International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

SINGLE AND MULTIVARIATE REGRESSION ANALYSIS BY GRADIENT DESCENT METHOD ON SOIL RADON DATA IN BRAHMAPUTRA VALLEY OF ASSAM

Subir Sarkar*(1), H K Sarma (2)

(1)*Department of Physics, Arya Vidyapeeth College, Guwahati – 781016, Assam, INDIA.*

(2)*Principal, Suren Das College, Hajo-781102, Kamrup, Assam, INDIA.*

ABSTRACT

The radium content, radon concentration and radon exhalation rate from the soil has been carried out in some selected places in the Brahmaputra valley of Assam using standard formulae. The radium content in the soil samples varies from 1.55 Bqkg^{-1}to 5.9Bqkg^{-1}with an average value 3.504 Bqkg^{-1}. The radon concentration varies from 96.68 Bq.m^{-3} to 247.38Bq.m^{-3} with an average value 162.5 Bq.m^{-3}. The radon exhalation rate from the soil samples in terms of area varies from 35.55 mBqm^{-2}h^{-1}to 128.66mBqm^{-2}h^{-1}with an average value 77.22 mBqm^{-2}h^{-1}and the same parameter in terms of mass varies from 0.95 mBqkg^{-1}h^{-1}to 3.64mBqkg^{-1}h^{-1}with an average value 2.16 mBqkg^{-1}h^{-1}. In the present study, we have explored the interdependence between the parameters by standard procedure and made single and multivariate regression analysis by gradient descent method, an approach adopted in Artificial Neural Network (ANN) analysis. The ensemble of the observed data agrees very well with those predicted from Linear Regression (LR) model and Artificial Neural Network model (ANN) model. The hypothesis function makes good prediction which has been validated by observed test data.

Keywords: Radium, Radon, Radon Exhalation Rate, Gradient Descent, single and multiple variable regression.

1. INTRODUCTION

Radon is a colourless, tasteless and odourless radioactive gas originating from radioactive decay of Uranium, present in the earth cast. Its half-life (3.82 days) helps radon to travel significant distance in the medium where it is formed and there is a fair chance that radon concentration will enhance once it is inside a room. The inhalation of radon and its daughter products when present in enhanced level inside a dwellings can cause significant health hazard. The radon concentration in the ground depends on radium content and emanation power of the soils and rocks.

Radium is a decay product of Uranium found in solid form under ordinary conditions of temperature and pressure. The rate at which radon escapes from the soil to the surrounding air is known as radon exhalation rate of the soil. This may be measured either per unit area or per unit mass. The knowledge of radon exhalation rate from soil and building material can help us to understand the relative contribution of these components on indoor radon concentration and consequent health hazard.

Radon exhalation is greatly influenced by amount of radium content in the bed-rock and soil. Since radon poses potential health hazard, so the study of radium content in the soil and the radon exhalation rate from ground and building materials should be studied in all possible ways – theoretically, experimentally and possibly through data analysis.

Radon exhalation rate from soil is one of the most important factors for evaluation of indoor radon level. Solid state Nuclear Track Detectors (SSNTD) have been widely used for the study of different aspects of radon emission from soil. We have carried out a study of estimating radon exhalation rate from soil in some selected places of Brahmaputra valley of Assam. We have calculated radium content and radon concentration in the soil using standard formulae[1]. The radium content in the soil samples varies from 1.55 $Bqkg^{-1}$ to 5.9 $Bqkg^{-1}$ with an average value 3.504 $Bqkg^{-1}$. The radon concentration varies from 96.68 $Bq.m^{-3}$ to 247.38 $Bq.m^{-3}$ with an average value 162.5 $Bq.m^{-3}$. The radon exhalation rate from the soil samples in terms of area varies from 35.55 $mBqm^{-2}h^{-1}$ to 128.66 $mBqm^{-2}h^{-1}$ with an average value 77.22 $mBqm^{-2}h^{-1}$ and the same parameter in terms of mass varies from 0.95 $mBqkg^{-1}h^{-1}$ to 3.64 $mBqkg^{-1}h^{-1}$ with an average value 2.16 $mBqkg^{-1}h^{-1}$[2]. Similar study done in the Karbi Anglong district of Assam is found to be in the range 348.37±17.4 $mBqm^{-2}\ h^{-1}$ (10.52±0.52 $mBqkg^{-1}\ h^{-1}$) and1864.2±92.7 $mBqm^{-2}\ h^{-1}$ (56.29±2.8 $mBqkg^{-1}\ h^{-1}$) [3]. Similar work was reported from southern Sakarya, Turkey. Mass and areal radon exhalation rates in soil samples vary from 35.76 ± 1.5 to $253.15 \pm 3.8\,\mathrm{mBqkg^{-1}h^{-1}}$ with an average value of $112.53 \pm 2.7\,\mathrm{mBqkg^{-1}h^{-1}}$ and 0.73 ± 0.2 to $5.18 \pm 0.6\,\mathrm{Bqm^{-2}h^{-1}}$ with an average value of $2.30 \pm 0.6\,\mathrm{Bqm^{-2}h^{-1}}$ respectively. The effective radium content was found to vary in the range 3.77 ± 0.5 to $26.69 \pm 1.3\,\mathrm{Bqkg^{-1}}$ with an average value of $11.86 \pm 0.9\,\mathrm{Bqkg^{-1}}$. They have observed good correlation between radon exhalation and radium concentration in the soil [4].Another result from Turkey reports effective radium contents are found to vary from 6.66 to 34.32 $Bqkg^{-1}$ with a mean value of 18.01 $Bqkg^{-1}$. The radon exhalation rates measured in terms of mass and area of soil samples are found to vary from 50.35-259.41 $mBqkg^{-1}h^{-1}$ with a mean value of 136.12 $mBqkg^{-1}h^{-1}$ and 1035.18-5333.39 $mBqm^{-2}h^{-1}$ with a mean value of $mBqm^{-2}h^{-1}$[5]. The result from Amara at Maysan, Iraq reportsradon concentrations values varies from 53.18-2047.51 Bq m^{-3} with an average 776.98±126.08 Bq m^{-3}, while the radium concentrations values ranges from 2.58-99.61 Bq kg^{-1} with an average 37.79±6.13 Bq kg^{-1}. The average values of exhalation rates in term of area, the effective dose equivalent and the annual effective dose is reported as 40.08±6.50 μBq m^{-2} h, 3.635±0.59 WLM/y and 24.50±3.97 mSv/y, respectively and good positive correlation exists between radium content and radon concentration[6]. Radon exhalation rates from stone and soil samples of Aravali hills in India was carried out and radon concentration was to found to varybetween 729 Bq m^{-3} to 1958 Bq m^{-3} with an average of 1440 ± 134 Bq m^{-3} in stone samples whereas it was observed between 806 Bq m-3 to 1325 Bq m-3 with an average of 1040 ± 101 Bq m-3 in case of soil samples [7].

2. Motivation

First of all, it is observed that there is a good correlation either between radium content in the soil and radon concentration or between radium content and radon exhalation rate from soil. But what would be the combined effect of radium content in the soil and radon concentration on the radon exhalation rate has not been observed.

Secondly, in radon study the data obtained from different locations are analysed separately. If different locations are situated on the same geological and geographical set up, what would be the outcome of the analysis when data from different locations are taken together and analysed on a single platform.

An attempt to predict the radon exhalation rate from soil with the input features like – radium content in the soil and radon concentration, by regression method or by artificial neural network (ANN) will give a better chance to validate the computer code before using it in more complex situation.

In this paper, we will address these aspects by considering the data obtained from different locations togetherand try to investigate (a) the dependence of radon concentration on the basis of radium content in the soil and (b) the dependence of radon exhalation rate from the soil both in terms of mass and area on the combined effect of radium content and radon concentration.

We will develop a computer code with the help of PYTHON languagefor developing the above models by using linear regression method with gradient descent approach. This method appears to be fundamental step of Artificial Neural Network (ANN) analysiswhich is designed to simulate the way the human brain analyses and process information[8].

2. Methodology

In the present study we will make use of the data obtained from our previous study[2]. From soil radon study, the data on radium content, radon concentration and radon exhalation rate from soil both in terms of mass and area are available. In the first model, we will try to predict the radon concentration for any value of radium content in the soil and in the second model, radon exhalation rate from soil both in terms of area and mass will be predicted on the basis of input data – radium content and radon concentration. We will adopt linear regression method. It is the most common predictive analysis for finding the relationship between one dependent variable and one or more independent variables. But instead of solving normal equations for finding the parameters, we will use gradient descent method, a method generally used in the study of Neural Network. In this framework, with each input feature, we associate a weight called parameters. By convention, $x_0 = 1$ and w_0 is called "intercept" term in linear regression terminology and "bias" term in Neural Network approach. Schematically, the operation of linear regression is followed as shown in Figure 1. The detail description will be discussed in the following section.

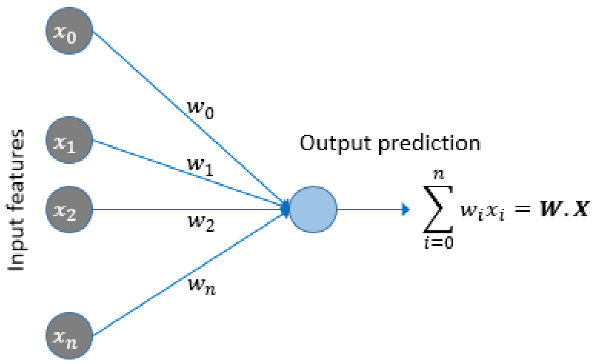


Figure 1-Model to predict output.

We will utilise the art of "supervised learning" in the present analysis. We shall also explain linear regression with gradient descent, followed in the present analysis.

2.1 Supervised Learning, Linear regression and Gradient descent

Let us start with defining "supervised learning". Given a set of variables $(\boldsymbol{x^{(i)}}, \boldsymbol{y^{(i)}})$***where,*** $\boldsymbol{i = 1, \ldots, n}$***; where*** $\boldsymbol{n}$

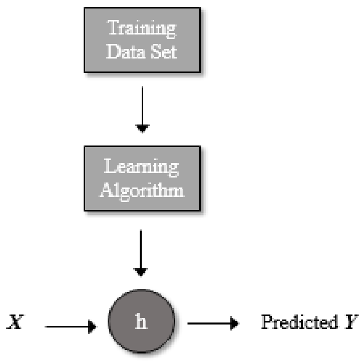


Figure 2- Model to predict output through hypothesis function.

stands for total number of dataset. Here, $x^{(i)}$ is the independent variable and $y^{(i)}$ is the dependent variable and there is no apparent functional relationship between them. If the said variables are correlated, how we can predict the values of *'y'* for a given values of *'x'*. The art of doing this practise is called "supervised learning". Here, $x^{(i)}$is called input variable or input feature and $y^{(i)}$ is called output variable or output feature. In supervised learning, we divide the whole dataset into two parts – training and testing.

When the target variable is continuous, the learning problem is called regression one. In regression problem under supervised learning, we begin with this assumption that a functional relation $h: X \to Y$ exists between the variables$x$ and y so that $h(x)$ is a good predictor for the corresponding value of y. This function h is called "hypothesis".

As an initial choice, we approximate a linear relation to exist between input and output variables so that we can consider a following hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \qquad (1)$$

Here, θ_i's are the parameters or weights parameterizing the space of linear function mapping from X to Y. To simplify the above expression, we consider $x_0 = 1$, called "intercept term" and write the above equation in the following way

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \boldsymbol{W^T X} \qquad (2)$$

Here, n is the number of input variables without considering x_0 and $\boldsymbol{W}$ and $\boldsymbol{X}$ are considered as the vectors containing weight and input features. Now, we have to estimate the parameters θ. For this we start with an "initial guess" for the values of θ so that $h(x)$ becomes close enough to the corresponding value of y at least for training data set. To carry out this process we define a function called "cost function" which measures how close the values of $h(x^{(i)})$ are to the corresponding values of $y^{(i)}$ for a particular values of θ's. We define "cost function" as follows:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \tag{3}$$

For estimating the values of θ which best suits the hypothesis function, we have to minimize $J(\theta)$. To do so, we start with an "initial guess " for θ and repeatedly changes it until $J(\theta)$ converges to a minimum value. We call this "Gradient Descent" approach. We use a search algorithm and repeatedly update θ:

$$\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta) \tag{4}$$

This update is made for all values of $j = 0, \dots, n$. Here, α is called "learning rate". Here, "learning rate" has to be carefully chosen since if it is too large, it will overshoot the minimum of the cost function else it will take too long time to reach to the point of convergence. This algorithm follows the direction of steepest decrease of J value.

To implement the algorithm, we have to first evaluate the partial derivative term on the right hand side of the eq.(4). For simplicity of mathematical operation, consider the case of a single training variable and we drop summation over J values. We have

$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{\partial}{\partial\theta_j}\frac{1}{2}\left(h_\theta(x) - y\right)^2$$

$$=\frac{1}{2}.2.\left(h_\theta(x) - y^{(i)}\right)\frac{\partial}{\partial\theta_j}\left(h_\theta(x) - y\right)$$

$$=\left(h_\theta(x) - y\right)\frac{\partial}{\partial\theta_j}\left(\sum_{i=1}^{m}\theta_i\, x_i - y^{(i)}\right)$$

$$=\left(h_\theta(x) - y\right)x_j$$

For a single training example, the update rule is

$$\theta_j := \theta_j + \alpha\left(y^{(i)} - h_\theta\left(x^{(i)}\right)\right)x_j^{(i)} \tag{5}$$

We follow the following steps for the minimisation of cost function and estimation of parameters.

Step (1) – We initialise the values of the parameters $\theta_0, \theta_1, \theta_2, \dots$ with some initial values.

Step (2) – Calculate the partial derivative of the cost function with respect to θs and update θ through equation (5). The direction of gradient gives the direction of steepest descent of cost function.

The process continues until the cost function becomes minimum. Further increase of parameters will increase the cost function again and we must stop updating θs before the cost function is further increased[9].

In this analysis, we have normalised the variables through the following relation

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{6}$$

IJARSE
ISSN: 2319-8354

Here, x is the original data and x_{norm} is the normalised data, x_{max} ,x_{min} are respectively minimum and maximum values of the data set. This is necessary because when variables are of different units and size,one variable can dominate over others causing imbalance in predictive analysis.

3. Our Studied areas

We conducted our study in some selected places of the Brahmaputra valley of Assam – Barpeta, Pathsala, Noonmati, Numaligarh and Duliajan. The global position and other details of the studied areas can be found elsewhere[2].

4. Soil Radon Study through Linear Regression with Gradient descent

4.1 Univariate Linear Regression model of Radium content and Radon concentration:

We consider radium content as the predictor variable (X) and radon concentration as the target variable (Y). We consider a mapping $f: X \rightarrow Y$ with a hypothesis function: $h(x) = \theta_0 + \theta_1 x_1$, where by convention $x_0 = 1$. We start with a training data set which is 80% of the whole data set, chosen randomly. We calculate cost function $J(\theta_0, \theta_1)$ given by equation (3) starting with an initial value of θ_0 and θ_1 chosen by a random seed and systematically update according to equation (5) over successive iteration until cost function converges to a minimum value. The result of the variation of the cost function for different values of θ_0 and θ_1 are shown inFigure 3.It is seen that only a particular combination of θ_0 and θ_1, cost function $J(\theta_0, \theta_1)$ will be minimum.

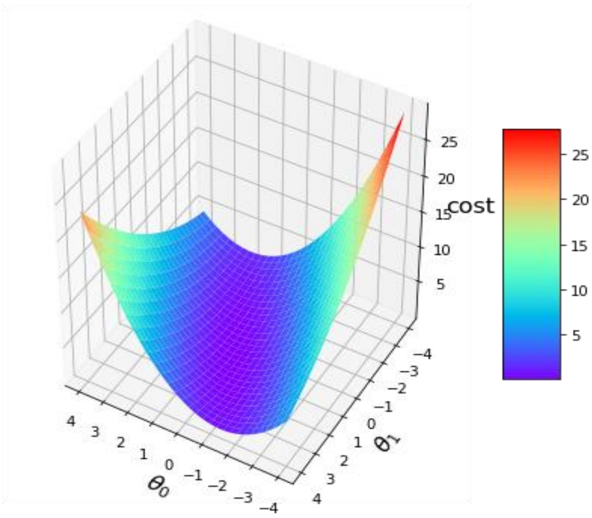


Figure 3- The variation of cost function with relative change of parameters.

We define loss parameter as difference between hypothesis function $h(x)$ for a particular value of θ_0 and θ_1 and the corresponding observed value of y. We start with an initial guess value for θ_0 and θ_1 chosen by a random seed and calculate loss for all input variables. We iterate the process over number of times and update θ_0 and θ_1 at each iteration until the loss function becomes minimum and fluctuation of θ_0 and θ_1 stabilises. This is shown

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

in the Figure 4. We have got the values of θ_0 and θ_1 as -0.0304 and 0.7229 respectively. So our model looks like

$$h(x) = -0.0304 + 0.7229x_1 \tag{7}$$

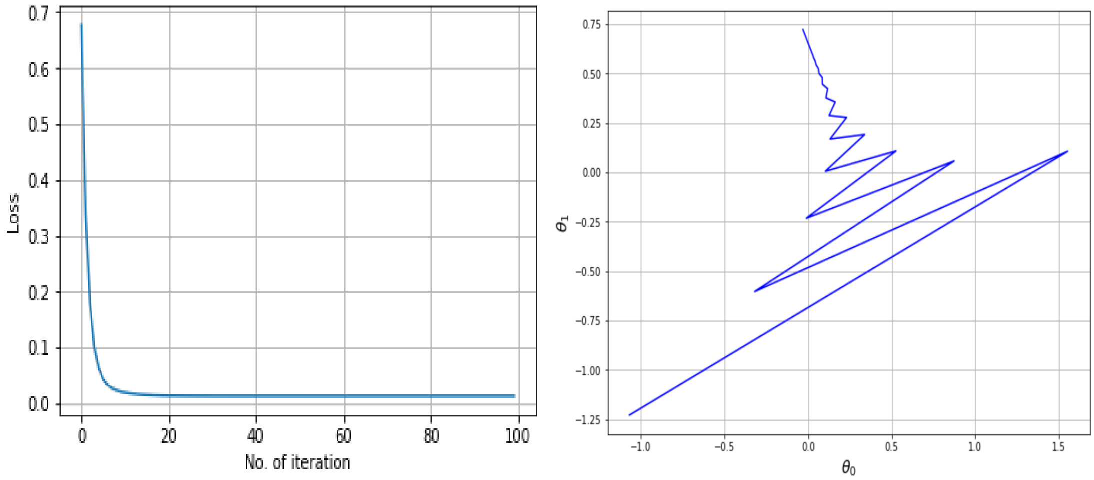


Figure 4-(a) The variation of Loss function with number of iterations, (b) The fluctuations of parameters at the beginning and final convergence to a stable value.

We have checked the model with our test data set and compared the value of $y$estimated by $h(x)$ with our actually observed data. The result is shown in Figure 5.

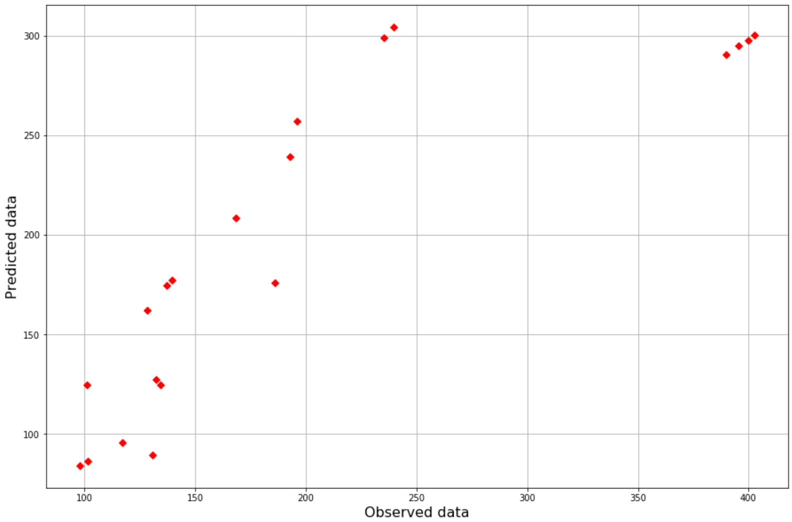


Figure 5- Examining observed test data set with the predicted data set.

 IJARSE
ISSN: 2319-8354

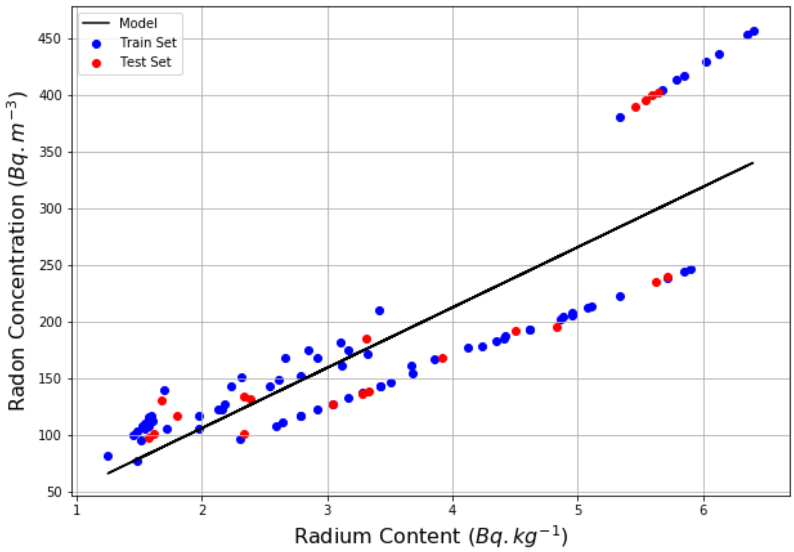


Figure 6- Comparing observed data (blue dot representing train data set and red dot representing test data set) with model fitted line.

In Figure 6, the black solid line represents the model fitted data points compared with the train data set shown by blue dot and test data set by red dot.The root mean square error, which measures the standard deviation of the difference between observed values and model predicted values, is found to be 51.1731.The r-squared value which is a measure of how close the observed data are to the fitted line is found to be 0.7229.The Pearson's correlation co-efficient (r) is found to be 0.85. The observed data shows three distinct patterns – one group is above the model fitted line and the other one below the line, both are highly correlated in appearance and in between these two, there a few slightly scattered group of data points lying over the fitted line.Whether radon concentration depends only on the radium content or there are other factors, as well, can be ascertained if similar analysis is performed for other locations as well. The apparent splitting of radium content and radon concentration level relationship may be the possible reason for higher value of root mean square error.

We have further used the observed data set to make neural network analysis with the help standard package called "Scikit learn v0.19.1".The programme "sklearn.neural_network.MLPRegressor" has been used with the following specifications - (hidden_layer_sizes=(500), activation='relu', solver='adam', learning_rate_init=0.001,early_stopping=True, max_iter=1000)[10]. The observed data has been compared with that obtained from Linear Regression (LR) model and from Artificial Neural Network (ANN) model. This has been show in

Figure 7.It is observed that ensemble of data sets – observed data, Linear Regression (LR) model predicted data and Artificial Neural Network (ANN) model predicted data match very well. The predicted data obtained from LR model and that obtained from ANN model almost overlap. Besides ensemble matching, there are a few points where there are some gap between predicted data and observed data. More data from those locations can give some insight of this disagreement. Unusually high value of radon concentration visible in Noonmati area has also been reflected in predicted models as well. It is to be noted that the places like Barpeta and Pathsala are non-industrial area whereas Noonmati, Numaligarh and Duliajan are industrial area where oil installations are there. It is seen that except Noonmati, the other two places – Numaligarh and Duliajan show almost same range

of radon concentrations as those of Barpeta and Pathsala. The higher value of radon concentration observed in Noonmati – minimum = 390.02 Bq.m^{-3}, maximum = 457.14 Bq.m^{-3} with an average value 418.29 Bq.m^{-3}, demands that the area needs to be investigated with more data covering wider area.It is observed that the range of radon concentration for Barpeta and Pathsala - minimum = 78 Bq.m^{-3}, Maximum = 380.95 Bq.m^{-3} with average value equal to 137.3 Bq.m^{-3}and that for Numaligarh and Duliajan - minimum = 96.68 Bq.m^{-3}, maximum = 247.38 Bq.m^{-3} with an average value 172.84 Bq.m^{-3}, indicates that the range is very close to each other for these places. It, thus, reveals that the category of places like industrial and non-industrial onesdo not have much significant impact over radon concentration.

It is also seen that predicted data obtained from LR analysis and ANN method, almost overlap. The PYTHON code used in the present study for Linear Regression analysis with gradient descent choosing random number as initial guess is very close to the Artificial Neural Network Analysis.

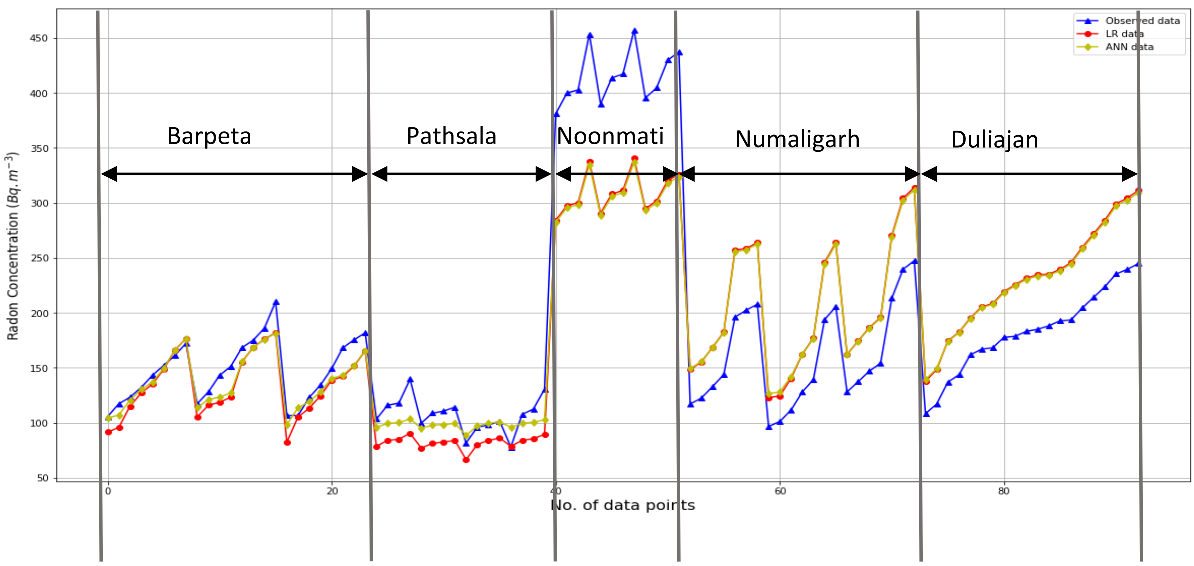


Figure 7- Comparing radon concentration observed in different locations with the predicted values estimated on the basis of Linear Regression (LR) model and that estimated on the basis of Artificial Neural Network (ANN) approach.

4.2Multivariate Linear Regression model to predict Radon exhalation rate:

Here, we choose radium content and radon concentration as the input variables and radon exhalation rate from soil as the target variable. So, we choose a linear mapping $f: X \rightarrow Y$ where X and Y are represented by $X = (x_0 \quad x_1 \quad x_2)$ and $Y = (y_0 \quad y_1 \quad y_2)$ with their respective weight given by $w = (w_0 \quad w_1 \quad w_2)$. Here, x_1 represents radium content and x_2 represents radon concentration. The weights or parameters w_1 and w_2 are associated with the variables x_1 and x_2 respectively. By convention, $x_0 = 1$ and w_0 is the intercept parameter. So our hypothesis function in this case is

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = \sum \theta_i x_i = \boldsymbol{w^T X},$$

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Here,***w*** and ***X*** represents weight and input feature matrices. So, like previous case,this time also, we divide the whole dataset in to training and testing part and run the code with training data set to calculate the parameters. We run the code to estimate model parameters for radon exhalation rate in terms of mass. We set the machine learning rate at 0.95 and iterate the code 100 times to look for convergence of the parameters. We see that loss function stabilises and corresponding parameters initially fluctuates and after several round of iterations of the code, fluctuations converges to a stable valueFigure 8. The values of the parameters are$\theta_0 = 0.0156, \theta_1 = 0.9864, \theta_2 = -0.0089$. With these values our hypothesis function looks like

$$h(x) = 0.0156 + 0.9864x_1 - 0.0089x_2 \qquad (8)$$

With this model, we have examined our test data set and compared with observed values. The result is shown inFigure 9.

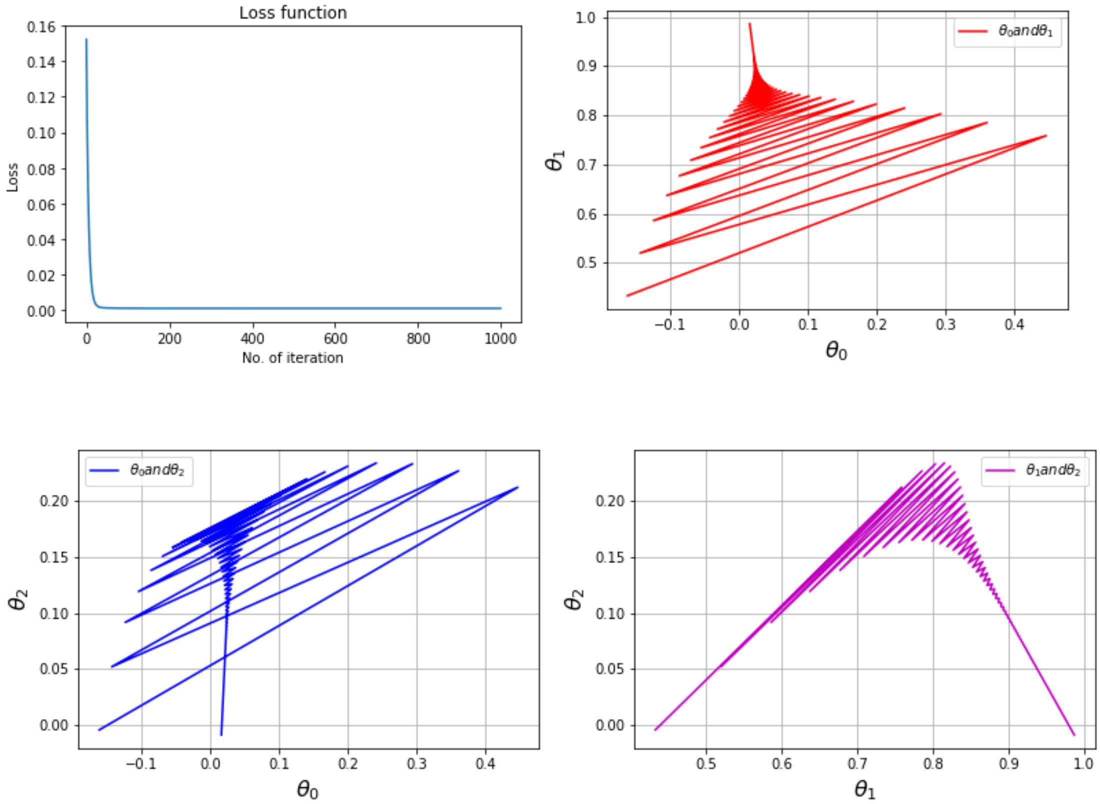


Figure 8- The variation of Loss function with the number of iterations and the initial fluctuations and final convergence of parameters to a stable value for determining radon exhalation rate in terms of mass.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

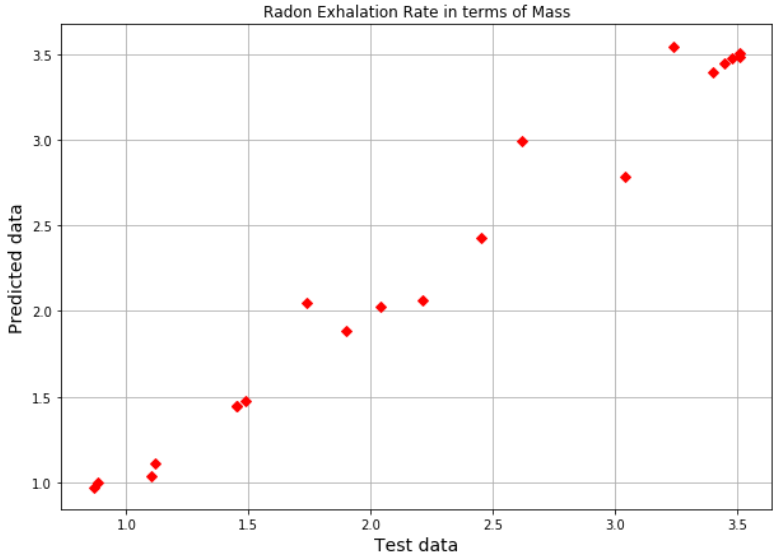


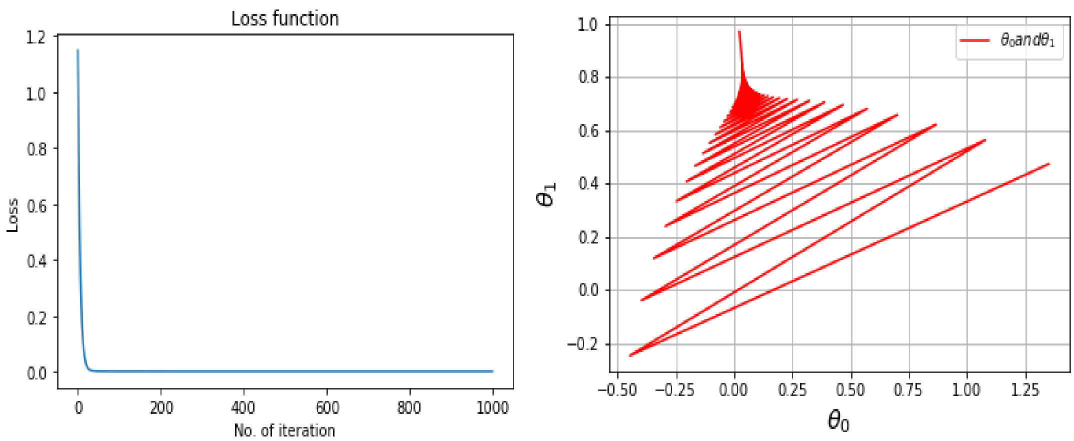
Figure 9- Examining the observed test data set with the predicted data set estimated from the model (Radon Exhalation rate in terms of mass).

It is seen that observed values are nearly equal to the predicted values. The root mean square error is 0.1231 and r-squared value is 0.5269.

Next we run the code to study radon exhalation rate in terms of area at the learning rate 0.95 and iteration 100. After several round of iterations, the loss function stabilises and the fluctuations of the parameters also converges to a stable value. The variation of loss function and fluctuation of the parameters are shown inFigure 10.The values of the parameters are$\theta_0 = 0.0242, \theta_1 = 0.9705, \theta_2 = -0.0044$. With these values our hypothesis function looks like

$$h(x) = 0.0242 + 0.9705x_1 - 0.0044x_2 \qquad (9)$$

With this hypothesis function, we have examined the test dataset with observed values and there is a good agreement between the two. The root mean square error is 6.4681 and r-squared value is 0.4995.



International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

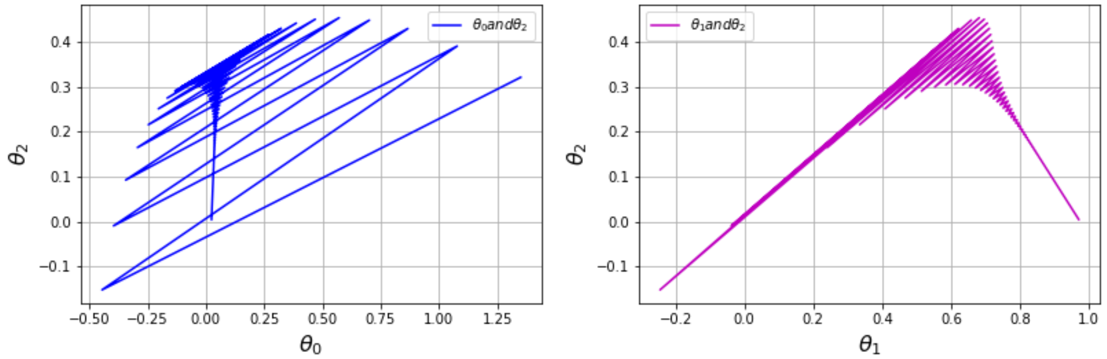


Figure 10- The variation of Loss function with the number of iterations and the initial fluctuations and final convergence of parameters to a stable value for determining radon exhalation rate in terms of area.

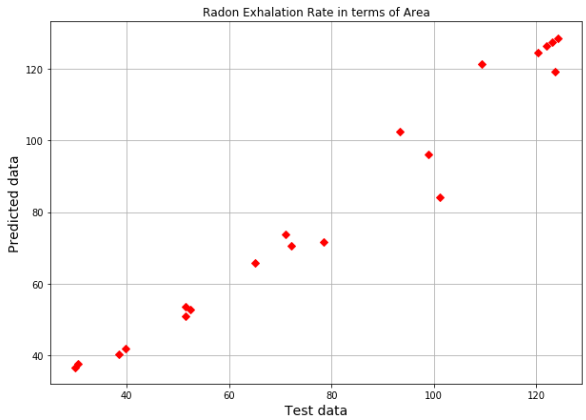


Figure 11- Examining the observed test data set with the predicted data set estimated from the model (Radon Exhalation rate in terms of area).

With the hypothesis functions represented by equations (8) and (9) for radon exhalation rate for area and mass, we have plotted a three dimensional surface which can be regarded as a model plane for estimating radon exhalation rate either in terms of area or in terms of mass when radium content and radon concentration are known. We have plotted the observed values of radon exhalation rate both for in terms of area and mass with reference to the radium content and radon concentration shown by blue triangle in both the graphs. We have also estimated radon exhalation rate both for area and mass with the given values of radium content and radon concentration using the hypothesis function represented by equations (8) and (9). The estimated values are shown by red dots in the same respective planes. The observed values and estimated values closely match each other and lie on either sides of the model plane. In this 3-Dimensioanl analysis also it is seen that for certain range of radium content, radon concentrations in the ground is different. The radium-radon relationship discussed in the last section has been reflected in the 2-D plane.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.08, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

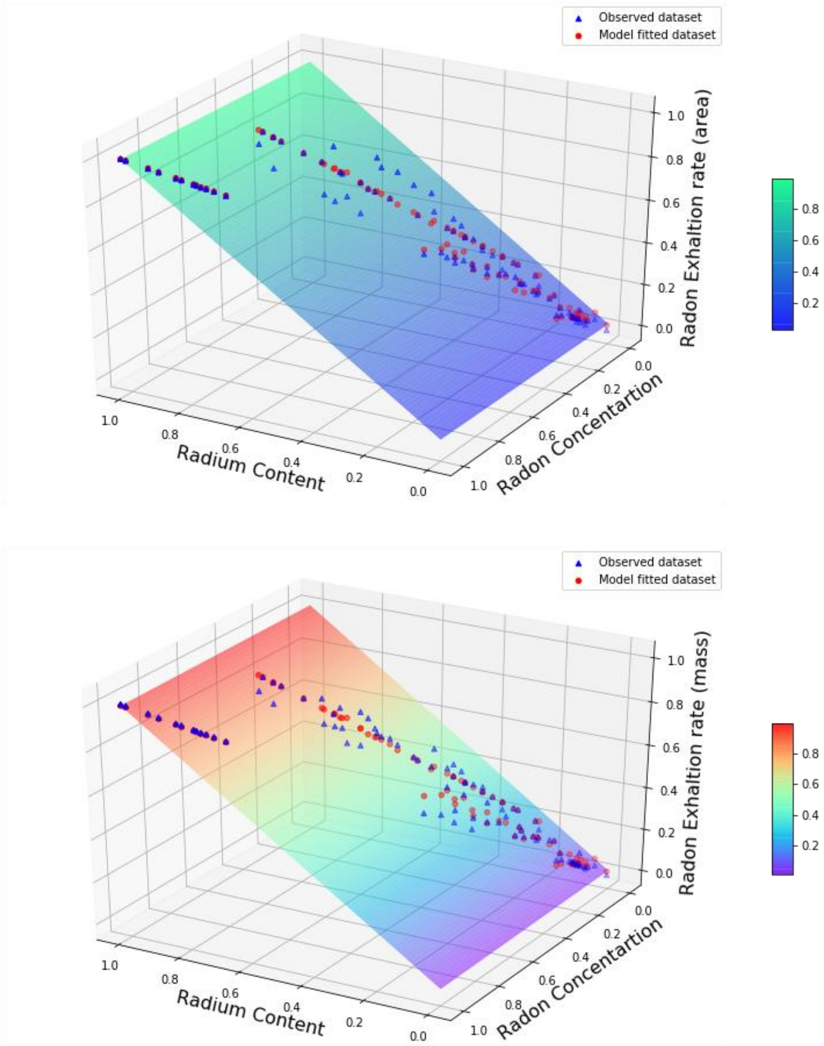


Figure 12- Model plane to represent Radon Exhalation rate in terms of (a) area and (b) mass. The observed data (blue triangle) and predicted data set on the basis of the model (red dot) are also plotted in the same space.

We have further used the observed data set to make neural network analysis with the help of standard package called "Scikit learn v0.19.1"[10]. The programme "sklearn.neural_network.MLPRegressor" has been used with the following specifications - (hidden_layer_sizes=(50), activation='relu', solver='adam', learning_rate_init=0.001,early_stopping=True, max_iter=1000)

The observed data has been compared with that obtained from Linear Regression (LR) model and from Artificial Neural Network (ANN) model. This is shown in Figure 13. It is observed that ensemble of data sets – observed data, Linear Regression (LR) model predicted data and Artificial Neural Network (ANN) model predicted data matches very well. The predicted data obtained from LR model and that obtained from ANN model almost overlap. Besides ensemble matching, there are a few points where there are some gap between predicted data and observed data. More data from those locations can give some insight of this disagreement.

5. CONCLUSION

It is thus seen that analysis of radon data together from different locations which stands on same geological and geographical set up appears to be a good idea. This can reveal unusual observations, present if any. The different values of radon concentrations at certainvalues of radium content is a clear evidence of the fact. More data in this range along with the critical consideration of geophysical condition of the soil may reveal some insight.

The 2D plane representing radon exhalation rate either in terms of area or mass gives a wider scope to estimate radon exhalation on the basis of radium content and radon concentration in the soil. More data from different locations with same geological and geographical set up is required to validate our model plane.

However, the computer code developed on the basis of Linear Regression with Gradient Descent has done a good job in predictive analysis which is comparable with Artificial Neural Network (ANN) approach.

6. ACKNOWLEDGEMENTS

One of the the authors Dr. Subir Sarkar, gracefully acknowledge the the valuable discussion and suggesstions put forward by Dr. Pradip Kumar Das, Faculty Member in the Department of Computer Science, IIT Guwahati. Thanks is due to his brilliant research team for valuable support.

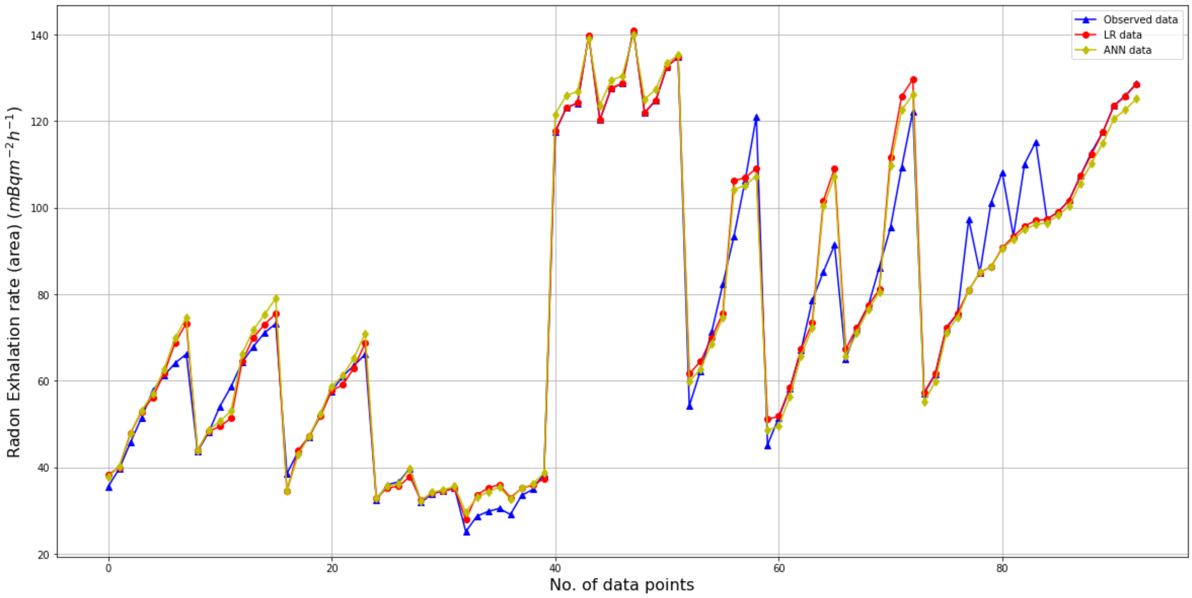


Fig 13 (a)

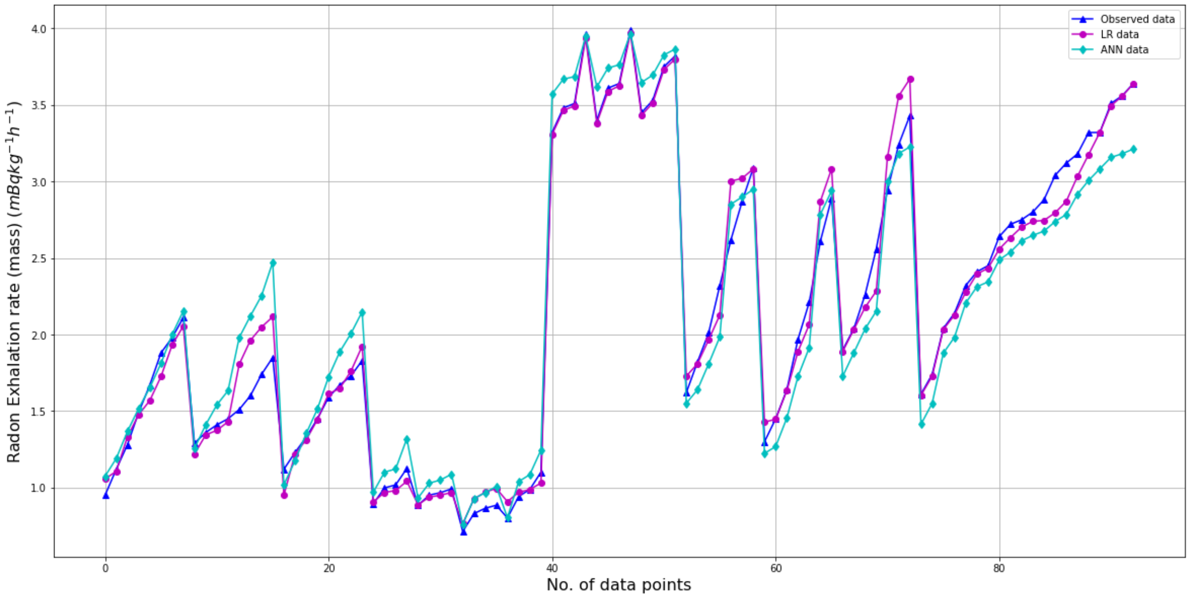


Figure 13- Comparison of Radon Exhalation rate in terms of (a) area and (b) mass observed in different locations are compared with that predicted from Linear Regression (LR) model and that estimated on the basis of Artificial Neural Network (ANN) approach.

REFERENCES

[1] Sarma H K, "Radon Activity and Radon Exhalation Rates From Some Soil Samples By Using SSNTD," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO 3297: 2007 Certified Organization), vol. Vol. 2, no. 10, 2013,.

[2] H. K. Sarma, "Radon estimation in different environments of Assam," http://shodhganga.inflibnet.ac.in:8080/jspui/handle/10603/115215, 2010.

[3] Kakati R K, "Radon Exhalation rate of soil and indoor radon concentration of various places of karbi Anglong District of Assam," IOSR Journal of Applied Physics (IOSR-JAP) e-ISSN: 2278 -4861, pp. 13-16, Vol. 6, Issue 4, Ver II (Jul-Aug. 2014).

[4] E. Tabar, H. Yakut and A. Kus, "Measurement of the radon exhalation rate and effective radium concentration in soil samples of southern Sakaraya, Turkey," Indoor and Built Environment, October 5, 2016.

[5] A. Kus, H. Yakut and E. Tabar, "Radon exhalation rates and effective radium contents of the soil samples in Adapazari, Turkey.," in AIP Conference Proceedings 1722, 030009, 2016.

[6] A. A. Abojassim, S. H. Kadhimand, A. H. A. Ali and A. A. Hashim, "Radon, radium Concentrations and radiological Parameters in Soil Samples of Amara at Maysan, Iraq," Asian Journal of Earth Sciences ISSN 1819-1886 DOI:10.3923/ajes.2017.44.49, October 21, 2017.

[7] Chauhan R P, "Radon exhalation rates from stone and soil samples of Aravali hills in India," Iran. J. Res., pp. 57-61, 2011; 9(1).

[8] "https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp," [Online].

[9] Ng. Andrew. [Online]. Available: http://cs229.stanford.edu/notes/cs229-notes1.pdf.

[10] Pedregosa et al.;. e. al., Scikit-learn: Machine Learning in Python,, [Online]. Available: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.