Genetic Algorithm and Statistical Applications in Mines for Radiation Safety Requirements

Ghada I. El-shanshoury[1] Eman Sarwat[2]

1. Nuclear and Radiological Regulatory Authority (NRRA), PO box 7551, Nasr City 11762, Egypt
2. Nuclear and Radiological Regulatory Authority (NRRA), PO box 7551, Nasr City 11762, Egypt

Abstract

Genetic algorithm and statistical probability distributions can give a good result for estimating the radon airborne levels into underground mines. The computer-aided algorithm for a regional mines controlling plan is presented. The mines are modeled and analyzed with the use of genetic algorithm and the total population will be distributed rationally according to the result to reach optimal values. Thus, offering an effective approach for regional radon condition improvement and pollutants control. Probability distributions are used for reducing the error rate of the radon prediction model. This is done by developing and converting the multiple regression model to probability multiple regression model using Cumulative Distribution Function (CDF) of suitable probability distributions. The CDF is used to convert the actual values to probability values for creating the probability model. Then the predicted probability values are converted to the original values using the inverse CDF (quantile function). The optimal results obtained from Genetic Algorithm have been used in the probability multiple regression model for estimating the radon levels in the entire mines. Accuracy measurements are calculated to evaluate the two investigated models. The results show that the probability multiple regression model diminishes the error rate nearly by 50% to 70%. The results give accurate prediction for determining the radon levels in mines.

Keywords: Radon level, genetic algorithm, multiple regression model, probability distributions

1. Introduction

One of the occupational risks of mining result from the exposure of miners to airborne radioactive gases radon 222 (^{222}Rn), thoron 220 (^{220}Rn) and their short-lived decay products (alpha emission). The inhalation of these radionuclides constitutes the most important occupational exposure in mines, especially in uranium mines. Exposure of miners to high concentration of radon and radon decay products has been correlated with the induction of lung cancer in several mining groups. Miners' deaths probably attributable to the inhalation of radon and radon decay products are recorded as far back as the sixteen century (Pergamon press 1986). A radon-222 gas concentration is commonly measured in picoCuries per liter of air (pCi/L). Therefore, one pCi/L represents the concentration of radon 222 atoms per liter of air that will result in 2.22 alpha emissions per minute. For health effects are primarily due to the Radon Decay Products (RDPs) and not to the radon-222 gas itself, a unique unit of measure exists for quantifying the amount of RDPs in the air. This unit of measure is the Working Level (WL) and was previously used to measure the occupational exposure of underground miners. The higher the WL, the higher the risk of adverse health effects (EPA 2003**.** The atmosphere underground is limited and confined, and can quickly become sub-standard or dangerous if contaminants are not controlled, extracted, or diluted to harmless levels. In the case of an underground mining operation or tunneling operation, the ventilation system should be designed so that it is possible to maintain a healthy and safe atmosphere underground at all times. Ventilation should provide adequate quantities of fresh air to mine workers. The ventilation system design should consider the physical parameters of the airways, the layout of the mine, and the hazards likely to be encountered underground (Work SAFE 2015).

Under the Egyptian program for radiation safety and control, airborne radioactivity measurements were carried out of airborne radon and thoron gases from mine walls, ceilings and floors (Ashraf et al. 2003). Radioactivity measurements have been conducted in many underground phosphate mines in Egypt. In this work, three Egyptian phosphate mines namely: Zogel Bohar mine, West Yonus mine and Safaga South mine are investigated. Each mine has one opening entrance.

Genetic Algorithm (GA) is often used for optimization problems in which the evolution of a population is a search for a satisfactory solution given a set of constraints.

In this paper, Cumulative Distribution Functions (CDF) describes the probability that a real-valued random variable *X* with a given probability distribution will be found to have a value less than or equal to *x* (Cumulative 2015). Five probability distributions are applied. The application of artificial intelligence techniques in optimization of radioactive material in mines is a promising area of research across the improvement of techniques that will result in significant time savings. The aim of this paper is estimating the radon airborne levels in a manner of which can get a good result closer to actual radon airborne levels into the underground mines. This target is conducted by using GA and CDF. The GA and CDF are applied to get the optimal values at the beginning and end of mines, and then reduce the error rate of radon prediction model. The paper is structured in the following way: Section 2 describes the methodology used including the model description, basic structure of genetic algorithm, probability distributions and models evaluation; section 3 represents results and discussion and finally Section 4 describes the conclusions.

2. Methodology

The collected data of radon and thoron from the mines under study lacked to the values at the beginning and the end of mines. Therefore, GA program is used for getting the optimal values of these missing or unmeasured points. The collected data of radon, thoron, ventilation and distances is considered to be as ascending or descending.

The main goal of the research is reduce the rates of errors, caused by the radon values that are predictable from multiple regression model, by a large margin. This is done by converting the multiple regression model to probability multiple regression model using the suitable probability distributions. The creation of probability multiple regression model based on converting the actual and genetic optimization values to probability values using CDF. This is followed by converting the probability values of the prediction model to the original values using the inverse CDF (quantile function).

2.1 Model Description

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data (Stat.yale. 2015).

Multiple regression model is applied for predicting Radon levels Rn (response variable) in the investigated mines. This model is based on three independent variables (predictor variable). These variables are Thoron levels (Th), Required Ventilation (V) and Distances (D). The Rn and Th levels are measured at different distances. The amount of air (V), which must be added to lower radon daughter in a given environment, is calculated at the volume of uncontaminated air of 5 m^3/s.

The multiple regression model takes the form:

$$R_n = b_0 + (b_1 * Th) + (b_2 * V) – (b_3 * D)$$

Where, b_0 is the regression intercept while b_1, b_2 and b_3 are regression coefficients. The Rn and Th measurements are calculated in WL. V is calculated in m^3/s and D is taken in meter (m).

The probability multiple linear regression model takes the form:

$$F(R_n) = b_0 + (b_1 * F(Th)) + (b_2 * F(V)) – (b_3 * F(D))$$

2.2 Genetic Algorithm

A GA begins with estimations and attempts to improve the estimations by evolution. A GA consists of five main parts: a representation of a guess called a chromosome, an initial pool of chromosomes, a fitness function, a selection function and a crossover operator and a mutation operator. A chromosome can be a binary string or a more complex data structure. The initial pool of chromosomes can be randomly produced or manually created. The fitness function measures the suitability of a chromosome to meet a specified objective: a chromosome is fitter if it corresponds to greater coverage. The selection function decides which chromosomes will participate in the evolution stage of the genetic algorithm made up by the crossover and mutation operators. The crossover operator exchanges genes from two chromosomes and creates two new chromosomes. The mutation operator changes a gene in a chromosome and creates one new chromosome (Praveen 2009). A basic algorithm for a GA is as follows:

Initialize (population)

Evaluate (population)

While (stopping condition not satisfied) do

{Selection (population)

Crossover (population)

Mutate (population)

Evaluate (population)}

The algorithm will iterate until the population has evolved to form a solution to the problem, or until a maximum number of iterations have taken place (Chou 2008).

2.3 Some Statistical Notation

F(x) = Cumulative Distribution Function (CDF).

x(F) = Inverse CDF (quantile function).

Φ(.) = CDF of standard normal distribution.

Φ^{-1}(.) = Inverse CDF of standard normal distribution.

x_i = observed value (actual value).

$\hat{x}_i$ = modeled value (predicted value).

$\overline{x}$ = the mean of observed value.

n = number of values.

2.4 Probability Distributions

Table 1. The CDF, parameters estimation and inverse CDF

Exponential distribution (1 parameter)		
CDF	Parameters estimation (MOM)	*x(F)*
$F(x)=1-\exp(-\lambda x)$	$\lambda=\frac{1}{\overline{x}}=\frac{n}{\sum_{i=1}^{n}x_i}$ where, λ is the inverse scale parameter. $\lambda > 0$ and $0\le x<+\infty$.	$x(F)=\left(-\ln(1-F(x))\right)/\lambda$
Gumbel minimum distribution (2 parameters)		
CDF	Parameters estimation (MOM)	*x(F)*
$F(x)=1-\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$	$\sigma=\frac{s\sqrt{6}}{\pi}$, $\mu=\overline{x}+\gamma\sigma$ where, σ & μ are scale and location parameters, respectively. $\sigma > 0$ and $-\infty<x<+\infty$. $\overline{x}$ & s are mean and standard deviation of the measured values respectively. $\gamma\cong 0.577216$ is Euler's constant.	$x(F)=\mu+\sigma\left(\ln(-\ln(1-F(x)))\right)$
Frechet distribution (2 parameters)		
CDF	Parameters estimation (LSM)	*x(F)*
$F(x)=\exp\left(-\left(\frac{\beta}{x}\right)^{\alpha}\right)$	$\alpha=\frac{\sum_{i=1}^{n}y\ln x_i-n\overline{y}\,\overline{\ln x}}{\sum_{i=1}^{n}(\ln x_i)^2-n\left(\overline{\ln x}\right)^2}$, $\beta=\exp\left(\overline{\ln x}-\frac{\overline{y}}{\alpha}\right)$	$x(F)=\beta\left(-(\ln(F(x))^{-1/\alpha}\right)$

	$\overline{\ln x} = \frac{\sum_{i=1}^{n} \ln x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, $y = -\ln(-\ln(F_i)$, $F_i = \frac{i}{N}$ where, β & α are scale and shape parameters, respectively. $\beta > 0$, $\alpha > 0$ and $0 < x < +\infty$.	
Exponential distribution (2 parameters)		
CDF	Parameters estimation (MLE)	*x(F)*
$F(x) = 1 - \exp(-\lambda(x-\eta))$	$\eta = mini\ x_i$, $\lambda = \frac{n}{\sum_{i=1}^{n}(x_i - \eta)}$ where, λ and η are the inverse scale and location parameters, respectively. $\lambda > 0$ and $\eta \le x < +\infty$.	$x(F) = \frac{(-\ln(1-F(x)))}{\lambda} + \eta$
Lognormal distribution (3 parameters)		
CDF	Parameters estimation (MLE)	*x(F)*
$F(x) = \Phi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)$	γ is the lower limit of lognormal distribution. $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_i - \hat{\gamma})$, $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\ln(x_i - \hat{\gamma}) - \mu\right)^2}$ where, σ is scale parameter, μ is shape parameter and γ is the location. $\sigma > 0$ and $\gamma < x < +\infty$.	$x(F) = \exp((\sigma * \Phi^{-1}(F(x)) - \mu) + \gamma$
Johnson SB distribution (4 parameters)		
CDF	Parameters estimation (MOM)	*x(F)*
$F(x) = \Phi\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)$ where, $z = \frac{x-\xi}{\lambda}$	The parameters estimated by Easy-Fit program. The model is characterized by the location parameter ξ, the scale parameter λ and the shape parameters γ and δ (asymmetry and kurtosis parameters, respectively). $\lambda > 0$, $\delta > 0$ and $\xi \le x \le \xi + \lambda$.	$x(F) = \frac{\lambda \exp\left(\frac{\Phi^{-1}(F(x)) - \gamma}{\delta}\right)}{1 + \exp\left(\frac{\Phi^{-1}(F(x)) - \gamma}{\delta}\right)} + \xi$

Probability distributions and methods of estimating its parameters have been selected through Easy-Fit program. The choice of the probability distributions has been in accordance with x^2 test.

Three methods are applied for estimating the parameters of probability distributions. These methods are: Method of Moments (MOM), Least Square Method (LSM) and Maximum Likelihood Estimation (MLE). The accurate results depend on selecting the suitable distribution. The CDF, parameters estimation and inverse CDF of the applied distributions are shown in Table 1 (Exponential 2015 - Johnson SB 2015).

2.5 Models Evaluation

Some of model accuracy is calculated as the goodness-of-fit criteria to judge the validity of the model to represent the observed data and to assess the results between actual and modeled data. The compared modeled data are multiple regression model and probability multiple regression model. The model accuracy is estimated using Root Mean Square Error (*RMSE*), Relative Root Mean Square Error (*RRMSE*), Root Mean Square Percentage Error (*RMSPE*), Mean Absolute Percentage Error (*MAPE*) as an error-measure and Coefficient of Determination (R^2) as a measure of goodness-of-fit.

The RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. Basically, the RMSE represents the sample standard deviation of the differences between predicted values and observed values (Rob 2006).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left((x_i - \hat{x}_i\right)^2}{n}} \tag{1}$$

The RRMSE may vary from 0 to positive infinity. The smaller the *RRMSE* is, the better the model performance. Sometimes it is expressed as percentage form (Ali 2014):

$$RRMSE = \frac{RMSE}{\overline{x}} * 100 \tag{2}$$

The measures based on percentage errors are the most common in forecasting domain. The group of percentage based errors includes the following errors (Maxim et al. 2013):

$$RMSPE = \sqrt{\underset{i=1,n}{mean}\left(100 * \left(\frac{x_i - \hat{x}_i}{x_i}\right)\right)^2} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} 100 * \left|\frac{x_i - \hat{x}_i}{x_i}\right| = \underset{i=1,n}{mean}\left(100 * \left|\frac{x_i - x_i}{x_i}\right|\right) \tag{4}$$

The R^2 is a number that indicates how well data fit a statistical model. It provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model (Draper 1998).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad 0 \leq x_i \leq 1 \tag{5}$$

Evaluation of the results is based on the lowest values of RMSE, RRMSE, RMSPE, MAPE and the highest value of R^2.

3. Results and Discussion

The sample reading of radon and thoron levels was measured at different distances in three Egyptian phosphate mines and the required ventilation are calculated. The study is based on intromission of the measured levels of R, Th, V and D into statistical minitab computer program for composition of the multiple regression model used for estimating the Rn levels. Rn is inserted as dependant variable and Th , V and D are inserted as independent variables. The multiple regression model is evaluated using some of the model accuracy to assess the results between actual data (observed values of Rn) and modeled data (predicted values Rn).

Radon value in three mines, Zogel Bohar Mine, West Yonus Mine and Safaga South Mine as shown in cases 1, 2 & 3 respectively, is optimized using GA. The objective function value *Rn* shown in (6, 8 & 10) are used as the fitness function, and the crossover probability and mutation probability values are both set to be 0.1. There are three variables (Th, V & D) in the equations, and each of them should be determined. The GA is a heuristic random calculating operation, and this feature requires repeated calculation in Matlab to seek the optimal solution. A relay evolutionary GA is used and a good result will be kept and used as initial population in the next evolution iteration, thus accelerating the calculation process. Finally, values of the three variables (Th, V & D) which corresponds to the optimal values of Rn are obtained in Tables (2, 7 & 12).

At attempt to reduce the prediction error is made by calculating the probability multiple regression model through converting Rn, Th, V and D values to probability values using CDF of probability distributions. The converted probability values, F(Rn), F(Th) , F(V) and F(D,) has been inserted into statistical minitab computer program for composition of the probability multiple regression model used for estimating the F(Rn) values. Consequantly, the inverse CDF is used to convert the probability values of predicted F(Rn) to the actual or original values. The probability multiple regression model is evaluated using the same model accuracy to assess the results between actual data and modeled data of Rn. To return to the actual values of F(Th) , F(V) and F(D), the quantile function of each variable (x(F)) is used.

Case 1: Zogel Bohar mine

Equation 6, which is conducted from the observed data, represents the multiple regression model. This equation is used for prediction of Rn values and for getting the optimized radon values and the corresponding predictors variables.

$$\hat{Rn} = -0.07653 + (0.02598 * Th) + (0.06557 * V) + (0.0000065 * D) \qquad (6)$$

Table 2. Optimized radon values and the corresponding variables in Zogel Bohar mine

Optimized Value	Radon Results	Variables		
	Rn	Th	V	D
Min Value	0.0302	0.013	1.6212	15
Max Value	0.1912	0.088	4.0066	420

The minimum and maximum objective function values (*Rn*) and the corresponding variables are shown in Table 2. The optimized values indicate that all values in the mine fall between these maximum and minimum values.

Table 3**.** Model accuracy measurements for multiple regression model

RMSE	RRMSE	RMSPE	MAPE	R^2
0.004	3.685	4.596	3.395	0.990

Table 3 gives the results of model accuracy which represented in RMSE, RRMSE, RMSPE, MAPE and R^2. These measurements are calculated from equations 1-5. The model accuracy measurements are calculated to assess the results between actual data and modeled data (The modeled data are calculated according to equation 6).

Table 4**.** The parameters estimation of probability distribution of Zogel Bohar mine

Distribution	Frechet 2 parameters	Gumbel minimum 2 parameters	Exponential 2 parameters	Johnson SB 4 parameters
Parameters estimation	Rn	Th	V	D
α	1.8927	-	-	-
β	0.06521	-	-	-
μ	-	0.06061	-	-
σ	-	0.01605	-	-
λ	-	-	0.95559	-
η	-	-	1.6212	-
γ	-	-	-	0.02633
δ	-	-	-	0.47932
λ	-	-	-	456.87
ξ	-	-	-	11.1

For generating the probability multiple regression equation, converting the actual and optimization values to probability values is necessary. This conversion is done by using CDF of suitable selecting distribution. In Zogel

Bohar mine Frechet, Gumbel minimum, Exponential and Johnson SB distributions are selected for calculating the probability values of Rn, Th, V and D, respectively. Table 4 shows the parameters estimation of each distribution which has been selected. By substitute these estimators in the CDF of its corresponding distribution, the probability values of each variable is obtained. Easy fit computer program is used to estimate the parameters.

Table 5. The minimum and maximum values of the probabilities

Probability values	F(Rn)	F(Th)	F(V)	F(D)
Minimum value	0.01367	0.05019	0.0000	0.012137
Maximum value	0.8776	0.9959	0.8977	0.8539

Table 5 gives the probability (CDF) minimum and maximum values corresponding to the optimization minimum and maximum radon values and other predictor variables in Table 2. All probability values of each variable in the mine fall between the following minimum and maximum values.

Equation 7, which is conducted from the probability values of observed data, represents the probability multiple regression model.

$$\hat{F}(Rn) = 0.007644 + (0.00305*F(Th)) + (0.99305*F(V)) - (0.016066*F(D)) \quad (7)$$

Equation 7 used for the prediction of the probability values *of* Rn by choosing any probability value between the range of maximum and minimum value of each variable. Substituting these values in equation 7, the probability of Rn is resulted (F(Rn)). For example: choosing any probability value of the predictors such as F(Th) of 0.3896, *F(V)* of 0.4483 and F(D) of 0.2567, the result of F(Rn) estimated value is equal 0.4499. To get the original value of F(Rn), the quantile function of the selected radon distribution must be applied (The numerical value of 0.4499 is 0.07343 *WL*) . The following equation is the quantile function of Rn distribution (See Table 1): $x(F)_{(Rn)} = \beta\left(-(\ln(F(Rn))^{-1/\alpha}\right)$.

For returning to the actual or original values of F(Th) , F(V) and F(D), the inverse CDF of the selected distribution for Th, V and D must be applied (See Table 1). The values of Th, V and D are 0.04928 *WL*, 2.2436 m^3/s and 100.15 *m*, respectively.

Table 6. Model accuracy measurements for probability multiple regression model

RMSE	RRMSE	RMSPE	MAPE	R^2
0.0012	1.1814	1.975	1.295	0.9993

Table 6 shows the values of model accuracy measurements to assess the results between actual data and probability modeled data (The modeled data are calculated according to equation 7).

Comparing the results of the model accuracy of multiple regression model (Table 3) and probability multiple regression model (Table 6) indicates that the error rate is reduced nearly by 60% in favor of probability multiple regression model (equation 7). The higher R^2 value is a good indicator for prediction. The results also indicate that the R^2 value of probability multiple regression model is higher than the R^2 value of multiple regression model. Therefore, the results gave more accurate prediction for determining the radon levels in this mine by using probability multiple regression model.

The prediction of radon levels in cases 2 and 3 are carried out in a similar manner as in case 1.

Case 2: West Yonus mine

Equation 8 represents multiple regression model for Rn of West Yonus mine.

$$\hat{Rn} = -0.1915+(0.308*Th)+(0.09878*V)-(0.0000964*D) \quad (8)$$

Table 7. Optimized radon values and the corresponding variables

Optimized Value	Radon Results	Variables		
	Rn	Th	V	D
Min Value	0.1395	0.034	3.2642	15
Max Value	0.4094	0.1449	6.029	407

The optimization minimum and maximum values of Rn and the corresponding variables Th, V and D are given in Table 7.

Table 8. Model accuracy measurements for multiple regression model

RMSE	RRMSE	RMSPE	MAPE	R^2
0.0037	1.553	1.459	1.257	0.9984

The accuracy measurements of multiple regression model is presented in Table 8.

Table 9. The parameters estimation of probability distribution of West Yonus mine

Distribution	Lognormal 3 parameters	Frechet 2 parameters	Lognormal 3 parameters	Johnson SB 4 parameters
Parameters estimation	***Rn***	***Th***	***V***	***D***
μ	-3.0928	-	-0.59061	-
σ	1.5719	-	1.4152	-
γ	0.13865	-	3.2382	-
α	-	2.0862	-	-
β	-	0.0523	-	-
γ	-	-	-	-0.02898
δ	-	-	-	0.60626
λ	-	-	-	452.85
ξ	-	-	-	-18.389

Table 9 shows the applied probability distributions and their estimators for West Yonus mine. The choices of the probability distributions is done according to x^2 test. These estimators are used in CDF calculation.

Table 10. The minimum and maximum values of the probabilities

Probability values	F(Rn)	F(Th)	F(V)	F(D)
Minimum value	0.01267	0.08580	0.015326	0.05899
Maximum value	0.8721	0.88752	0.87338	0.94869

The probability values of *Rn*, *Th*, *V* and *D* are ranging between the following minimum and maximum values as shown in Table 10.

Equation 9 represents the probability multiple regression model for F(Rn) of West Yonus mine.

$$\hat{F}(Rn) = 0.00214-(0.00005*F(Th))+(0.9666*F(V))+(0.03064*F(D)) \quad (9)$$

The inverse CDF of *Rn* values for this mine is calculated according the following equation (inverse CDF of 3 parameter Lognormal distribution):

$$x(F)_{(Rn)} = \exp((\sigma * \Phi^{-1}(F(Rn)) - \mu) + \gamma .$$

Table 11. Model accuracy measurements for probability multiple regression model

RMSE	*RRMSE*	*RMSPE*	*MAPE*	R^2
0.0024	1.006	0.686	0.508	0.9994

The accuracy measurements of probability multiple regression model is presented in Table 11.

The comparison of the results of the model accuracy of multiple regression model (Table 8) and probability multiple regression model (Table 11) shows that the error percentage is reduced nearly by 55% in favor of probability multiple regression model (equation 9). The R^2 value of probability multiple regression model is slightly higher than the R^2 value of multiple regression model. Therefore, the results of probability multiple regression model gave more accurate prediction for determining the radon levels in this mine.

Case 3: Safaga South mine

Equation 10 represents the multiple regression model for Rn of Safaga South mine.

$$\hat{R}n = - 0.854+(0.193* Th) +(0.189* V) – (0.000009*D) \quad (10)$$

Table 12. Optimized radon values and the corresponding variables

Optimized Value	Radon Results	Variables		
	Rn	Th	V	D
Min Value	0.8147	0.0240	8.8052	15
Max Value	1.3018	0.0585	11.3691	473

Table 13. Model accuracy measurements for multiple regression model

RMSE	RRMSE	RMSPE	MAPE	R^2
0.0044	0.4275	0.4277	0.4012	0.998

Table 14. The parameters estimation of probability distribution of West Yonus mine

Distribution	Exponential 1 parameters	Frechet 2 parameters	Exponential 1 parameters	Johnson SB 4 parameters
Parameters estimation	Rn	Th	V	D
λ	0.97105	-	0.10041	-
α	-	5.6869	-	-
β	-	0.03466	-	-
γ	-	-	-	-0.23462
δ	-	-	-	0.63401
λ	-	-	-	524.408
ξ	-	-	-	-24.721

Table 15. The minimum and maximum values of the probabilities

Probability values	F(Rn)	F(Th)	F(V)	F(D)
Minimum value	0.54666	0.00031	0.58693	0.03433
Maximum value	0.71751	0.95032	0.68068	0.94743

Equation 11 represents probability multiple regression model for F(Rn).

$$\hat{F}(Rn) = -0.488 + (0.00488 * F(Th)) + (1.767 * F(V)) + (0.001771 * F(D)) \tag{11}$$

The *x(F)* of the exponential distribution for radon level is calculated from the following equation:

$$x(F)_{(Rn)} = \left(-\ln(1-F(Rn))\right)/\lambda$$

Table 16. Model accuracy measurements for probability multiple regression model

RMSE	RRMSE	RMSPE	MAPE	R^2
0.00163	0.1578	0.1864	0.1027	0.9998

The results of accuracy measurements of the models application (equations 10 & 11) are represented in Tables 13 &16. It is obvious that the errors of probability multiple regression model are less than the errors of multiple regression model at the rate of 70%.

4. Conclusion

Genetic Algorithm (GA) and statistical probability distributions are applied for estimating the radon airborne levels in three Egyptian phosphate mines. The prediction of radon airborne levels in a manner of which can get a good result closer to actual radon airborne levels into the underground mines is estimated. The possibility of applying GA techniques for finding the minimum and maximum values for radioactive material paths to improve radon prediction model have been demonstrated. A population of 100 generation is used to get the optimal results of radon airborne levels. Developing and converting the multiple regression model to probability multiple regression model is conducted. The development is carried out to improve the prediction results of radiation levels of airborne radon. The development depends on converting the optimized and actual data to probability values using the CDF of the suitable distribution. Comparison between the two statistical models is carried out to define the optimal model for predicting the radon levels in the investigated mines. The result of this development reduce the error rate of prediction by a large margin. From the study the following conclusion is reached:

1) The GA is used independently for any problem and shows great importance for users.

2) The probability distributions give a good results when the data is considered to be ascending or descending.

3) The choice of the suitable distribution is very important to get a good results.

4) The probability multiple regression model reduces the error rate of radon prediction results by 50% to70% when compared with the multiple regression model. The values of radon prediction is almost identical with the actual values.

5) The range between the maximum and minimum radon optimized values included all values in the entire mines. The value of radon level at any distance can be estimated. This reduces the time, effort and calculate the exposure dose for miners at any proposed distance.

References

Pergamon press (1986). Radiatin Protection of Workers in Mines, Reprot of Committee 4 of the International Commission on Radiological Protection.

EPA Assessment of Risks from Radon in Homes (2003). Office of Radiation and Indoor Air, US Environmental Protection Agency. Available: www.epa.gov/radiation/.../assessment/402-r-03-003.pdf.

Work SAFE (2015). Approved Code of Practice Ventilation in Underground Mines and Tunnels. Available: www.business.govt.nz/worksafe/.../ventilation-in-underground-mines-an.

Ashraf E. Khater, M.A. Hussein and Mohamed I. Hussein (2003). Occupational Exposure of Phosphate Mine Workers: Airborne Radioactivity Measurements and Dose Assessment, Journal Environmental Radioactivity, 75, 47-57.

Cumulative Distribution Function, fromWikipedia (2015). Available: https://en.wikipedia.org/wiki/Cumulative_distribution_function

Stat.yale.edu. (2015). Multiple Linear Regression. Available: http://www.stat.yale.edu/Courses/1997- 98/101/linmult.htm.

Praveen Ranjan Srivastava and Tai-hoon Kim (2009). Application of Genetic Algorithm in Software Testing**,** International Journal of Software Engineering and Its Applications, 3(4), 87-95.

Chou T.Y., Liu T. K., Lee C. N., Jeng C. R. 2008. Method of Inequality-Based Multiobjective Genetic Algorithm for Domestic Daily Aircraft Routing, IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans, 38(2), 299-308.

Exponential Distribution, from Wikipedia (2015). Available: http://en.wikipedia.org/wiki/Exponential_distribution.

Bolarinwa, I.A. and Alhassan, B.G. (2013). An Investigation of the Performance of Kolmogorov Smirnov and Anderson-Darling Goodness of Fit Statistics on Pareto and Gumbel Minimum Densities, International Journal of Advanced Scientific and Technical Research, 1(3), 597-614.

Kamran Abbas, Yincai Tang (2013). Estimation of Parameters for Frechet Distribution Based on Type-II Censored Samples Caspian Journal of Applied Sciences Research 2(7), 36-43.

Distribution, MLE (2015). 'MLE For Two-Parameter Exponential Distribution'. Stats.stackexchange .com. Available:http://stats.stackexchange.com/questions/124398/mle-for-two-parameter-exponential-distribution.

Rodrigo J. Aristizabal, (2012). Estimating the Parameters of the Three-Parameter Lognormal Distribution, FIU Electronic Theses and Dissertations, Florida International University.

Johnson SB Distribution-Ntrand (2015). Ntrand.com. Available: http://www.ntrand.com/johnson-sb-distribution/.

Rob J. Hyndman and Anne B. Koehler (2006). Another look at measures of forecast accuracy, International Journal of Forecasting 22 (4), 679–688.

Ali M. H. and Abustan I. (2014). A new novel index for evaluating model performance. Journal of Natural Resources and Development, 04, 1-9.

Maxim V. Shcherbakov, Nataliya L. Shcherbakov, Timur A. Janovsky and Valeriy A. Kamaev (2013). A Survey of Forecast Error Measures, World Applied Sciences Journal, (24), 1-176.

Draper, N. R.; Smith, H. (1998). Applied Regression Analysis. Wiley-Interscience. ISBN 0-471-17082. Available*: https://e*n.wikipedia.org/wiki/Coefficient_of_determination.