WiFi Fingerprinting Signal Strength Error Modeling for Short Distances

Vahideh Moghtadaiee
School of Surveying and Geospatial Engineering
University of New South Wales
Sydney, Australia
v.moghtadaiee@student.unsw.edu.au

Andrew G. Dempster
School of Surveying and Geospatial Engineering
University of New South Wales
Sydney, Australia
a.dempster@unsw.edu.au

Abstract—**With increasing user demands on Location-based Services (LBS) and Social Networking Services (SNS), indoor positioning has become more crucial. Because of the general failure of GPS indoors, non-GNSS navigation technologies are essential for such areas. Wireless Local Area Networks (WLAN) have widely been employed for indoor localisation based on the Received Signal Strength (RSS)-based location fingerprinting technique. The fingerprinting technique stores the location-dependent characteristics of a signal collected at known locations ahead of the system's use for localisation in a database. When positioning, the user's device records its own vector(s) of signal strength and matches it against the pre-recorded database of vectors by applying pattern matching algorithms. Location is then calculated based on the best matches between the new and stored vectors. We examined the relationship between the measured Manhattan Distance (MD), Euclidean Distance (ED), and other vector distances over the geometric distance between Reference Points (RPs) in a fingerprint database. The correlation between geometric and vector distance was poor. However, because "nearest neighbor" algorithms are used, only short vector distances are important. Furthermore, the measured RSSs varied much more as a function of distance (due to fast fading) than it did as a function of time at a single test point. Hence, the difference between variances measured at two test points was not a good indicator of the measured difference in signal strength. This led to the current investigation of very short geometric distances. In this paper, a new algorithm is applied to examine data from locations at very short ranges from each other and to investigate the relationship between vector distance and the geometric distance in closer areas in order to observe the nature of the relationship between short-range fingerprints. The experimental test bed was carried out in a large furnished office. Two west-east and south-north lines in a cross shape with 4m length are considered. We find that even at short distances, variation due to fading dominates and using other vector distances instead of MD or ED can help decrease the effect of such variation in positioning.**

Keywords- indoor positioning; fingerprinting; error modeling

I. INTRODUCTION

Indoor positioning has become highly important because of the difficulties satellite-based technology such as the Global Positioning System (GPS) experiences operating in such areas, e.g. low received signal power and low visibility of satellites. Non-satellite-based technologies, therefore, are important for indoor localisation. Utilizing signals of opportunity is a reassuring alternative to GPS due to much higher power levels and wider coverage in indoor environments [1].

Many studies have usefully employed wireless networks for indoor localisation based on the Received Signal Strength (RSS)-based location fingerprinting technique [2]. Unlike almost all other radio-navigation techniques, fingerprinting is not geometrical. In other words, the position solution does not rely on the angle to or distance from the transmitters. Instead, it requires a survey of Radio Frequency (RF) signal strength vectors to be made ahead of the system's use for localisation. The fingerprinting technique stores the location-dependent characteristics of a signal collected at Reference Points (RPs) in a database and applies pattern matching algorithms to find the best match between the fingerprint of the user and the database, and eventually estimates the position of the user based on good matches. The matching methods are based on deterministic [3] and probabilistic [4] algorithms which have been used in Wi-Fi [2], FM radio [5], and mobile phone [6] networks. The measurements of the Received Signal Strength (RSS) values at one location can vary considerably, but in deterministic location fingerprinting the average value is stored for the post processing and position determination stage.

This investigation of short-range fingerprint statistics arose from an earlier work [7] in which we found that there was not a good relationship between geometric distance and the signal strength metric used to indicate it. For short distances, this relationship was better. We therefore here investigate even shorter distances in order better to understand this relationship.

The rest of the paper is organised as follows. Section II presents the related work on fingerprinting technique and investigation of the relationship between real distance and vector distance between fingerprints. It also explains the motivation to examine shorter distances and other vector distances. The short distance experiment and a new method to evaluate the variation of vector distance versus real distances for short distances are described in Section III. The effect of using various vector distances on this relationship is also investigated in this section. Finally, the conclusions of the work are discussed in Section IV.

II. Related Work

If the propagation environment in which the system operates is known, the absolute distance between the transmitter and receiver can be calculated in an accurate manner. However, the point of fingerprinting is that it does not require knowledge either of the transmitters' location, or the characteristics of the environment. Only the measurements which imply the characteristics of the environment, that is the RSSs, are needed.

In training stage of fingerprinting method, when recording the database of fingerprints associated with RPs, many individual RSSs are recorded, and these can vary significantly. A typical fingerprint is the average of the recorded RSSs. The fingerprint can also include information about the distribution, either a histogram for each transmitter or a more simplified parameter such as variance.

Once the database of fingerprints exists, a device calculates position in positioning stage of fingerprinting technique by recording a fingerprint and matching to the database. This usually consists of measuring a distance between the recorded RSS fingerprint and each RP fingerprint in the database. We will refer to this distance as the "vector distance" which has units related to dBm (as opposed to "geometric distance" in meters between the Test Point (TP) and a RP).

Assume we have a set of n RPs in a desired area, the positions of which are known as $\mathrm{loc_i} = (\mathrm{x_i}, \mathrm{y_i})$ and are stored in the database along with the RSS vector of all the APs at all RPs. The vector distance between fingerprints of the TP and a RP defined as:

$$VD = (\sum_{i=1}^{P} |RSS_{RP}(i) - RSS_{TP}(i)|^{q})^{1/q} \qquad (1)$$

where P is the number of APs in Wi-Fi positioning, i.e. the number of elements in the fingerprint vectors. $\mathrm{RSS_{RP}}$ and RSS_{TP} are the RSS vector at one RP and the TP respectively. The index q defines the type of the vector distance measures. The most popular vector distance measures are Manhattan Distance (MD) and Euclidean Distance (ED), the L1 (q=1) and L2 (q=2) norms. The minimax or infinity norm is $q=\infty$.

Once this vector distance is calculated, different matching algorithms can be applied to provide location with respect to the RPs. The most popular pattern matching methods in deterministic approach of fingerprinting are Nearest Neighbor (NN) and K-Weighted Nearest Neighbors (KWNN). NN method simply selects the RP with shortest vector distance, while KWNN algorithms calculates the weighted average of the positions of the K nearest neighbors as a position estimation of a TP. This algorithm gives improved results [2], [5], [8].

Existing data from location fingerprinting experiments is used to help gain some insight into the nature of errors arising in this process. It should be noted that it is possible for two or more remote locations to have near-identical sets of RSS values, and a location estimate may consequently be totally inaccurate. Hence, the investigation of the relationship between real distance and vector distance is significant. This relationship was first introduced for MD in Wi-Fi positioning in our previous work [7], while most of the studies are based on the relationship between geometric distance from an AP and the RSS from that AP [9]. The relationship between real distance and RSS distance also is investigated based on ED for both Wi-Fi and FM based positioning in our previous work for a set of RPs in an indoor environment [10]. More information on the layout and the RPs distribution can be found in [8]. Fig.1 illustrates the comparison between MD and ED distribution versus the real distances between 119 RPs for the data recorded for [8].

Fig. 1 indicates that the inferred relationship between the vector distance and the real distance is not as strong as we might like. By studying the vector distance between the RP fingerprints and comparing them with real distance, it can be seen that there is a definite trend, i.e. that they are related. However, the spread that is shown in the figure is due to the power measurements varying much more erratically than would allow good prediction, due to the specifics of the indoor environment, signal fading, and features such as walls between the RPs attenuating signals.

Comparison between using MD and ED for calculating the vector distances between the points in this figure demonstrates that the spread of the signal strength measurements can interestingly decrease when utilising ED. In addition, ED gives lower vector distance values as could be expected based on (1) when q equals 2. The better the standard deviation with respect to the linear regression is the more considerable the characteristics of the linear regression (slope and intercept) are. In other words, even though higher slope of MD induces that MD works better, since the MD linear regression does not represent a good estimation of MD especially in short distances the slope function cannot be a good criteria in this case. Hence, ED seems to be a better estimator since the standard deviations with respect to the linear regressions are 15.97dBm and 8.08dBm for MD and ED, respectively. It can be also seen that the linear regression fits the variation of signal strength over the nearest neighbors or short distances much better for ED than for MD.

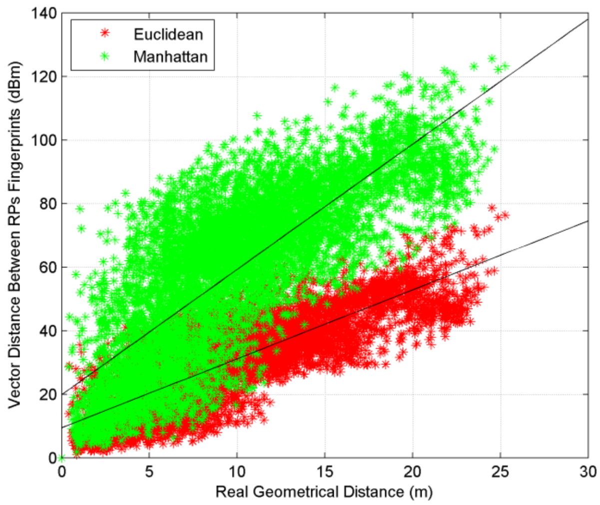

Figure 1. Distribution of MD and ED between 119 RPs shown in [8] versus the real distance between the RPs. Linear regressions are also shown.

The significant point here is that we cannot estimate the user position by the linear regression in Fig. 1. The reason for this is that while this figure gives a good indication of how vector distance between RP fingerprints indicates real distance, the positioning algorithms tend to match to nearest neighbors so behavior of distant RPs is not relevant. Furthermore, the presence of walls in an indoor environment makes the nearest RPs more influential. Hence, the investigation of the points with short distances is more essential. For this purpose, we examine the relationship between vector distance and the geometric distance in closer areas. Furthermore, comparison of results between MD and ED motivates us to further investigate using other vector distance measures based on (1) and to examine whether utilizing another *q* decreases the effect of high signal strength variation, especially over short distances.

III. A SHORT DISTANCE EXPERIMENT

A. Data Collection

This paper uses the same data recorded for [11]. The experimental test bed was located on the fourth floor of Electrical Engineering Building at the University of New South Wales (UNSW), Sydney, Australia. In a large furnished office (about $45m^2$), two west-east and south-north lines in a cross shape were marked on each axis at [-200 -100 -50 -20 -10 -5 -1 0 1 5 10 20 50 100 200] centimeters so that a local coordinate system was created.

The intersection of the west-east and south-north lines called "Origin" in this paper. The distances were selected so that they range from well below the signal wavelength (12cm) to above the accuracy of fingerprinting systems (1.2 - 1.5m). The experiment was carried out when few people were in the vicinity. The researcher recorded data for 2 minutes at each point from west to east and then south to north as in Fig. 2, always facing north. There is a limit to how accurately a user can hold a position like this, but it is important to replicate real operating conditions as far as possible.

Two separate sets of data were recorded. The first used a palm sized notebook computer equipped with internal Wi-Fi receiver running the UNSW School of Computer Science & Engineering (CSE) developed location-based RSSI collection software for a Linux platform called Kismet. The second set of data used an iPAQ pocket PC running Ministumbler. MiniStumbler detecting 802.11b, 802.11a and 802.11g WLANs. It sends out a probe request about once a second, and reports the responses. The output file provides SSID, BSSID, maximum signal, minimum noise, maximum SNR, channel number, IP Address of AP, IP Subnet and Mask. As discussed in [11] bimodal problems were observed in the Kismet data, which makes it less reliable. Therefore, we used just the Ministumbler data here in this work.

B. Analysis Results

In this investigation, there was no positioning phase so all of the reference points used are effectively our test points with the same RSS vectors. There are 7 APs observed and 54 samples were recorded at all points. The RSS data were simply processed to provide mean signal strength for each of the APs at each point. The standard deviation of the measurements is 2.4dBm.

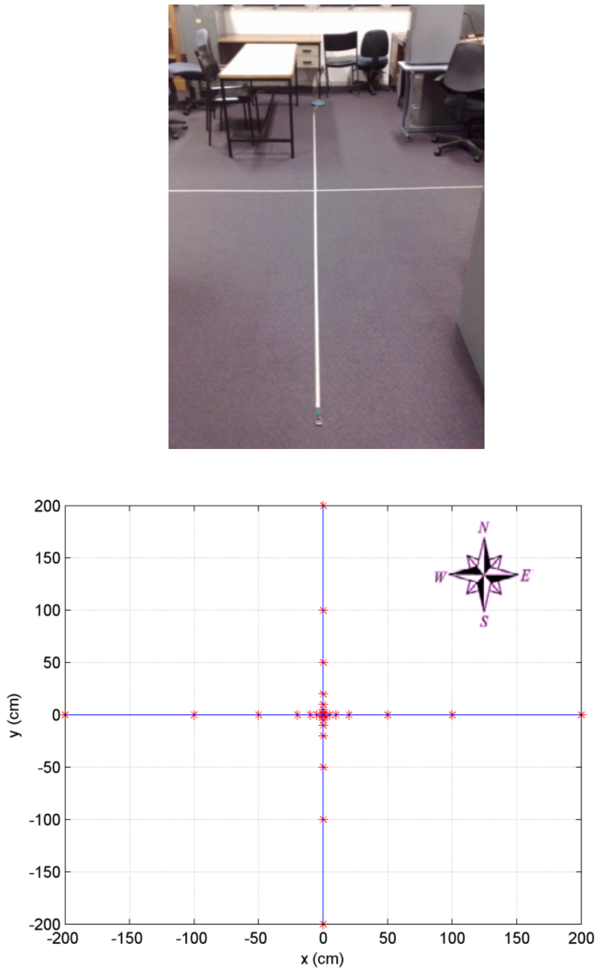

Figure 2. Test bed- the crosses indicate the points where data is recorded

Different vector distances (with various *q* values) and geometric distance were calculated between all pairs of test points. The results based on different values of *q* are shown in Table I. The standard deviation with respect to the linear fit is 4.23dBm for MD and 1.98dBm for ED which again demonstrates that ED is a better estimator than MD. However, this Table shows that by considering other values of *q* the fitted standard estimation is minimum when *q* equals 4, indicating that using vector distances with $q = 4$ gives the best estimator here (although the most popular vector distance is ED). However, it should be noted that the fitted standard deviations are almost the same for the higher values of *q*, even when *q* is infinity.

Furthermore, as can be seen in Table I, the linear fits have intercepts of 10.22dBm and 4.88dBm and slopes of 0.0590dBm/cm and 0.0266dBm/cm based on MD and ED respectively. ED, hence, has lower intercept and shows better definition of the "vector distance" between points compared to MD. However, based on this Table, *q* equals 4 can be again selected as the best estimator due to the fact that even though the slopes and intercepts decrease as *q* increases, the difference when using *q* equals 4 and other higher signal distance measures (i.e. *q* with higher values) is not considerable.

The considerable point here is that q equals infinity (infinity norm) can also be as good as q equals 4, as the values of slope, intercept and fitted standard deviation are similar$q \geq 4$. Therefore, it may be a more appropriate decision to accept infinity norm as the best estimator and then easily use the following equation instead of using (1):

$$VD = max_i(|RSS_{RP}(i) - RSS_{TP}(i)|), i = 1,2, \dots, P \qquad (2)$$

Utilising (2) means that we just need to simply pick the maximum signal strength difference received from all APs between two points and consider it as a vector distance between those two points. Fig. 3 illustrates this result for MD (q = 1), ED (q = 2), and $q = \infty$.

TABLE I. LINEAR FITS RESULTS WHEN USING DIFFERENT VECTOR DISTANCES

q	Slopes (dBm/cm)	Intercepts (dBm)	Fitted std. (dBm)
1/3	2.421	394.77	198.83
1/2	0.3641	60.29	28.3
1	0.059	10.22	4.23
2	0.0266	4.88	1.98
3	0.0216	4.08	1.75
4	0.0199	3.82	1.7
5	0.0191	3.7	1.71
6	0.0187	3.64	1.72
7	0.0184	3.61	1.72
8	0.0182	3.58	1.73
9	0.0181	3.57	1.73
10	0.018	3.56	1.74
50	0.0177	3.52	1.76
Inf	0.0176	3.52	1.76

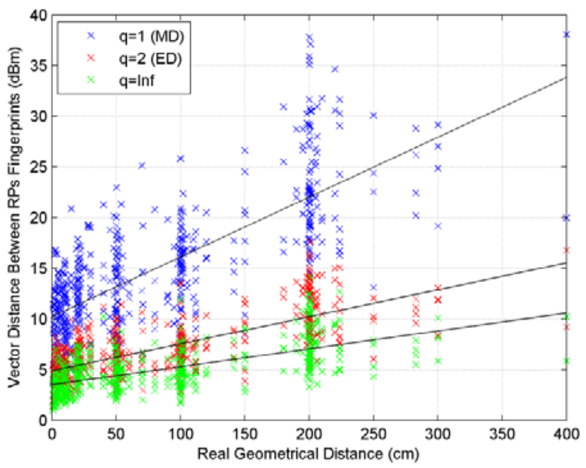

Figure 3. Vector distances for q=1, 2, ∞ vs. real geometric distance for Ministumbler data (with linear fits)

In order to better investigate the nearest areas for one point and to identify different behaviour of the signal in various distance regions, a new method is proposed. In this method different areas defined by circles are considered around one point and then the variation of vector distances over real distances between that point and the other points inside each circle are examined. It can indicate how different the vector distances between the points are even with very short real geometric distances. As can be seen in Fig. 4, in our experiment, the radii of the circles are from 1cm to 400cm because they are the minimum and maximum distance between the points, respectively.

By putting the origin of theses circles on every desired point, we can see how a vector distance changes as real distance varies from that point. Fig. 5 shows how this method works. It depicts the variation of ED over real distances between the origin and the RPs inside the different radius of circles. The linear regression has been shown over the points within the circles. For clarity, the variation based on only four circles with radii of 50, 100, 150, and 200cm are displayed in this figure. The slopes of the linear regression defined as:

$$slope = \frac{VD}{RD} \qquad \text{(dBm/cm)} \qquad (3)$$

where VD is vector distance and RD is real geometric distance. The slopes, intercepts, and the fitted standard deviation with respect to the linear regression of the distribution of points in Fig. 5 based on different radii are reported in Fig. 6 for MD, ED, and $q = \infty$. For the Origin we just considered radii from 1cm to 200cm since there is no point beyond 200cm distance for the Origin.

The first point we notice in Fig. 6 is that even for very close points the variation of vector distance is considerable. This is explained mostly by the random nature of electromagnetic propagation effects such as fading, shadowing and especially multipath. To the best of our knowledge, there seems not to be a propagation model that helps us in predicting the behaviour of the vector distance. Existing indoor path loss models are based on an ensemble average of signal strengths at radius r from an AP [12-14]. Taking the derivative of this does not give us any indication of the rate of change of signal strength to expect at radius r.

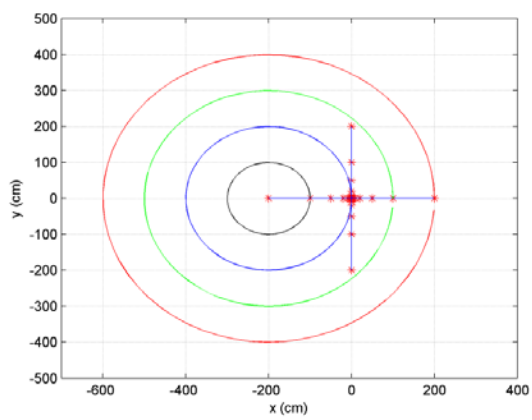

Figure 4. Different neighbourhood for one point defined by circles with radius from 1cm to 400cm (just four circles are shown)

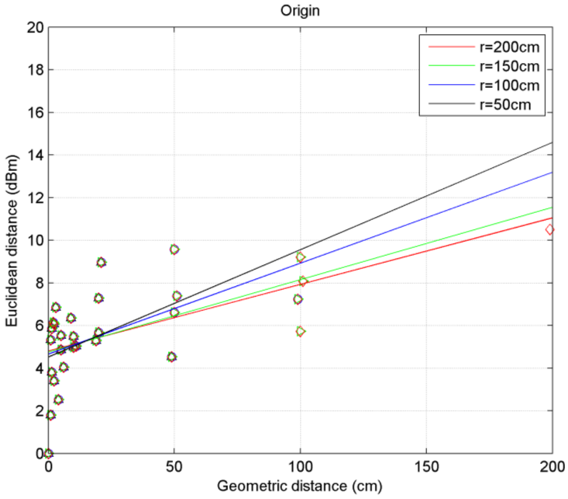

Figure 5. The variation of the ED over geometric distance between the Origin and the RPs inside the different radii of the circles. Linear regressions are shown when different radii are chosen.

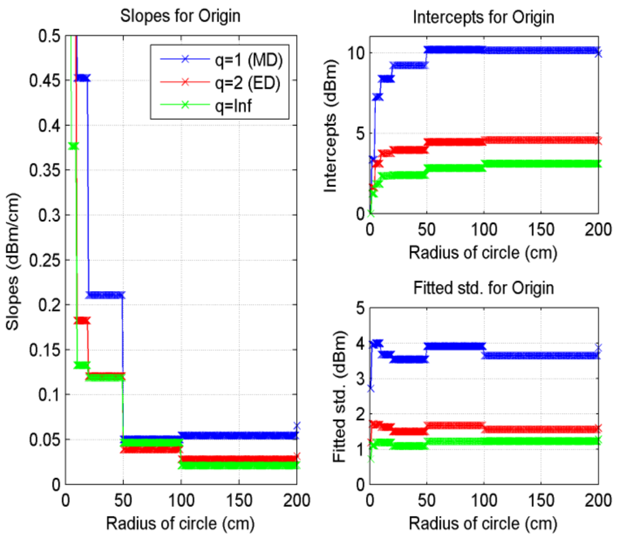

Figure 6. The linear regression slopes (left), intercepts (right-up), and fitted standrad deviation with respect to linear regression (right-down) at the Origin for different radii of the circles from 1cm to 200cm.

In addition to what is explained above, Fig. 6 shows that the slope decreases with increasing the radius of circles around the Origin. Regarding (3), the fact is that although the variations of received vector distance between two points are large, when the distance is greatly increased i.e. by 200 times, the slope of lines is decreased. As such for a small circle with radius of 1cm the slope is 9.25dBm/cm, 4.14dBm/cm, and 2.73dBm/cm while for a large circle with radius of 200cm the slope is 0.065dBm/cm and 0.031dBm/cm, and 0.022dBm/cm for MD, ED, and $q = \infty$). Furthermore, Fig. 6 shows that by increasing the radius of circles, the intercept level of the experiment is increasing and its increasing rate is higher for smaller circles within 50cm.

It can be also seen from this figure that using different vector distances when q=1, 2, ∞ leads to the same trends in slopes, intercepts, and fitted standard deviation, however MD results in higher values as expected from the earlier work. Using the infinity norm can decrease the high variation of vector distances between the points with short distances since the slopes and the intercepts are lower in this case compared to those in MD and ED. The fitted standard deviation results also confirm that the infinity norm of power differences is a better estimator. In fact, by comparing the results of using all q values we can have the best estimator of the vector distances over the real distances with respect to the linear regression when q is ∞.

The mean value of the linear regression slopes, intercepts, and fitted standard deviation at different radii of the circles for all the points on the experiment is shown in Fig. 7. It can be realised that by increasing the radius of the circle the mean value of slopes decreases and for large radii the mean value becomes almost flat. However, by increasing the radius of circles, the mean value of intercepts is incrementing but with different rates for different q values. It is noticeable that the MD results in higher mean of slopes and intercepts than the other values of q when $q \geq 1$ (which can be also seen in Fig. 3) as such for a radius of 100cm the MD, ED, $q = \infty$ produces the mean slope of 0.1dBm/cm, 0.05dBm/cm, and 0.03dBm/cm, the mean intercept of around 7.19dBm, 3.43dBm and 2.45dBm and the mean fitted standard deviation of 3.16dBm, 1.50dBm, and 1.23dBm, which again confirms advantages of utilising $q = \infty$.

In addition, the mean value of fitted standard deviation in the three cases shown demonstrates that by using the infinity norm, not only the calculations get simpler, but also the relationship between the points at short distances can be improved. It also shows that we can get a better estimation of the real distance based on the vector distance when considering the nearby points (less than 50cm distance), while the estimation gets worse by increasing the neighborhood area and including more distant points.

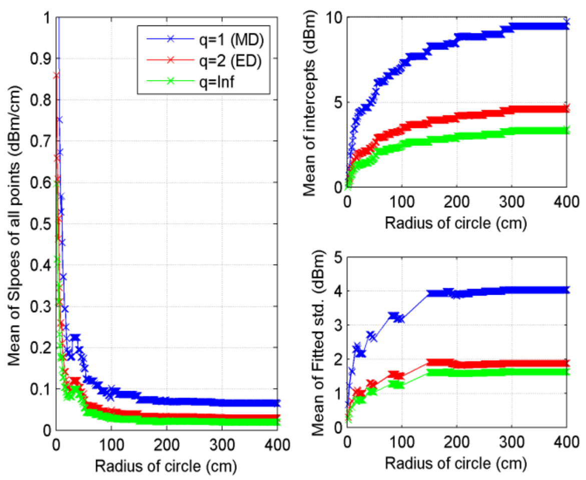

Figure 7. The mean of linear regression slopes (left), intercepts (right-up), and fitted standrad deviationd with respect to linear regression (right-down) at different radii of the circles for all of the points.

IV. CONCLUSION AND FUTURE WORK

Comparing the relationship between MD and ED over the geometric distance between points in a fingerprint database leads us to further investigate using other vector distance measures based on (1). The results show that the infinity norm of power differences seems to be a better estimator since its standard deviation with respect to the linear regression is minimum. It also simplifies (1) to (3) which is much easier to use and has less complexity in the calculations.

Short-distance fingerprint analysis using different vector distances is investigated by a new method since the positioning algorithms tend to match to nearest neighbors. The presence of walls in an indoor environment also makes the nearest RPs more influential. We find that even at short distances, variation due to fading is significant. This is explained mostly by random nature of electromagnetic propagation effects such as fading, shadowing and especially multipath. Wi-Fi signals have short wavelengths (12cm), so short-distance variation due to fading is more serious. However, using another vector distance measure ($q = \infty$ or infinity norm of power differences) instead of MD or ED can help decrease the vector distance variance over the real distance and improve the relationship between the points in short distances. Therefore, we expect several-nearest-neighbor algorithms can be able to supply reasonable results and better positioning results with less error can be obtained at the end, which is an issue to work on in future. The effect of using different numbers of APs is also of interest as well.

The existing propagation models cannot help us predict the behavior of the vector distance since they are based on an ensemble average of signal strengths at radius r from an AP. Taking the derivative of this does not give us any indication of the rate of change of the signal strength to expect at radius r. Developing the signal strength model for fingerprinting can help in improving the positioning accuracy.

REFERENCES

[1] J. Raquet and R. K. Martin, “Non-GNSS radio frequency navigation,” in *Proc. of IEEE ICASSP*, 2008, pp. 5308–5311.

[2] B. Li, Y. Wang, H.K. Lee, A.G. Dempster and C. Rizos, "Method for yielding a database of location fingerprints in WLAN," *Proc. IEE Communications*, vol. 152, no. 5, pp. 580–586, Oct 2005.

[3] P. Bahl and V. N. Padmanabhan, “RADAR: An In-Building RF-based User Location and Tracking System,” Proc IEEE Conference on Computer Communications (INFOCOM), 2000.

[4] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, J. Sievanen, “A Probabilistic Approach to WLAN User Location Estimation”, *Int J Wireless Info Networks*, vol 9, no 3, July 2002, pp155-164.

[5] V. Moghtadaiee,A. G. Dempster, S. Lim, “Indoor Localization Using FM Radio Signals: A Fingerprinting Approach,” *Proc. IEEE Indoor Positioning Indoor Navigation*, Guimarães, Portugal, Sep. 2011, pp. 1-7.

[6] B. Li, A. G. Dempster, J. Barnes, C. Rizos, D. Li, “Probabilistic algorithm to support the fingerprinting method for CDMA location”, Int. Symp. On GPS/GNSS, Hong Kong, 8-10 Dec 2005, paper 9C-05, CD-ROM procs.

[7] A. G. Dempster, B. Li, I. Quader, “Errors in Determinstic Wireless Fingerprinting Systems for Localisation,” *Proc. Of International Symposium on Wireless Pervasive Computing*, 2008, pp. 111-115.

[8] V. Moghtadaiee, A.G. Dempster, S. Lim, “Indoor Positioning Based on FM Signals and Wi-Fi Signals,” *Proc. of IGNSS*, Sydney, Australia, 2011.

[9] B. Li, A. G. Dempster, C. Rizos, J. Barnes, “Hybrid method for localization using WLAN”, Spatial Sciences Conference, Melbourne, Australia, 12-16 September, pp. 341-350, CD-ROM procs

[10] V. Moghtadaiee, A. G. Dempster, B. Li, “Accuracy Indicator for Fingerprinting Localization Systems,” *Proc. Of IEEE/ION PLANS*, Myrtle Beach, South Carolina, Apr. 2012, pp. 1204-1208.

[11] A. G. Dempster, B. Li, I. Quader, “Short Baseline Propagation Characteristics of Determinstic Wireless Fingerprinting Systems for Localisation,” *Proc . Of IEEE Consumer Communications and Networking Conference,* 2009, pp. 1-5.

[12] S. S Saunders and A. Aragon-Zavala, “Antennas and Propagation for Wireless Communication Systems”, 2nd ed., Wiley, Chichester, 2007.

[13] A. Oka and L. Lampe, "Distributed target tracking using signal strength measurements by a wireless sensor network," *IEEE J. Sel. Areas Communications*, vol. 28, no. 7, pp. 1006-1015, Sep. 2010.

[14] G. Wang and K.Yang, "A New Approach to Sensor Node Localization Using RSS Measurements in Wireless Sensor Networks," *IEEE Trans. Wireless Communcations*, vol. 10, no. 5, pp. 1389–-1395, May 2011.