Estimating Propagation Parameters using a Modified EM Algorithm for Mobile Location Estimation

Kenneth Man-Kin Chu, Joseph Kee-Yin Ng
Department of Computer Science
Hong Kong Baptist University Hong Kong
Kowloon Tong, Hong Kong
Tel: (852)-3411-7864
Email: {mkchu, jng}@comp.hkbu.edu.hk

November 1, 2004

Abstract

Recently, mobile location estimation is drawing considerable attention in the field of wireless communications. Among different mobile location estimation methods, the one which estimates the location of mobile stations with reference to the wave propagation model is drawing much attention on the grounds that it is applicable to different kinds of cellular network. However, the signal propagation models require estimation of propagation parameters. We have revised the existing EM algorithm which is designed for signal propagation models. We have classified the EM algorithm into two cases — Sufficient Data Collection and Insufficient Data Collection. We experimented our modified EM algorithm with 192,177 sets of real life data collected from a major mobile phone operator of Hong Kong. Results show that the modified EM algorithm improves the existing parameters estimation algorithm in terms of estimation accuracy, stability in different types of landscapes.

1 Introduction

Using the signal propagation model for wireless location estimation is one of the most promising approach since signal strength is an intrinsics of wireless communication and is well handled in all cellular systems. However, there are many factors affecting the electromagnetic wave propagation, they can be generally classified into three attributes — reflection, diffraction and scattering. In most cellular radio systems operate in urban areas, the electromagnetic waves travel along different paths of various lengths due to multiple reflections from various objects such as high rise buildings. This phenomenon causes multipath fading at a specific location which increases the difficulties in predicting the signal strength at the specific location.

Signal propagation models are used to predict the mean signal strength for an arbitrary distance between transmitter and receiver (T-R Seperation). The propagation model which is used to estimate the coverage area of transmitter is called *large-scale* propagation model. The T-R seperation for this kind of propagation model is usually around several hundreds meters to several thousand of meters. Whereas, another propagation model is called *small-scale* fading, which characterize the rapid fluctuations of received signal strengths over short travel distance (a few wavelengths) or short time duration (on the order of seconds). In this technical report, only the large-scale fading is discussed.

For the rest of this report, in Section 2, we first give an overview of the propagation model and existing propagation parameters estimation methods. These are followed by the description of our modified EM algorithm in Section 3. The experiments and results of validating the modified EM algorithm are described in Section 4. Finally, we summarize our work in Section 5.

2 Background

2.1 Signal Propagation Model

The fundamental signal propagation model is the free space propagation model which is given by Friis [1],

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{1}$$

where, P_t is the transmitted power, P_r is the received power, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the T-R seperation, L is the system loss factor not related to propagation ($L \geq 1$), and λ is the wavelength in meters. The free space propagation model is derived from the first principles and it shows that the receive power decays with a distance at a rate of 20dB/decade when there is no obstacles between the transmitter and the receiver. In real environment, a single Line of Sight (LOS) path between the transmitter and receiver seldom occurs especially in urban environment. Thus, the free space propagation model cannot be applied in this situation. Many propagation models based on the free space propagation model have been proposed like 2-ray Model, Okumura Model and HATA Model [2, 5, 7]. However, these propagation models are non-directional. From our observations, the directive gain of the directional antenna affects the Received Signal Strength (RSS) to a certain extent. Furthermore, the environmental factor is another attributes which should be considered. Unfortunately, most of the existing propagation models have not dealt with these two factors. In view of this, our group had proposed a statistical Directional Propagation Model (DPM) [3] to incorporate these two parameters for mobile location estimations.

The DPM model enhanced the free space signal propagation model by relating the directive gain of the directional antenna to it. In brief, the DPM is defined as follows,

$$\overline{pl} = \beta_0 + \beta_1 g_a + \beta_2 g_e + \beta_3 \log h + (\beta_4 + \beta_5 \log h + \beta_6 e) \ln d \tag{2}$$

where, $\overline{pl}$ is the mean propagation loss. It is defined as the difference between the RSS, s and the transmit power, p in decibel. d is the distance in meter between the Mobile Station (MS) and the Base Station (BTS)[1]. g_a and g_e are the azimuth and elevation directive gain of an antenna type t respectively. e is the environment index which is the building density in specific directions of the BTS and h is the height in meter of the BTS. $\beta = [\beta_0\beta_1\beta_2\beta_3\beta_4\beta_5]^T$ are the coefficients needed to be estimated.

The mean propagation loss ($\overline{pl}$) is assumed to follow a Gaussian distribution and its probability density function (p.d.f.) is defined as follows,

$$f\left(\overline{pl}|d, g_a, g_e, h, e, t, \theta\right) = \frac{1}{\sqrt{2\pi}\sigma^{(t)}} exp\left[-\frac{1}{2}\left(\frac{pl - \overline{pl}^{(t)}}{\sigma^{(t)}}\right)^2\right] \tag{3}$$

where θ is the set of propagation parameters, $\sigma^{(t)}$ is the standard deviation and exp is the exponential function with base e.

2.2 Propagation Parameters Estimation

Using the Ordinary Least Square (OLE), the propagation parameters, $\widehat{\beta^{(\mathbf{t})}}$ [2] can be estimated by [3],

$$\widehat{\beta^{(\mathbf{t})}} = \left[\left(\mathbf{X}^{(\mathbf{t})^T}\mathbf{X}^{(\mathbf{t})}\right)^{-1}\mathbf{X}^{(\mathbf{t})^T}\right]\overrightarrow{y}^{(\mathbf{t})} \tag{4}$$

where,

$$\widehat{\beta^{(\mathbf{t})}} = \begin{bmatrix} \widehat{\beta_0^{(t)}} \\ \widehat{\beta_1^{(t)}} \\ \widehat{\beta_2^{(t)}} \\ \vdots \\ \widehat{\beta_6^{(t)}} \end{bmatrix} \qquad \overrightarrow{y}^{(t)} = \begin{bmatrix} \overline{pl}_1^{(t)} \\ \overline{pl}_2^{(t)} \\ \overline{pl}_3^{(t)} \\ \vdots \\ \overline{pl}_n^{(t)} \end{bmatrix}$$

[1]Both MS and BTS can be transmitter and receiver at the same time.

[2]The estimated variable $\mathbf{X}$ is represented by $\hat{\mathbf{X}}$

[3]The transpose of Matrix $\mathbf{X}$ is represented by $\mathbf{X}^{\mathbf{T}}$

$$\mathbf{X^{(t)}} = \begin{bmatrix} 1 & g_{a1}^{(t)} & g_{e1}^{(t)} & \log h_1^{(t)} & \ln d_1^{(t)} & \log h_1^{(t)} \ln d_1^{(t)} & e_1^{(t)} \ln d_1^{(t)} \\ 1 & g_{a2}^{(t)} & g_{e2}^{(t)} & \log h_2^{(t)} & \ln d_2^{(t)} & \log h_2^{(t)} \ln d_2^{(t)} & e_2^{(t)} \ln d_2^{(t)} \\ 1 & g_{a3}^{(t)} & g_{e3}^{(t)} & \log h_3^{(t)} & \ln d_3^{(t)} & \log h_3^{(t)} \ln d_3^{(t)} & e_3^{(t)} \ln d_3^{(t)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & g_{an}^{(t)} & g_{en}^{(t)} & \log h_n^{(t)} & \ln d_n^{(t)} & \log h_n^{(t)} \ln d_n^{(t)} & e_n^{(t)} \ln d_n^{(t)} \end{bmatrix}$$

where n is the number of measurements and $\widehat{\sigma^{(t)}}$ is given as follows,

$$\widehat{\sigma^{(t)}} = \sqrt{\frac{\mathbf{e^{(t)}}^T \mathbf{e^{(t)}}}{n}} \tag{5}$$

where, $\mathbf{e^{(t)}} = \left(\mathbf{I} - \mathbf{X^{(t)}} \left(\mathbf{X^{(t)}}^T \mathbf{X^{(t)}}\right)^{-1} \mathbf{X^{(t)}}^T\right) \overrightarrow{y}^{(t)}$.

In additional to the OLE, the propagation parameters can be estimated by using the Expected-Maximization (EM) algorithm proposed by Teemu Roos, Petri Myllymaki, and Henry Trri [6]. The EM algorithm estimates the propagation parameters by finding the local maximum of the likelihood function from the incomplete data. It consists of two steps, the expectation step and the maximization step. During the expectation step, unknown parameters are estimated based on the measurement and the current parameters. This estimated complete data is called the ”hidden data”. While in the maximization step, the parameters are estimated by maximizing the ”hidden data” using the Maximum Likelihood method. These two steps are repeated until the estimated parameters converge.

The EM algorithm is used because the RSS obtained (or measured) is either rounded off to finite accuracy or truncated. These two problems occurred because the RSS is mapped to an RXLEV in accordance with the GSM specifications [4] as shown in table 1. However, we found that the EM algorithm in [6] has

Table 1: Mapping between RXLEV and RSS

RXLEV	**RSS (dBm)**
0	< -110
1	-110 to -109
2	-109 to -108
$\vdots$	$\vdots$
62	-49 to -48
63	> -48

some problems. It assumed that the actual RSS can be any value exceeds the boundary when the observed RSS is equal to the threshold value. However, according to the GSM specification, the range of received signal strength is between -29dBm and -114dBm only. In view of these problems, we revised the EM algorithm in [6] accordingly.

3 Modified EM Algorithm

3.1 Sufficient Data Collection

We classify the observed RSS and actual RSS according to the range of received signals into three types which are defined as follows.

$$\begin{cases} o_i^{(t)} - \varepsilon_l \leq s_i^{(t)} < o_i^{(t)} + \frac{\varepsilon_n}{2} & \text{(C1: } s_i^{(t)} \leq -110\text{)} \\ o_i^{(t)} - \frac{\varepsilon_n}{2} \leq s_i^{(t)} < o_i^{(t)} + \frac{\varepsilon_n}{2} & \text{(C2: } -110 < s_i^{(t)} < -48\text{)} \\ o_i^{(t)} - \frac{\varepsilon_n}{2} \leq s_i^{(t)} < o_i^{(t)} + \varepsilon_u & \text{(C3: } s_i^{(t)} \geq -48\text{)} \end{cases} \tag{6}$$

where, $o_i^{(t)}$ is the ith observed RSS from a BTS of antenna type t at a location. s_i^t is the ith actual RSS from a BTS of antenna type t at a location. While ε_l, ε_n and ε_u are the lower boundary error, normal error and upper boundary error respectively, their values are shown as follows.

$$\begin{cases} \varepsilon_l = 4\text{dBm} \\ \varepsilon_n = 1\text{dBm} \\ \varepsilon_u = 19\text{dBm} \end{cases} \tag{7}$$

The mean observed RSS, $\overline{o}$, and the mean actual RSS, $\overline{s}$, are defined as follows [4].

$$\overline{o} \overset{def}{=} E\left(o_i^{(t)}\right) \tag{8}$$

$$\overline{s} \overset{def}{=} E\left(s_i^{(t)} | o_i^{(t)}\right) \tag{9}$$

So, $\overline{o}$ and $\overline{s}$ can be represented as follows,

$$\overline{o} - l \leq \overline{s} < \overline{o} + u \tag{10}$$

where

$$l = \frac{2C_1\varepsilon_l + C_2\varepsilon_n}{2(C_1+C_2)} \qquad u = \frac{C_2\varepsilon_n + 2C_3\varepsilon_u}{2(C_2+C_3)} \tag{11}$$

$C_1 =$ Total no. of observations from a BTS of antenna type t at a location satisfies C1 in Eq. (6)
$C_2 =$ Total no. of observations from a BTS of antenna type t at a location satisfies C2 in Eq. (6)
$C_3 =$ Total no. of observations from a BTS of antenna type t at a location satisfies C3 in Eq. (6)

[4] $E(X)$ denotes the expectation of X.

During the expectation step of the EM algorithm, $\overrightarrow{y}^{(t)}$ is defined as,

$$\overrightarrow{y}^{(t)} \stackrel{def}{=} \begin{bmatrix} p_1^{(t)} - E\left(\overline{s_1^{(t)}}|\widehat{\theta^{(t)}}'\right) \\ p_2^{(t)} - E\left(\overline{s_2^{(t)}}|\widehat{\theta^{(t)}}'\right) \\ \vdots \\ p_i^{(t)} - E\left(\overline{s_i^{(t)}}|\widehat{\theta^{(t)}}'\right) \\ \vdots \\ p_j^{(t)} - E\left(\overline{s_j^{(t)}}|\widehat{\theta^{(t)}}'\right) \end{bmatrix} \tag{12}$$

where, $p_i^{(t)}$ is the transmitted power of a BTS of antenna type t at location i. j is number of location of antenna type t. $\widehat{\theta^{(t)}}'$ is the estimated propagation parameters in the previous maximization step.

$$E\left[\overline{s_i^{(t)}}|\widehat{\theta}^{t'}\right] = \frac{\left[exp\left(-\frac{1}{2}(a_i^{(t)})^2\right)\right] - \left[exp\left(-\frac{1}{2}(b_i^{(t)})^2\right)\right]\widehat{\sigma}^{t'}}{\sqrt{2\pi}\left[\Phi\left(b_i^{(t)}\right) - \Phi\left(a_i^{(t)}\right)\right]} + \widehat{\mu_i^{(t)}}' \tag{13}$$

where, $a_i^{(t)}$ and $b_i^{(t)}$ are given by,

$$a_i^{(t)} \stackrel{def}{=} \frac{(\overline{o_i^{(t)}} - l_i^t) - \widehat{\mu_i^{(t)}}'}{\widehat{\sigma}^{t'}} \tag{14}$$

$$b_i^{(t)} \stackrel{def}{=} \frac{(\overline{o_i^{(t)}} + u_i^t) - \widehat{\mu_i^{(t)}}'}{\widehat{\sigma}^{t'}} \tag{15}$$

Φ is the cumulative distribution function of a Gaussian distribution with zero mean and unity variance, $\widehat{\mu_i^{(t)}}'$ is the mean signal strength value given by Eq 2 using the estimated propagation parameters, $\widehat{\theta}^{t'}$ in the previous maximization step. $\widehat{\sigma}^{t'}$ is the standard derivation of antenna type t in the previous maximization step and is given as follows,

$$\widehat{\sigma}^{t'} = \sqrt{\frac{\mathrm{SESE}_t}{j}} \tag{16}$$

where, SESE_t is defined as,

$$\mathrm{SESE}_t \stackrel{def}{=} \sum_{i=1}^{j} E\left[(\overline{s_i^{(t)}} - \widehat{\mu_i^{(t)}})^2|\widehat{\theta^{(t)}}'\right] \tag{17}$$

and,

$$\begin{aligned} E\left[\left(\overline{s_i^{(t)}} - \widehat{\mu_i^{(t)}}\right)^2|\widehat{\theta^{(t)}}'\right] = \;& \frac{(\widehat{\sigma_i^{(t)}}')^2\left(a_i^{(t)}\left[exp\left(-\frac{1}{2}a_i^{(t)}\right)^2\right] - b_i^{(t)}\left[exp\left(-\frac{1}{2}b_i^{(t)}\right)^2\right]\right)}{\sqrt{2\pi}\left[\Phi\left(b_i^{(t)}\right) - \Phi\left(a_i^{(t)}\right)\right]} \\ & + \frac{2\widehat{\sigma_i^{(t)}}'\left(\widehat{\mu_i^{(t)}}' - \widehat{\mu_i^{(t)}}\right)\left(\left[exp\left(-\frac{1}{2}a_i^{(t)}\right)^2\right] - \left[exp\left(-\frac{1}{2}b_i^{(t)}\right)^2\right]\right)}{\sqrt{2\pi}\left[\Phi\left(b_i^{(t)}\right) - \Phi\left(a_i^{(t)}\right)\right]} \\ & + (\widehat{\sigma_i^{(t)}}')^2 + (\widehat{\mu_i^{(t)}}' - \widehat{\mu_i^{(t)}})^2 \end{aligned} \tag{18}$$

After calculating the $\overrightarrow{y}^{(t)}$ and $\widehat{\sigma^{(t)}}$, the propagation parameters $\theta^{(t)}$ are maximized using Eq 4. The updated propagation parameters are then looped back to the expectation step until they are converged.

3.2 Insufficient Data Collection

Parameters estimation involve a lot of training data. However, we may not be able to collect enough data for this purpose. To solve this problem, we would like to introduce the hidden RSS. Generally, a MS can received signals from 9 BTSs only. However, when there a lot of surrounding BTSs, the MS may receive more than 9 BTSs. These RSSs can not be collected directly but they should not be greater than the minimum RSS measured by the MS at a location. We called these RSS as hidden RSS. Thus, in addition to the observed RSS, we could used the hidden RSS to increase the training sample size.

In brief, the hidden RSS is given by,

$$h_i^{(t)} \leq min(o_i^{(t)} + \frac{\varepsilon_n}{2}) \tag{19}$$

where, $min(o_i^{(t)})$ is the minimum of the ith observed RSS from a BTS of antenna type t at a location. h_i^t is the ith hidden RSS from a BTS of antenna type t at a location.

The mean hidden RSS, $\overline{h}$ is defined as,

$$\overline{h} \stackrel{def}{=} E\left(h_i^{(t)} | min(o_i^{(t)})\right) \tag{20}$$

and,

$$\overline{h} \leq \overline{min(o)} + \frac{\varepsilon_n}{2} \tag{21}$$

Similar to the Sufficient Data Collection case, in the expectation step, $\overrightarrow{y}^{(t)}$ is defined as,

$$\overrightarrow{y}^{(t)} \stackrel{def}{=} \begin{bmatrix} p_1^{(t)} - E\left(\overline{s_1^{(t)}} | \widehat{\theta^{(t)}}'\right) \\ \vdots \\ p_i^{(t)} - E\left(\overline{s_i^{(t)}} | \widehat{\theta^{(t)}}'\right) \\ p_{i+1}^{(t)} - E\left(\overline{h_{i+1}^{(t)}} | \widehat{\theta^{(t)}}'\right) \\ \vdots \\ p_j^{(t)} - E\left(\overline{h_j^{(t)}} | \widehat{\theta^{(t)}}'\right) \end{bmatrix} \tag{22}$$

and,

$$E\left[\overline{h_i^{(t)}} | \widehat{\theta^t}'\right] = -\frac{exp\left(-\frac{1}{2}c_i^{(t)}\right)^2 \widehat{\sigma^t}'}{\sqrt{2\pi}\Phi\left(c_i^{(t)}\right)} + \widehat{\mu_i^{(t)}}' \tag{23}$$

where, $c_i^{(t)}$ is given by,

$$c_i^{(t)} \stackrel{def}{=} \frac{\overline{(min(o_i^{(t)})} + \frac{\varepsilon_n}{2}) - \widehat{\mu_i^{(t)}}'}{\widehat{\sigma^t}'} \tag{24}$$

$\widehat{\sigma^t}'$ is given by Eq 16 and Eq 17, and $E[(\overline{s_i^{(t)}} - \widehat{\mu_i^{(t)}})^2 | \widehat{\theta^{(t)}}']$ is given by,

$$\begin{aligned} E\left[\left(\overline{s_i^{(t)}} - \widehat{\mu_i^{(t)}}\right)^2 |\widehat{\theta^{(t)}}'\right] = & -\frac{(\widehat{\sigma_i^{(t)}}')^2 c_i^{(t)} exp\left(-\frac{1}{2}(c_i^{(t)})^2\right)}{\sqrt{2\pi}\Phi\left(c_i^{(t)}\right)} - \frac{2\widehat{\sigma_i^{(t)}}'\left(\widehat{\mu_i^{(t)}}' - \widehat{\mu_i^{(t)}}\right) exp\left(-\frac{1}{2}c_i^{(t)}\right)^2}{\sqrt{2\pi}\Phi\left(c_i^{(t)}\right)} \\ & + (\widehat{\sigma_i^{(t)}}')^2 + (\widehat{\mu_i^{(t)}}' - \widehat{\mu_i^{(t)}})^2 \end{aligned} \tag{25}$$

Similar to the Sufficient Data case, the propagation parameters $\theta^{(t)}$ are maximized using Eq 4 and then repeated the expectation step with the updated propagation parameters.

3.3 Location Estimation

To perform location estimation, first, we find the posterior p.d.f. of each location with reference to these serving and neighboring BTSs. The estimated location is then the maximum posterior p.d.f. of these locations.

Given a set of propagation loss ($\overline{\mathbf{pl}}$) from a set of receiving BTSs ($\mathbf{bs}$) and the estimated propagation parameters $\hat{\theta}$, $\overline{\mathbf{pl}}$ is given by,

$$\overline{\mathbf{pl}} = \mathbf{p} - \overline{\mathbf{o}} \tag{26}$$

and the posterior p.d.f of a location (l) is as follows,

$$p\left(l|\overline{\mathbf{pl}}, \mathbf{bs}, \hat{\theta}\right) = \frac{g\left(\overline{\mathbf{pl}}|l, \mathbf{bs}, \hat{\theta}\right)\pi\left(l\right)}{\int g\left(\overline{\mathbf{pl}}|l', \mathbf{bs}, \hat{\theta}\right)\pi\left(l'\right)dl'} \tag{27}$$

where, $\mathbf{p}$ and $\overline{\mathbf{o}}$ are the transmitted power and the observed RSS respectively, and $\pi\left(l\right)$ is the prior p.d.f of location l.

Furthermore, for Sufficient Data Collection case, we have

$$g\left(\overline{\mathbf{pl}}|l, \mathbf{bs}, \hat{\theta}\right) = \prod_{i=1}^{n} \int_{\overline{pl}_i - u_i}^{\overline{pl}_i + l_i} g_i\left(\overline{pl}_i|l, bs_i, \hat{\theta}\right) d\overline{pl}_i \tag{28}$$

where n is the number of receiving BTSs and pl_i is the propagation loss with respect to bs_i.

For Insufficient Data Collection case,

$$g\left(\overline{\mathbf{pl}}|l, \mathbf{bs}, \hat{\theta}\right) = \prod_{i=1}^{n_1} \int_{\overline{pl}_i - u_i}^{\overline{pl}_i + l_i} g_i\left(\overline{pl}_i|l, bs_i, \hat{\theta}\right) d\overline{pl}_i \prod_{j=1}^{n_2} \int_{\overline{pl}_j - \frac{\varepsilon_n}{2}}^{\infty} g_j\left(\overline{pl}_j|l, bs_j, \hat{\theta}\right) d\overline{pl}_j \tag{29}$$

where n_1 and n_2 is the number of observed receiving BTSs and hidden BTSs respectively.

$g_i\left(\overline{pl}_i|l, bs_i\hat{\theta}\right)$ is defined as

$$g\left(\overline{pl}|l, bs, \theta\right) \stackrel{def}{=} f\left(\overline{pl}|d\left(l, l_{bs}\right), g_a\left(l, l_{bs}, t_{bs}\right), g_e\left(l, l_{bs}, t_{bs}\right), h_{bs}, e\left(l, l_{bs}\right), t_{bs}, \theta\right) \tag{30}$$

where $f\left(\overline{pl}|d\left(l, l_{bs}\right), g\left(l, l_{bs}, t_{bs}\right), h_{bs}, e\left(l, l_{bs}\right), t_{bs}, \theta\right)$ has been defined in Eq. (3) with $d\left(l, l_{bs}\right)$ is the distance between the location l and BTS's location l_{bs}. $g_a\left(l, l_{bs}, t_{bs}\right)$ and $g_e\left(l, l_{bs}, t_{bs}\right)$ are the azimuth and elevation directive gain with respect to the location l respectively and, BTS's location l_{bs} and the antenna type t_{bs}. h_{bs} is the height of the BTS and $e\left(l, l_{bs}\right)$ is the environment index between l and l_{bs}.

4 Results

We obtained supports from a major mobile operator of Hong Kong in conducting this research. We conducted field tests in 52 different areas in Hong Kong and classified them into four types of environments, which are seashore, suburban, urban and metropolitan areas. In our field tests, we first planned some check points in the districts for our study. We call each of these points *Marker Location*. Then we collect the geographical location of the serving or involved BTSs in the districts on the days of field tests and the received signal strength from the marker locations from the operator. We used 30% of the data for parameters estimation and used the rest of the data for location estimation. Table 2 shows the results of the DPM using different parameters estimation methods. The EM*, EM and OLE in Table 2 denote the Modified EM algorithm, the EM algorithm in [6] and the OLE algorithm respectively. As observed from Table 2, the EM* and EM perform much better than the OLE method in different kinds of environment for location estimation. When comparing EM* and EM, we found that the EM* give a slightly better estimation than the EM. Moreover, the variance of the EM* is smaller than the EM in all aspect of areas which proved that the EM* is a more stable algorithm.

5 Conclusions

We have revised the EM algorithm for propagation parameters estimation. The modified EM algorithm is based on the EM algorithm proposed in [6]. In addition, we have provided the Insufficient Data Collection mode which can be applied for parameters estimation when the amount of data collected is not enough for estimation. We used the DPM for validating the modified EM algorithm. From our experiments, we proved that the modified EM algorithm yields better results than the EM algorithm in [6].

Table 2: Estimation results using different parameters estimation methods

Environment	Models	Avg. Err.	Min. Error	Max. Err.	Variance
Seaside	DPM [EM*]	475.46 m	11.18 m	3370.21 m	282129.92
	DPM [EM]	478.01 m	11.18 m	3370.21 m	282326.62
	DPM [OLE]	500.19 m	3.61 m	3175.96 m	232713.37
Suburban	DPM [EM*]	445.57 m	37.47 m	2051.01 m	154725.03
	DPM [EM]	458.68 m	37.47 m	2302.40 m	178398.44
	DPM [OLE]	461.03 m	26.86 m	2799.15 m	180202.05
Urban	DPM [EM*]	305.29 m	7.37 m	1730.46 m	93357.31
	DPM [EM]	304.63 m	7.37 m	1730.46 m	94061.88
	DPM [OLE]	363.02 m	12.50 m	2502.28 m	173112.37
Metropolitan	DPM [EM*]	124.87 m	5.39 m	909.47 m	10143.43
	DPM [EM]	123.43 m	5.39 m	909.47 m	10520.63
	DPM [OLE]	220.45 m	6.82 m	1285.24 m	62189.91
Overall	DPM [EM*]	325.41 m	5.39 m	3370.21 m	142738.08
	DPM [EM]	328.18 m	5.39 m	3370.21 m	149183.93
	DPM [OLE]	376.73 m	3.61 m	3175.96 m	169308.30

Acknowledgment

The work reported was supported in part by the Innovation and Technology Fund of the Innovation and Technology Commission of the Hong Kong SAR Government under ITS/02/22.

References

[1] H.T. Friis. A note on a simple transmission formula. In *Proceedings IRE*, volume 34, 1996.

[2] M. Hata. Empirical formula for propagation loss in land mobile radio services. In *IEEE Transactions on Vehicular Technology*, volume 29, pages 317–325, August 1980.

[3] Kenneth M.K. Chu, Karl R.P.H. Leung, Joseph Kee-Yin Ng, Chun Hung Li. Locating mobile stations with statistical directional propagation model. In *Proceedings of the 17th International Conference on Advanced Information Networking and Applications (AINA2004)*, pages 230–235, Fukuoka, Japan, March 29-31 2004.

[4] Technical Specification Group GSM/EDGE Radio Access Network. *3GPP TS 05.08 v8.16.0*. 3rd Generation Partnership Project (3GPP), 2003-04. Radio subsystem link control.

[5] T.S. Rappaport. *Wireless Communications: Principles & Practice*. Prentice Hall PTR, Upper Saddle River, New Jersey, 2002.

[6] Teemu Roos, Petri Myllymaki, Henry Tirri. A statistical modeling approach to location estimation. In *IEEE Transactions on Mobile Computing*, volume 1, pages 59–69, January-March 2002.

[7] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda. Field Strength and Its Variability in VHF and UHF Land-Mobile Radio Service. Technical report, Review of the Electrical Communication Laboratory, September-October 1968.