# Neural Network Regression for Modelling the Effects of Selected Soil Physico-Chemical Properties on Adsorption

**C. G. Udomboso[1]; N. Nzelu[2]; B. I. Olu-Owolabi[3]**

[1,2]Department of Statistics,
University of Ibadan,
Ibadan, Nigeria.
e-mail: cg.udomboso@gmail.com[1]; ngozinzelu@gmail.com[2]

[3]Department of Chemistry,
University of Ibadan,
Ibadan, Nigeria.
e-mail: iromidayobamidele@yahoo.co.uk[3]



***Abstract*— Heavy metals in soils have been known as soil pollutants, to constitute serious economic importance as their accumulation has led to reduced agricultural production and quality of life. In the present paper, we studied the adsorption behaviour of selected heavy metals in soils, due to some physico-chemical properties. The soil under study was obtained from the River Benue Basin in the middle belt region of Nigeria. The heavy metals considered included lead (Pb), zinc (Zn), copper (Cu), and cadmium (Cd), while the physico-chemical properties included hydrogen ion concentration (pH), percentage goethite, percentage humic acid, time, and sorbate concentration. Estimation of the effects was carried out using the statistical neural network at $\alpha = 0.05$, while the cubic spline was used to interpolate within values and extrapolate forecasted values. Results show that rates of adsorption differ across properties. In all physical properties, except humic acid, Cd is most adsorped at AIC of 0.067, 0.079, 0.002, and 21.137 (all at *p<0.05*). For humic acid, most adsorped is Zn at AIC of 5.692 (*p<0.05*). These call for effective soil management system in Nigeria, which is expected to yield reliable data on soil behaviour, as well as concerted effort in eradicating (or reducing) the presence of these pollutants.**

**Keywords**-- *Soil, heavy metals, physico-chemical properties, adsorption, neural network.*

## I. Introduction

The determination of concentration of heavy metals and their accumulation in soil is an important issue due to their consequence, especially in the area of agricultural production and human health as well as its detriment effects on soil ecosystem [1].

Soil contaminated with lead (Pb), copper (Cu), zinc (Zn), cadmium (Cd) and heavy metals have serious consequences to the environment because it can change the soil quality, which is likely to affect the normal use of soil endangering public health and living environment. Their fate in the polluted soil is a subject of study because of the direct potential toxicity and indirect threat to human health via the contamination of groundwater and the accumulation in food crop [8].

Heavy metals are natural components which occurred in high concentration under natural conditions with a subsequent range of natural concentrations in soil, sediments, water and organisms [5]. They can remain in soils for a very long period of time. At a certain level, they are essential to a human body, but they also can cause toxic effects if exceeded the limit value. Sources of these elements in soils mainly include natural occurrence derived from parent materials and human activities; industrial wastes, commercialization, biological agent, radioactive pollutants, urban effluent, fertilizer application.

Apart from that, the physic-chemical properties of soil also affect the concentration of heavy metals in soil. Hydrogen ion concentration (pH), Goethite percentage, humid acid, time, sorbate concentration and other organic matters are one of most parameters controlling the accumulation and the availability of heavy metals in soil environment. It is necessary then to evaluate the relationship among these parameters and heavy metal accumulation in soil.

The Neural Network procedure, as powerful tools, can assist in soil modelling and predictions. In recent times, the Neural Network has been applied widely to investigate heavy metal concentration, accumulation and distribution in soils. It has been widely used in the field of soil science for prediction of soil hydraulic properties [11,10], generation of digital soil maps [9] and modelling of the behaviour of trace metals. There is anticipation of metal concentration in soil from reflectance spectroscopy using back propagation Neural Network (BPNN) and multiple linear regressions (MLR) [6], while there has been works on the comparative analysis of rainfall prediction using statistical Neural Network and Classical Regression [2,4].

An artificial neural network technique was used to develop a model to predict the constituents of the heavy metal in the soils such as Mercury, Cadmium, and Iron [7].The aim of their paper was to design artificial neural network as an alternative accurate tool to estimate concentration of Cadmium in contaminated soils.

The aim of this paper was to design artificial neural network as an alternative accurate tool to estimate concentration of Cadmium in contaminated soils. We aim to determine by means of Artificial Neural Network methods the best model for modeling soil adsorption of these metals in the soil sample collected from River Benue Basinin Adamawa State. It is hoped that this preliminary study would provide a scientific basis for contamination control and further monitoring of the heavy metal accumulation in soil.

An Artificial Neural Network (ANN) is network of Artificial Neurons and hence constitutes crude approximations of parts of real brains. They may be physical device or simulated on conventional computers from a practical point of view. An ANN is just a parallel computational system consisting of many simple processing elements connected together in a specific way in order to perform a particular task. One should never lose sight of how crude the approximations are and how over simplified our ANN are compared to real brains.

The history of neural networks begins in the early 1940’s and thus nearly simultaneously with history of programmable electronic computers. In 1943 Warren Mcculloch and Walter Pitts introduced models of neurological networks; recreated threshold switches based on neurons and showed that even simple networks of this kind are able to calculate nearly any logic or arithmetic function. In 1947 Wallter Pitts and Warren Mcculloch indicated a practical field of application (which was not mentioned in their work from 1943), namely the recognition of special patterns by neural network.

## II. Materials and Methods

### A. Data Description

Data used in this study are physio-chemical properties of soils used from four sites along the River Benue bank in Adamawa State. Adamawa is one of the largest states in Nigeria and occupies about 36,917 square kilometres. It is bordered by the states of Borno to the northwest, Gombe to the west and Taraba to the southwest. Its eastern border also forms the national eastern border with Cameroon. Topographically, it is a mountainous land crossed by large river valleys - Benue, Gongola and Yedsarem.

The four sites were coded as A1, A2, A3, and NB. From each site, physical properties observed included the hydrogen ion concentration (pH), goethite, humic acid, time, and sorbate, while the chemical properties observed include lead (Pb), copper (Cu), zinc (Zn), and cadmium (Cd).

In measuring the physical properties, the following levels were used. For pH, the measurements were taken at concentration levels 3, 4, 5, 6, 7, and 8, while for goethite, the concentrations were at 0, 2, 4, 6, 8, and 10. For humic acid, the concentrations were at 0, 0.5, 1, 2, 4, 6, 8, 10, while for time, the measurements were in hours of 0.5, 1, 3, 6, 12, and 16.

The concentrations of the sorbate were at 0, 25, 50, 75, 100, 125, 150, 200, 300, 400, and 500.

For all four sites, this gives a total of 24 sample sizes for pH, goethite and time, 32 sample sizes for humic acid, and 44 sample sizes for sorbate, Samples obtained were anaysed in the laboratory.

### B. Statistical Neural Network

An Artificial Neural Network (ANN) is composed of a network of largely interconnected neurons working together to solve a specific problem. It consists of input and output layers with at least one hidden layer in between them. The numbers of nodes in input and output layers are decided by the number of input and output parameters whereas the number of hidden layers and number of nodes in each hidden layer is decided by the complexity of the multivariable relationship to be developed.

In other to predict the best Neural Network model for modelling soil adsorption of these metals, Artificial Neural Network (ANN) Model proposed by Anders (1996) was used. The model is given below, using the logistic transfer function

$$y = f(\mathrm{X}, \mathrm{w}) = \alpha \mathrm{X} + \sum_{h=1}^{H} \beta_h g\left(\sum_{i=0}^{I} \gamma_{hi} x_i\right) \quad (1)$$

The introduction of the stochastic term makes equation a statistical model, known as the Statistical Neural Network, given as,

$$y = f(\mathrm{X}, \mathrm{w}) + e_i = \alpha \mathrm{X} + \sum_{h=1}^{H} \beta_h g\left(\sum_{i=0}^{I} \gamma_{hi} x_i\right) + e_i \quad (2)$$

where

$\mathrm{X} = (x_0 \equiv 1, x_1, \dots, x_I)$, $\boldsymbol{y}$ is the dependent variable, w=$(\alpha, \beta, \gamma)$ with $\alpha$ being the weight of the input unit,$\beta$ is the weight of the hidden unit,$\boldsymbol{\gamma}$ is the weight of the output unit, and $e_i$is the error term that is normally distributed ($e_i \sim N(0, \sigma^2 I_n)$).

For the hidden neurons, a different number of neurons in the hidden layer was fixed and through evaluation of their model performance, pick the one that gives the maximum performance among others to be the best model. The independent variables include hydrogen ion concentration (pH), percentage humid acid, time, and sorbate concentration, while the dependent variables were the heavy metals (Pb, Zn, Cu, Cd). The hidden neurons used include 1, 2, 3, and 4 at 1000 iterations each.

The transfer function used in this study is the hyperbolic tangent sigmoid (TANSIG), written in the following form,

$$tansig = f(x) = \frac{2}{1-e^{-2x}} - 1 \quad (3)$$

Specifically, the model is expressed as

$$y = \alpha_0 + \alpha_1 x_1 + \sum_{h=1}^{H} \beta_h \left(\frac{2}{1-e^{-2(\gamma_{h0}+\gamma_{h1}x_1)}} - 1\right) + u \quad (4)$$

at a simple 1 hidden unit, we can write

$$y = \alpha_0 + \alpha_1 x_1 + \beta_1 \left(\frac{2}{1-e^{-2(\gamma_{10}+\gamma_{11}x_1)}}\right) + u \quad (5)$$

Therefore, for simple model involving Pb as the dependent variable, and pH as the independent variable, we can write

at 1 hidden unit,

$$Pb = \alpha_0 + \alpha_1 pH_1 + \beta_1 \left(\frac{2}{1-e^{-2(\gamma_{10}+\gamma_{11}pH_1)}}\right) + u \quad (6)$$

For lack of space, we can go on to write for other models at different hidden units.

*C. Model Performance Measures*

The model checking criteria used in this study were the Coefficient of Determination (R-squared) and the Akaike Information Criterion (AIC).

$$AIC = MSE + 2\sigma^2 \frac{k}{n} \quad (4)$$

where k is the number of parameters and n is the total number of observation.

While the best model is determined by the smallest AIC, the p-values are reported for the t-statistic that determines the significance of the independent variables for prediction in the best model.

## III. Results And Discussion

In Table 1.1, the effect of pH on Pb and Zn is best predicted at hidden neuron 2, with AIC of 33.2999 ($p < 0.05$) and 10.296 ($p < 0.05$) respectively. While the effect of ph on Cu and Cd is best predicted at hidden neuron 3, having AIC of 0.4664 ($p < 0.05$) and 0.067 ($p < 0.05$) respectively. On the other hand, the prediction of Goethite is in Table 1.2, where the effect on Pb and Cu is best at hidden neuron 3, having AIC of 4.333 ($p < 0.1$) and 7.895 ($p < 0.05$) respectively. Its effect on Zn and Cd is best at hidden neuron 4, with AIC of 7.866 ($p < 0.05$) and 0.079 ($p < 0.05$) respectively.

As for Humic Acid in Table 1.3, its effect is best predicted at hidden neuron 3 for Pb, Cu, and Cd, having AIC of 86.461 ($p < 0.05$), 87.207 ($p < 0.05$), and 15.691 ($p < 0.05$) respectively. While, its effect on Zn is best predicted at hidden neuron 4 with AIC of 5.692 ($p < 0.05$).

In Table 1.4, the effect of time on Pb, Zn, and Cu is found to be predicted best at hidden neuron 4, with AIC of 0.850 ($p < 0.05$), 17.865 ($p < 0.05$), and 240.405 ($p < 0.05$) respectively. The effect on Cd is seen to be best predicted at hidden neuron 3, having AIC of 0.002 ($p < 0.05$).

Finally, in Table 1.5, the effect of Sorbate is best predicted at hidden neuron 3 for Pb, Zn, and Cu, having AIC of 1356.200 ($p < 0.1$), 3.463 ($p < 0.1$), and 257.244 ($p < 0.05$) respectively. On the other hand, the effect on Cd is best at hidden neuron 4 with AIC of 21.137 ($p < 0.05$).

Table 1.1 pH on Physico-Chemical Properties

| Hidden Neurons | Lead (Pb) | | Zinc (Zn) | | Copper (Cu) | | Cadmium (Cd) | |
|---|---|---|---|---|---|---|---|---|
| | AIC | p_value | AIC | p_value | AIC | p_value | AIC | p_value |
| 1 | 2873.200 | 0.142 | 173.458 | 0.109 | 18.449 | 0.336 | 4.851 | 0.349 |
| 2 | 33.300 | 0.014 | 10.296 | 0.026 | 0.507 | 0.043 | 0.991 | 0.036 |
| 3 | 48.524 | 0.017 | 36.707 | 0.049 | 0.466 | 0.041 | 0.067 | 0.031 |
| 4 | 183.834 | 183.834 | -15.286 | 1.000 | 1.022 | 0.061 | 0.085 | 0.035 |

Table 1.2 Goethite on Physico-Chemical Properties

| Hidden Neurons | Lead (Pb) | | Zinc (Zn) | | Copper (Cu) | | Cadmium (Cd) | |
|---|---|---|---|---|---|---|---|---|
| | AIC | p_value | AIC | p_value | AIC | p_value | AIC | p_value |
| 1 | 10.232 | 0.089 | 111.642 | 0.090 | 148.309 | 0.100 | 3.063 | 0.233 |
| 2 | -0.184 | 1.000 | -5.626 | 1.000 | 15.208 | 0.030 | 0.380 | 0.072 |
| 3 | 4.333 | 0.057 | 15.501 | 0.032 | 7.895 | 0.022 | 0.355 | 0.070 |
| 4 | -0.886 | 1.000 | 7.866 | 0.023 | -2.509 | 1.000 | 0.079 | 0.032 |

Table 1.3 HumicAcid on Physico-Chemical Properties

| Hidden Neurons | Lead (Pb) | | Zinc (Zn) | | Copper (Cu) | | Cadmium (Cd) | |
|---|---|---|---|---|---|---|---|---|
| | AIC | p_value | AIC | p_value | AIC | p_value | AIC | p_value |
| 1 | 3319.500 | 0.174 | 820.688 | 0.154 | 1033.500 | 0.183 | -143.761 | 1.000 |
| 2 | -163.495 | 1.000 | 19.645 | 0.022 | 102.804 | 0.053 | -10.271 | 1.000 |
| 3 | 86.461 | 0.026 | 68.473 | 0.042 | 87.207 | 0.049 | 15.691 | 0.032 |
| 4 | 135.237 | 0.032 | 5.692 | 0.012 | 221.722 | 0.079 | -3.476 | 1.000 |

Table 1.4 Time on Physico-Chemical Properties

| Hidden Neurons | Lead (Pb) | | Zinc (Zn) | | Copper (Cu) | | Cadmium (Cd) | |
|---|---|---|---|---|---|---|---|---|
| | AIC | p_value | AIC | p_value | AIC | p_value | AIC | p_value |
| 1 | 1.713 | 0.128 | 58.025 | 0.262 | 575.356 | 0.199 | 0.003 | 0.121 |
| 2 | 2.767 | 0.128 | 76.066 | 0.321 | 274.396 | 0.130 | 0.002 | 0.100 |
| 3 | -0.011 | 1.000 | -0.180 | 1.000 | -144.899 | 1.000 | 0.002 | 0.088 |
| 4 | 0.850 | 0.088 | 17.865 | 0.128 | 240.405 | 0.120 | 0.006 | 0.184 |

Table 1.5Sorbate on Physico-Chemical Properties

| Hidden Neurons | Lead (Pb) | | Zinc (Zn) | | Copper (Cu) | | Cadmium (Cd) | |
|---|---|---|---|---|---|---|---|---|
| | AIC | p_value | AIC | p_value | AIC | p_value | AIC | p_value |
| 1 | 3697.800 | 0.101 | 58.736 | 0.243 | 1801.900 | 0.081 | 47.872 | 0.047 |
| 2 | 4099.800 | 0.107 | 4.611 | 0.059 | 377.223 | 0.037 | 87.637 | 1.000 |
| 3 | 1356.200 | 0.060 | 3.463 | 0.051 | 257.244 | 0.030 | 9.424 | 1.000 |
| 4 | 112.994 | 1.000 | -0.537 | 1.000 | 438.066 | 0.040 | 21.137 | 0.031 |

## IV. CONCLUSION

The outcome of this study shows that the level of contamination of the soils along the River Benue bank of Adamawa State is high. This is readily seen in the low hidden neurons that easily predicts the effects of the physio-chemical properties. Hence, this study generally concludes that the Artificial Network Model methods can be a strong tool for monitoring of current environmental quality of agricultural soils in terms of heavy metal accumulation and for predicating the best model for modelling future soil contamination.

**Acknowledgement**

The authors acknowledged the input of the Department of Chemistry, University of Ibadan, in carrying out this experiment, as well as giving access to the data generated.

## References

[1] Adriano D.C. 2001. *Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability, and Risks of Metals*. 2nd Edition, Springer Verlag, NY.

[2] Amahia & Udomboso (2011). On the comparative analysis of rainfall prediction using statistical Neural Network and Classical Regression. Department of Statistics, university of Ibadan.

[3] Anders U. (1996): Statistical Model Building for Neural Networks. *AFIR Colloqium. Nunberg, Germany*.

[4] Ganthimathi, A. &Anbarasi, A. (2015) Environmental Impact Assessments Of Heavy Metal On Soil Using Ann Modeling Techniques Department of Civil Engineering, Kumaraguru College of Technology, Coimbatore – 30. *International Journal of Advances in Mechanical and Civil Engineering,* ISSN: 2394-2827 Volume-2,

[5] Hutton, M., Symon,C., (1986). The Quantities of Cadmium, Lead, Mercury and Arsenic Entering the U.K. Environment from Human Activities. Sci. *Total Environ* 57, pp129-150.

[6] Kemper, T. and Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science & Technology*, 36, pp. 2742-2747.

[7] Luma N. M. Tawfiq and Farah. F. Ghazi (2015). Using Artificial Neural Network Technique For The Estimation Of Cd Concentration In Contaminated Soils. *International Journal of Innovations in Scientific Engineering*. http://www.irapub.com (IJISE), Vol. No. 1, e-ISSN: 2454-6402, p-ISSN: 2454-812X

[8] Martinez C.E., Motto H.L. (2000): Solubility of lead, zinc of copper added to mineral soils. *Environmental Pollution*, 107: 153–158

[9] McBratney, A.B., Minasnya, M.L., and Rossel, R.V. Spectral Soil Analysis and Inference Systems: A Powerful Combination for Solving the Soil Data Crisis. *Geoderma*. 2006. 136 (1-2) 272-278.

[10] Minasny, B. Tranter, G. McBratney, A. Brough, D. and Murphy, B., 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma*, 153, pp. 155-162.

[11] Schaap M.G., Leij F.J. and van Genuchten M.T. 1988. Neural Network Analysis for Hierarchical Prediction of Soil Hydraulic Properties. *Soil Science Society of America Journal* Vol. 62, No. 4.