Indian Journal of Science and Technology, *Vol 9(6), DOI: 10.17485/ijst/2016/v9i6/75195, February 2016*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Multi Linear Regression Model for Mobile Location Estimation in GSM Network

Longinus S. Ezema* and Cosmas I. Ani

Department of Electronic Engineering, University of Nigeria Nsukka (UNN), Enugu, Nigeria; ezemms@yahoo.com, cosmas.ani@unn.edu.ng

Abstract

Background/Objectives: Recently, several researchers are directed in the area of mobile location estimation in GSM network. The major interests in the research include improving the accuracy of location estimation. **Methods/Statistical analysis:** Therefore, this paper present the first application of Multiple Linear Regression (MLR) analysis for mobile location estimation in a GSM network and without pre-processing or manipulating the Location Dependent Parameter (LDP) - Received Signal Strength Indicator (RSSI). The proposed model was developed and evaluated using Received Signal Strength (RSS) and geographical coordinates obtained from drive tests. **Findings:** The results show that, 67% of the calls, the positioning error is less than 64 m and 95% of the calls will result in positioning error less than 115 m while the maximum error is 275m for the urban area. **Application/Improvements:** Results show improved accuracy in location estimation. This model can be adopted for any mobile location application including the emergency call services (E-911) that requires very high accuracy level.

Keywords: GSM Network, GPS, Mobile Location Estimation, Multi-Linear Regression Model, Received Signal Strength,

1. Introduction

The deployment and advancement of mobile location system in various countries is largely due to ever- increasing Location Based Services (LBS), such as safety and security, tourism, position based billing, tracking of people and materials, WEB information based on location, entertainment, and improvement of handover paging and cellular network among others[1]. The Federal Communication Commission (FCC) in the USA has been regulating the system to ensure compliance and that any Mobile Station (MS) is located once the number dialled 911 emergency numbers. Their regulations has come in phases with accuracy requirement of phase II currently in use while researchers are currently on their toes to meet the accuracy requirement of phase III realised this February 2015 and expected to take effect in six-year time.

Mobile positioning involves a variety of technologies, which are divided into two main categories: Network-based and handset or MS-based location estimation[2]. In handset-based, the MS receives the Location Dependent Parameters (LDP) from the Base Station (BS) and calculates its position and report the result back to the network. The major disadvantages of this technique are that it requires modifications to the MS, SIM card, or both to enable it calculate its location, and it is also power consuming. The MS categories are Global Positioning System (GPS), Assisted GPS (AGPS) and Differential GPS (DGPS). By contrast, the network-based estimation requires less or no modification to the existing network and is relatively less complex. Here the nearby BS receives mobile station LDP and computes the position of the MS. Although the MS-based method provides better accurate estimate than the network method, the later has the advantages of being applied with all model of MS and where GPS cannot be used. Therefore, the application of MS based techniques is essential in a positioning system for emergency 911 calls[3,4].

A mobile station technique based on modifications to of MS, SIM cards, or both, implies that subscribers must either upgrade their handsets or renew SIM cards to

**Author for correspondence*

enable them use location services. By contrast, a network based can only upgrade the operator's network if it has to but not a handset. Even the GPS technology proven to be more mature, accurate and desirable system introduced to assist in mobile location is very inaccurate in populated and indoor area. Furthermore, from an engineering point of view, it is a serious engineering and economic concern to modify all handset and upgrade millions of handsets that are already in operation to enable location estimation capability. Thus, it is necessary to develop positioning techniques for unmodified mobile stations[6,11].

GSM network was designed without mobile positioning in mind, the LDP that can be measured are usually scarce. The LDP used in this paper if Received Signal Strength (RSS) because it is considered as the simplest and cheapest method amongst the wireless distance estimation techniques. Might not be the best in terms of positioning accuracy. The RSS can be applied with little or no modification to the existing network and for the fact that GSM MS is designed to send the RSS to the serving base station to aid in handoff decision of the network on a regular basis[8,9,11,12].

As environments are different, so are the signal propagation characteristics and as a result it is not wise to use a single generic signal attenuation model to characterise all environments and use it for all cases in the localisation process. Even if outdoor environments are dealt with, it will very likely have areas that possess urban, suburban or even rural signal propagation characteristics. Therefore, if environment type in which the MS in consideration resides can be estimated, that will provide us with valuable information to be used in the localisation process. Standard location techniques, based on different type of measurements or Location Dependent Parameters (LDP), have been applied to both indoor and outdoor positioning, and cellular networks (GSM, CDMA, TETRA etc.)[3].

2. Materials and Methods

In this section, a close look at the proposed process design in Figure 1 is presented with more emphasis on the localisation model.

2.1 Site Survey

In mobile network deployment, the first step is site survey. Site survey is a task-by-task process that enables the surveyor study the propagation environment and facilities

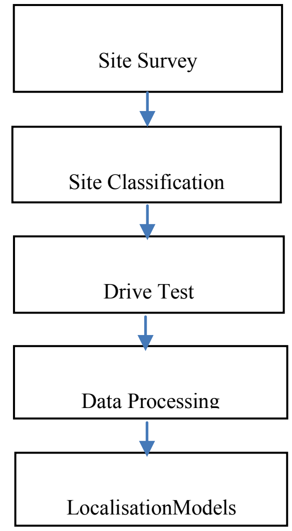

Figure 1. Proposed process design.

to understand the Radio Frequency (RF) behaviour, discover the area under RF coverage, check for overshoot and determine the positioning of BSs to ensure optimum operation through random collection of received signal strength (RSSI)[13,16]. The measurement of real-world interference, blockage and (RSSI) at a site cannot be by proxy, only on-site measurements and surveys can give the complete picture. Drive test route is determined based on cluster under consideration and to ensure that the route should pass through the urban area and no route should be passed twice[16]. The site survey also helps to classify the site.

2.2 Site Classification

The site under study is classified as an urban area. Some of the parameters which were considered for the classification are; the population of people which directly correlate with the number of mobile phones in the area, the base station cluster, the height of the buildings (which average three storey is approximately 12metres), the separation distance of the buildings and other signal barriers in the area. Also considered is the road network for the smooth drive test. Achara layout in the city of Enugu, Enugu State was adopted as a suitable urban area and the site for study.

2.3 Drive Test

The drive test was conducted for existing Mobile Telephone Network (MTN) in the Achara layout in the city of Enugu for the purpose of this research by MERIT Technologies Nigeria. A vehicle carrying the drive test set-up traveled a predetermined route within the area at a relative steady speed of 30Km/h. Due to the line-of-sight limitation of GPS system, the external GPS receiver antenna was fixed on top of the car roof at a height of 1m from the ground. TEMS software was installed in the GPS system for data collection. Figure 2, Shows the TEMS software as configured for data collection. Signal strength of MTN's GSM network was recorded in the TEMS software at different locations for every 1second along the drive test routes. The measurement recorded the received signal strength of mobile links between the mobile station/receiver and the base stations and the corresponding GPS coordinates at each point of measurements. The measurement will also include the serving and neighboring base station with timestamps to indicate when measurements were taken.

2.4 Modelling Approach

Multiple linear regression attempts to model the relationship between two or more explanatory/independent variables (received signal strength in this case) and a response or dependent variable (MS coordinates) by fitting a linear equation to observed data. Before attempting to fit a linear or multiple linear models to observed data, it is necessary to establish if there is a relationship between the variables that makes up the equation. This does not necessarily imply that one variable causes the other. The method of least-squares adopted here is the most used method for fitting a regression line. Minimising the summed squared of residuals of actual and estimated location is the system behind all regression techniques, then the regression coefficients that achieve this minimization are obtained by least squared algorithm. Theoretically, these coefficients are determined by differentiating summed squares of residuals with respect to each parameter and then setting the result equal to zero.

It is assumed that an MS is located at (x_M, y_M) and that location information $(x_i, y_i)_{i=1}^{m}$ as well as the RSSI of several

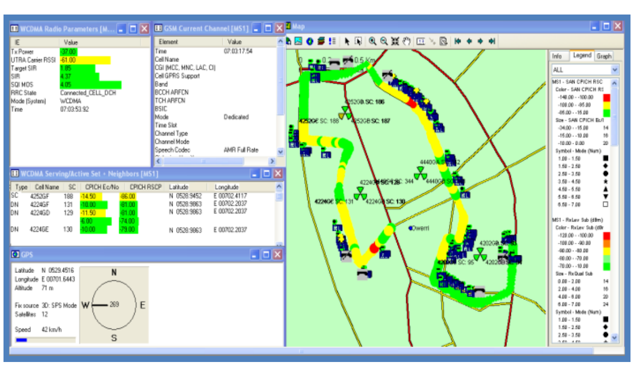

Figure 2. Drive test routes in TEMS Software.

BSs within the target area are stored. The location equation can be formulated as in Equations (1a) and (1b)[14,15].

$$\{x_i\}_{i=1}^{m} = b_0 + \sum_{j=1}^{N} b_j R_{ij} \tag{1a}$$

$$\{y_i\}_{i=1}^{m} = c_0 + \sum_{j=1}^{N} c_j R_{ij} \tag{1b}$$

Where, b_0 and c_0 are constants (intercepts) and b_j and c_j are the vectors (multiple) regression coefficients (slopes) of x – y coordinates, R_{ij} is the RSSI of j^{th} BS at location i, and m is the number of training data length. Equation (1a) and (1b) can be written in matrix form as in Equations (2a), (2b), (3a) and (3b).

$$\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{279} \end{bmatrix} = \begin{bmatrix} 1 & R_{1,1}\, R_{1,2} & \cdots & R_{1,12} \\ 1 & R_{2,1}\, R_{2,2} & \cdots & R_{2,12} \\ & \vdots & & \vdots \\ & \vdots & & \vdots \\ 1 & R_{279,1}\, R_{279,2} & \cdots & R_{279,12} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_{12} \end{bmatrix} \tag{2a}$$

$$\underline{X} = \underline{KB} \tag{2b}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_{279} \end{bmatrix} = \begin{bmatrix} 1 & R_{1,1}\, R_{1,2} & \cdots & R_{1,12} \\ 1 & R_{2,1}\, R_{2,2} & \cdots & R_{2,12} \\ & \vdots & & \vdots \\ & \vdots & & \vdots \\ 1 & R_{279,1}\, R_{279,2} & \cdots & R_{279,12} \end{bmatrix} \times \begin{bmatrix} c_0 \\ c_1 \\ . \\ . \\ c_{12} \end{bmatrix} \tag{3a}$$

$$\underline{Y} = \underline{KC} \tag{3b}$$

Vectors B and C can be expressed using least square method as in Equations (4a) and (4b)[15].

$$\underline{B} = \left(\underline{K^T K}\right)^{-1} K^T X \tag{4a}$$

$$\underline{C} = \left(\underline{K^T K}\right)^{-1} K^T Y \tag{4b}$$

Having obtained the constants and the regression coefficients from Equations (4a) and (4b), the MS coordinates can be estimated by using these coefficients vectors are shown below.

$$\left.\begin{aligned} X_M &= \left[1\underline{R}\right].B \\ \\ Y_M &= \left[1\underline{R}\right].C \end{aligned}\right\} \tag{5}$$

Through the above mathematical processes expressed as in Equations (1a) and (1b), Equations (6a) and (6b) below are developed as the model equations for mobile location estimation in the area under consideration.

$$\text{Longitude} = x = \begin{pmatrix} 729253 - 0.067R_1 - 1.98R_2 + 3.2R_3 + \\ 0.74R_4 - 0.62R_5 - 2.64R_6 + 0.61R_7 + 0.14R_8 \\ -0.51R_9 - 5.56R_{10} - 1.64R_{11} + 0.32R_{12} \end{pmatrix} \tag{6a}$$

$$\text{Latitude} = y = \begin{pmatrix} 625330 + 3.62R_1 + 0.24R_2 + 0.41R_3 - \\ 0.87R_4 + 0.53R_5 + 1.22R_6 - 1.01R_7 - 2.78R_8 \\ +2.78R_9 - 1.73R_{10} + 0.52R_{11} - 0.04R_{12} \end{pmatrix} \tag{6b}$$

2.4.1 Goodness-of-Fit

How good an analysis is can be determined by certain parameters which can be defined as follow;

Total sum of squares (SST);

$$SST = \sum_{i=1}^{n} \left(x_i - \overline{X_m}\right) \tag{7}$$

The explain sum of squares (SSE)

$$SSE = \sum_{i=1}^{n} \left(\hat{x}_i - \overline{X_m}\right)^2 \tag{8}$$

The residual sum of squares or sum of squared residuals (SSR)

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_{x_i}^{\;2} \tag{9}$$

$$SST = SSE + SSR \tag{10}$$

The above equation means that the total variation in (x_i) is the sum of the total variation in $(\hat{x}_i)$ and in $(\hat{\varepsilon}_{x_i})$.

It also shows that;

$$\frac{SSR}{SST} + \frac{SSE}{SST} = 1 \tag{11}$$

Then R-squared can be defined as;

$$R_x^2 = \frac{SSE}{SST} = 1 - SSR/SST \tag{12}$$

The same applies to y (latitude) as the R-squared expressions in equation 2 and 3 show.

$$R_x^2 = \frac{\left(\sum_{i=1}^{n}\left(x_i - \overline{X_m}\right)\left(\hat{x}_i - \overline{\widehat{X_m}}\right)\right)^2}{\left(\sum_{i=1}^{n}\left(x_i - \overline{X_m}\right)^2\right)\left(\sum_{i=1}^{n}\left(\hat{x}_i - \overline{\widehat{X_m}}\right)^2\right)} \tag{13}$$

$$R_y^2 = \frac{\left(\sum_{i=1}^{n}\left(y_i - \overline{Y_m}\right)\left(\hat{y}_i - \overline{\widehat{Y_m}}\right)\right)^2}{\left(\sum_{i=1}^{n}\left(y_i - \overline{Y_m}\right)^2\right)\left(\sum_{i=1}^{n}\left(\hat{y}_i - \overline{\widehat{Y_m}}\right)^2\right)} \tag{14}$$

3. Simulation Result and Discussion

In order to simulate the developed model, 40% of the measured sample data (RSS) were applied to it and the results obtained are the longitude and the latitude of each measurement points. There are two hundred and seventy nine (279) measurement points and four (4) base stations with three sectors each. At each point is a GPS measured coordinate for validation. The performance measure of the model was determined using methods such as the Euclidian distance, Root Mean Square Error (RMSE) and the Cumulative Distribution Function (CDF).

The term positioning error is generally considered as the Euclidean distance between the estimated location and true MS location. It can be represented in two dimensions as,

$$d_i = \sqrt{\left(x - x_e\right)^2 + \left(y - y_e\right)^2}, \tag{15}$$

Where (x_e, y_e) is the estimated location and (x, y) the corresponding true location (GPS measured coordinate). A general impression of the error distribution can be obtained quickly by looking up the CDF graph in Figure 3, while the Root Mean Square Error (RMSE) is gotten from the following equation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i}, \tag{16}$$

Where n is the number of positioned samples and d is the positioning error of sample i

With the cumulative distribution function, one can find out any desired percentile by simple visual estimation. Numerical values are used to simplify the representation as shown in Table 1, and Table 2. However, RMSE, 65% and 95% are perhaps the most commonly used measures. From Figure 3, it is seen that the MLR model has a good performance when all the hearable sectors are considered.

Figure 4 Shows that the performance of the MLR model deteriorates when the number of cell sectors is reduced.

The scenario considered in this simulation can be experienced in suburban area or rural area where there is likely few hearable BSs available for the MS. It is unrealistic in an urban area. Despite these the positioning accuracies still very good relative to the FCC standard (see Table 2.), and therefore, can be applied for E-911 emergency services. Thus, the result achieved can be greatly improve if the area under consideration is dense urban as it is very possible to have more hearable sectors than obtained for this research.

The result proves that the model can be applied for any mobile location services including the emergency calls (E-911) with high accuracy requirement.

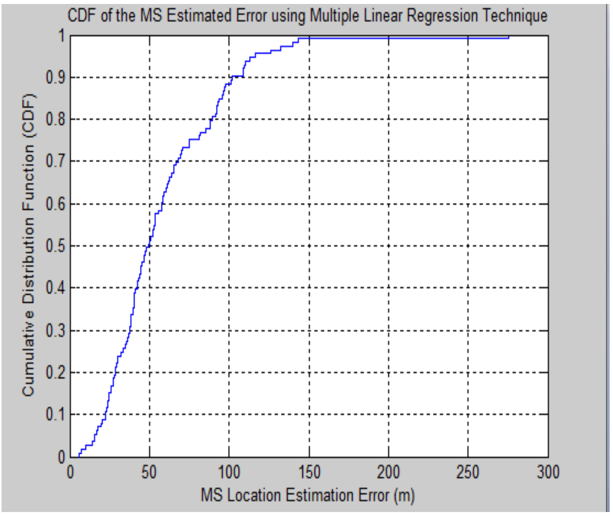

Figure 3. CDF of the MS estimated error using the proposed model.

Table 1. Summary statistics of simulation for different number of sector

No. of Sector	RMSEr (m)	67% (m)	95% (m)	STD (m)	Min (m)	Max (m)
12	63	64	115	37	6	275
8	73	83	140	41	10	271
4	94	103	181	59	5	383

Table 2. Compare simulation result of the developed MLR model over fcc maximum error standard for network based-techniques

	FCC(m)	MLR(m)	Improvement
65%	100	64	36%
95%	300	115	61%

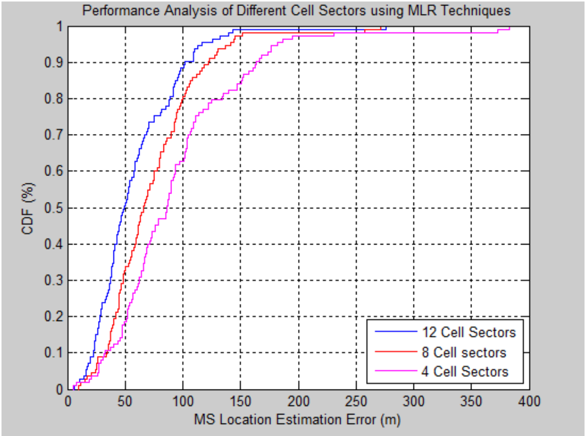

Figure 4. Performance analysis of different cell sectors using MLR technique.

3.1 Goodness-of- Fit Result

There is a mean residual ($\bar{\varepsilon}$) for the analysis and a residual (ε_i) for each observation. If $\bar{\varepsilon} > 0$, then Xm is below x and Ym is blow y, which means that for the analysis or observation x and y are under predicted. If $\bar{\varepsilon} < 0$, then Xm > x and Ym > y, which means that for the analysis or observation, x and y, are over predicted. The result shows that the mean error term of latitude ($\bar{\varepsilon}_y$) is 1.2, which means that the analysis of the estimated latitude is underestimated or under predicted and the longitude value is overestimated as the error term or residual ($\bar{\varepsilon}_x$) is –0.6.

The value $R^2 = 0.755$ and 0.91 means that the models (the twelve R-variables) explains 75.5% and 91% of the observed variation in the GSM's latitude and longitude position respectively.

4. Conclusion

The proposed multiple-linear regression model based mobile location estimation for outdoor application in GSM network deployed in an urban gave high estimation accuracy. From the simulation result, the Root Mean Square Error (RMSR) of 58 m, 67% of estimation show an error of less than 64m, and 95% of error is 116m were achieved in GSM network. The achieved location accuracy having met the FCC requirement for a network-based method can be implemented for an emergency application like E-112 and E911 as well as other location-based services. Aside the accuracy, this technique being a network-based with no modifications to both the network and the MS is seen as a great added advantage.

5. References

1. Campos RS. Localização de Terminais Móveisutilizando Correlação de Assinaturas de Rádio-Frequência Revista Telecomunicaões. 2010; 13(1).
2. Turkyilmaz F, Guir AG, Tugcu T. Environmental-aware location estimation in cellular networks, satellite networks research laboratory (SATLAB). Istanbul, Turkey. EURASIP Journal on Advance Signal Processing. 2008.
3. Ezema LS, Ani CI, Ezeh GN. Mobile location estimation in GSM/UMTS. International Journal of Emerging Technology and Research. 2014; 1(3):63–70.
4. Renato PMP, Camila FVV, Gustavo HRL. Mobile location estimation for emergency calls using data correlation method. International Workshop on Telecommunication (IWT); 2013.
5. Brian F, Dirk H, Dieter F. Gaussian processes for signal strength-based location estimation; 2006. p. 39–47.
6. Zhi-li W, Chun-hung L, Jesoph KN, Karl RPHL. Location estimation via support vector regression. IEEE Transactions on Mobile Computing. 2007; 6(3):311–21.
7. Saveeda P, Vinothini E, Swathi V, Ayyappan K. Received Signal Strenght (RSS) calculation for GSM cellular system at BSNL pondicherry using modified HATAModel. IJSETR. 2013; 2(1):43–7.
8. Hong Y-S. An indoor location tracking using wireless sensor networks cooperated with relative distance fingerprinting. Indian Journal of Science and Technology. 2015 Jan; 8(1):517–23.
9. Roos T, Petri M, Henry TA. Statistical modelling approach to location estimation. IEEE Transaction on Mobile Computing. 2002; 1(1):59–69.
10. Samiei M, Mehrjoo M, Pirzade B. Advances of Positioning Methods in Cellular Networks. 1st International Conference on Communications Engineering; Iran. 2002. p. 174–83.
11. FCC, FCC Enhanced 911 Services Rev; 2004. Available from: http://www.fcc.gov/911/enhanced
12. Khalid AM. Location aware system using mobile station in GSM Network [MSc Thesis]. Kuala Lumpur: University Malaya; 2007.
13. Hussain SA, Emran M, Salman M, Shakeel U, Naeem M, Ahmed S, Azeem A. Positioning a mobile subscriber in a cellular network system based on signal strength. International Journal of Computer Science. 2007; 34(2):1–6.
14. Cheung KW, So HC, Ma WK, Chan YT. Least sqaures algorithms for time-of-arival-based mobile location. IEEE Transaction on Signal Processing. 2004; 52(4):1121–8.
15. Walpole RE, Myers RH. Probability and Statistics for Engineers and Scientists. 3rd ed. New York: Macmillian; 1985.
16. Ezema LS, Mbonu WE, Nwogu UO, Owuamanam C. Plan, design and simulation of university network. International Journal of Advanced Computer Research. 2014; 4(3):834–42.
17. Subramanian JV, Sadiq MAK. Implementation of artificial neural network for mobile movement prediction. Indian Journal of Science and Technology. 2014 Jun; 7(6):858–63.
18. Baghyasree T, Janakiramam K, Parkavi D. Efficient remote video monitoring for industrial environment based on LPC2148 and GSM. Indian Journal of Science and Technology. 2014 Sep; 7(9):1333–41.