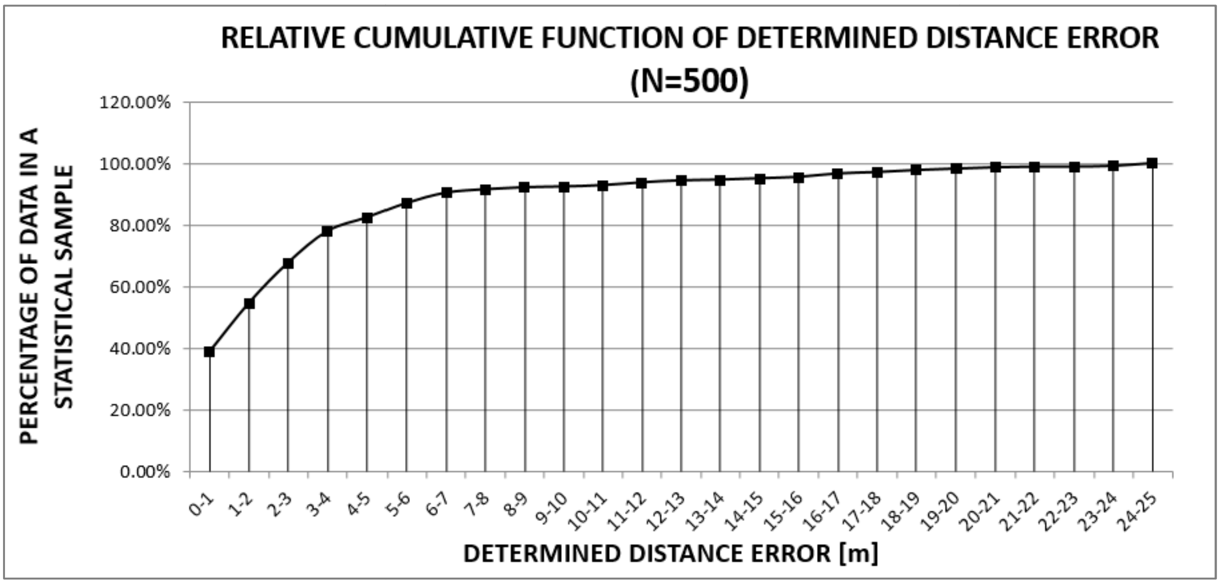


Figure 3-6 Relative cumulative function of determined distance errors (statistical sample of 500 random points).

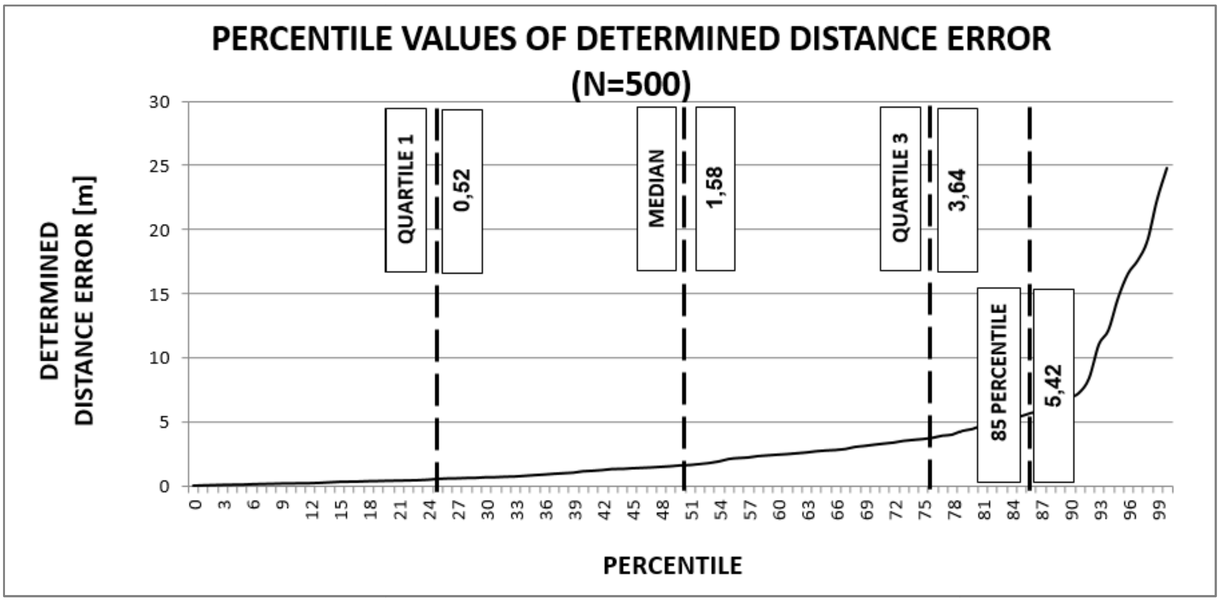


Figure 3-7 Percentile function of determined distance errors (statistical sample of 500 random points).

Results of performed descriptive statistical analysis on third data subset (statistical sample collected inside London city centre area) are showed on Figures 8 to 10. Measured values were first grouped into 25 statistical classes. After that, absolute and relative frequencies of GPS positioning error values were determined. Histogram of relative frequencies of GPS positioning error values (Figure 3-8) shows that majority of measured error values are present in the range from 0.03 to 24.99 meters. Error values less than 1 meter are observed in only 4.02% of cases, while error values greater than 10 meters occur very often (in 55.45% of cases).

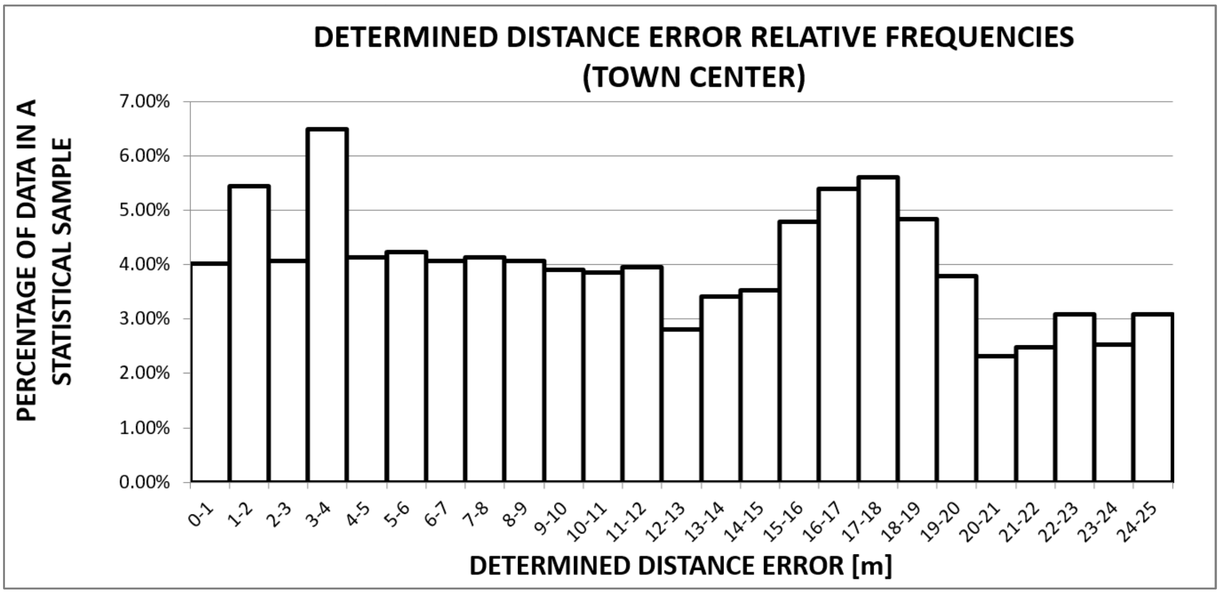


Figure 3-8 Histogram of determined distance errors relative frequencies (statistical sample collected inside London city centre area).

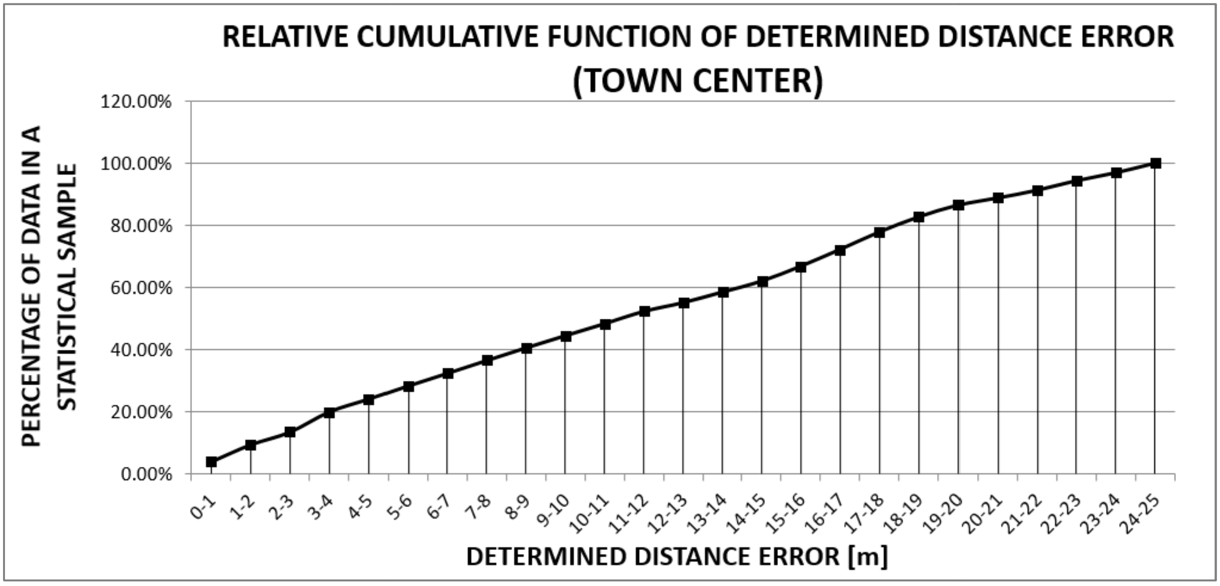


Figure 3-9 Relative cumulative function of determined distance errors (statistical sample collected inside London city centre area).

In the second phase of analysis, absolute and relative cumulative function of determined distance error values was determined. The measured positioning error values were than classified in ascending order (from minimal to maximal value) and percentile values of observed characteristics were also calculated.

Relative cumulative function of determined GPS positioning error values in London city centre area also indicates that occurrences of deviations greater than 10 meters occur very often (Figure 3-9). Percentile function of absolute GPS positioning error values shows nearly constant linear growth of deviation values over whole range of data between minimal and maximum determined error value (Figure 3-10).

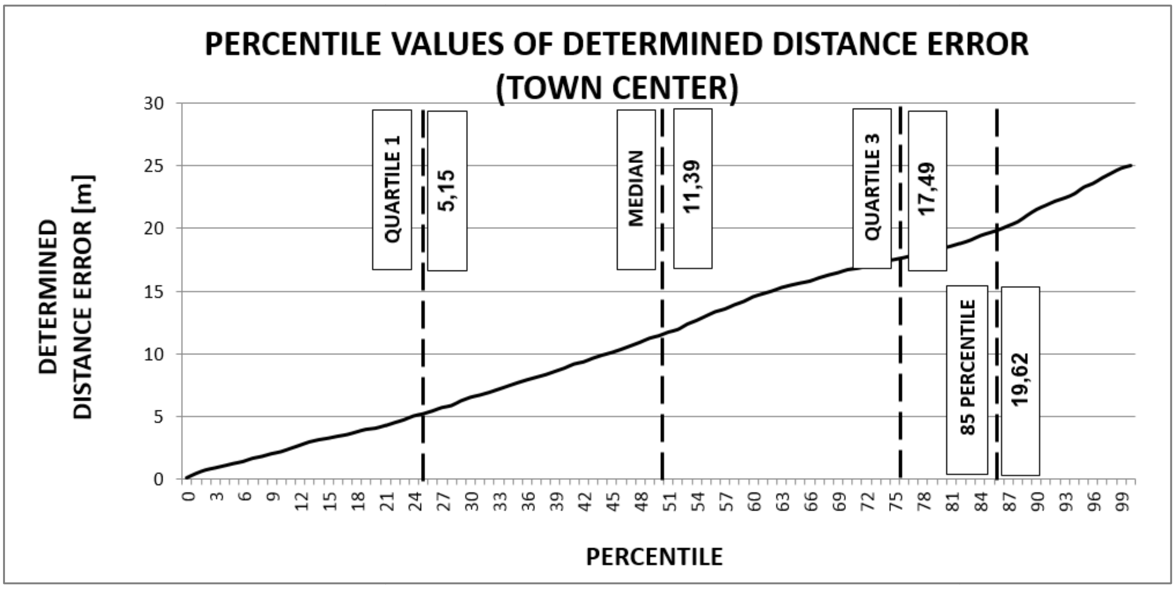


Figure 3-10 Percentile function of determined distance errors (statistical sample collected inside London city centre area).

4. POSITIONING ERROR PROBABILITY ANALYSIS

In the final phase of statistical analysis, three separate statistical test were performed to in order to determine optimal statistical distribution suitable for description of measured GPS positioning error empirical frequencies. On each dataset we have tested the hypothesis that measured error values can be accurately approximated by Gamma probability distribution. Since the results of performed statistical tests showed that initial hypothesis was rejected, we have conducted a series of additional statistical tests on different types of probability distributions. Additional statistical tests were performed with EasyFit – Distribution Fitting Software. Final results of the performed statistical tests showed that 3-parameter Dagum probability distribution can be used for accurate approximation of measured positioning error values on both datasets (suburban and urban environment).

Three separate statistical tests (Kolmogorov-Smirnov test, Anderson-Darling test and χ02-test) were performed in order to determine optimal statistical distribution suitable for description of measured GPS positioning error empirical frequencies. Results of conducted statistical tests show that initial hypothesis on Gamma distribution of measured error values is rejected due to excessive deviations between empirical and theoretical frequencies.

Since initial hypothesis was rejected, we have performed a series of additional statistical tests in order to determine probability distribution which is suitable for approximation of measured distance error values. Final results of performed tests have shown that Dagum probability distribution can be used for accurate description of empirical distance error values both in urban and suburban environments.

Results of conducted statistical tests show that minimal deviations between empirical and theoretical frequencies are determined for 3-parameter Dagum probability distribution. Dagum distribution is a family of curves which includes different probability density functions depending on selected shape, scale and location parameter values. The 3-parameter Dagum distribution is the 4-parameter generalized beta II distribution with shape parameter q=1. It is known under various other names, such as the Burr III, inverse Burr, beta-K, and 3-parameter kappa distribution. It can be considered a generalized log-logistic distribution. Some distributions which are special cases of the 3-parameter Dagum are the inverse Lomax (a=1), Fisk (p=1), and the inverse paralogistic (a=p) (Kleiber and Kotz (2003)). Hypothesis on Dagum probability distribution of measured distance error values is accepted for both area types (inside and outside London city centre area).

4.1 Full Dataset

Results of statistical tests performed on full dataset which includes 19656 vehicle location points show that initial hypothesis on Gamma distribution of distance errors is rejected. Relatively large deviations between empirical and theoretical frequencies are determined in second statistical class (absolute distance error values fom 1 to 2 meters). Comparation of empirical and theoretical frequencies derived from Gamma probability distribution are shown on Figure 4-1.

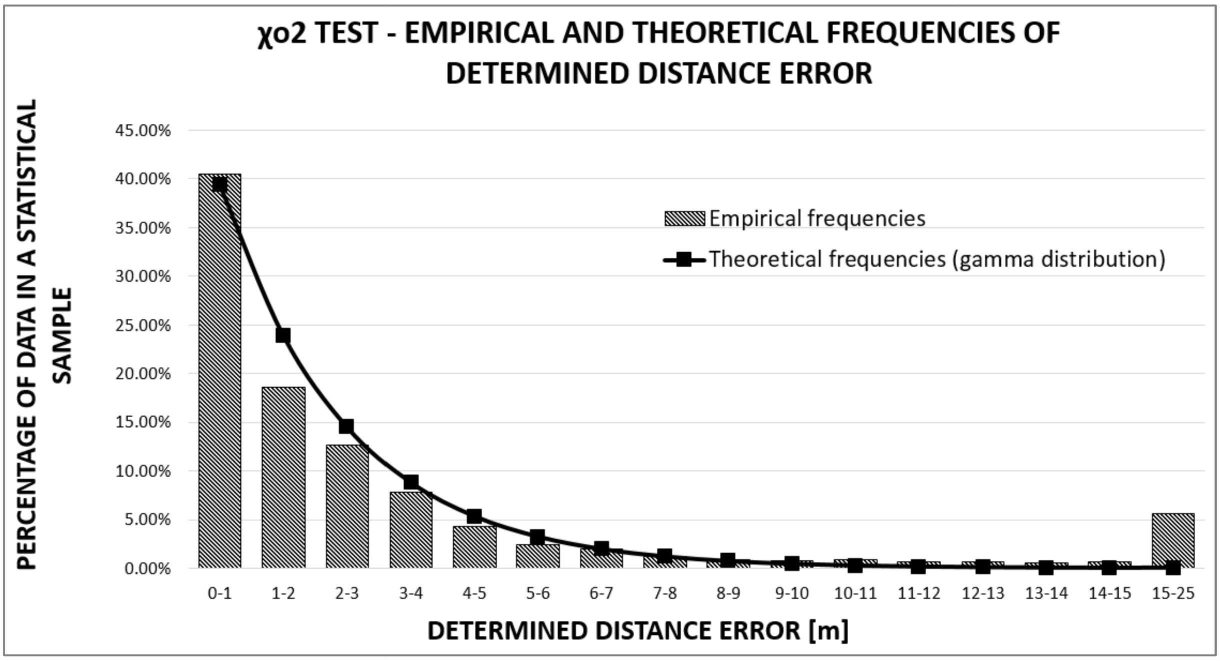


Figure 4-1 Comparison of empirical and theoretical frequencies of determined distance errors on full dataset (results for Gamma probability distribution hypothesis test).

Additional statistical tests performed in EasyFit – Distribution Fitting Software showed that determined distance errors can be precisely described by three-parameter Dagum probability distribution.

4.2 Random Points

Full dataset of measured vehicle position points includes total of 19656 pairs of coordinates. Due to the fact that inappropriate statistical sample size can have significant negative influence on results obtained from perfomed statistical tests which can result in false rejection or acceptance of initial hypothesis, it was necessary to reduce initial statistical sample in order to increase statistical tests confidence level. This was achieved by random selection of partial data subsets (random vehicle position points) from the original collected statistical sample. Extreme distance error values (errors larger than 25 meters) were also removed from further analysis. In that way, two new data subsets were formed which include 1000 and 500 randomly chosen vehicle position points, respectively. Initial hypothesis on Gamma probability distribution of distance errors was then tested on both subsets of data. Results of these tests have showed that initial null hypothesis is rejected regardless of the sample size (Figure 4-2). To determine optimal probability distribution, empirical frequencies of distance errors were compared with theoretical frequencies of various statistical probability distributions in EasyFit – Distribution Fitting Software. Obtained results in both cases (statistical samples of 1000 and 500 points) indicate that best fit with empirical data is achieved by three-parameter Dagum probability distribution (Figure 4-3). Optimal values of continuous shape, scale and location parameter of Dagum probability distribution for statistical sample of 1000 random points amounts to α = 1.3956, β = 2.2611 and γ = 0.64849, respectively. Results of statistical tests performed on statistical sample of 500 random points have showed somewhat different optimal values of shape, scale and location parameter (α = 1.7335, β = 3.5209 and γ = 0.43951).

Dagum cumulative distribution function of determined distance errors on statistical sample of 500 random points is shown on Figure 4-3. Form the cumulative Dagum distribution function, it is evident that 65 percent of measured GPS positioning error values lie in the range from 0 to 2.5 meters, 90 percent of all measured error values lie in the range between 0 and 8 meters, 95 percent in the range between 0 and 12 meters, while 99 percent of all measured positioning error values lie in the range from 0 to 24 meters (Figure 4-4).

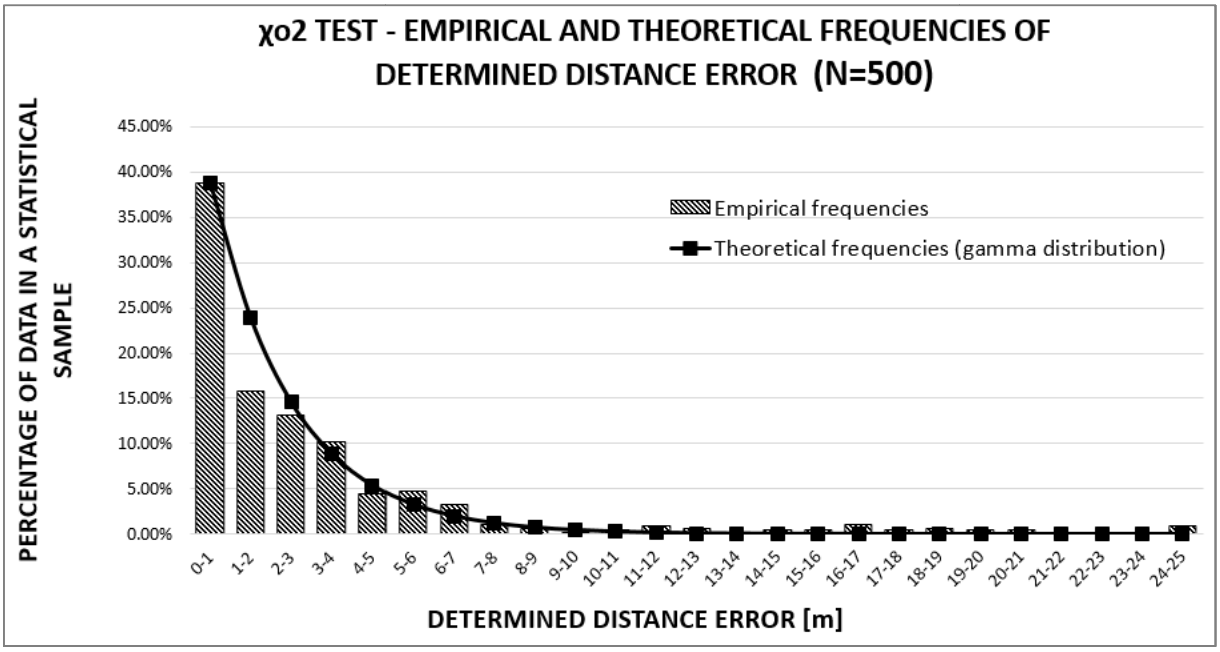


Figure 4-2 Comparison of empirical and theoretical frequencies of determined distance errors on statistical sample of 500 random points (results for Gamma probability distribution hypothesis test).

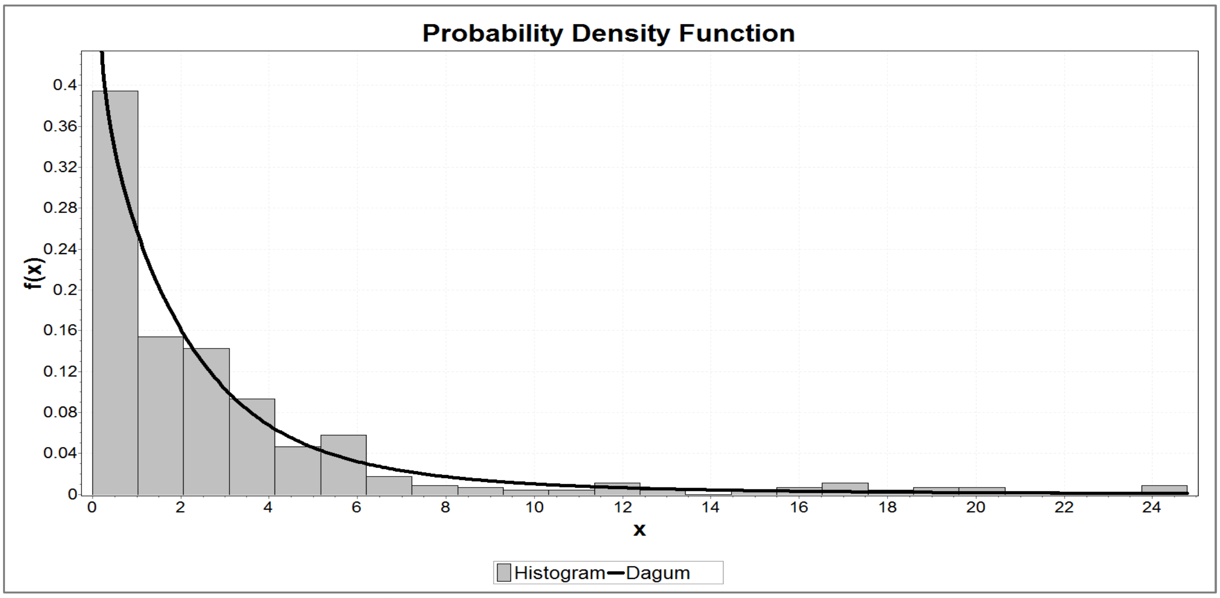


Figure 4-3 Comparison of empirical and theoretical frequencies of determined distance errors on statistical sample of 500 random points (results for three-parameter Dagum probability distribution hypothesis test).

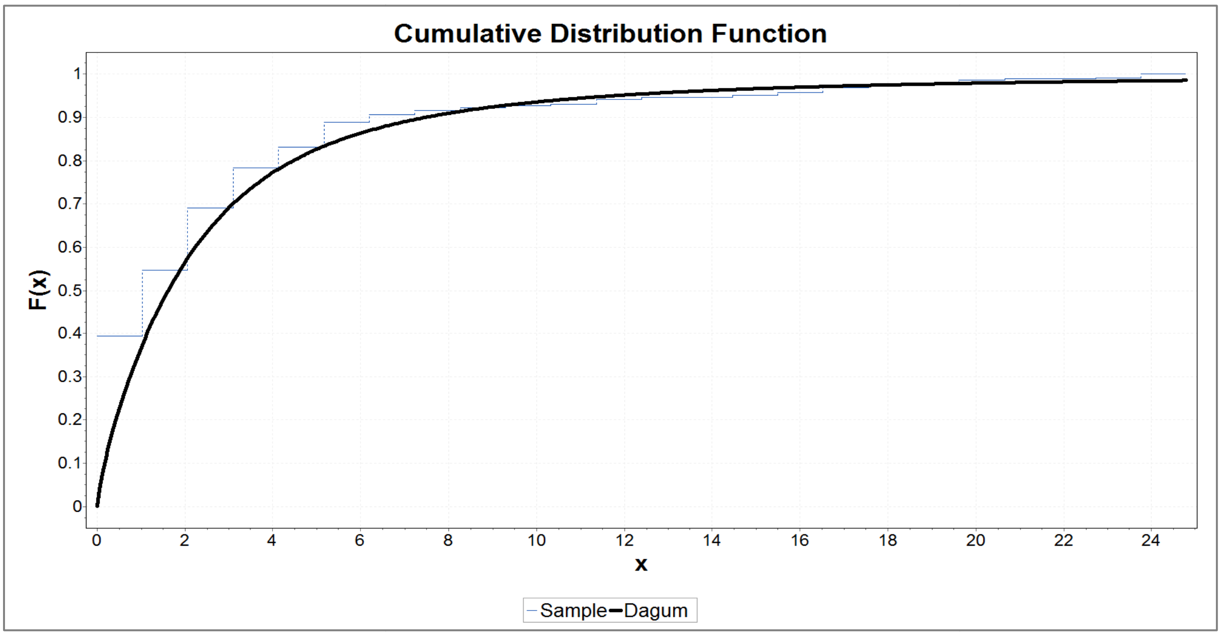


Figure 4-4 Dagum cumulative distribution function of determined distance errors on statistical sample of 500 random points (results for three-parameter Dagum probability distribution hypothesis test).

Determined deviations of empirical distance error relative frequencies (empirical error occurrence probabilities) from the selected three-parameter Dagum probability distribution theoretical occurrence probabilities (statistical sample of 500 random points) are visible from the probability difference function shown on Figure 4-5.

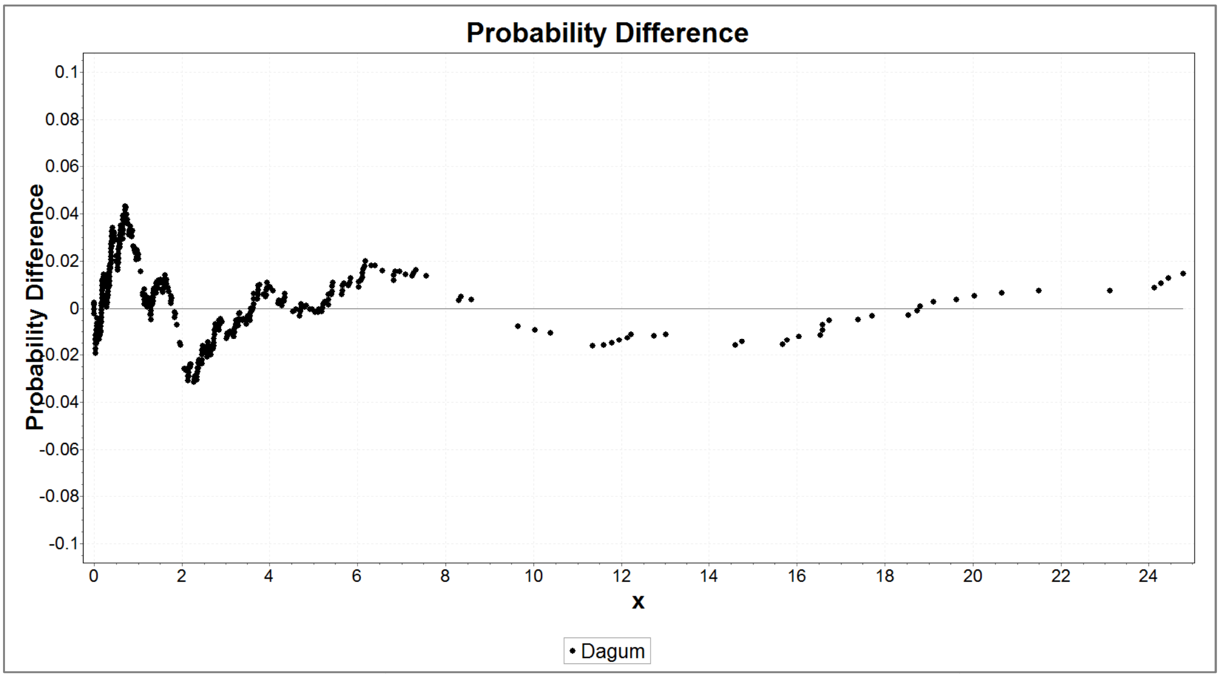


Figure 4-5 Dagum probability difference function of determined distance errors on statistical sample of 500 random points (results for three-parameter Dagum probability distribution hypothesis test).

4.3 City Centre Only (High Multiphath)

Third data subset includes only vehicle location points which were recorded inside London city centre area. This sample was seperatly analysed in order to determine level of influence that multipath effect has on resulting distance error values (influence of “urban canyons”) and also to consider the possibility of approximation of this influences by statistical probability distributions. Statistical tests have been performed on two partial subsets of data. First partial data subset includes full statistical sample collected in London city centre area (3360 vehicle location points), while second partial data subset completely excludes all extreme distance error values (errors larger than 25 meters). Results of the preformed statistical tests have shown that initial null hypothesis on Gamma probability distribution of distance errors is rejected. From the Figure 4-6 it is obvious that significant deviations between empirical distance error frequencies and theoretical Gamma distance error distribution is present over whole range of data. If extreme distance error values are excluded from analysis, empirical frequencies of determined distance errors in London city centre area appear to be almost uniformly distributed. On the other hand, if extreme distance error values remain in the analysis, results show empirical distribution which is more similar to the Gamma probability distribution. However, determined distance errors are then distributed over much broader range of values. Determined distance values in the London city area lie in the range from 0 to 220 meters which is not surprising if we consider the adverse influences of multipath effect on positioning accuracy of GNNS receivers.

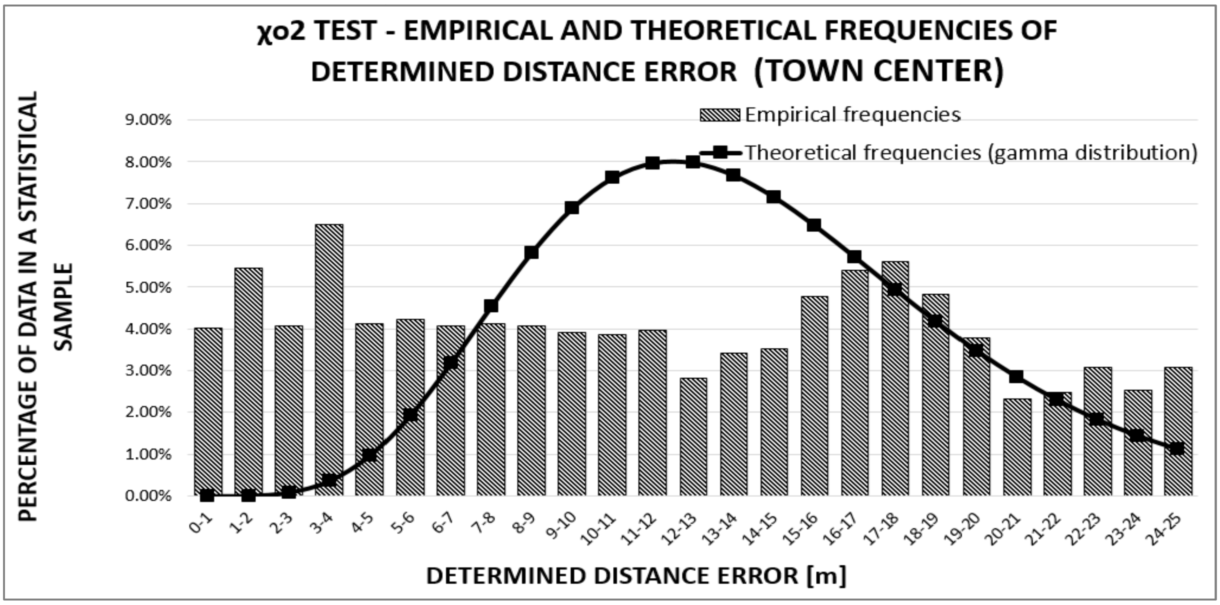


Figure 4-6 Comparison of empirical and theoretical frequencies of determined distance errors on statistical sample collected in city centre (results for Gamma probability distribution hypothesis test).

After additional statistical tests were performed in EasyFit – Distribution Fitting Software, it was determined that good approximation of the empirical distribution of distance errors in London city centre area can be achieved by three-parameter Dagum probability distribution (Figure 4-7 and Figure 4-10). In case where extreme distance error values were excluded from further analysis, optimal values of continuous shape, scale and location parameter of Dagum probability distribution for statistical sample sample collected in city centre amounted to $\alpha = 217.01$, $\beta = 24.642$ and $\gamma = 0.00417$, respectively. On the other hand, in case were all distance error values were included, considerably different optimal values of shape, scale and location parameter were determined ($\alpha = 3.3019$, $\beta = 47.683$ and $\gamma = 0.28259$).

Dagum cumulative distribution function of determined distance errors for statistical sample collected in London city centre area (partial data subset with excluded extreme distance error values) is shown on Figure 4-8. Form the cumulative Dagum distribution function, it is evident that 65 percent of measured GPS positioning error values lie in the range from 0 to 15.25 meters, 90 percent of all measured error values lie in the range between 0 and 22 meters, 95 percent in the range between 0 and 23 meters, while 99 percent of all measured positioning error values lie in the range from 0 to 24.2 meters.

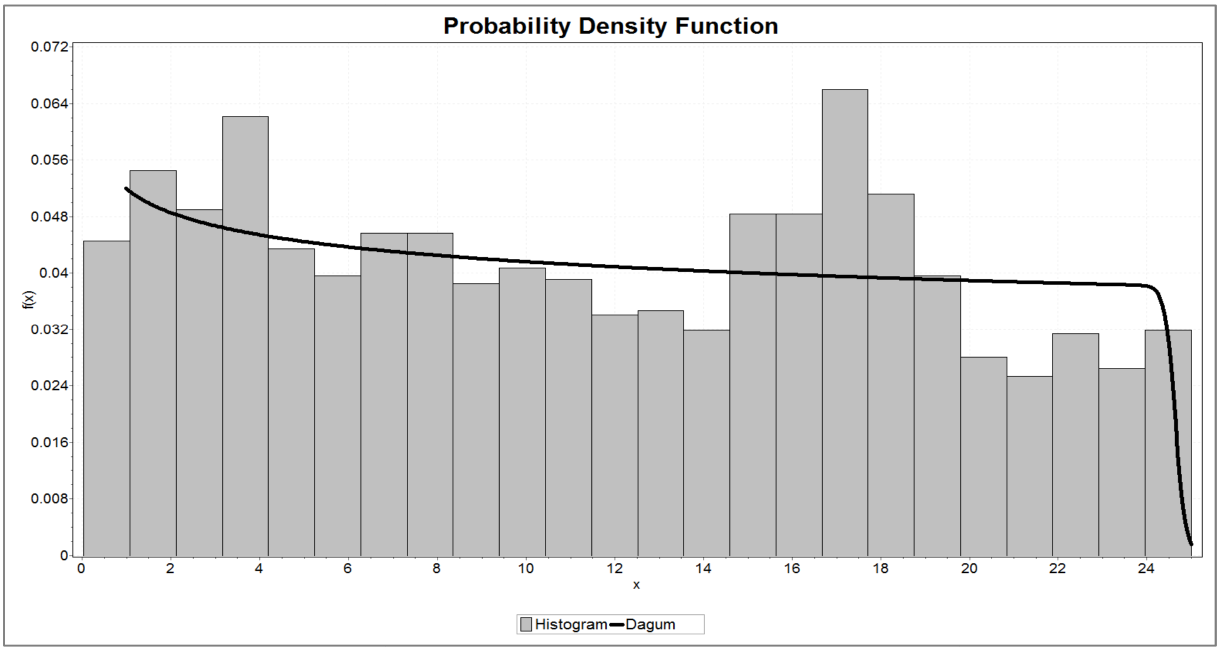


Figure 4-7 Comparison of empirical and theoretical frequencies of determined distance errors on statistical sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values excluded).

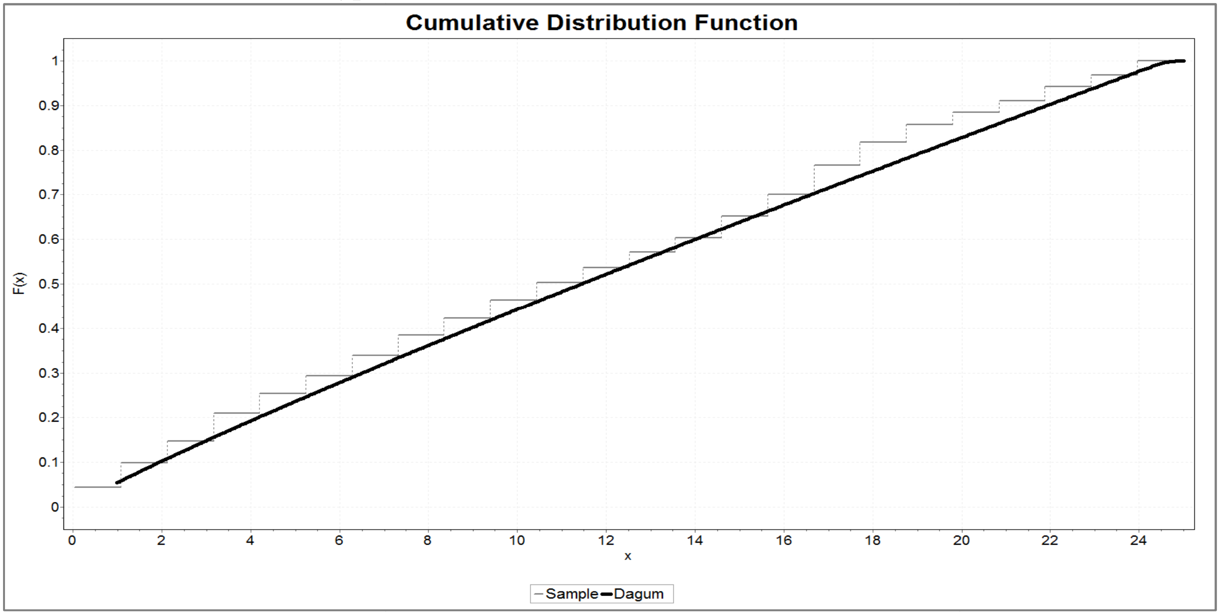


Figure 4-8 Dagum cumulative distribution function of determined distance errors on statistical sample sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values excluded).

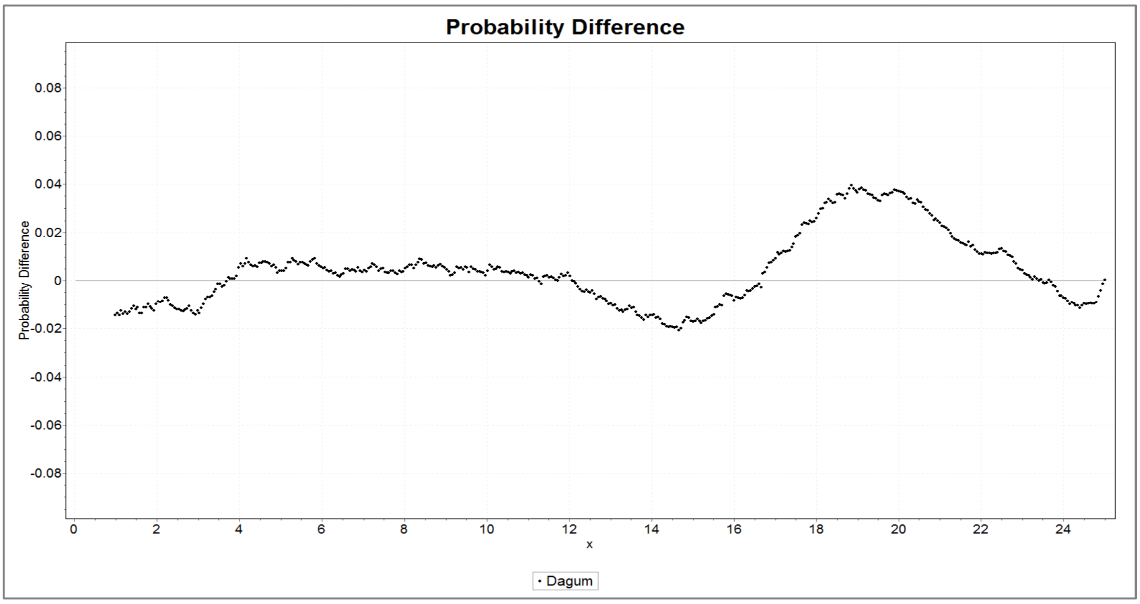


Figure 4-9 Dagum probability difference function of determined distance errors on statistical sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values excluded).

Determined deviations of empirical distance error relative frequencies (empirical error occurrence probabilities) from the selected three-parameter Dagum probability distribution theoretical occurrence probabilities (statistical sample collected in London city centre area with excluded extreme distance error values) are visible from the probability difference function shown on Figure 4-9.

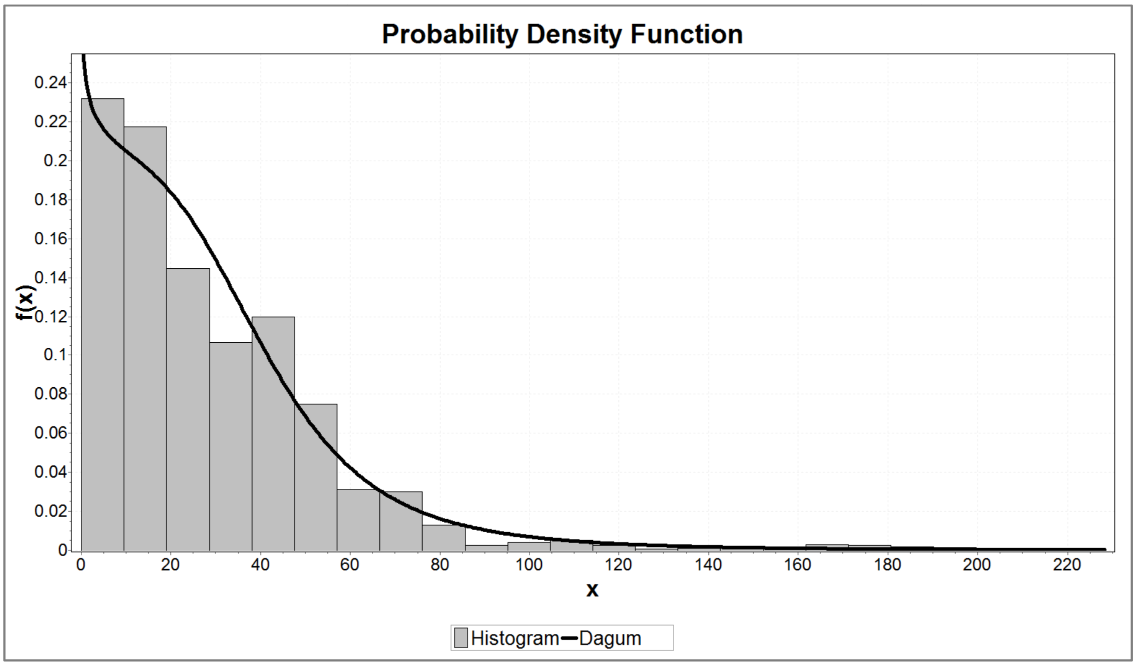


Figure 4-10 Comparison of empirical and theoretical frequencies of determined distance errors on statistical sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values included).

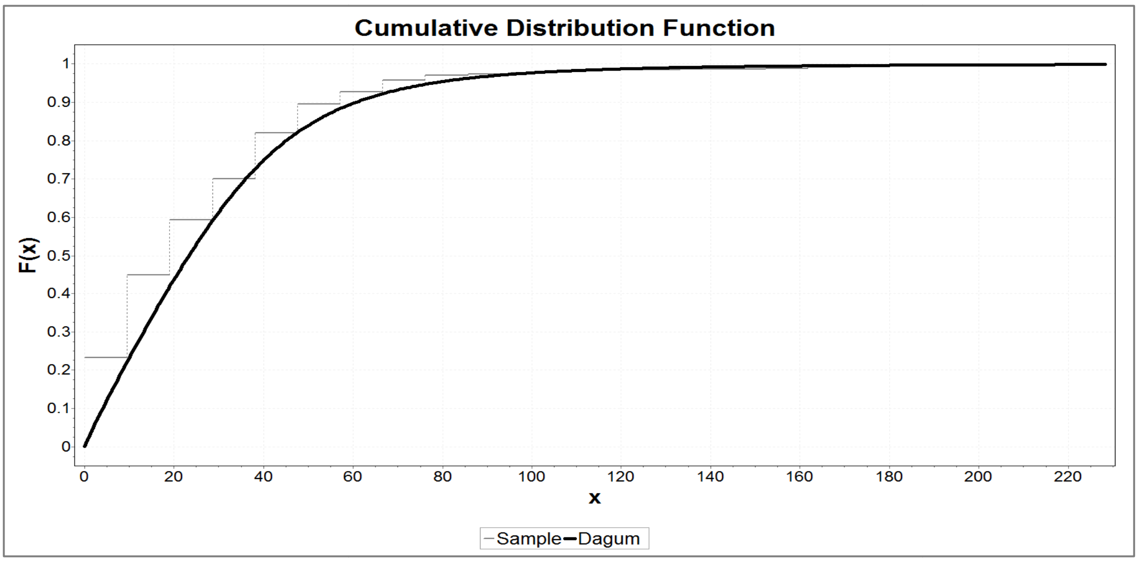


Figure 4-11 Dagum cumulative distribution function of determined distance errors on statistical sample sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values included).

Dagum cumulative distribution function of determined distance errors for statistical sample collected in London city centre area (partial data subset with included extreme distance error values) is shown on Figure 4-11. Form the cumulative Dagum distribution function, it is evident that 65 percent of measured GPS positioning error values lie in the range from 0 to 30 meters, 90 percent of all measured error values lie in the range between 0 and 60 meters, 95 percent in the range between 0 and 80 meters, while 99 percent of all measured positioning error values lie in the range from 0 to 170 meters.

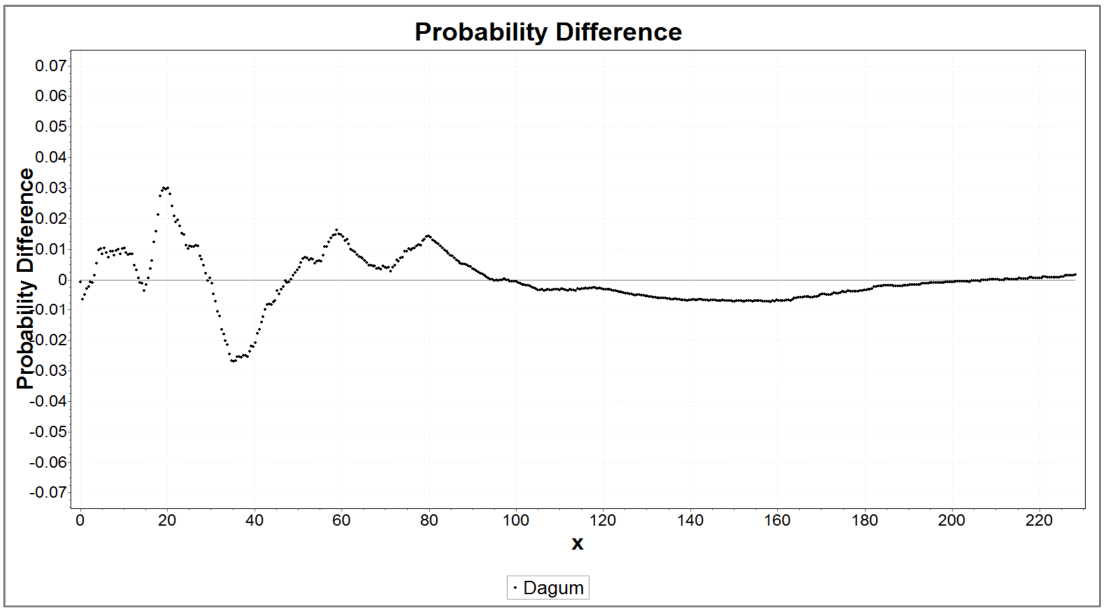


Figure 4-12 Dagum probability difference function of determined distance errors on statistical sample collected in city centre (results for three-parameter Dagum probability distribution hypothesis test, extreme distance error values included).

Determined deviations of empirical distance error relative frequencies (empirical error occurrence probabilities) from the selected three-parameter Dagum probability distribution theoretical occurrence probabilities (statistical sample collected in London city centre area with included extreme distance error values) are visible from the probability difference function shown on Figure 4-12.

5. CONSLUSION

Collected data set on empirical observations of distance errors determined in real conditions have been used for statistical analysis of positioning error occurrence probability. Descriptive statistical methods have been used to describe main quantitative features of two collected data sets through the relevant statistical mean parameters such as arithmetic mean, median, mode and dispersion measures values such as variance and standard deviation.

Based on conducted statistical analysis on full dataset (all vehicle location points inside and outside London city area), it is determined that absolute positioning error values lie in the range from 0.00018 to 24.99 meters with weighted arithmetic mean value of 3.21 meters. Median value is somewhat lower from arithmetic mean value and it amounts to 1.42 meters. Average deviation from arithmetic mean value is 4.77 meters. Variation coefficient shows 148.60 percent of dispersion of measured deviations from arithmetic mean value.

Results of performed descriptive statistical analysis on second data subset (statistical sample of 500 random points) show that absolute GPS positioning error values lie in the range from 0.00034 to 24.78 meters with weighted arithmetic mean value of 3.17 meters. Median value amounts to 1.58 meters. Average deviation from determined arithmetic mean value is 4.50 meters. Variation coefficient shows 142.20 percent of dispersion of measuered deviations from arithmetic mean value.

Performed descriptive statistical analysis on third data subset (statistical sample collected inside London city centre area, with excluded extreme distance error values) show that the arithmetic mean value of positioning errors is 11.61 meters. Positioning error median value is slightly lower and amounts to 11.39 meters, while the 85 percentile value of positioning error is 19.62 meters. Empirical frequencies of distance errors are almost uniformly distributed over whole range of data. The obtained results also show that the 65 percent of all measured positioning error values lie in the range between 0 and 15.25 meters, 90 percent of all measured error values lie in the range between 0 and 22 meters, 95 percent in the range between 0 and 23 meters, while 99 percent of all measured positioning error values lie in the range from 0 to 24.2 meters.

Analysis of determined distance errors for statistical sample collected in London city centre area (partial data subset with included extreme distance error values) has shown significantly different results. From performed statistical analysis it is evident that 65 percent of measured GPS positioning error values lie in the range from 0 to 30 meters, 90 percent of all measured error values lie in the range between 0 and 60 meters, 95 percent in the range between 0 and

80 meters, while 99 percent of all measured positioning error values lie in the range from 0 to 170 meters.

After descriptive statistical analysis of collected data was performed, the methods of inductive statistics were used to form the null hypothesis on Gamma distribution of measurement errors in determined vehicle positions by GNSS system. The initial hypothesis on Gamma distribution of measurement errors was rejected on all datasets based on the results of the performed statistical tests (Chi-squared, Anderson–Darling and Kolmogorov–Smirnov test). Using several statistical software tools, a new statistical hypothesis on three-parameter Dagum distribution of measurement errors in determined vehicle positions by GNSS system was formed. Performed statistical tests have shown that the three-parameter Dagum distribution is the most appropriate statistical distribution for the description of positional error occurrence probability in the vehicle GNSS positioning process. Hypothesis on Dagum probability distribution of measured distance error values is accepted for both area types (inside and outside London city centre area).

Optimal values of continuous shape, scale and location parameter of Dagum probability distribution for statistical sample of 1000 random points amounts to $\alpha = 1.3956$, $\beta = 2.2611$ and $\gamma = 0.64849$, respectively. Results of statistical tests performed on statistical sample of 500 random points have showed somewhat different optimal values of shape, scale and location parameter ($\alpha = 1.7335$, $\beta = 3.5209$ and $\gamma = 0.43951$). In case where extreme distance error values were excluded from further analysis, optimal values of continuous shape, scale and location parameter of Dagum probability distribution for statistical sample sample collected in city centre amounted to $\alpha = 217.01$, $\beta = 24.642$ and $\gamma = 0.00417$, respectively. On the other hand, in case were all distance error values were included, considerably different optimal values of shape, scale and location parameter were determined ($\alpha = 3.3019$, $\beta = 47.683$ and $\gamma = 0.28259$). In order to confirm obtained results future research should be conducted in other areas of the city as well as in other types of urban and suburban environment.

Due to the fact that hypothesis on three-parameter Dagum probability distribution of determined distance errors was accepted in both analysed area types (outside and inside London city centre) it is obvious that area type influence factor (which includes influences of multipath effect) can be determined by comparison of shape, scale and location parameter values of Dagum probability distribution in specific conditions of the observed road network, traffic flow characteristics and other relevant influential factors. Optimal Dagum curve for different area types can be determined by varying continuous shape, scale and location parameter values.

Future research should be focused on upgrading defined statistical probability model to enable its usage in different types of enviroment. To determine the optimal Dagum curve that gives the best fit in different types of environment, it is necessary to conduct detailed analysis of changes in shape, scale and location parameter values of Dagum probability distribution depending on specific conditions of the observed road network, traffic flow characteristics and other relevant influential factors. It is also necessary to consider the possibility of introducing additional calibration parameters in basic probability model in order to increase its accuracy.

REFERENCES

Filjar, R., Ševrović, M. and Dadić, I. (2012). Impact assessment of urban GPS positioning error on intelligent transport system's road use charging service. International conference on applied internet and information technologies, Zrenjanin, Serbia.

Hughes, I., Hase T. (2010). Measurements and their Uncertainties: A Practical Guide to Modern Error Analysis, Oxford University Press.

Kleiber, C. and Kotz, S. (2003) Statistical Size Distributions in Economics and Actuarial Sciences, Wiley-Interscience.

Peck, R., Devore, J. L. (2011). Statistics: The Exploration and Analysis of Data, Cengage Learning, 816.

Spilker, J.J., (1996). Global positioning system: theory and applications. American Institute of Aeronautics and Astronautics, Volume 1, Volume 163.

Taylor, G., Blewitt G. (2006). Intelligent Positioning and GIS-GPS Unification, John Wiley & Sons, Ltd.

Thomas, M. et al. (2011). Global Navigation Space Systems: reliance and vulnerabilities. The Royal Academy of Engineering, London, UK.

U-Blox A.G. (2007). GPS, Essentials of Satellite Navigation, Compendium. U-blox A.G.

Walck, C. (2007). Hand-book on statistical distributions for experimentalists, University of Stockholm.