FILTERING THE RESULTS OF ZIGBEE DISTANCE MEASUREMENTS WITH RANSAC ALGORITHM

Joanna JANICKA and Jacek RAPIŃSKI

University of Warmia and Mazury in Olsztyn, Institute of Geodesy, Oczapowskiego 2, 10-719 Olsztyn, Poland

Corresponding author's e-mail: joannasuwm@gmail.com, jacekrapinski@uwm.edu.pl

ABSTRACT

Positioning systems are usually based on the satellite observations. However the traditional satellite positioning has some outage, it cannot be used inside buildings or underground. But there are some new technologies that allow for indoor positioning. Among those methods, there is a conception of the RF ranging technology. There are two basic methods of obtaining the distance in RF networks – time of flight (and its variations) and RSSI-based algorithms. The new method to obtain a distance between nodes in RF network is a new ranging approach based on the phase shift measurement.

Analyzing the results of the measurements performed with this new technology, it was observed that a large number of them significantly differ from the reference value. Thus, in order to obtain correct distance filtering of the measurement results must be introduced to "smooth" the results. There is a necessity to find an algorithm that filters the observations. The authors present a proposition to use the RANSAC algorithm in the filtration process. In this paper the strategy of the use of RANSAC algorithm was presented.

1. INTRODUCTION

Nowadays people have increasing demands regarding positioning. Traditional satellite positioning systems based on the GNSS observations satisfies the need for localization in the outdoors environment. However this is not enough in modern world. People want to know their position not only outside of the buildings but also inside. The traditional satellite positioning systems cannot be used in such places. The alternative in those places can be for example inertial measurement systems INS or pseudolite positioning systems. Furthermore, there are also other technologies, which allow performing indoor positioning. In the last few years the concept of the RF ranging technology is developed.

Positioning with the use of this technology can be based on the distance measurement (range-based positioning), or on the radio signal strength indicator (range-free positioning). In the range-based positioning there are several different methods to obtain a distance between nodes in RF networks: techniques that translate the radio signal strength into distance (RSS), time of flight (TOF), time of arrival (TOA) or differenced time of arrival (TDOA) described in (Chen et al., 2012; Lourenco et al., 2013) or angle of arrival (AOA) (Peng and Sichitiu, 2006). The new ranging approach based on the ZigBee protocol and phase shift measurement that is implemented in Atmel AT86RF233 chip shows promising results in this area (Rapinski and Smieja, 2015; Smieja, 2014).

Analyzing the results of the measurement it was observed that a large number of measure distances are disturbed and significantly different from the reference value (test value). Thus to obtain the correct distance, filtering of the measurement results must be introduced in order to "smooth" the results. The authors present a RANSAC algorithm that allows to find the correct solution in noisy data.

2. EXPERIMENT SETUP DESCRIPTION

The effectiveness of the proposed algorithm was tested using a Low Power, 2.4 GHz Transceiver for ZigBee, RF4CE, IEEE 802.15.4, 6LoWPAN, and ISM Applications AT86RF233. Its main purpose is to provide wireless communication, yet it has a build-in ranging functionality (Atmel, 2013).

For the test purposes the REB233SMAD evaluation kit was used. The evaluation kit consists of two devices, both equipped with the AT86RF233 chip, ATxmega256A3 microcontroller, AS222-92 RF-switch and two SMA mounted pole antennas. Communication between devices and a PC computer is provided by Dresden Electronic USB level shifter. Two-antenna design allows measuring a distance between each pair of antennas. The distance measurement for each pair is supplemented with a distance quality factor (DQF) expressed in %. This parameter gives an estimate of the environmental conditions by evaluating and weighting antenna diversity results. High values of DQF correspond to minimal interferences and multipath while low values
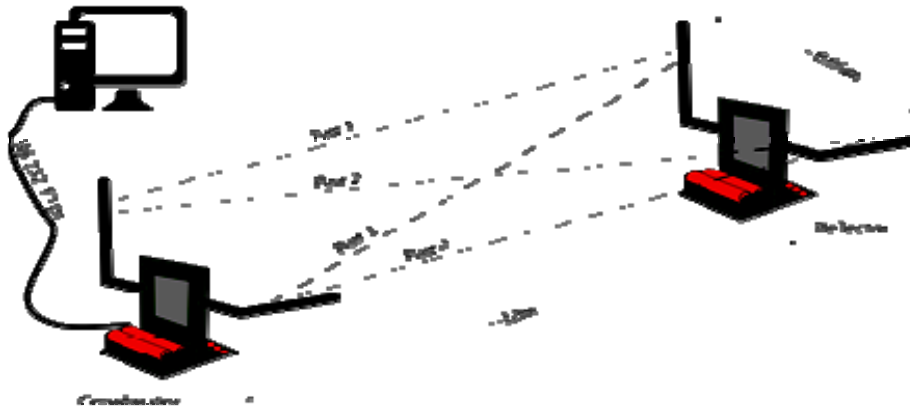
Fig. 1   The scheme device.

indicate severe multipath or interference during ranging (Atmel, 2013). The distance between nodes is from few meters up to 300 meters, while the distance between antennas in single device is about 5 cm. Therefore four measured distances are considered as the same distance with expected accuracy of 10 cm. From the communication network point of view, the ranging procedure between two nodes is mostly based on IEEE 802.15.4. Ranging is initialized by user from a PC, and results are transmitted and recorded by a PC. As a result of a ranging procedure four distances along with its DQF's are stored in a file for each measurement epoch. Figure 1 presents the scheme of the device.

Thus as it is shown in Figure 1 four pairs (distance values with DQF factor) of the results for one distance are obtained. The tests consist of four sets of data measured at four baselines with different lengths. The baselines were established in the moderate multipath environment (In the city on the street during moderate traffic). The baselines were measured using a laser range-finder. The shortest one was 10.31 meters and the longest was about 62 meters. At each baseline 300 samples were collected. Because each distance is measured four times, it means that more than thousand observations are obtained for each distance.

3.   RANSAC ALGORITHM.

3.1.   GENERAL RANSAC ALGORITHM

The RANSAC algorithm was first introduced by Fischler and Bolles (1981) as a method to estimate the parameters of a certain model, starting from a set of data contaminated by large amounts of outliers. It is an iterative, non-deterministic algorithm which uses least-squares to estimate model parameters. The basic premise of RANSAC is the presence in the data set of both observations that fit the model (inliers) and those which differ from the values (outliers). The sources of data that do not fit into the model are gross errors (measurement errors), noise or other disturbances. The input data of the algorithm are: a set of data and a mathematical model that will be matched to the data

set. The advantage of this method is that the percentage of outliers which can be handed by RANSAC can be larger than 50 % of the entire data set (Murray and Torr, 1997). Such a percentage, known also as the "breakdown points", is commonly assumed to be the practical limit for many other commonly-used techniques (such as a robust estimation method).

Information, which is used in the process of RANSAC estimation:

- the minimum number of points (observations) required to determine the model parameters,

- the minimum number of iterations,

- parameters determining the extent to which data are correct (inliers),

- the size of the data set, which completes the process of the iteration.

The RANSAC algorithm is essentially composed of two steps that are repeated in an iterative process:
1. Hypothesis
2. Tests

- Hypothesis. The first minimal sample set (mss) is randomly selected from the input dataset and the model parameters are computed using only the elements of the mss. The mss is the amount of data (observations) required to clearly define the model (the minimum number of observations that is required to describe the model is equal to the number of model parameters). The minimum number is determined by the selected function describing the model. Thus, the first phase starts with selecting a necessary and minimum number of observations of the data set. Based on these selected observations, the output model (hypothetical) is estimated. All of the remaining data are tested in terms of fit to the hypothetical model.

- **Testing.** In the second step, RANSAC iteratively checks which observations of the entire dataset are consistent with the hypothetical model. This requires determining the value of the parameter

m_d specifying the maximum distance from a test point to a hypothetical model. If it fits the criterion of m_d, the point is treated as just another hypothetical inlier. The minimum percentage of observations that must be the correct data in the whole data set is also defined (for example, the model can be regarded as properly defined if 80 % of the observations are those that are not burdened with outliers). The estimated model is correct if it has a sufficient number of points that have been classified as correct observations (inliers). The best set of observations which is selected from the entire dataset is called the consensus set (*cs*). Defining an iteration as a single process of random selection of *mss* and fit testing, the number of iterations is determined by the following formula

$$T = \frac{log\,\varepsilon}{log\,(1-q)} \qquad (1)$$

where ε is the probability of incorrect identifications of the model and q – is calculated on the basis of the following equation

$$q = \left(\frac{N_i}{N}\right)^k \qquad (2)$$

where N_i is the number of points that belong to the consensus set, N the total number of points and k minimal number of data that are necessary to the model be clearly defined.

If we want to obtain an error-free selection of points (*mss*) with a probability 1-ε, we need to perform at least T iterations.

3.2. ALGORITHM FOR DATA PROCESSING

Filtering the results of PMU ZigBee distance measurement with RANSAC algorithm is composed of two steps. In the first step the a'priori parameters are defined: *t, n, m* and *T*, where, *t* is the parameter, that defines the minimum value of DQF, *n* the minimal number of observations that must be correct in the whole data set, *m* the parameter, that specifies the maximum distance from a test observation to a hypothetical model and *T* the number of iterations.

The observations are selected with specified DQF threshold t. In the next step one distance (mss) is randomly selected from the dataset satisfying threshold t. In this particular case, the minimal sample set is also a hypothetical model.

In the test step RANSAC iteratively checks which observations of the selected dataset are consistent with the hypothetical model. It means that the distances with specified DQF factor are testing relative to the hypothetical model. Thus the residuals are calculated.

If the calculated residuals are greater than parameter *m* or the number of the distances is smaller

than *n* or both conditions are not met another distance is selected as a minimal sample set and the iterative process is repeated. If the residuals are smaller than the parameter *m*, and the number of the distances satisfies the condition of the required minimum number *n,* then the weighted mean is calculated. Weighted mean is also called *this model.* In the next step the standard deviations are checked. If the standard deviation of *this model* is greater than the standard deviation of *best model* the algorithm is move to the next iteration, in the opposite case *best model* becomes *this model* and next iteration is performed. The algorithm is repeated *T* times and results with the *best model*.

4. THE RESULTS

The measurement test was performed using two nodes placed on the known baseline and there are four separate measurements at different distances. The reference values (test value) of the distances that were measured (with its mean and standard deviations) and the distances obtained in the process of filtering with RANSAC (with mean and standard deviations of the best set) algorithm are presented in Table 1.

The statistics of performed measurements are depicted in Figure 3 as a set of four boxplots (representing each pair of antennas) at four distances.

The results of the filtering with the RANSAC algorithm are presented in Figures 4, 5, 6 and 7. In all figures the green line shows the reference distance (test value), the red line presents the weighted mean of the entire set of data for the measured distance and the blue line represents the value obtained after the RANSAC filtering.

Figure 4 presents the results of the first distance (at 10.31 m) measurement and calculations. The green line represents the test value of the distance that was measured. The blue line shows the value after using the RANSAC algorithm. In this case these two lines are almost overlapping. Figure 5 shows the results of the second distance measurement (at 12.76m) and calculation results.

The performance of the RANSAC filtering for the second distance is similar to the first distance. Thus the green line and the blue line are almost overlapping. Also in this case the RANSAC filtering allows obtain the correct value. The third distance measurement (at 21.32 m) and calculation results are presented in Figure 6.

Also, in the third measured distance the RANSAC filtering allows the value almost the same as the reference to be obtained. Therefore, the green and the blue lines overlap. Figure 7 presents the fourth distance measurement (at 62.00 m) and calculation results.

The fourth distance difference between the reference value and the obtained after RANSAC filtration is greater than the previous cases, however,

Table 1 The reference distance (test value) and the RANSAC filtering value.

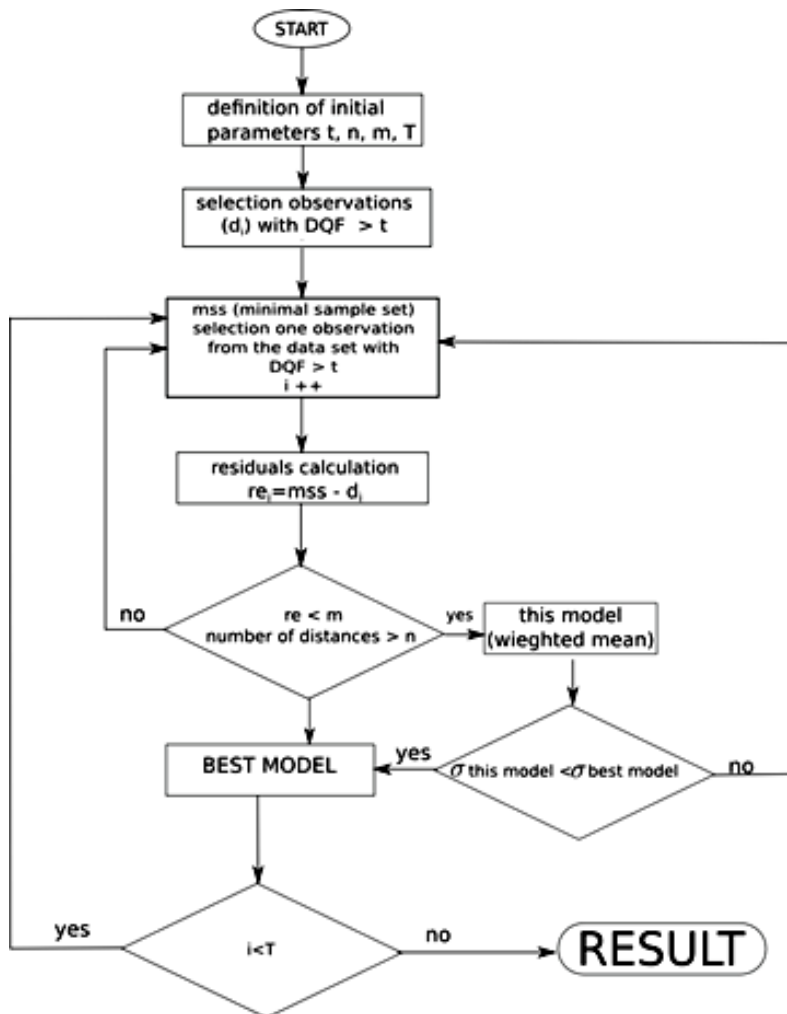|	Test value [cm]	Mean (raw data) [cm]	Standard deviation (raw data) [cm]	RANSAC filtering value [cm]	Mean (best set) [cm]	Standard deviation (best set) [cm]	Relative error [%]
d_1	1031	1834	1323	1035	1355	112	0.39
d_2	1276	2259	1564	1280	1744	102	0.31
d_3	2132	2392	1080	2134	2158	137	0.09
d_4	6200	6078	1102	6208	6254	148	0.13

Fig. 2   The block diagram.

it is still a small difference not exceeding 10 cm. The result of the calculations using the RANSAC algorithm confirms the effectiveness of the proposed method. The filtration of the measured distances allows obtain the distance values close to the true value. The value of relative error for each distance is presented in Table 1.

5. CONCLUSIONS

The non-line-of-sight propagation of the radio signal leads to a certain uncertainty in ranging. To avoid this issue the REB233SMAD evaluation kit is equipped with two antennas. This provides a certain redundancy of measurement data, which can be used to mitigate the multipath effect. In this paper this task was accomplished using the RANSAC method.

In each variant of calculations, the desired results were achieved. Therefore, the effectiveness of
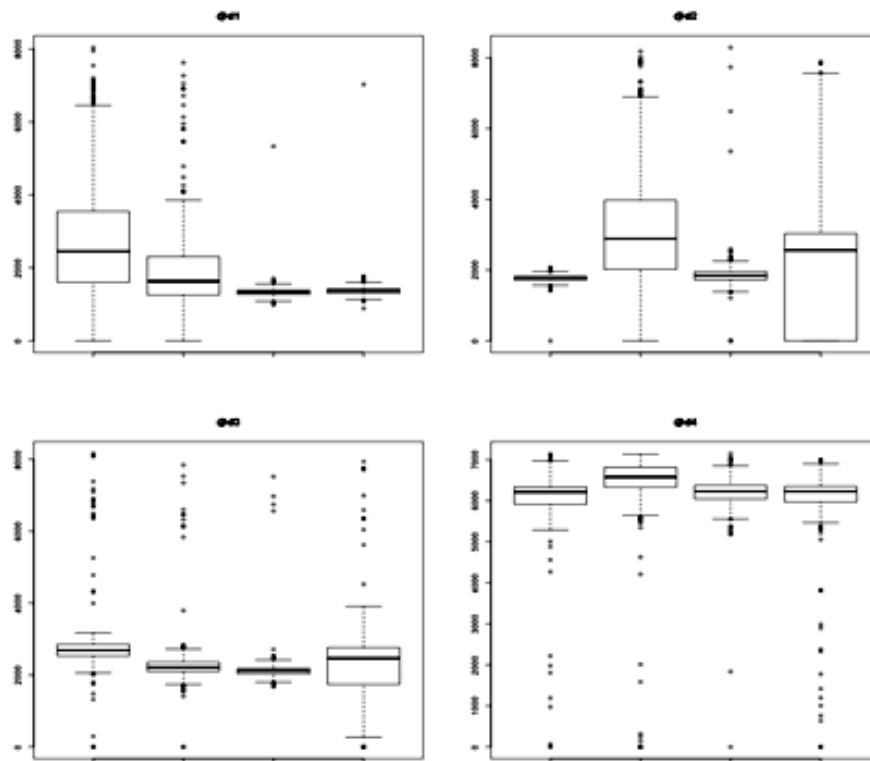
Fig. 3   The statistics of the measurement results (in cm).

filtering of the results of ZigBee distance measurements with RANSAC is confirmed. Regardless the measured distance the results were correct. The main goal of the study was to confirm the possibility to obtain a correct value of the distance when the total number of the distances significantly different from the correct value is greater than 50 %. Calculations performed for the four different distances prove the effectiveness of the proposed algorithm. Despite the advantages of this method, it has some flows. It requires many iterations and many operations which sometimes (especially in the case of large sets) take a longer time than the standard procedure. There is also a risk that the algorithm will not select the optimally points for the best solution. To high value of the ε parameter (probability of incorrect identification of the model) can lead to multiple selection of data points not belonging to the model. In this case the solution is not reliable. On the other hand low value of the ε parameter causes the increase in the number of itarations, which is undesirable from the computational efficiency point of view.

The main benefit of the proposed approach is that the RANSAC algorithm applied to ZigBee ranging can estimate a correct distance. The biggest disadvantage is the computational cost of the algorithm.

REFERENCES

Atmel AVR2150: 2013, RTB Evaluation Application. User's Guide, 1st Edition.

Chen, Q., Liu, H., Yu, M. and Guo, H.: 2012, RSSI ranging model and 3D indoor positioning with ZigBee network. In: Position Location and Navigation Symposium (PLANS) 2012 IEEE/ION, 1233–1239. DOI: 10.1109/PLANS.2012.6236979

Fischler, M. and Bolles, R.: 1981, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24, No. 6, 381–395.

Lourenco, P., Batista, P., Oliveira, P., Silvestre, C. and Chen, P.: 2013, A received signal strength indication-based localization system. 21st Mediterranean Conference on Control Automation (MED), 1242–1247. DOI: 10.1109/MED.2013.6608878

Murray, D.W. and Torr, P.H.S.: 1997, The development and comparison of robust methods for estimating the fundamental matrix. International Journal of Computer Vision, 24, No. 3, 271–300. DOI: 10.1023/A:1007927408552

Peng, R. and Sichitiu, M.: 2006, Angle of arrival localization for wireless sensor networks. In: Sensor and Ad Hoc Communications and Networks. SECON '06.2006 3rd Annual IEEE Communications Society on Sensor, 1, 374–382. DOI: 10.1109/SAHCN.2006.288442

Rapinski, J. and Smieja, M.: 2015, ZigBee ranging using phase shift measurements. Journal of Navigation, 68, No. 4, 665–677. DOI: 10.1017/S0373463315000028

Smieja, M.: 2014, The use of RF communication devices in ranging applications. International Conference on Environmental Engineering (ICEE), Selected papers. DOI: 10.3846/enviro.2014.245
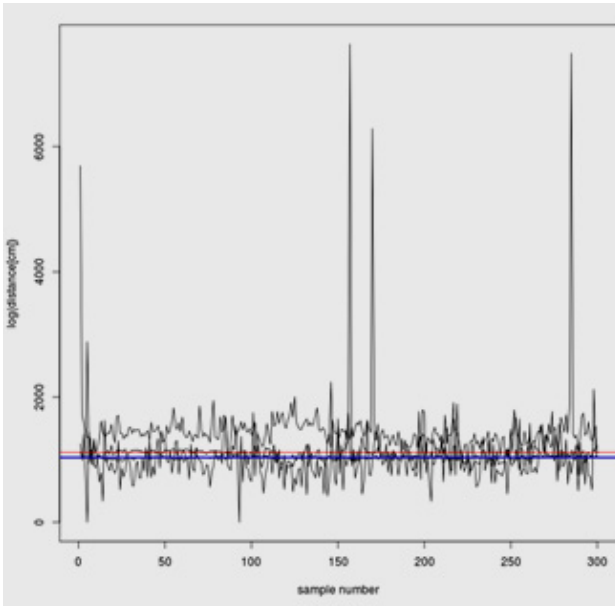
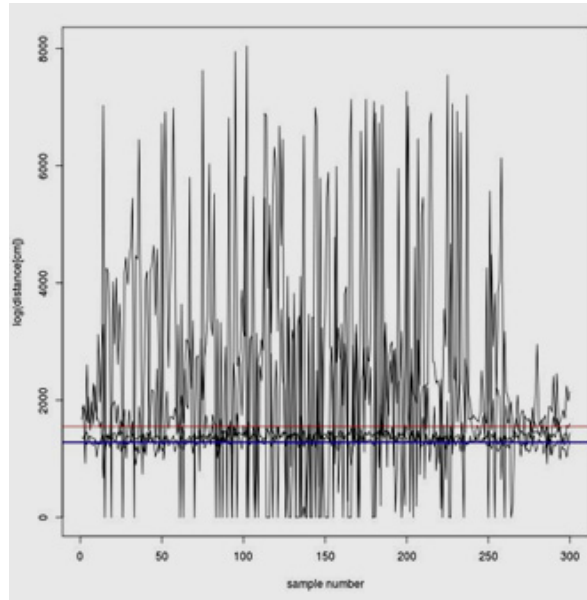**Fig. 4**   The results of the d_1 distance measurement.

Fig. 5   The results of the d_2 distance measurement.
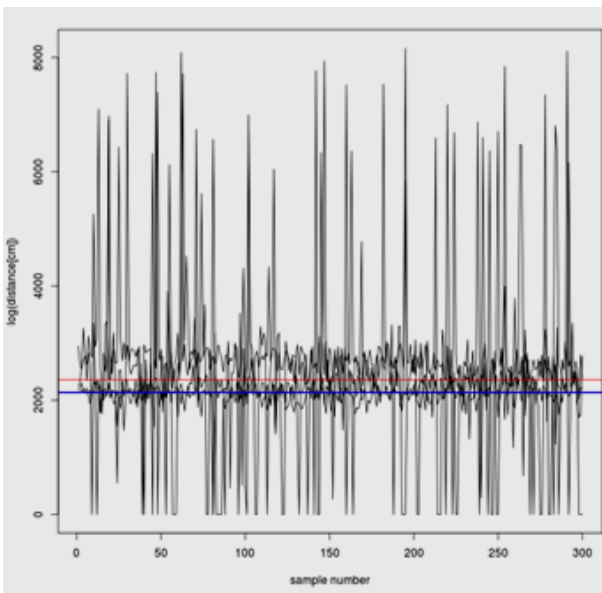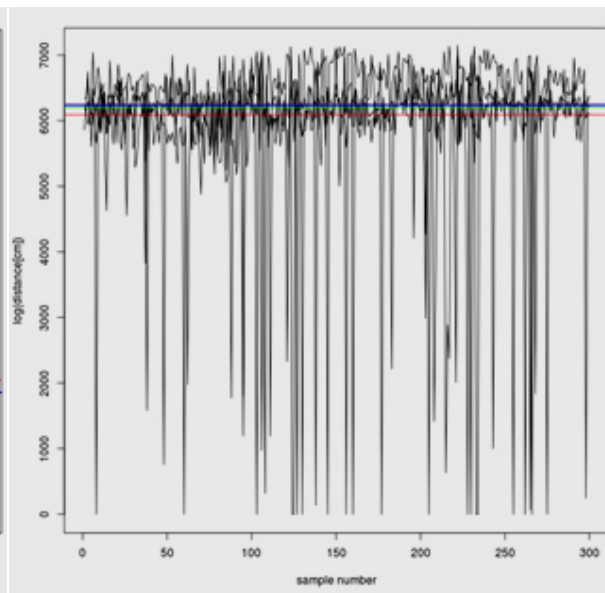

Fig. 6   The results of the d_3 distance measurement.

Fig. 7   The results of the d_4 distance measurement.