IOSR Journal Of Applied Physics (IOSR-JAP)
e-ISSN: 2278-4861.Volume 13, Issue 1 Ser. II (Jan. – Feb. 2021), PP 27-38
www.Iosrjournals.Org

Linear, Lasso and Ridge Regression Methods: A Quantitative Case Study with Spatial Electromagnetic Field Strength Signals

[1]Ojuh, O. Divine and [2] Isabona Joseph
[1]Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State
[2]Department of Physics, Federal University Lokoja, Kogi State

Abstract
Regression-based approximation of practical-based measurement observation is of great interest in physical sciences and telecommunication system engineering. This paper considers and appraise the performance application of three least square regression methodsfor electromagnetic field strength data estimation analysis. The three regression methods, which includes the ordinarylinear regression, ridge regression and lasso regression, algorithms were implemented with matlab 2018a software and compared to examine their regression fittingefficacies and scalability on different electromagnetic field strength data sample sizes. It was observed in terms of first order statistics that ridge regression outperformed slightly the lasso regression and the ordinarylinear regression at lower field strength samples sizes.For example, in terms mean square error, while the ridge method attained 3.33 and 2.25 dB values in site 1 and site 2 for field strength data size of 30, the lasso and the ordinary linear methods scored 3.70 and 3.71dB in site 1and 3.27 and 3.26 dB values in site 2 respectively. However, the regression fitting accuraciesof the three regression methods behaved same as the field strength data sizes increases to 120.This could also imply that ridge regression is a better regression method for prognostic estimation of field data of smaller sizes.

Keywords: *LTE, Field strength, Ordinary least square regression, Ridge regression, Lasso regression.*

Date of Submission: 10-02-2021 Date of Acceptance: 24-02-2021

I. Introduction

Regression trend analysis remained a distinctive statistical method of conducting prognostic datamodelling and analysis in many scientific and none scientific fields such as physical science, biological sciences, social sciences, chemical sciences and engineering. In these various fields, the analyst may perhaps be engrossed in modelling and estimating the trend pattern in training data set or predicting responses for future data set. However, identifying the right regression model to choose from a host of available ones plus the ability select their input variables in order make good prognostic estimations is a challenging task owing to complex relations between models and input variables [1][2]. For example, in cellular telecom networks, field drive tests or work tests are routinely conducted to examine and characterize quality of the field signal strength in temporal and spatial domain.The resultant outcome are usually employed for useful decision making process during the re-planning and optimisation of the networks.

There exist numerous regression models have been engaged by different researchers for diverse investigative studies in literature. In [3], the influence of temperature on myocardial infarction in Wales and England using time series linear regression is presented. In [4], a penalized partial least squares regression methodwith specific usagefor b-spline transformations application and functional data is presented. In [5]-[12], the researchers employed least square regression and absolute least deviation regression methods to model and perform predictive analysis of signal attenuation and propagation loss pattern in telecommunication networks and wireless local area networks.

This study focused on the application of the linear, ridge and lasso regression modelling techniques to estimate measured field strength data variability trend. The lasso and ridge models have been employed to measured field strength data of different sizes, ranging from 30 to 120 acquired from two transmitting eNodeB LTE cell sites in a typical urban environment.

Ordinary Linear, Ridge and Lasso Regression

The introduction of the least squares method, which isformally accredited to Gauss (1795) [13],grew out of the fields of geodesy and astronomy as scientists were trying to provide clearand perfect interpretation of how the celestial bodiesbehave. Till date, its applications in different fields knows no bound.

Given a measured field strength data sample (x_i, y_i), i=1,2,…,n, where y_i and $x_i = (x_{i,1},..xi_{,p})^T$ indicate the response and the predictable variables. Let $\beta = (\beta_1,..\beta_p)^T$ be the modelling parameters. In ordinary linear least square regression, the modelling parameter, β can be estimated by minimizingthe loss residual function as indicated in equation (1);

$$Linear \quad : \min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{i=1}^{p} \beta_j x_{ij} \right)^2 \quad (1)$$

The method of the ridge regression was first introduced by Hoerl and Kennard, (1970)[14]. In ridge regression, the loss functions are augmented minimizing the squared sum residuals at the expense of adding biasregularization parameter λ, subject to the constraints. That is:

$$Ridge \quad : \min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{i=1}^{p} \beta_j x_{ij} \right)^2 \text{subject to} \quad \|\beta\|_2^2 \leq t \quad (2)$$

The term 'lasso' stands for Least Absolute Shrinkage and Selection Operator;it was introduced in a publication by Tibshirani, (1996) [15]. Lasso regression, is somewhat conceptually and hypothetically similar to ridge regression, as a penalty is added for non-zero coefficients. However, lasso penalizes the sum of their absolute values, unlike ridge regression which penalizes sum of squared coefficients. Thus, the lasso regression can be expressed by:

$$Lasso \quad : \min_{\beta} \sum_{i=1}^{n} \left(yi - \sum_{i=1}^{p} \beta_j x_{ij} \right)^2 \text{subject to} \quad \|\beta\|_1 \leq t \quad (3)$$

From equation (2) and (3), it is clear that the ridge regression method utilizes $\|\beta\|_2^2$ penalty, while the Lasso utilizes penalty $\|\beta\|$.

In terms of Lagrangian formulation, equations (2) and (3) can be written as:

$$Ridge \quad : \min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} \beta_j^{\,2} \quad (4)$$

$$Lasso \quad : \min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad (5)$$

Empirical Data

In this paper, we will be employing the field strength data for least square, ridge and lasso regression analysis. The field strength was obtained based on local mean power measured at the receiver with dedicated TEMS investigation drive test tools. The power measurement campaign was conducted at 2600 MHz band different routes round two LTE eNodeB transceivers, in the waterline area of Port Harcourt, Nigeria. The heights of the two eNodeB transceivers were 30 m and 32m the ground surface, respectively. In order facilitate a relatively steady speed of about 36 km/h, the drive test measurements were conducted at weekends. On this speed, temporal and spatial resolution at about 0.25λ was objectified between measurement intervals. The field strength is directly related to the acquired local mean power at the receiver based on the following expressions [16]:

$$\Pr = \frac{E_f^2}{120\pi} \cdot \frac{c^2 . G_r}{4\pi F} \quad (6)$$

$$P_r^{dB} = (P_t - 30) + E[dB(\mu V/m)] - (120 + G_t^{dB} - 2.14) + 10 + K \quad (7)$$

where

$$K = 10 \log 10 \left(\frac{c^2 .}{480 \pi^2 f^2} \right)$$

(8)

$$E[dB(\mu V/m)] = P_r^{dB} - P_t - 30 + 120 + G_t^{dB} + 2.14 - G_t^{dB} - 10 - K$$

(9)

$$E[dB(\mu V/m)] = P_r^{dB} - P_t + G_t^{dB} - G_t^{dB} - K + 82.14$$
(10)

Empirical Results and Analysis

To build the linear, lasso and ridge regression models for the field strength analysis, we useMatlab 2018a software. To evaluate and compare the results of linear, lasso and ridge regression performance, this study use five errors estimation metrics. They include mean absolute error (MAE), maximum absolute error (MaxMAE), standard deviation error, root mean squared error (RMSE), mean absolute percentage error (MAPE) and coefficient of coefficient (R). To determine the effect of each regression method on data size, the field strength data was also divided into four sizes each (i.e., N=30, 60, 90,120). The graphs in figures 1 to 25 are provided display the correlation fits of each regression method to the measured field strength data as a function between the measurement distances from the eNodeB transmitters. From the results summary in tables 1 and 2, it was observed in terms of first order statistics that ridge regression performed slightly better ordinary least square, andlasso regression egression at lower field strength samples sizes. For example, in terms MAE, while the ridge method attained 3.33 and 2.25 dB values in site 1 and site 2 for field strength data size of 30, the lasso and the ordinary least square methods scored 3.70 and 3.71dB in site and 3.27 and 3.26 dB values in site 2 respectively. However, the regression fittingaccuraciesof the three regression methods behaved same as the field strength data sizes increases to 120 data sizes.This could also imply that ridge regression is a better regression method for prognostic estimation of field data of smaller sizes.

Table 1: Computed fitted correlation values using the three Regression Methods for site 1

Regression Type	Estimation Parameters	N=30	N=60	N=90	N=120
Least Square	Estimation Accuracy	MAE=3.71 MAPE=6.33 RMSE=4.81 MaxErr =10.84 R=0.9947	MAE=3.71 MAPE=6.32 RMSE=4.81 MaxErr =13.06 R=0.9932	MAE=4.99 MAPE=8.66 RMSE=6.25 MaxErr =20.15 R=0.9881	MAE=4.33 MAPE=7.70 RMSE=5.66 MaxErr =20.12 R=0.9895
	Regression model parameters	70.39 -0.04	70.39 -0.04	62.28 -0.012	62.31 -0.01
Ridge	Regression model parameters	MAE=3.33 MAPE=5.30 RMSE=4.59 MaxErr =10.83 R=0.9947	MAE=3.71 MAPE=6.32 RMSE=4.82 MaxErr =13.29 R=0.99432	MAE=4.99 MAPE=8.66 RMSE=6.26 MaxErr =20.15 R=0.9881	MAE=4.33 MAPE=7.70 RMSE=5.66 MaxErr =20.13 R=0.9895
	Regression parameters	74.05 -0.076	69.94 -0.045	62.22 -0.01	62.24 -0.01
Lasso	Estimation Accuracy	MAE=3.70 MAPE=6.32 RMSE=4.81 MaxErr =10.76 R=0.9947	MAE=3.70 MAPE=6.32 RMSE=4.81 MaxErr =13.24 R=0.9932	MAE=4.98 MAPE=8.64 RMSE=6.26 MaxErr =20.31 R=0.9881	MAE=4.34 MAPE=7.70 RMSE=5.66 MaxErr =20.27 R=0.9941
	Regression model parameters	70.1 -0.04	70.39 -0.046	62.11 -0.01	62.31 -0.01

Table 2: Computed fitted correlation values using the three Regression Methods for site 2

Regression Type	Estimation Parameters	N=30	N=60	N=90	N=120
Least Square	Estimation Accuracy	MAE=3.27 MAPE=6.12 RMSE=4.17 MaxErr =10.35 R=0.9941	MAE=3.48 MAPE=6.87 RMSE=4.53 MaxErr =10.85 R=0.9922	MAE=2.94 MAPE=5.97 RMSE=3.99 MaxErr =10.69 R=0.9933	MAE=2.79 MAPE=5.96 RMSE=3.72 MaxErr =10.63 R=0.9935
	Regression model parameters	64.23 -0.05	58.60 -0.02	58.60 -0.023	56.99 -0.01
Ridge	Estimation Accuracy	MAE=3.25 MAPE=6.07 RMSE=4.17 MaxErr =10.22 R=0.9941	MAE=3.47 MAPE=6.84 RMSE=4.53 MaxErr =10.82 R=0.9921	MAE=2.94 MAPE=5.97 RMSE=3.99 MaxErr =10.62 R=0.9941	MAE=2.79 MAPE=5.96 RMSE=3.73 MaxErr =10.6 R=0.9933
	Regression model parameters	63.99 -0.05	58.37 -0.02	57.69 -0.02	56.85 -0.01
Lasso	Estimation Accuracy	MAE=3.26 MAPE=6.09 RMSE=4.17 MaxErr =10.32 R=0.9941	MAE=3.48 MAPE=6.85 RMSE=4.53 MaxErr =10.78 R=0.9921	MAE=2.94 MAPE=5.97 RMSE=3.99 MaxErr =10.60 R=0.9933	MAE=2.79 MAPE=5.96 RMSE=3.73 MaxErr =10.80 R=0.9933
	Regression model parameters	64.46 -0.05	58.41 -0.02	57.73 -0.02	56.81 -0.01

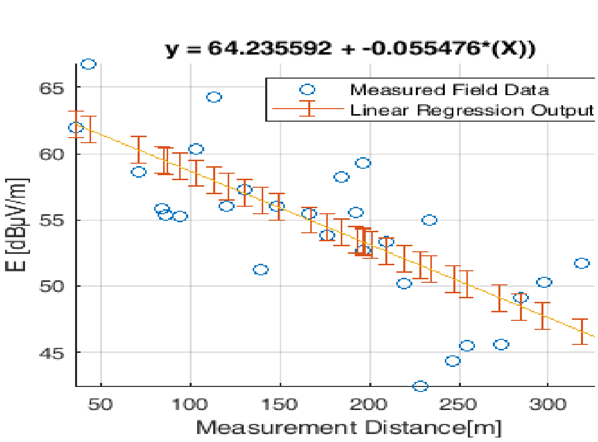

Figure 1 (a):linear regression fitting of E- field strength data size of 30, for site 1

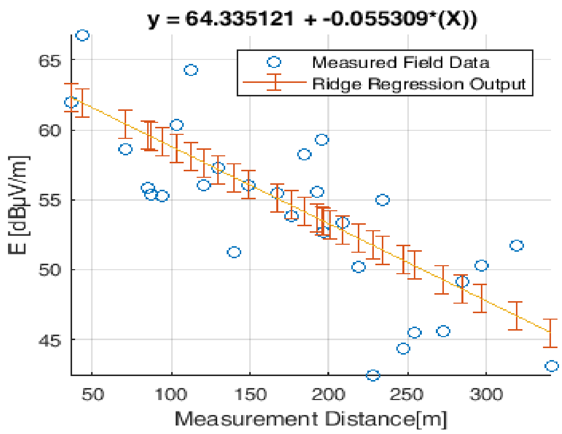

Figure 2 (a): Ridge regression fitting of E- field strength data size of 30, for site 1

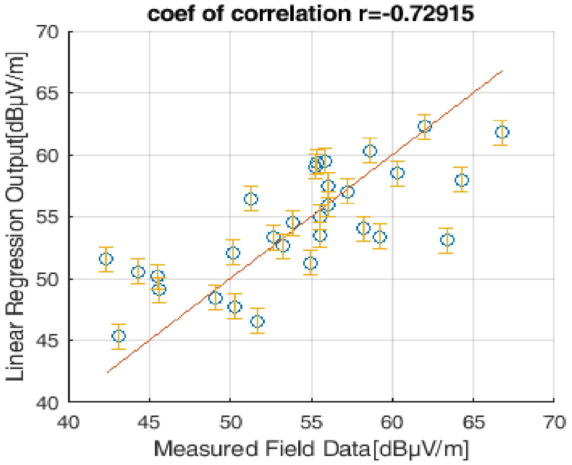

Figure 1 (b):Linear correlation fitting of E- field strength data size of 30, for site 1

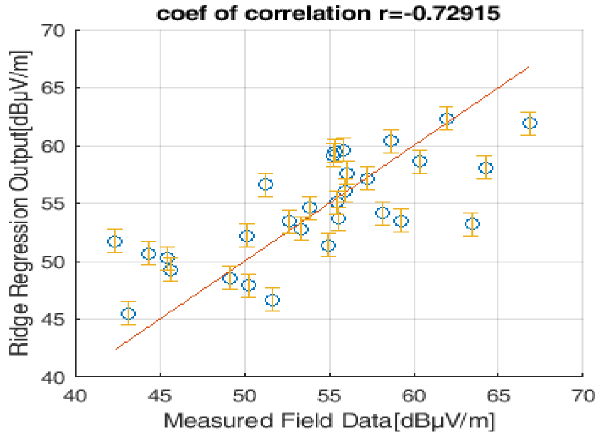

Figure 2 (b): Ridge correlation fitting of E- field strength data size of 30, for site 1

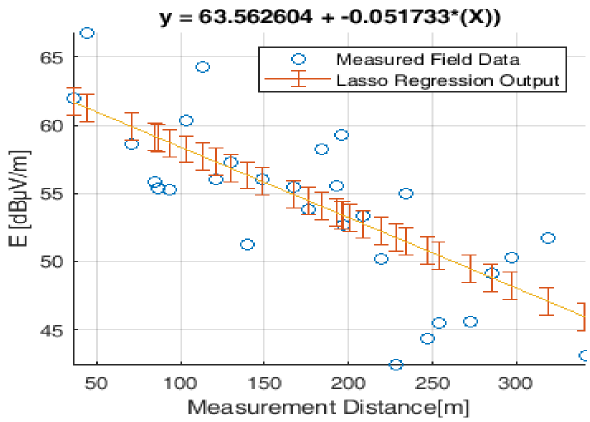

Figure 3 (a):Lasso regression fitting of E- field strength data size of 30, for site 1

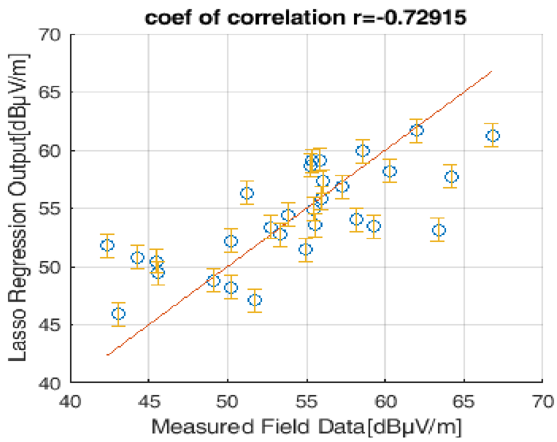

Figure 3 (b):Lasso correlation fitting of E- field strength data size of 30, for site 1

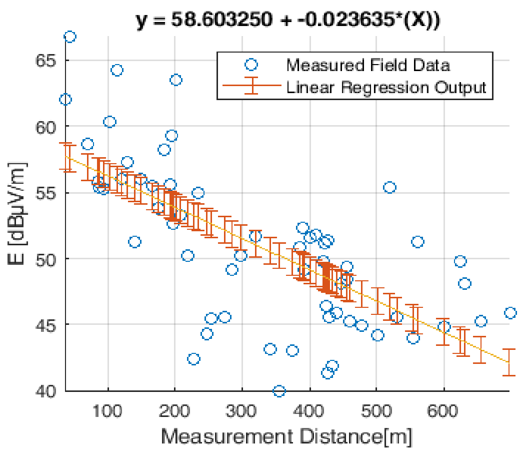

Figure 4 (a):Linear regression fitting of E- field strength data size of 60, for site 1

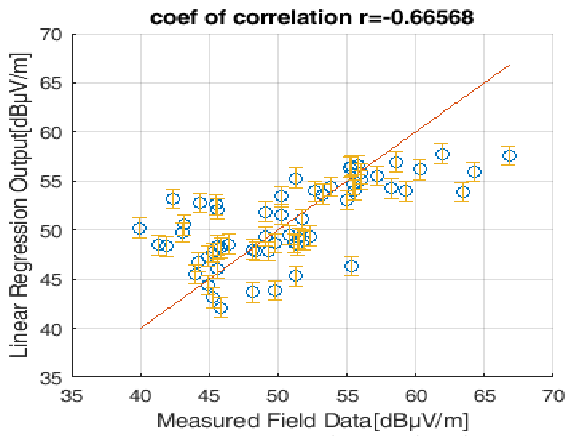

Figure 4 (b):Linearcorrelation fitting of E- field strength data size of 60, for site 1

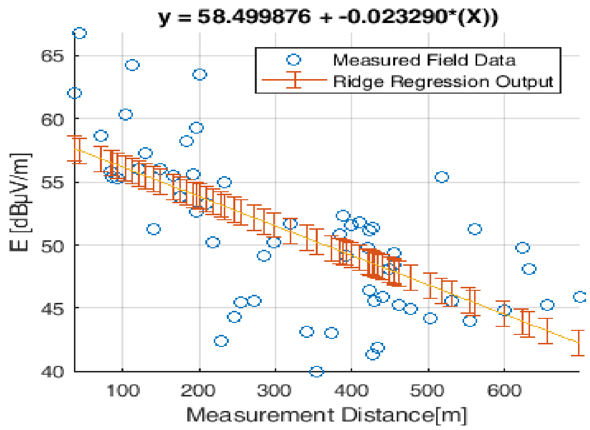

Figure 5 (a):Ridge regression fitting of E- field strength data size of 60, for site 1

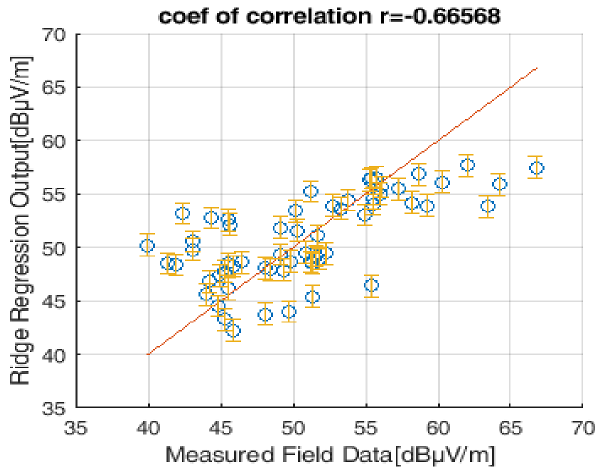

Figure 6 (b):Ridge correlation fitting of E- field strength data size of 60, for site 1

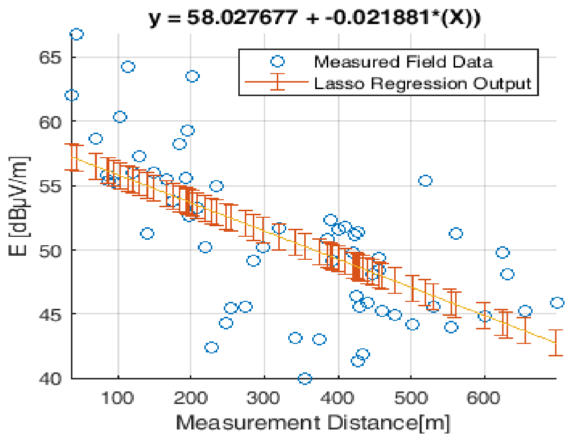

Figure 7 (a):Ridge regression fitting of E- field strength data size of 60, for site 1

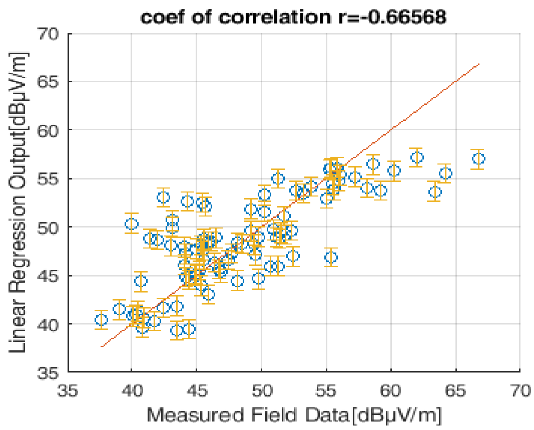

Figure 8 (b):Linear correlation fitting of E- field strength data size of 90, for site 1

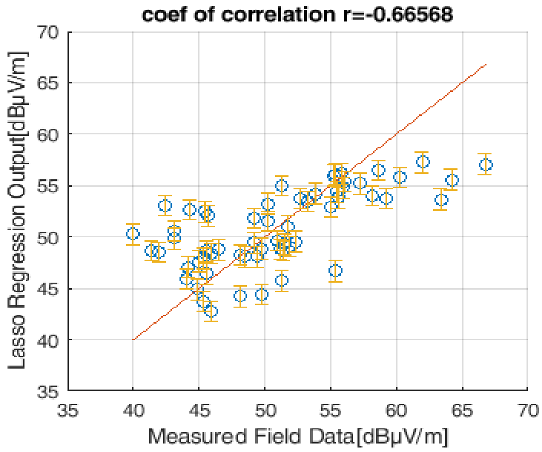

Figure 7:Lasso correlation fitting of E- field strength data size of 60, for site 1

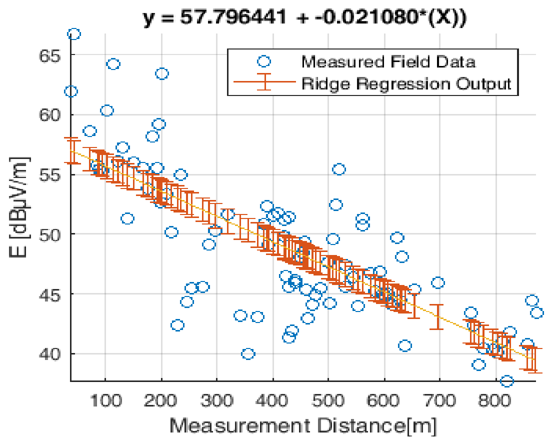

Figure 9 (a):Ridge regression fitting of E- field strength data size of 90, for site 1

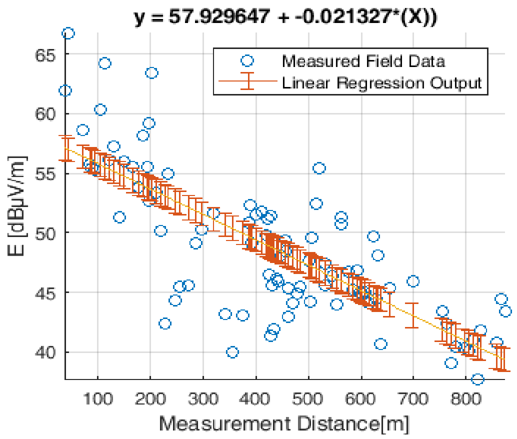

Figure 8 (a):Linear regression fitting of E- field strength data size of 90, for site 1

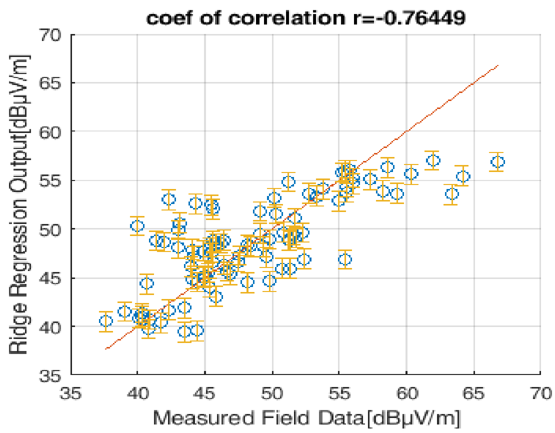

Figure 9 (b):Ridge correlation fitting of E- field strength data size of 90, for site 1

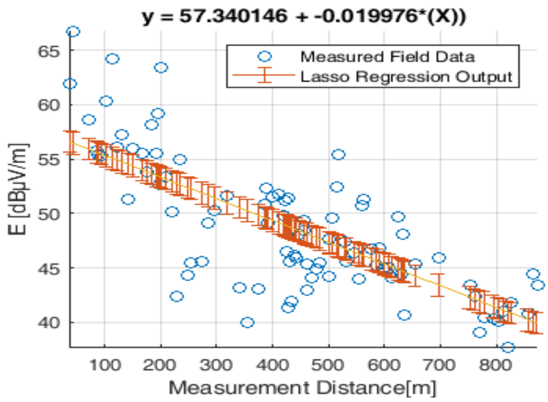

Figure 10 (a):Lasso regression fitting of E- field strength data size of 90, for site 1

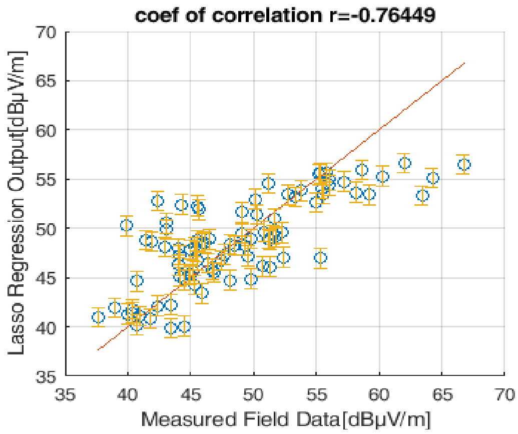

Figure 10 (b):Lasso correlation fitting of E- field strength data size of 90, for site 1

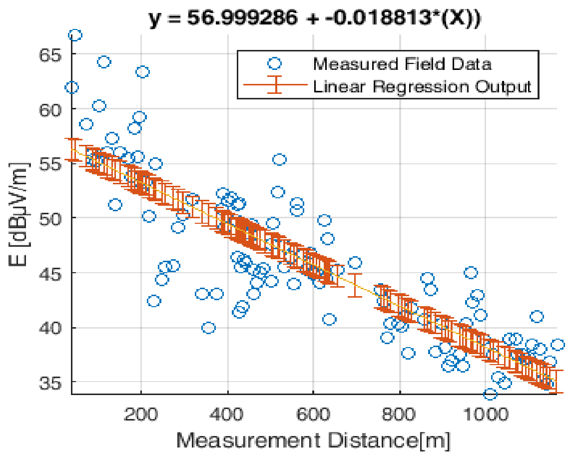

Figure 11 (a):Linear regression fitting of E- field strength data size of 120, for site 1

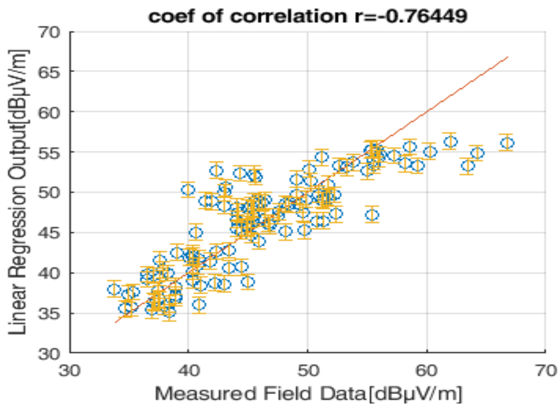

Figure 11 (b):Linear correlation fitting of E- field strength data size of 120, for site 1

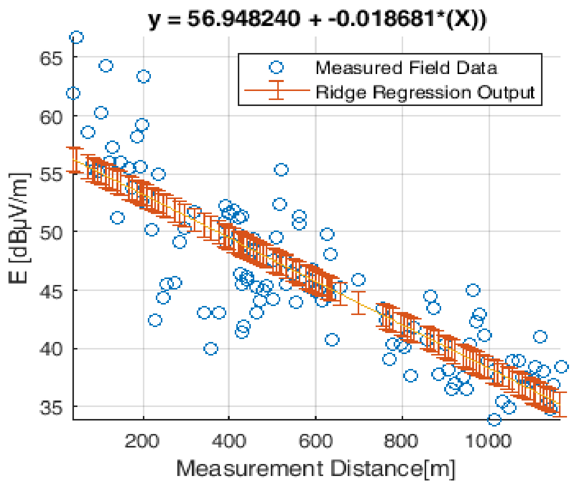

Figure 12 (a):Ridge regression fitting of E- field strength data size of 120, for site 1

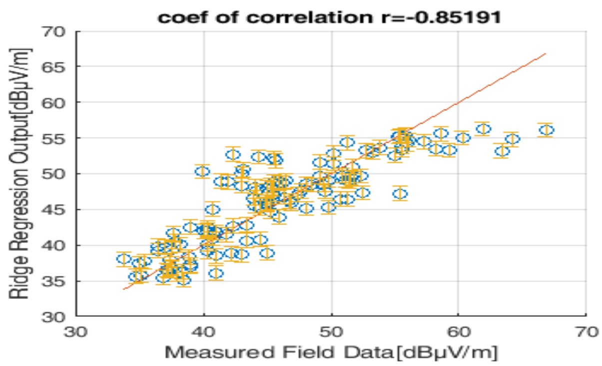

Figure 12 (b):Ridge correlation fitting of E- field strength data size of 120, for site 1

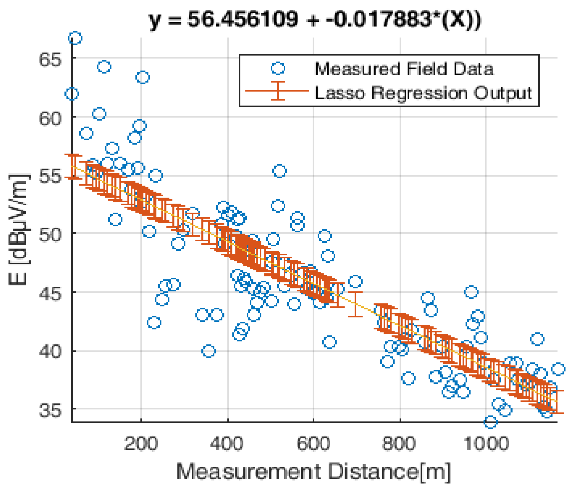

Figure 13 (a):Lasso regression fitting of E- field strength data size of 120, for site 1

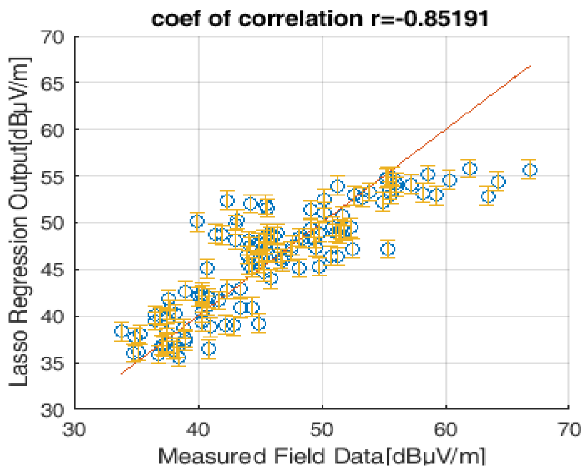

Figure 13 (b): Lasso correlation fitting of E- field strength data size of 120, for site 1

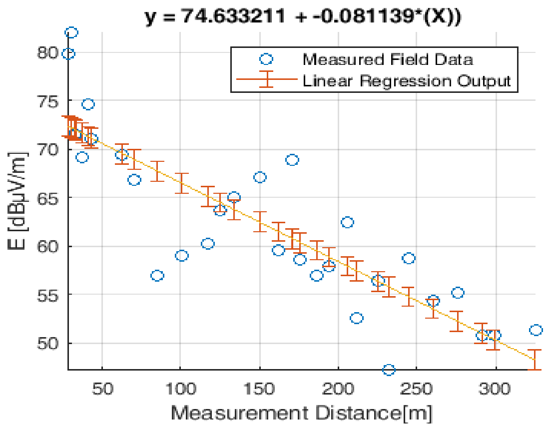

Figure 14 (a):Linear regression fitting of E- field strength data size of 30, for site 2

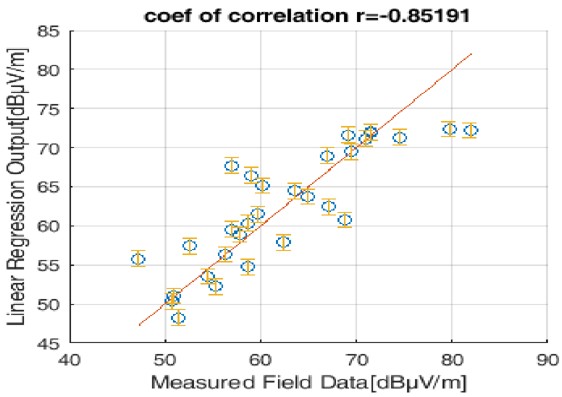

Figure 14 (b):Linear correlation fitting of E- field strength data size of 30, for site 2

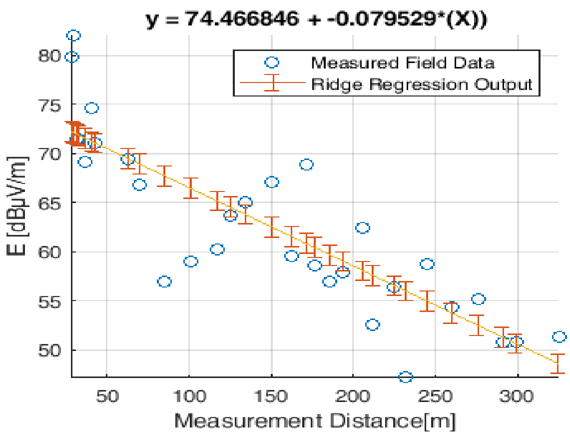

Figure 15 (a):Linear regression fitting of E- field strength data size of 30, for site 2

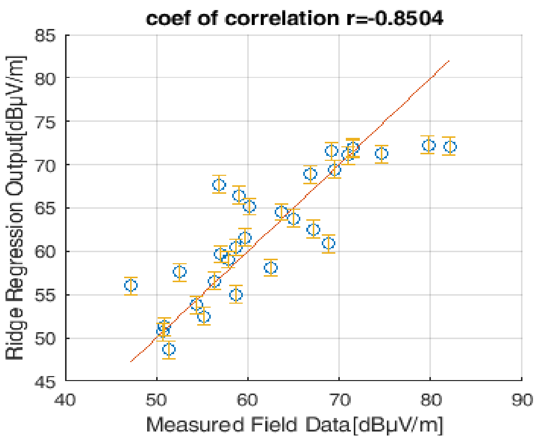

Figure 15 (b):Ridge correlation fitting of E- field strength data size of 30, for site 2

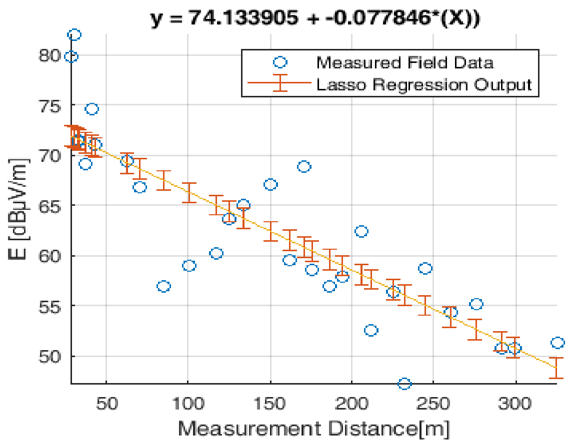

Figure 16 (a):Lasso regression fitting of E- field strength data size of 30, for site 2

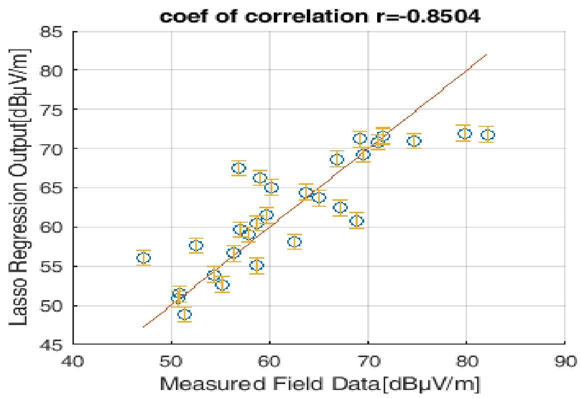

Figure 16 (b):Lasso correlation fitting of E- field strength data size of 30, for site 2

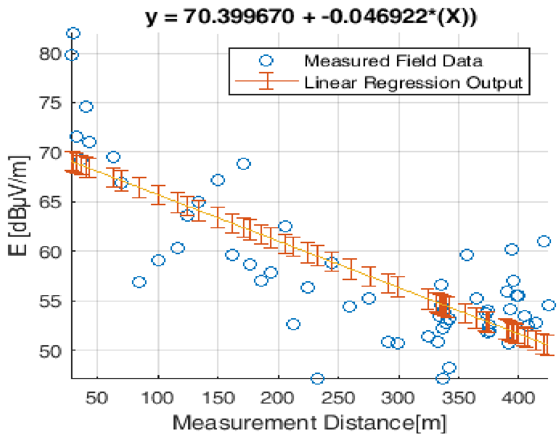

Figure 17 (a):Linear regression fitting of E- field strength data size of 60, for site 2

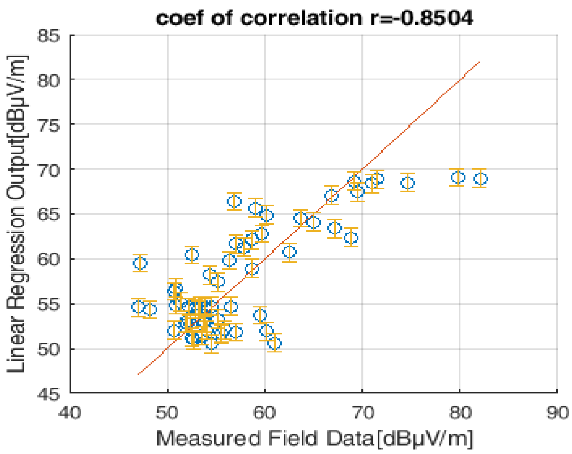

Figure 17 (b):Linear correlation fitting of E- field strength data size of 60, for site 2

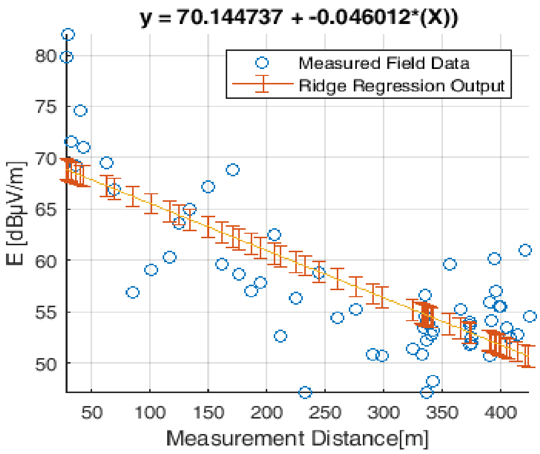

Figure 18 (a):Ridge regression fitting of E- field strength data size of 600, for site 2

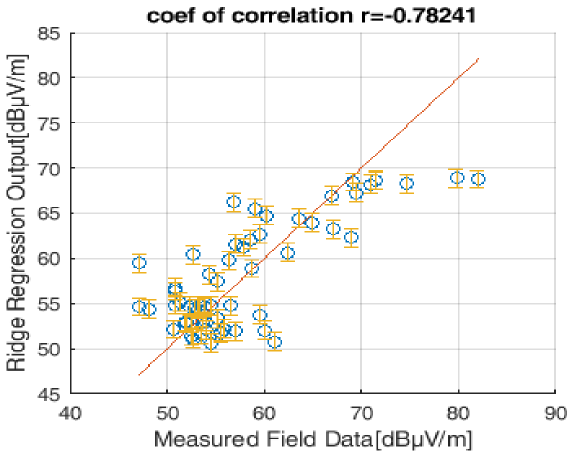

Figure 18 (b):Ridge correlation fitting of E- field strength data size of 60, for site 2

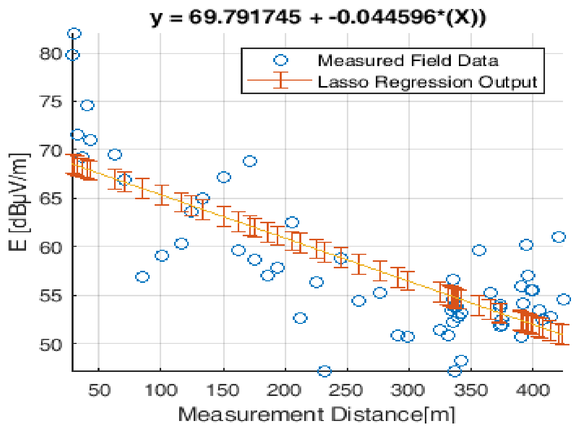

Figure 19 (a):Lasso regression fitting of E- field strength data size of 60, for site 2

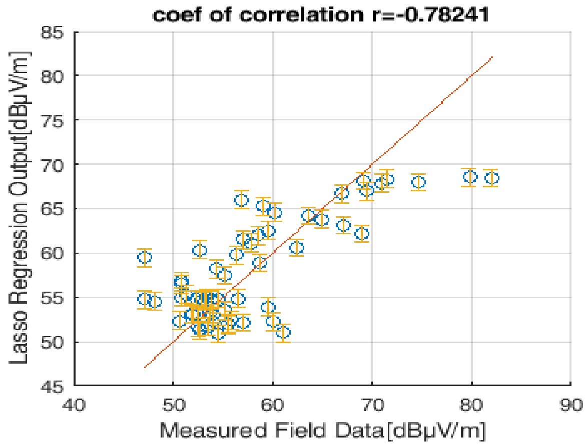

Figure 19 (b):Lasso correlation fitting of E- field strength data size of 60, for site 2

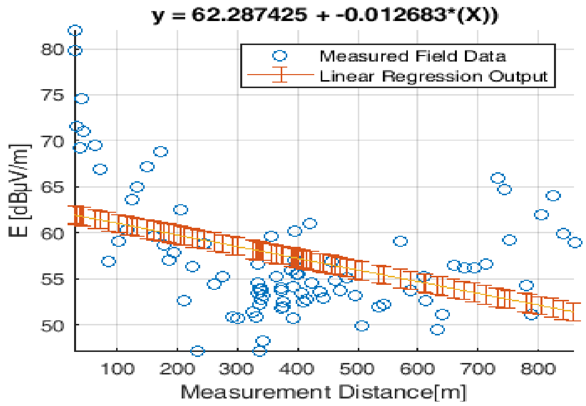

Figure 20 (a):Linear Regression fitting of E- field strength data size of 90, for site 2

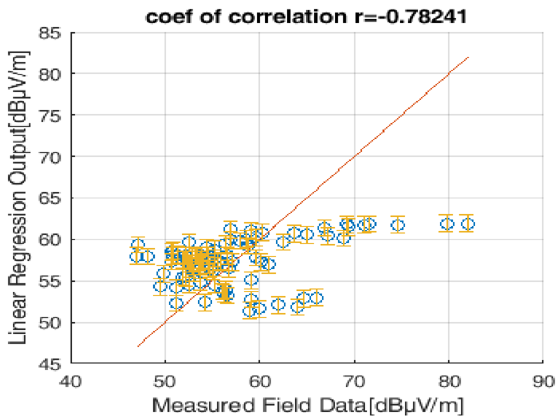

Figure 20 (b):Linear correlation fitting of E- field strength data size of 90, for site 2

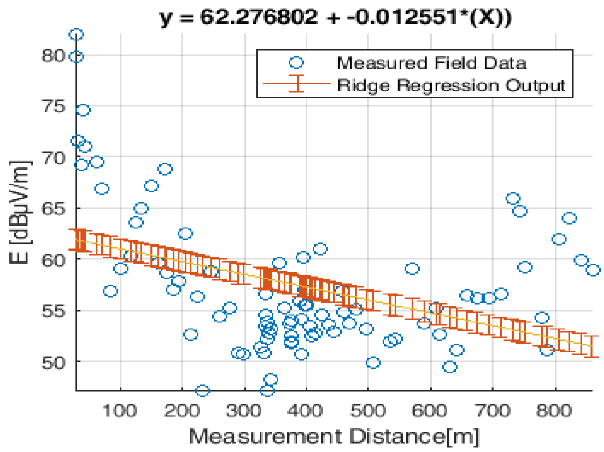

Figure 21 (a):Ridge regression fitting of E- field strength data size of 90, for site 2

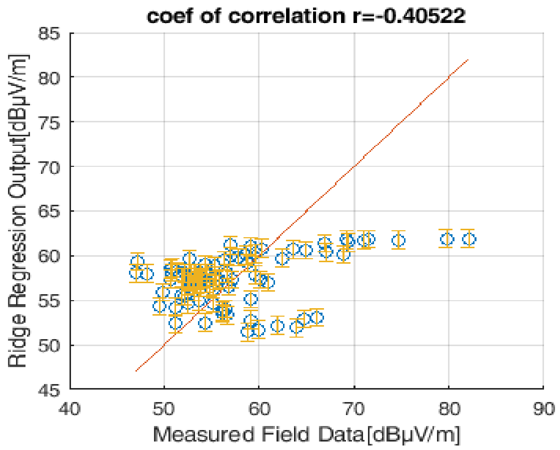

Figure 21 (b):Ridge correlation fitting of E- field strength data size of 90, for site 2

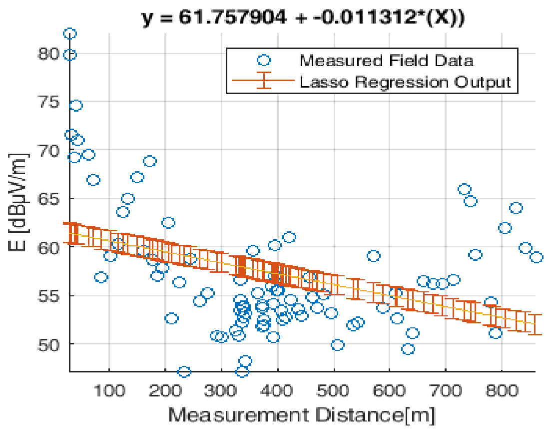

Figure 22 (a):Lasso Regression fitting of E- field strength data size of 90, for site 2

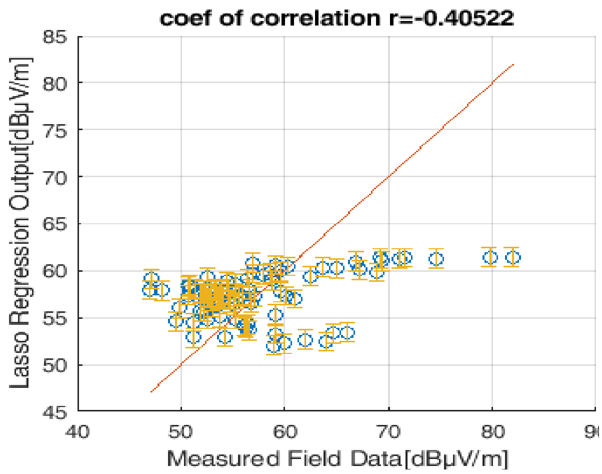

Figure 22 (b):Lasso correlation fitting of E- field strength data size of 90, for site 2

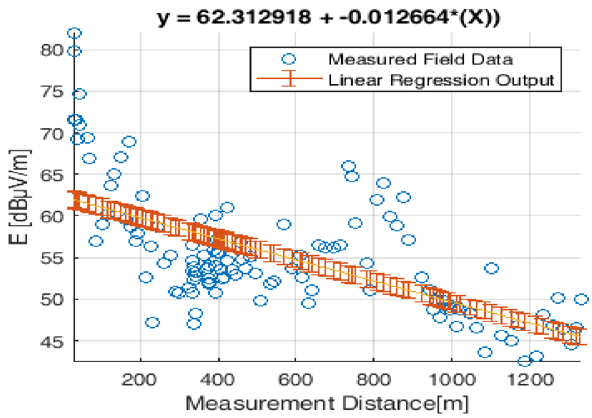

Figure 23 (a):Linear Regression fitting of E- field strength data size of 120, for site 2

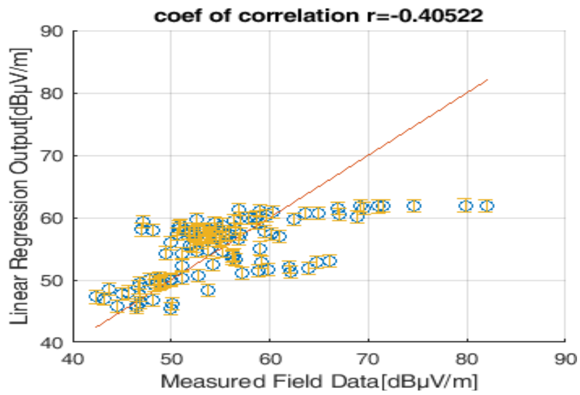

Figure 23 (a):Linear correlation fitting of E- field strength data size of 120, for site 2

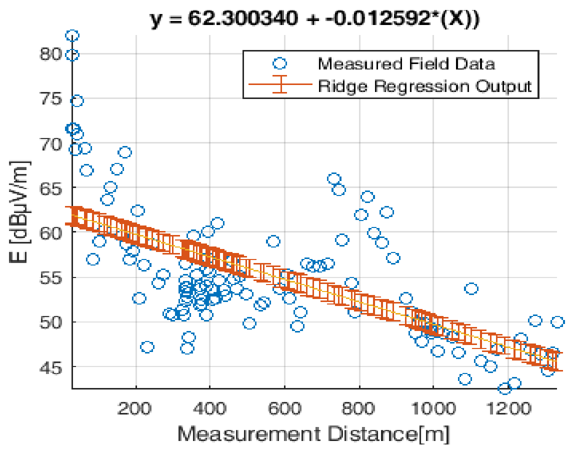

Figure 24 (a):Ridgecorrelation fitting of E- field strength data size of 120, for site 2

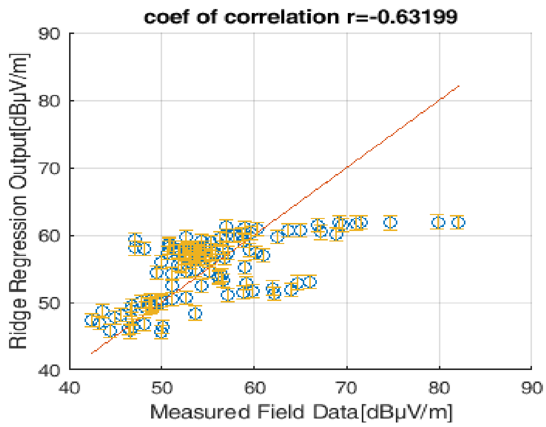

Figure 24 (b):Ridge correlation fitting of E- field strength data size of 120, for site 2

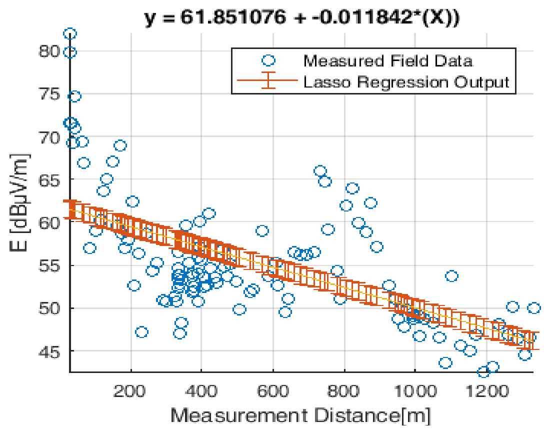

Figure 25 (a):Lasso Regression fitting of E- field strength data size of 120, for site 2

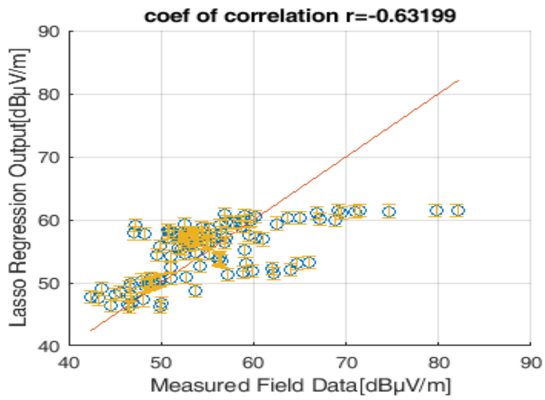

Figure 25 (b): Lasso correlation fitting of E- field strength data size of 120, for site 2

II. Conclusion

Many Regression methods have been applied to quantitatively study and estimate trends in field strength data set. This study focused on the application of linear, ridge and lasso regression modelling techniques to estimate measured field strength data variability trend. Specifically, the linear, ridge and lasso methods have been employed to model and estimate the measured field strength, ranging from 30 to 120 data sizes acquired from two transmitting eNodeB LTE cell sites in a typical urban environment. The estimation results revealed that the ridge regression methodrelatively outperform the lasso and ordinary least square regression model, especially at lower field strength data sizes.

References

[1]. T. Hastie, R. Tibshirani, and J Friedman. The elements of statistical learning: data mining, inference, and prediction. Springer, 2009. Second edition

[2]. Isabona, J, and Enagbonma, O(2014) A Least Absolute Deviation Tuning Method to reduce Signal Coverage Loss Prediction Error in Electromagnetic Wave Propagation Channel optimization, International Journal of Basic and Applied Sciences, Vol. 1, Issue 1, PP 44-55.

[3]. Bhaskaran K, Hajat S, Haines A, Herrett E, Wilkinson P, Smeeth L. Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry. BMJ 2010;341:c3823.

[4]. Nicole Krr, Anne-Laure Boulesteix, and Gerhard Tutz. Penalized partial least squares with applications to b-spline transformations and functional data. Chemometrics and Intelligent Laboratory Systems, 94:60–69, 2008.

[5]. Isabona, J, and Azi. S.O, (2012). ‘‘Optimised Walficsh-Bertoni Model for Pathloss Prediction in Urban Propagation Environment’’, International Journal of Engineering and Innovative Technology (IJEIT) Vol. 2, Issue 5, pp 14-20.

[6]. Isabona, J, Konyeha. C. C, Chinule. C. B, Isaiah G. P. (2013) ‘‘Radio Field Strength Propagation Data and Pathloss calculation Methods in UMTS Network’’, Advances in Physics Theories and Applications, vol.21.pp 54-68.

[7]. Isabona, J and Konyeha. C.C (2013) ‘‘Urban Area Path loss Propagation Prediction and Optimisation Using Hata Model at 800MHz’’, IOSR Journal of Applied Physics (IOSR-JAP), Vol. 3, Issue 4, pp.8-18.

[8]. Isabona, J and Isaiah. G.P (2013) ‘‘CDMA2000 Radio Measurements at 1.9GHz and Comparison of Propagation Models in Three Built-Up Cities of South-South, Nigeria’’, American Journal of Engineering Research (AJER), Vol. 2, Issue-05, pp-96-106.

[9]. Isabona, J, and Azi. S.O, (2013) ‘‘Enhanced Radio Signal Loss Prediction with Correction Factors for Urban Streets in the IMT-2000 Band’’, Elixir Space Science, vol. pp.15958-15962.

[10]. Isabona, J, and Babalola, M (2013) ‘‘Statistical Tuning of Walfisch-Bertoni Pathloss Model based on Building and Street Geometry Parameters in Built-up Terrains’’. American Journal of Physics and Applications, vol. 1, pp. 10-16.

[11]. Isabona, J and Azi, S (2013). ‘‘Measurement, Modeling and Analysis of Received Signal Strength at 800MHz and 1900MHz in Antenna Beam Tilt Cellular Environment’’, Elixir Computer Science and Engineering. 54 (2013) 12300-12303.www.elixirjournal.com.

[12]. Isabona, J, and K.Obahiagbon (2014) ‘‘RF Propagation Measurement and Modelling to Support Adept Planning of Outdoor Wireless Local Area Networks in 2.4 GHz Band,American Journal of Engineering Research, vol. 3, Issue-1, pp 58-67.

[13]. Stigler, Stephen M. (1981). "Gauss and the Invention of Least Squares". Ann. Stat. **9**(3): 465–474. doi:10.1214/aos/1176345451.

[14]. A. E. Hoerl and R.W. Kennard, Technometrics, Vol. 12, No. 1 (Feb., 1970), pp. 55-67.

[15]. R. Tibshirani. Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc B, 58:267–288, 1996.

[16]. Isabona, J (2020). Wavelet Generalized Regression Neural Network Approach for Robust Field StrengthPrediction in Open and Shadow urban Microcells, Wireless Personal Communications,Vol. 114 (3), pp.3635–3653

Ojuh, O. Divine. “Linear, Lasso and Ridge Regression Methods: A Quantitative Case Study with Spatial Electromagnetic Field Strength Signals.” *IOSR Journal of Applied Physics (IOSR-JAP)*, 13(1), 2021, pp. 27-38.