Journal of the Nigerian Association of Mathematical Physics
Volume 47 (July, 2018 Issue), pp221 – 228
© J. of NAMP

GEOESPATIAL DISTRIBUTION AND PREDICTION OF TOTAL PETROLEUM HYDROCARBON CONTENT IN SOIL USING GEOSTATISTICS METHOD

[1]Okonofua E. S., [2]Babatola J. O., [3]Ojuri O. O., and [4]Ekun M. O.

[1&4]Department of Geomatics, University of Benin, PMB 1154, Edo state Nigeria,
[2&3]Department of Civil Engineering, Federal University of Technology, Akure, PMB 704, Ondo State Nigeria

Abstract

The aim of this study is to apply Geospatial Information System for the spatial distribution and prediction of Total Petroleum Hydrocarbon contaminant in Ologbo oil exploration field. 1 km^2 of the study area was digitized, georeferenced and gridded at 100 m interval using Google Earth and ArcGIS desktop 9.2. Soil samples were recovered at the intersections using calibrated hand auger at depths 0-15 cm, 15-30 cm and 30-60 cm respectively. Handheld GPS was used to acquire the coordinates of the soil sampling locations while the recovered samples were preserved and taken to the laboratory for TPH analysis using Gas Chromatogram with Flame Ionization Detector (GC-FID). Across the depth of exploration (0-60cm), TPH in the soil ranged from 2-863 mg/kg, with higher values obtained at subsurface depth but decreases as the depth of investigation increased. These values were exported into ArcGIS 9.2 software where it was used to generate the TPH pollution map and also predict pollution for unsampled area within the 1 km^{2}at the depths of exploration. The correlation distances resulted from modeling the Semivirogram are 323.67m, 561.22 m, and 485.78 m for the depths 0-15cm, 15-30cm and 30 - 60cm respectively. The spatial distribution map of TPH at depth 0-15 cm shows high TPH Concentration (Strong pollution) in the south eastern area especially around grid lines D6 – D8. The predicted concentration maps shows similarity in the spatial distribution of TPH in the entire region. Concentration of TPH contaminants were strongest at the first depth (0-15 cm) and gradually grew weaker with change in layer until the last layer (30-60 cm) shows the weakest of all. Cross validation also indicated that predicted data tallied well with ground data. The TPH values obtained in some parts of the study area were far above permissible limit of TPH in soil recommended by World Health Organization hence there is the urgent need to carryout bioremediation at the site.

***Keywords**:* Contamination, Prediction, Geospatial Information System, Soil, Treatment

1.0 Introduction

Crude oil exploration and exploitation within a sustained process is essential for the proper wellbeing of nations and entire populace that depends on it as main source of economic earning. However, the petroleum industry responsible for the mining of crude oil have been perceived in bad light due to the hazard it portends to public health, damages to the environment, land, water, soil and forest [1]. Prior to this period of large-scale urbanization, crude oil contaminated land were abandoned to allow for self-recovery which usually takes longer period; but due to current development demand for land, there is the need to hasten remediation and reuse of contaminated lands [2].

A great task and crucial issue in environmental protection practice is remediation of crude oil polluted soil. Today, hundreds of hectares of lands in the Niger Delta region are exposed to contamination by crude oil and proposing appropriate remediation methods or even determining the extent of pollution is almost near impossible. This is majorly responsible for the delay in cleanup of most contaminated sites [3]. Some techniques for the prediction and assessment of pollutant distribution in soil horizontal and vertical profiles are particularly cumbersome. Knowledge of the depth and spread including the degree of pollution must be ascertained before proposing adequate remediation methods.

Corresponding Author**:** Okonofua E.S., Email: ehizonomhen.okonofua@uniben.edu, Tel: +2348134826148, +2348073522637

However, soil samples recovered from crude oil sites for analysis are usually limited due to either human or financial constraints; this often affects the results of such research. The main priority of any remediation effort is to utilize method(s) that is cost friendly, rugged and with sufficient accuracy in minimizing risk of spread of pollutants into the environment.

Geospatial Information System (GIS) is a power based technology and methods used for collection, management, analysis, modeling and presentation of wide range of spatial data. Recently, geospatial information relating to different events have been on the increase in diverse and sophisticated forms. GIS is now often combined with other analytical methods and models such as probability, statistical and data harvesting methods to compliment the inherent abilities of GIS in evaluating the spatial patterns of events including their attributes [4]. GIS and geostatistical methods aid in determining pollution levels in soil volumes. Previous studies by [5, 6, 7]; revealed that GIS techniques and geostatistical methods produced useful results when applied in evaluation of contaminated sites.

Although soil contamination is three dimensional (3D) in phenomenon, nearly all geostatistical methods focus only on the horizontal dimension (2D). Current geostatistical technologies now have improved prediction tools which make it relevant in depth contamination prediction. This application incorporates location, spatial relationship and classical statistics into the estimation process [8]. The technology utilizes the theory of regionalized variables in ascertaining contamination prediction. The theory stipulates that regionalized variables exhibit statistically measurable degree of continuity within a limited region. In that region, a statistical relationship between the value of a pair of regionalized variables and their distance apart can be determined. At greater distance, the difference should be statistically independent of each other. If the spatial variation of a regionalized variable can be determined, then that information can be used to predict the values at unknown locations. A regionalized variable has a spatial variation that is unknown but its variability with respect to distance is statistically measurable within a finite area [9]. Therefore, this work seeks to investigate the possibility of predicting crude oil contamination spread in Ologbo oil field using GIS and geostatistical models.

2.0 Materials and Methods

2.1 Site Location

The Project site is located in Ologbo, Ikpoba Okha local government area of Edo state which lies between longitude 05° 38' 36.47"E to 05°4' 26.56" E and latitude 06° 4' 28.17"N to 06° 4' 33.79"Nwhich is 32 km south-west of Benin City. The location of the project is almost 18km from NPDC link road which is off Benin-Sapele Road. Within this location, soil and sometimes water are contaminated almost regular from pipes transporting products to the flow station.

2.2 Sample Recovery and Treatment

To evaluate the baseline concentration level of Total Petroleum Hydrocarbon (TPH) content available in the soil which will be used for spatial modeling, systematic soil sampling was carried out in the field. The coordinate of the sampling positions was determined and registered with the aid of handheld global positioning system GPS) receiver (Garmin GPS 72). The satellite imagery of the study area was obtained and a reference area of 1km^2 was gridded at 100m interval and divided into five zones where soil samples was collected from the grid intersection points with the aid of depth calibrated augers at depth of 0-15cm, 15-30cm and 30-60cm respectively. Subsurface depth of 15cm was exceeded so as to recover samples at greater depth in order to examine the maximum vertical depth of TPH contaminants in accordance with [10].Recovered samples was placed in plastic bags and tightly sealed, and transported to the laboratory where the soil was characterized and analyzed. The results from the laboratory were compared with WHO standard for crude oil limit in soils.

This sample treatment involves; sample preservation, sample extraction and clean-up in order to obtain reliable values for analysis. Samples were placed in plastic bags and put into a glass jar with seal. Each sample was labelled differently and stored in a refrigerator at 4^{o}C. Sample extraction was carried out using extraction procedure detailed in USEPA method 3540 and ASTM method D5369 with little adjustments on flask size, choice of solvent, volume of solvent and extraction time.

2.3 Geospatial and Geostatistical Operations

The step by step approach in determining the spatial dependence of the field data is shown in figure 1. It begins with statistical analysis to get an idea of the distribution of the field data with the following underlying assumptions: the data needs to have normal distribution otherwise transform, it must be stationary otherwise treat local variance separately and must not have trends or else remove trend.

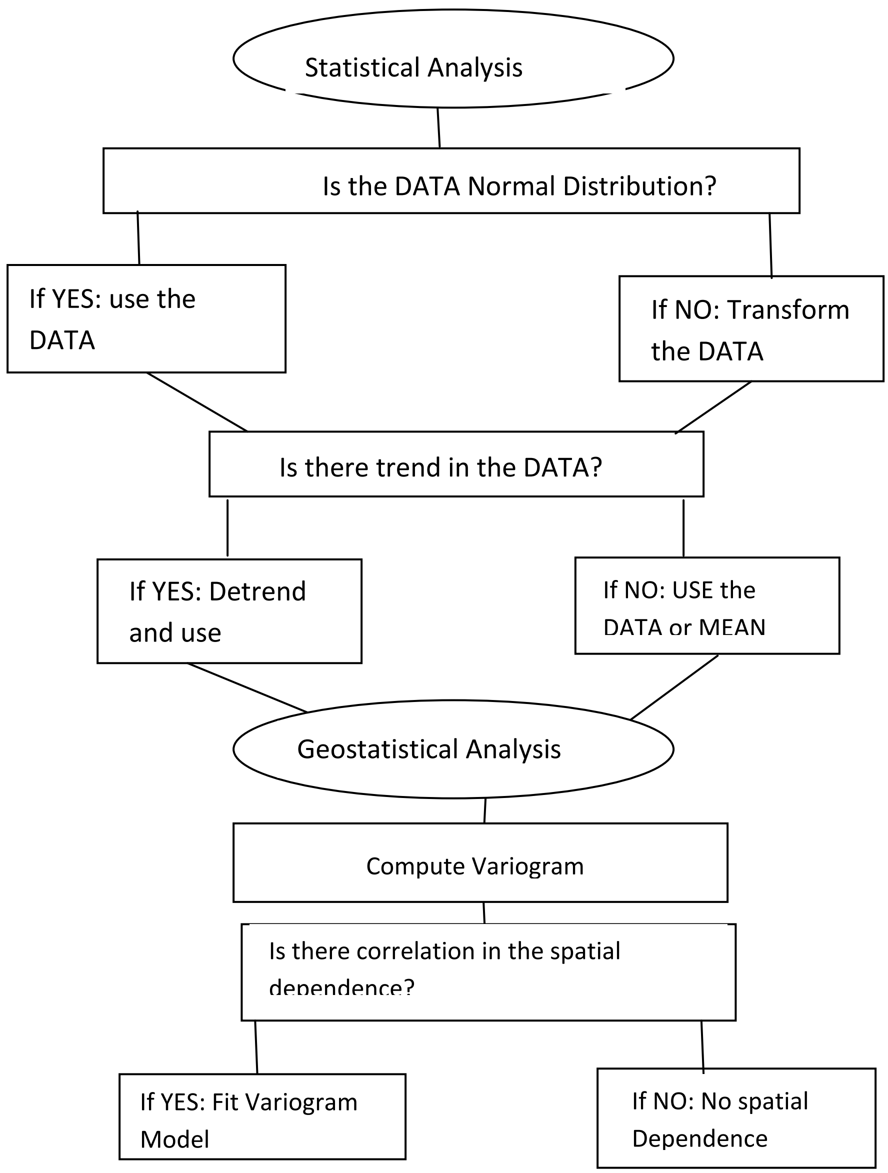

Fig 1: Steps to Perform Geostatistical Analysis

1. Data transformation: To check for normality, histogram plots were generated for every depth (0-15cm, 15-30cm and 30-60cm). The data set which fell short of normality were transformed using lognormal transformation a tool geostatistical extension of the ArcGIS. This is to ensure that data distribution did not deviate too severely for normality [11].
2. Semivirogram/variogram models fittings: Following normality and transformation of data, variography (i.e. variogram model fitting) was carried out. According to [12], the semi-variogram function used for characterization of spatial correlation is expressed as:

$$r\,(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \qquad (1)$$

Where N(h) is the number of data pairs at each step width (h), and Z(x_i) and Z(x_i) and Z(x_i+h) are the values of the variable Z at location xi and (x_i+h), respectively, I is the position of soil samples r(h) is for the vertical coordinate mapping, that is semi-variance diagram [13, 14]. The experimental variograms were fitted with four models: Exponential, Spherical, Gaussian and stable, and the basic parameters of variogram: Nuggets (Co), sill (C) and range (R) were determined. These parameters aided in the recognition of measurement error, determination of spatial dependency and the distance at which spatial dependency seized to exist.

3. Spatial dependency determination: Before optimal interpolation, the spatial dependencies of TPH in soil were evaluated using the variogram Nugget/sill ratio. The ratio of 25% (0.25) and 75 (0.75) are two thresholds for the relative strength index of spatial auto-correlation [15].
4. Kriging: The TPH concentrations in soil were estimated in unsampled locations using ordinary kriging. According to [14], the kriging action takes place through the equation:

$$Z(x) = \sum_{i=1}^{n} \lambda_i \, Z(x_i) \qquad (2)$$

Where λ_i equals variable amount in measured points. These predictions were subjected to cross validation process, before the production of predicted concentration maps.

3.0 Result and Discussion

i. Statistical Analysis and Data Transformation: TPH contaminants across the various depths in the study areas have lower medians than mean values. This indicates a positively skewed data as the distribution tends to the right. This is as a result of a high frequency of low concentration and some exceptional high concentration in the datasets. This is a common distribution in soil contamination. Skewness and kurtosis characterize the degree of symmetry of a distribution around its mean and give information about the normality of a variable.

Table 1: Statistics of Raw and Transformed Datasets at Various Depths

	Raw Data		Log Transformation	
Depth (cm)	**Skewness**	**Kurtosis**	**Skewness**	**kurtosis**
0 -15	1.96	7.62	-0.43	2.75
15-30	2.31	9.77	-0.27	2.39
30-60	2.57	11.68	-0.28	2.54

The summary of kurtosis and skewness presented in Table 1 shows that kurtosis and skewness for all the samples are higher than that of the normal distribution indicating positive skewness. The skewed distributions of the raw data requires a transformation hence, lognormal transformation was applied to all samples to achieve datasets distribution closer to normality before generating geostatistical prediction.

From the comparison between skewness and kurtosis of the datasets before and after the transformation as shown in Table 1, it is clear that the raw data set failed to meet the normality test. Because normality of data is necessary for optimal geostatistical methods like variogram and kriging models, and also because environmental data are usually skewed; histogram of raw data set were generated for the TPH raw data at every depth of sample collected. The results (histograms) are presented in Figure 2. Lognormal transformation was applied on the datasets to get distribution closer to normality before fitting variogram models to generate kriging prediction.

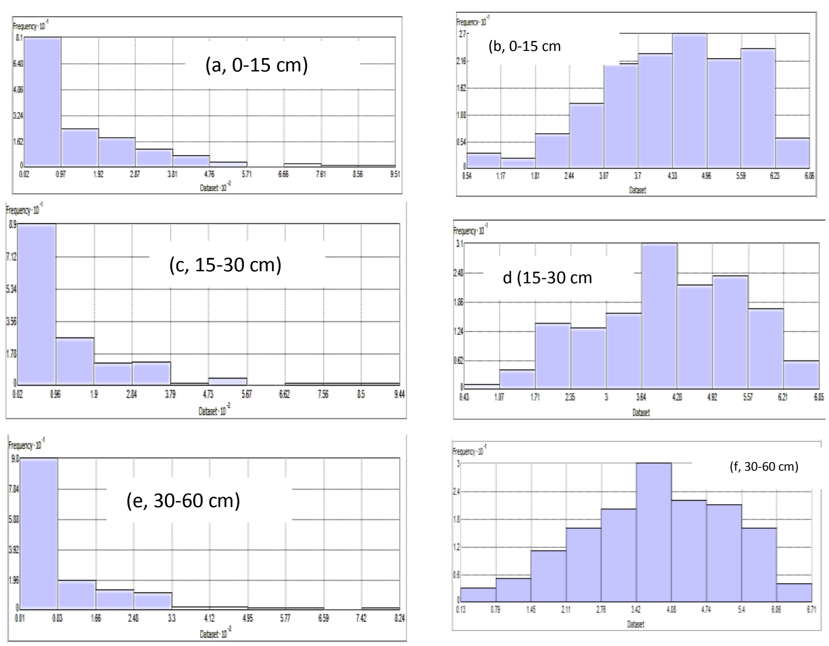

Fig 2: Distribution of TPH contaminants at examined depth; a, c, e are before transformation while b, d, f are after transformations.

ii. Variograms and Data Dependency: The best fitted mathematical models to the experimental semivariograms were applied to predict the spatial correlation of TPH at the examined sampled depth. The parameters of the Semivirogram model are presented in Table 2. The best fit model was generated using the Exponential model which is the best fitting descriptor of the data sets as shown in Figure 3.

Calculating the Semivirogram in different directions proved that all concentration of TPH in all depths did not exhibit geometrical anisotropy properties, all depths of sampling presented a nugget variance and sill. The nugget variance (C0) is the value of Semivirogram at a lag distance equal to zero. The sill (C) is the maximum variance at which the Semivirogram model flattened or takes the horizontal shape. The distance at which the spherical Semivirogram reaches to the sill is called the range (a). The range gives the correlation distance of the soil heavy metals. After this distance there is no relation between the variable. The correlation distances resulted from modeling the Semivirogram range from 323.67m, 561.22 m, and 485.78cm for the depths 0-15cm, 15-30cm and 30 -60cm respectively.

Table 2: Results of the best fitted Semivirogram models for TPH concentration in soil.

Depth (cm)	Fitted Model	Range (m)	Nugget (C0)	Still (C)	(C0/C)%
0-15	Exponential	323.67	0.00	1.55	0
15-30	Exponential	561.22	0.12	1.99	6
30-60	Exponential	485.78	0.06	2.21	3

The Nugget/Sill (C0/C) ratio can be regarded as a criterion to classify the spatial dependence of soil properties. If the ratio is less than 25%, the variable has strong spatial dependence; between 25% and 75%, the variable has moderate spatial dependence, and greater than 75%, the variable shows only weak spatial dependence [13, 15]. From Table 2 it is found that, all sampled depth has marked strong spatial dependence with C0/C ratio less than 10%.

iii. Predicted Contamination Maps: The fitted semivariograms were used to generate the predicted contamination concentration maps of the study area. Figure 4 shows the generated spatial distribution of TPH within the study area. The figures clearly showed that all sampled depths have a similar geographical distribution. The spatial distribution map of TPH at depth 15 cm shows high Concentration (Strong pollution) in the south eastern area especially around grid lines D6 – D8. Part of the south- eastern region also shows moderate pollution level while most regions in the North are seen to have lower TPH pollution for all the depths.

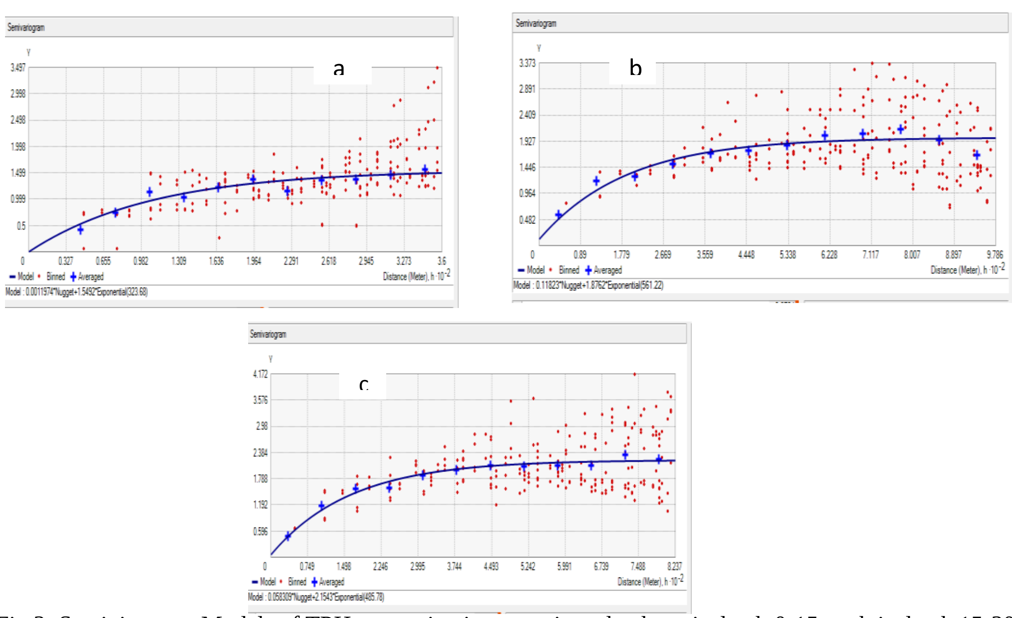

Fig 3: Semivirogram Models of TPH contamination at various depths; a is depth 0-15cm, b is depth 15-30cm while c is depth 30-60cm.

The predicted concentration maps shows similarity in the spatial distribution of TPH in the entire region. Concentration of TPH contaminants were strongest at the first depth (0-15 cm) and gradually grew weaker with change in layer until the last layer (30-60 cm) shows the weakest of all, with little deviations from the pattern at some points vertically down the layer.

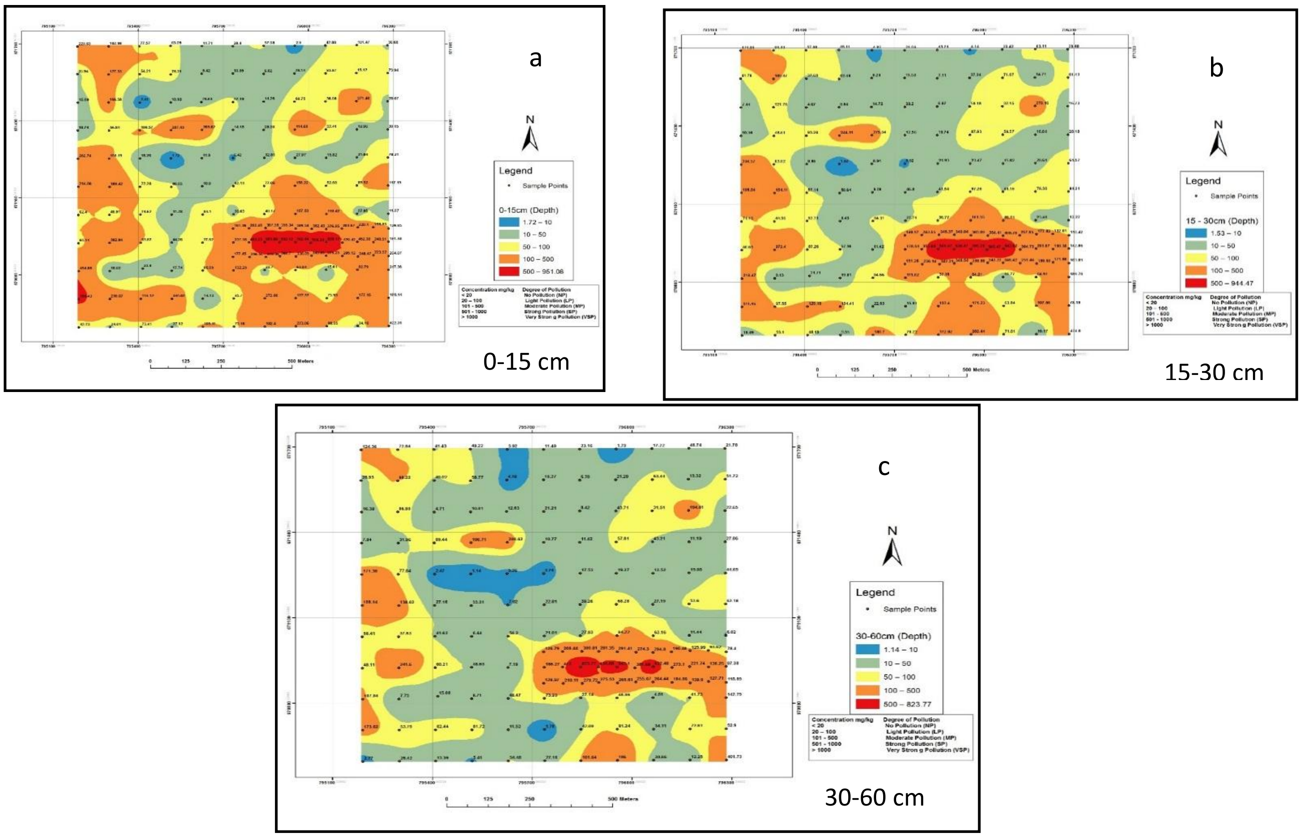

Fig 4: Predicted TPH contamination concentration at various Depths; a is depth 0-15cm, b is depth 15-30cm while c is depth 30-60cm.

Cross validation was carried out to give an idea of how well the model used predicts the unknown values at various locations. For all points, cross validation sequentially omits a point, and predicts its values using the remaining data after which the predicted point is compared to the actual measured value. In addition to visualization of the prediction error by graphs, a number of statistical measures can be used to assess the performance of the model. Some useful data obtained from cross validation are presented in table 3:

Table 3: Parameters Measured during Cross Validation

Depth (cm)	Sample Number	Mean	Root-Mean-Square	Mean Standardized	Root-Mean-Square Standardized	Average Error
0-15	121	0.6593	28.1581	0.00162	0.9811	31.5641
15-30	121	0.7522	24.7349	0.00761	0.9140	27.2163
30-60	121	0.6467	26.5514	0.00891	0.9862	25.6382

Conclusion: The major conclusions drawn from this study are:

1. TPH concentration in the study area especially a depth 0-15 cm is quite high (an average of 372 mg/kg). This value is higher than the USEPA allowable limit for TPH in soil hence there is the need for urgent remediation.
2. Geostatistics operations which include kriging, variogram and interpolation were used and compared for the capability to predict and describe the spatial distribution of TPH in the soil. Cross validation (table 3) also indicated that the predicted data tallied well with the ground data.

Reference:

[1]. **Choi, Y., Song, J. (2016):** "Sustainable Development of Abandoned Mine Areas Using Renewable Energy Systems: A Case Study of the Photovoltaic Potential Assessment at the Tailings Dam of Abandoned Sangdong Mine", Korea. Sustainability, 8, 1320

[2]. **Cai, L., Xu, Z., Ren, M., Guo, Q., Hu, X., Hu, G., Wan, H., and Peng, P. (2012):** "Source identification of eight hazardous heavy metals in agricultural soils of Huizhou, Guangdong Province, China". Ecotoxicol. Environ. Saf. 78, 2–8.

[3]. **Dublin-Green, W.F., Nwankwo, J.N., and Ilechukwu, D.O. (2009):** "Effective Regulation and Management of HSE Issues in the Petroleum Industry in Nigeria". SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production**.** Caracas, Venezuela, 7-110 June 2009. Paper No. SPE 40/26.

[4]. **Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W. (2005):** "Geographic Information Systems and Science", 2nd ed.; Wiley: Chichester, UK.

[5]. **Henshaw, S. L., Curriero, F. C., Shields, T. M., Glass, G. E., Strickland, P. T., and Breysse, P. N. (2004):** "Geostatistics and GIS: tools for characterizing environmental contamination". Journal of medical systems, 28(4), 335-348.

[6]. **Largueche, F. Z. B. (2006):** "Estimating soil contamination with Kriging interpolation method". American Journal of Applied Sciences, 3(6), 1894-1898

[7]. **Henriksson, S., Hagberg, J., Bäckström, M., Persson, I., and Lindström, G. (2013):** "Assessment of PCDD/Fs levels in soil at a contaminated sawmill site in Sweden –A GIS and PCA approach to interpret the contamination pattern and distribution". Environmental pollution, 180, 19-26

[8]. **Goovaerts, P., Fischer, M. M., and Getis, A. (2009):** "Handbook of applied spatial analysis: Geostatistical software software tools, methods and applications". Springer Science & Business Media

[9]. **Houlding, S. (2012).** "3D geoscience modeling: computer techniques for geological characterization". Berlin: Springer Science & Business Media.

[10]. USEPA (United States Environmental Protection Agency) Method 3540C Soxhlet extraction. Revision 3. December, 1996.

[11]. **Bohling, G. (2005):** "Introduction to Geostatistics and Vriogram Analysis", http//people.ku.edu/~gbohling/gpe940, accessed on 15th May, 2013

[12]. **Chamannejadian, A., Moezzi, A. A., Sayyad, G.A., Jahangirl, A. and Jafarnejadi, A. (2011):** "Spatial Distribution of Lead in Calcareous Soils and Rice Seeds of Khuzestan, Iran, Malaysia", Journal of Soil Sciences, 15, 115-125

[13]. **Liu, X., Wu, J. and Xu, J. (2006):** "Characterizing the Risk Assessment of Heavy Metals and Sampling Uncertainty Analysis in Paddy Field by Geostatistics and GIS". Environmental Pollution, Vol. 141, pp 257-265

[14]. **Jiachun, S., Haizhen, W., Xingmei, L. and Haiping, Z. (2007):** "Spatial Distribution of Heavy Metals in Soils: A Case Study of Chagxing, China", Environmental Geology, 52, 1-10.

[15]. **Chang, Y. H., Scrimshaw, M.D., Emmerson, R.H.C. and Lester, J.N. (1998):** "Geostatistical Analysis of Sampling Uncertainty at the Tollesbury Managed Retreat Site in Blackwater Estuary", Essex, U.K; Kriging and Co-kriging Approach to Minimize Sampling Density, Science Total Environment, Vol. 221, pp. 43-57.