A Thesis

entitled

Correction Model Based ANN Modeling Approach for the Estimation of Radon

Concentrations in Ohio

by

Pavan Yerrabolu

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master
of Science Degree in Electrical Engineering

Dr. Vijay Devabhaktuni, Committee Chair

Dr. Ashok Kumar, Committee Member

Dr. Mansoor Alam, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

August 2012

Copyright 2012, Pavan Yerrabolu

This document is copyrighted material. Under copyright law, no parts of this document may be reproduced without the expressed permission of the author.

An Abstract of
Correction Model Based ANN Modeling Approach for the Estimation of Radon
Concentrations in Ohio

by

Pavan Yerrabolu

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master
of Science Degree in Electrical Engineering

The University of Toledo
August 2012

According to National Cancer Institute, radon is one of the major causes for lung cancer related deaths after smoking in the United States. In order to prevent deaths due to radon inhalation there is a need to determine the level of radon concentration in each locality, *e.g.*, zip-codes. However, factors like inapproachability hinder the process of estimating radon concentration in some places. In such places the radon concentrations could be estimated using several interpolation techniques. In this thesis, a new approach that improves the accuracy of the neural model with the help of sensitivity based correction model for modeling and estimating radon concentrations in Ohio is proposed. The results are compared with commonly used techniques such as kriging, radial basis function (RBF), inverse distance weighting (IDW), global polynomial interpolation (GPI), local polynomial interpolation (LPI) and the recently developed conventional ANN modeling approach. Further, model accuracies of all the above interpolation schemes are evaluated based on the ranked performance measures criteria with emphasis on the extreme-end (peak-end, low-end), and mid-range radon concentrations. The results demonstrate the effectiveness of the proposed approach in estimating the radon concentrations. The prediction accuracy of the proposed approach is found to be improved by 70-80% compared to the other techniques.

For my parents, fiancé, and friends.

Acknowledgements

I would like to express my sincere gratitude to my advisor Dr. Vijay Devabhaktuni and co-advisor Dr. Ashok Kumar for their guidance and encouragement throughout my research. I am grateful to ODH and USEPA for providing the radon data for research purposes to University of Toledo.

I would also like to thank my colleague Lakshman Mareddy for patiently helping me out whenever I was stuck in my research. I am thankful to Deepak Bhatt and Arjun Akkala for giving valuable inputs during my research which made my work easier.

Special thanks to all my family especially my elder sister for supporting me morally throughout my research study. Without their support and motivation this research work wouldn't have been possible.

Above all, I am thankful to Swetha Vanam, my fiancé, for being by my side through my tough times and for inspiring me whenever I lost hope. I am also thankful to my friends in Toledo on whom I could count on whenever I had a problem.

Contents

Abstract	iii
Acknowledgements	v
Contents	vi
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
1 Introduction	1
1.1 Problem Statement.....	1
1.2 Motivation.....	3
1.3 Organization of Thesis.....	4
2 Literature Survey	6
2.1 Role of Interpolation in Environmental Studies	6
2.2 Traditional Interpolation Techniques	8
2.3 Review of Artificial Neural Networks.....	10
3 Artificial Neural Networks Approach	12
3.1 Introduction to ANN.....	12
3.2 Structure of MLP3	13
3.3 Data Preparation	15

3.4	Methodology	16
3.5	Results and Discussions	18
4	Correction Model based ANN Approach	21
4.1	Proposed Methodology	21
4.2	Ranked statistical performance measures for IAQ models.....	26
4.2.1	Mean of training and validation data (M_v M_v).....	28
4.2.2	Standard deviation of training and validation data (SD_v SD_v).....	28
4.2.3	Least square slope (m) and intercept (b) regression statistics.....	28
4.2.4	Paired peak accuracy (A_p).....	29
4.2.5	Geometric mean bias (MG).....	29
4.2.6	Geometric mean variance (VG)	29
4.2.7	Fractional bias (FB)	30
4.2.8	Fractional variance (FV).....	30
4.2.9	Normalized Mean Square Error ($NMSE$).....	30
4.2.10	Factor of Two ($Fa2$)	31
4.2.11	Coefficient of correlation (r)	31
4.2.12	Spearman correlation coefficient (ρ)	31
4.2.13	Model comparison measure (MCM_2)	31
4.2.14	Robust highest concentration (RHC).....	32
4.2.15	Revised Index of Agreement (IOA_r)	32
4.2.16	Scatter plots	33
4.2.17	Quantile-Quantile ($Q-Q$) plots.....	33

4.3 Results and Comparison	33
5 Conclusions and Future Work	46
5.1 Conclusions.....	46
5.2 Future Work.....	47
References	48
Appendices	59
A Source Code for Correction Model Based ANN modeling approach	59
B Screenshots of the <i>Neuromodeler</i> Software	62
B.1 <i>Neuromodeler</i> Software Main Window	62
B.2 New Model Creation Window in the <i>Neuromodeler</i> Software.	63
B.3 Screenshot of the training window of the <i>Neuromodeler</i> software, while training the MLP3 of the Conventional ANN technique	64
B.4 Screenshot of the testing window of the <i>Neuromodeler</i> software, while validating the MLP3 of the Conventional ANN technique	65
B.5 Screenshot of the <i>Neuromodeler</i> Software comparison window.....	66
B.6 Screenshot of interpolated radon concentration surface by ArcGis spatial analyst.	67

List of Tables

2.1	A list of available interpolation techniques commonly used in environmental studies. ...	7
3.1	Training and Validation Errors of the Conventional ANN Model with the number of hidden neurons fixed at 10..	19
3.2	Training and Validation Errors of ANN Models trained using Quasi-Newton Algorithm.	20
4.1	Ranked statistical performance measures for IAQ models.....	27
4.2	Training and Validation Errors of candidate correction models.	34
4.3	Comparison of Validation Errors of the Conventional ANN and Correction Model Based ANN Models.....	34
4.4	Performance measures computed for traditional interpolation techniques, Conventional ANN and Correction model based ANN model.....	37
4.5	Performance measures computed for traditional interpolation techniques, Conventional ANN and Correction model based ANN model (Cont.).	38

List of Figures

1-1 Comparison of deaths per year in US due to radon and other causes (Source: http://www.epa.gov/radon/pubs/citguide.html)	2
1-2 Ohio Map showing the GM of radon concentrations(http://aprg.utoledo.edu/radon/).3	
3-1 Structure of MLP3	14
3-2 Plot of Sigmoid function.....	15
3-3 Proposed MLP3 structure for modeling radon concentrations.	16
4-1 Candidate correction model with latitude as the output.	23
4-2 Candidate correction model with longitude as the output.....	23
4-3 Flow-chart illustrating the Correction model based ANN correction approach.....	26
4-4 Scatter plot of kriging technique	38
4-5 Scatter plot of GPI technique	39
4-6 Scatter plot of LPI technique..	39
4-7 Scatter plot of IDW technique.....	40
4-8 Scatter plot of RBF technique.....	40
4-9 Scatter plot of Conventional ANN technique.	41
4-10 Scatter plot of Correction model based ANN technique.....	41
4-11 $Q-Q$ plot of kriging technique.....	42
4-12 $Q-Q$ plot of GPI technique.....	42

4-13 $Q-Q$ plot of LPI technique.....	43
4-14 $Q-Q$ plot of IDW technique.....	43
4-15 $Q-Q$ plot of RBF technique	44
4-16 $Q-Q$ plot of Conventional ANN technique.....	44
4-17 $Q-Q$ plot of Correction model based ANN technique	45
B-1 Screenshot of the main window of the <i>Neuromodeler</i> software.....	62
B-2 Screenshot of the new neural model creation window in the <i>Neuromodeler</i> software.....	63
B-3 Screenshot of the training window of the <i>Neuromodeler</i> software, while training the MLP3 of the Conventional ANN technique.	64
B-4 Screenshot of the testing window of the <i>Neuromodeler</i> software, while validating the MLP3 of the Conventional ANN technique.	65
B-5 Screenshot of the <i>Neuromodeler</i> software window, showing the validation data (magenta line) and the predicted data (yellow line)	66
B-6 Estimated radon concentrations obtained for Ohio by using four conventional interpolation techniques.....	67

List of Abbreviations

ANN	Artificial Neural Network
AFB	Absolute Fractional Bias
EPA	Environmental Protection Agency
Fa2.....	Factor of Two
FB.....	Fractional Bias
FV	Fractional Variance
GM.....	Geometric Mean
GPI.....	Global Polynomial Interpolation
IAQ	Indoor Air Quality
IDW.....	Inverse Distance Weighting
IOA _r	Index of Agreement
KBNN....	Knowledge Based Neural Network
LPI.....	Local Polynomial Interpolation
MG	Geometric Mean Bias
MLP.....	Multi Layer Perceptron
MCM ₂	Model Comparison Measure
NMSE....	Normalized Mean Square Error
ODH.....	Ohio Department of Health
ORIS.....	Ohio Radon Information System
Q-Q	Quantile-Quantile
RBf.....	Radial Basis Function
RHC	Robust Highest Concentration
US.....	United States
USEPA....	United States Environment Protection Agency

Chapter 1

Introduction

1.1 Problem Statement

Radon is a naturally occurring radioactive, tasteless, and odorless gas which is the decay product of uranium in the soil. Radon enters a house or a building through many possible ways such as openings around floor drains or pipes, or through cracks in the floor or walls and it builds up to dangerous levels in closed spaces. Radon has adverse effects on humans when exposed to higher levels for prolonged durations of time, due to its radioactive nature [1]. In Unites States, radon accounts for about 21,000 lung cancer deaths every year, out of which 2,900 occur among people who have never smoked (EPA's 2003 Assessment of Risks from Radon in Homes) [2].Fig 1-1 shows the comparison of deaths per year in US due to radon and other causes.

Fig 1-1: Comparison of deaths per year in US due to radon and other causes (Source: <http://www.epa.gov/radon/pubs/citguide.html>).

Owing to the serious effects of radon on human health, efforts are being taken by the United States Environmental Protection Agency (USEPA) to identify radon concentrations levels. In Ohio, under the direction of the Ohio Department of Health (ODH) radon concentration levels for more than 200,000 homes/schools have been gathered so far, from the year 1989 [3-7] and the data is being managed using different database management systems [8-12]. The University of Toledo, under the aegis of Ohio Air Quality Development Authority and ODH developed one such database known as Ohio Radon Information Systems (ORIS) [13, 14]. The ORIS also helps in determining the best mitigation system based on pre-mitigation level and post-mitigation level of radon concentrations stored in its database [14]. These mitigation levels and radon concentrations of various zip codes are reported by mitigation testers and contractors all over Ohio. However, radon data is not available for some zip codes owing to factors like inapproachability (refer Fig 1-2). At present out of 1492 zip codes in Ohio, data is available for 1261 zip codes [15]. For the remaining zip codes we need to estimate radon concentrations using interpolation techniques based on the available data.

Fig 1-2: Ohio Map showing the GM of radon concentrations

(<http://aprg.utoledo.edu/radon/>).

1.2 Motivation

Artificial Neural Network (ANN)'s have emerged as a powerful tool in signal processing, pattern recognition, and other applications [16-18]. They have the advantage of faster learning capability as they are structurally parallel-distributed systems [19]. ANN based modeling has been proposed in [15] to estimate the radon concentrations in the zip codes where data is not available. This approach proved to be more accurate than the traditional interpolation techniques, such as Kriging, GPI, LPI, RBF, and IDW. The accuracy of the neural models depends on the adequacy of training data. A well-distributed, sufficient, and accurately measured training data is needed to obtain an accurate neural model [20-22]. As mentioned earlier, training data *i.e.*, radon concentration is obtained through measurements at various zip codes by mitigation testers and contractors allover Ohio. Therefore, obtaining

a huge training dataset could be very expensive as more testers and testing equipment is required. However, even though sufficient and uniformly distributed data is provided, ANN's have the drawback of poor extrapolation capability due to limited generalization of the model [22].

To enhance the generalization ability and extrapolation capability of the neural model, more advanced ANNs known as knowledge-based neural networks (KBNNs) have been developed. KBNN's incorporate the existing engineering knowledge in the form of another neural model into the neural network structure [15, 18, 21]. Various KBNN models including prior knowledge input method, source difference method and space-mapped neural network were applied to estimate radon concentrations [22]. These KBNN models provided accurate radon estimates compared to ANN's but at the cost of increased structural complexity.

To overcome the drawbacks of the above techniques, this thesis presents a new correction model based ANN approach which predicts the radon concentrations more accurately than all the previously used techniques.

1.3 Organization of Thesis

The remainder of the thesis is organized as follows. Chapter 2 discusses the role of interpolation in environmental studies and the traditional interpolation techniques commonly used. A review of ANN's and their recent applications in various fields of science and engineering is provided.

Chapter 3 gives a detail introduction to ANNs, discusses the structure of Multi Layer Perceptrons (MLP), and gives an introduction to radon data preparation. Various performance measures used to assess spatial interpolation techniques are discussed. The

ANN based methodology in estimating radon concentrations is discussed in detail and the results are compared with traditional interpolation techniques.

Chapter 4 discusses the implementation of correction model based ANN approach in estimating radon concentrations, and the comparison of results with ANN based methodology as well as traditional interpolation techniques is provided. Indoor Air Quality (IAQ) performance measures used to select the best model in predicting the radon concentrations are discussed and the selection of best model is described. Finally, Chapter 5 gives the conclusions of the work done and future work.

Chapter 2

Literature Review

2.1 Role of Interpolation in Environmental Studies

Spatially continuous data (or spatially continuous surfaces) play an important role in environmental sciences and management. Environmental analysts usually require spatially continuous data over the region of interest to make effective decisions, and scientists need spatially continuous data across a region to make accurate interpretations [23]. However, such spatially continuous data cannot be acquired manually over the area of interest as the geography of certain areas is inapproachable by humans. Hence, data is collected at discrete locations (*i.e.* grids/points). In order to generate spatially continuous data, the value of an attribute at unsampled location needs to be estimated. In such a case, spatial interpolation methods are used to estimate the values of an environmental variable at unsampled location using data collected from discrete locations.

Spatial interpolation techniques are based on the concept that points which are close together in space tend to have similar value attributes. This is known as positive spatial autocorrelation. This principle is used in generating spatially continuous data from discrete data points. In the present case of radon concentration, the data provided by the ODH is

used to create radon concentration maps of the entire state of Ohio. This data is used in quantitative research to reduce cancer deaths due to radon gas inhalation. Therefore, it can be concluded that when an interpolation technique output is the basis for other research studies its accuracy becomes a critical factor. There are many performance measures which define the accuracy of an interpolation technique and these are discussed in Chapter 3 and 4 of this document.

Spatial interpolation techniques have been applied in various fields such as environmental studies, mining engineering, soil sciences, civil engineering and probability and statistics [23]. Table 2.1 shows the different interpolation techniques used in environmental studies, their principle of working, advantages, and disadvantages [24].

Table 2.1: Interpolation techniques used in environmental studies.

Interpolation Technique	Principle	Advantages	Disadvantages
GPI	Captures global patterns in the data in fitting a polynomial	Interpolates global variations effectively	Exponential increase in prediction errors with increasing complexity in data
LPI	Similar to GPI, but the curve is fitted to a subset of data	Interpolates local variations effectively	Misses the global variations in data
Trend Surface Analysis	Separates the data based on regional and local trends	Removes broader variations prior to further analysis	Edge effects caused by spatial autocorrelation
IDW	Linear combination of data points, inversely weighted as per distance	Simple to use, and works well with any kind of data	Spatial arrangement of samples is not taken into consideration
Kriging	Similar to the principle of IDW and also includes spatial arrangement of data	Best in linear unbiased spatial prediction; and no edge-effects during polynomial fitting	Requires sophisticated programming
RBF	Fits a series of data points but the curve is not smooth	Requires few samples	Fails at extrapolation

2.2 Traditional Interpolation Techniques

GPI fits a smooth surface over the sampled data points using least-squares regression. The disadvantage of GPI is that the presence of outliers *i.e.*, extreme data values in the data leads to poor interpolation. This technique is more useful in determining the overall/global trend of the data and it lacks minor details. Global interpolation is used for:

1. Fitting a smooth surface to the sample points when the surface slowly varies over the area of interest.
2. Eliminating or studying the effects of global trends.

Recently, GPI has been employed in the field of image rectification [25] and climate characterization [26].

Unlike GPI, the LPI technique fits many polynomials for a subset of data each of which is optimized for a neighborhood. Although it is more flexible than GPI, LPI is not an exact interpolator like IDW. This technique captures local variations in data effectively and misses the global trend which is a disadvantage of this technique. Therefore, this technique is best suited when the data is in the form of clusters with no significant global trends. Recent applications of LPI include hydrologic modeling of daily precipitations [27], modeling seasonal rainfall erosivity [28], estimating air temperatures [29], and mapping the spatial distribution of soil contaminants [30].

IDW technique weighs the points nearer to the prediction location more than those that are far away, hence the name inverse distance weighting. In other words, as the distance of a surrounding point to the predicted location approaches zero, its relative weight in determining the value at predicted location becomes one. An advantage of IDW is that it can

efficiently estimate extreme changes in data such as: cliffs, fault lines etc. This technique works best when the training data is evenly distributed and is moderately dense. A disadvantage of IDW is that the presence of data clusters and/or outliers leads to ineffective interpolation. Another drawback is that this technique cannot estimate above maximum or below minimum values. Recently, IDW has been applied in the field of environmental sciences to map soil nitrogen, phosphorous and potassium [31], to map soil properties in Iran [32], to determine spatial variability of methane flux [33] and the variation in marine temperature and salinity [34].

Interpolation using radial basis functions is conceptually similar to exactly fitting a fairly stiff rubber sheet that bends and folds to the sample data points. This is achieved mathematically by approximating the relation between input and output data by using a linear combination of different radial basis functions. The five important radial basis function used by RBF are thin-plate spline, spline with tension, completely regularized spline, multiquadric function, inverse multiquadric function. The RBF's create a rubber like membrane that fits into each of the measured data points minimizing the area of curvature [35]. This technique is considered as a form of ANN where the radial basis functions functionality is similar to the activation function in a neural network.

Disadvantage of this technique is that the radial basis functions work effectively only when the data is smooth *i.e.*, there should not be significant changes in the data values [36]. Most recently, a predictive control based on RBF is being used to control the supercritical main stream temperature of a power plant [37]. Other recent applications of RBF include in the field of photovoltaic's [38], electric power components [39], health sciences [40] and antennas [41].

Kriging provides estimates for accuracy in predictions apart from predicting the values at unknown locations. It is similar to IDW, but the weights are determined based on spatial autocorrelation (function of distance) between data points. Recently kriging has been used to estimate the daily particulate matter PM 2.5 concentration [42], to develop Carbon: Nitrogen ratio surface [43], in remote sensing image analysis and geostatistics[44].

2.3 Review of Artificial Neural Networks

ANN's due to their faster learning capability and efficient generalization of the model are being used in computer aided modeling applications. When the data pattern is complex, which is practically the case in environmental data, the traditional interpolation techniques fail to derive the input-output relationship efficiently. In such cases, ANN's have been applied and the prediction accuracy has been found to be much better than the traditional interpolation techniques. The application of ANN's in the field of environmental studies is dated back to early 1900's. The remaining section discusses the most recent applications of ANNs and its growing importance.

Recently, ANN's were used to predict the mean monthly maximum and minimum temperatures to lake-basin scale in an arid region in India. These predictions were used to downscale the surface temperature for the Pichola lake region in Rajasthan State in India [45]. In Iran, ANN's has been used to forecast evapotranspiration, which is the sum of evaporation and plant transpiration from earth's surface [46]. The data provided a better understanding of the loss of water through soil and future groundwater level could be estimated. ANN's are also being used to determine the amount of suspended sediment concentration, an important factor in the design and management of water resource projects [47]

In industrial countries air pollution has become a major problem and is being considered as a hindrance to socio-economic development. ANN's have been proved to be effective in estimating the air quality by learning the air pollutants concentrations from the previous years [48]. For example, recently in Northern Italy, ANN's were used to estimate the secondary pollutants concentrations namely VOC, NO_x, NH₃, primary PM and SO₂ [49].

Recent applications of ANN in the field of material sciences include parametric modeling and optimization of LASOX cutting of mild steel [50], modeling adhesive wear resistance of Al-Si-Mg-/SiC PM compacts fabricated by hot pressing process [51], investigating boiling point heat transfer characteristics in TiO₂ nanofluids [52] and studying the synthesis of TiO₂ nanoparticles in different thermal conditions and modeling its photocatalytic activity [53].

Recently, ANN's were used to model and optimize the production of biogas [54]. They are being applied for quantitative analysis of radioactive sources [55]. Most recent applications of ANN's include in the field of metallurgy [56], polymer sciences [57], mineral processing [58], environmental research and risk assessment [59], mechatronics and automation [60].

The use of ANN based modeling techniques to predict atmospheric pollutant concentrations are increasing. Lately, [15, 22, 61-64] have used ANN's to predict particulate matter concentrations and compared the prediction accuracy with other traditional interpolation techniques. Recently, the use of additional knowledge input to aid the learning process of ANN's in predicting particulate concentrations has been demonstrated in [22]. From, [15] it was concluded that the ANN based modeling approach provided much better prediction accuracy compared to traditional interpolation techniques.

Chapter 3

Artificial Neural Network Approach

In this chapter we discuss a conventional ANN approach using quasi-newton algorithm to predict radon concentrations in the state of Ohio. A simple MLP3 network is used to develop the artificial neural network model. This chapter illustrates the procedure of data preparation, methodology of training and validation of the ANN model and comparison of the proposed methodology with traditional interpolation techniques. At first, the introduction to ANN and the structure of the most commonly used ANN known as Multi layer perceptrons (MLP) is discussed.

3.1 Introduction to ANN

ANN is a mathematical or computational model inspired by biological neural networks. ANN's consists of information processing units known as neurons which are organized in layers. Generally there are three different types of neurons: input neurons, output neurons and hidden neurons. The input neurons are present in the input layer and receive information from outside the neural network, whereas the output neurons present in the output layer send processed information to an external recipient. Hidden neurons are the

ones that receive information from other neurons and their output serves as input to other neurons. They are present in the intermediate layers between input and output layers known as hidden layers. Generally ANN has an input layer, an output layer and one or more hidden layers. The neurons in these layers are interconnected through links, each of which has a weight assigned to it. For an artificial neuron, the inputs are multiplied by the weights assigned to it and then computed by a mathematical function known as activation function which determines the output of the neuron. By adjusting the weights of the artificial neuron we can obtain the desired output from a specified input. The process of adjusting the weights is known as training or learning. Different training algorithms are used to adjust the weights of the artificial neurons. ANNs combine artificial neurons in specific order to process information.

Of late, the most commonly used ANN's are the multilayer feed-forward networks known as multilayer perceptrons or MLP networks. MLP networks are characterized by multiple hidden computational layers [65, 66] and multiple hidden neurons in each layer. The universal approximation theory states that MLP networks with a single hidden layer consisting of finite number of hidden neurons are universal approximators. Therefore, the structure of ANN used in this research is MLP3 neural network. In the next section the structure of MLP3 used in this research is discussed.

3.2 Structure of MLP3

As the name indicates a MLP3 neural network has three layers: input layer containing the input neurons, the hidden layer containing the hidden neurons and the output layer containing the output neurons. The structure of a typical MLP3 is shown in the next page.

Fig 3-1: Structure of MLP3.

The training algorithm used in this research to train the weights of MLP3 is Quasi Newton. The activation function for the input neurons is the relay function whereas for the output neurons it is a linear function of the weighted sum of inputs given to the output neurons. The hidden layer neurons are activated by sigmoid function because it is simple to compute [67]. The sigmoid function is given by

$$Z(\gamma) = \frac{1}{1 + e^{-\gamma}}, \quad (3.1)$$

where γ is the weighted sum of inputs to the hidden neuron and $Z(\gamma)$ is the output of the hidden neuron. The plot of sigmoid function is shown in the next page.

Fig 3-2: Plot of Sigmoid function.

3.3 Data Preparation

As mentioned in Chapter 1, radon concentration levels gathered from more than 200,000 homes/schools across Ohio are being stored in a database known as ORIS which is developed and maintained by the University Of Toledo. The data is provided by university researchers, government agencies, and commercial testing companies and has been published in a series of technical papers and reports published by

The radon data reported consists of the following information: the zip code of the tested building, type of room where the test was conducted, type of radon detection device used, and the season of the year when the test was performed. Out of the above information, during training of ANN or while using ArcGis spatial analyst software, the input data consists of only the latitude and longitude of a zip code and corresponding geometric mean of the radon concentration at that zip code. The procedure to obtain the geometric mean of the radon concentration at a particular zip code is explained in [24]. The radon data used in

this research was obtained during the year 2010 and had been reformatted by Arjun Akkala [24].

3.4 Methodology

In the case of radon modeling, where the model inputs, latitude and longitude are represented by $\mathbf{x}$, and the corresponding model output by y , such that

$$y = f(\mathbf{x}), \quad (3.2)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represents the input vector, n is the total number of inputs, which is two for radon modeling as latitude and longitude are the inputs to the neural network and f represents the functional relationship between $\mathbf{x}$ and y . Since there is only one output *i.e.*, radon concentration, y is a scalar. In conventional ANN modeling approach a MLP3 neural network (refer Fig 3-3) is trained using the Quasi-Newton algorithm to obtain the ANN model f_{ann} that approximates f .

Fig 3-3: Proposed MLP3 structure for modeling radon concentrations.

To train the MLP3 neural network, the available dataset containing latitude, longitude and radon concentration is randomly divided into training and validation datasets represented by

$$\{(\mathbf{x}_i, y_i), i = 1 \dots N_t\}, \quad (3.3)$$

$$\{(\mathbf{x}_i, y_i), i = 1 \dots N_v\}, \quad (3.4)$$

where y_i is the desired output for the corresponding input vector $\mathbf{x}_i$ and N_t, N_v represent the number of training and validation samples respectively. Initially, 90% of the given data is selected as training data. The ANN is trained iteratively using the training data which involves the adjustment of weights in the neural network. In such a case, eq. (3.2) can also be represented as

$$y = f_{\text{ann}}(\mathbf{x}, \mathbf{w}), \quad (3.5)$$

where $\mathbf{w}$ is the ANN weight vector. Once the ANN is trained using the Quasi-Newton algorithm the final step is the validation of the model which is done by using the validation data. The validation of the model is based on certain quality measures. The quality of the developed neural network is assessed using the quality measures *i.e.*, user defined error E_{user} , average validation error E_{avg} and worst-case error E_{worst} . The E_{user} is user defined and is set as 5%. E_{avg} and E_{worst} are expressed as

$$E_{\text{avg}} = \sqrt{\frac{1}{N_v} \sum_{i=1}^{N_v} \left(\frac{d_i - f_{\text{ann}}(\mathbf{x}_i, \mathbf{w})}{d_i} \right)^2} * 100, \quad (3.6)$$

and

$$E_{\text{worst}} = \max_i \left(\left| \frac{d_i - f_{\text{ann}}(\mathbf{x}_i, \mathbf{w})}{d_i} \right| \right) * 100, \quad (3.7)$$

where $f_{\text{ann}}(\mathbf{x}_i, \mathbf{w})$ is the output of ANN and d_i is the desired output corresponding to the input vector $\mathbf{x}_i$, and N_v represents the number of validation samples. The above procedure is repeated by varying the percentages of training and validation data and in each case the average validation error E_{avg} and the worst-case error E_{worst} are calculated. Then, the percentage of training data for which the neural network model gives least validation average error is identified.

Now, with the percentage of training data fixed, the number of hidden neurons is varied in steps of 10. In each step, the neural network model is trained and validated as described earlier and the average validation error E_{avg} and the worst-case error E_{worst} are calculated. The neural network model which gives the least average validation error, E_{avg} , is identified as the best model for the conventional ANN approach. Finally, the performance of the conventional ANN approach is assessed based on whether the condition $E_{\text{avg}} \leq E_{\text{user}}$ is satisfied.

3.5 Results and Discussion

The given radon concentration data is randomly divided into training and validation data. Initially, 90% of the given data is selected as the training data and the remaining 10% as the validation data. The number of hidden neurons in the conventional ANN model is fixed at 10. The model is then trained and validated using the training and validation data respectively. The percentage of training and validation data is varied and the average validation error E_{avg} and E_{worst} shown in eq. (3.6) and eq. (3.7) are calculated in each case. Table 3.1 shows the change in the validation error with the change in percentage of training and validation data. From Table 3.1 it is clear that the MLP3 network performs best when

the training data is 90% of the given radon concentration as it gives least validation error in that particular case.

Table 3.1: Training and Validation Errors of the Conventional ANN Model with the number of hidden neurons fixed at 10.

% Training Data	% Validation Data	Training Error (%)	Validation Error	
			E_{avg} (%)	E_{worst} (%)
90	10	10.81	12.13	72.19
80	20	11.06	13.26	73.15
70	30	11.69	12.66	72.98
60	40	12.14	12.85	79.19
50	50	12.41	12.36	86.72

With the training and validation data fixed at 90% and 10% of the total dataset the number of hidden neurons in the conventional ANN model is varied. The average validation error E_{avg} and E_{worst} shown in eq. (3.6) and eq. (3.7) are calculated in each case. The neural model which gives the least average validation error, E_{avg} , is identified as the best model for the conventional ANN approach. Table 3.2 shows the training and validation error with varying number of hidden neurons. From Table 3.2, it is clear that the neural model which consists of 10 hidden neurons gives the least validation error. Therefore, the number of hidden neurons in the conventional ANN model is fixed at 10 and the E_{avg} of this model is 12.13%.

Table 3.2: Training and Validation Errors of ANN Models trained using Quasi-Newton Algorithm.

No. of Hidden Neurons	Training Error (%)	Validation Error	
		E_{avg} (%)	E_{worst} (%)
10	10.81	12.13	72.19
20	10.59	12.89	78.11
30	10.71	12.91	71.89
40	10.63	12.74	77.31
50	10.66	12.60	72.26
60	10.54	12.86	72.43
70	10.69	12.60	73.96
80	10.46	12.74	73.81
90	10.21	13.15	72.54

From Table 3.2 it is clear that the condition $E_{\text{avg}} \leq E_{\text{user}}$ is not achieved by the conventional ANN approach. Owing to the limited generalization capability, the conventional ANN modeling approach is not suitable for extrapolation. To overcome these drawbacks, in the next chapter a new correction model based ANN approach is proposed.

Chapter 4

Correction Model Based ANN Approach

Due to the inadequacy of training data the conventional ANN modeling accuracy is limited and the condition $E_{\text{avg}} \leq E_{\text{user}}$ is not achieved. Owing to the limited generalization capability, the conventional ANN modeling approach is not suitable for extrapolation. In such a case, the next best alternative can be the KBNN's.

Various KBNN models like prior knowledge input method, source difference method and space-mapped neural networks were applied to estimate radon concentrations [22]. However, using uranium as the knowledge input in addition to latitude and longitude tends to increase the structural complexity of the ANN. To overcome these issues, the correction model based ANN approach is proposed.

4.1 Proposed Methodology

In this approach, apart from the original less-accurate neural model f_{ann} a set of neural models $f_{\text{ann},j}$ referred to as candidate correction models are defined [68] to improve the accuracy of f_{ann} such that condition $E_{\text{avg}} \leq E_{\text{user}}$ is achieved. Let us first define the structure of

f_{ann} and $f_{\text{ann},j}$. The structure of f_{ann} is problem dependent whereas, that of $f_{\text{ann},j}$ is not known initially. Let the desired neural model be

$$y = f_{\text{ann}}(x_1, x_2, x_3, \dots, x_n, \mathbf{w}), \quad (4.1)$$

and the corresponding set of candidate correction models be

$$\begin{aligned} x_{1,c} &= f_{\text{ann},1}(y, x_2, x_3, \dots, x_n, \mathbf{w}_1), \\ x_{2,c} &= f_{\text{ann},2}(x_{1,c}, y, x_3, \dots, x_n, \mathbf{w}_2), \\ &\vdots \\ &\vdots \\ &\vdots \\ x_{t,c} &= f_{\text{ann},t}(x_1, x_2, \dots, x_{n-1}, y, \mathbf{w}_n), \end{aligned} \quad (4.2)$$

where $x_1, x_2, x_3, \dots, x_t$ are the scalar inputs, n is the total number of inputs, $f_{\text{ann},i}$, $x_{i,c}$, and $\mathbf{w}_i$ represents the i^{th} candidate correction model, $1 \leq i \leq t$, its corresponding output, and its corresponding weight vector respectively. From eq. (4.2), it is clear that the number of candidate correction models is equal to the total number of inputs in the original less accurate neural model. In the case of radon modeling, as mentioned earlier two inputs are fed to the neural model *i.e.*, latitude and longitude, so we have two candidate correction models (refer Fig 4-1 and Fig 4-2). The eq. (4.2) can be rewritten as

$$x_{1,c} = f_{\text{ann},1}(y, x_2, \mathbf{w}_1), \quad (4.3)$$

$$x_{2,c} = f_{\text{ann},1}(x_{1,c}, y, \mathbf{w}_2), \quad (4.4)$$

where $x_{1,c}$ is the output of the first candidate correction model *i.e.*, latitude, $x_{2,c}$ is the output of the second candidate correction model *i.e.*, longitude, y is the radon concentration, x_1, x_2 are the latitude and longitude inputs, and $\mathbf{w}_1, \mathbf{w}_2$ are the ANN weight vectors of the first and second candidate correction models respectively.

Fig 4-1: Candidate correction model with latitude as the output.

Fig 4-2: Candidate correction model with longitude as the output.

The first step is to determine the best correction model out of the two candidate correction models. The error criterion employed in selecting the correction model is the quality measure, E , mentioned in eq. (3.6). First, one of the candidate correction models, $f_{\text{ann},1}$ is considered. As shown in Fig 4-1, the inputs to the first candidate correction model are radon concentration and longitude with latitude being the output. Accordingly, the dataset is

modified by swapping the latitude and radon concentration columns. The modified dataset is randomly divided into training data and validation data. Using the training data, the $f_{\text{ann},1}$ is trained by adjusting $\mathbf{w}_1$ iteratively using *Neuromodeler* [69] software. This model gives an output $x_{1,c}$ (which closely approximates x_1). Then, quality measure E_1 of the $f_{\text{ann},1}$ is evaluated. This process is repeated for the $f_{\text{ann},2}$ and E_2 is calculated. Out of the two candidate correction models the one with the least validation error is considered as the correction model $f_{\text{ann},j}$ according to

$$j = \arg \left[\min_i (E_i) \right], \quad (4.5)$$

where $i=1, 2$.

As the final step, once the correction model $f_{\text{ann},j}$ is determined, it is paired with the less-accurate original neural model f_{ann} . A root-finding method is iteratively implemented to improve the accuracy of neural model f_{ann} until the condition $E_{\text{avg}} \leq E_{\text{user}}$ is achieved.

In this thesis we define a new root finding technique which has the advantages of Newton's method and the steepest-descent method [70]. In this technique, the sensitivity information, *i.e.*, partial derivatives of the correction model is used to determine the step-size and update direction. The chain rule of calculus is applied to the correction model $f_{\text{ann},j}$ to get the partial derivatives of the correction model $f_{\text{ann},j}$ with respect to output y

$$\frac{\partial x_{j,c}}{\partial y} = \sum_{k=1}^r \frac{\partial x_{j,c}}{\partial z_k} \frac{\partial z_k}{\partial \gamma_k} \frac{\partial \gamma_k}{\partial y}, \quad (4.6)$$

and

$$z_k = \frac{1}{1 + e^{-\gamma_k}}. \quad (4.7)$$

In eq. (4.6) and eq. (4.7), γ_k is the weighted sum of all inputs to $f_{\text{ann},j}$, z_k is the sigmoid activation function and r is the number of hidden neurons in $f_{\text{ann},j}$.

The output y from f_{ann} serves as the initial solution or starting point to the root finding algorithm based on “ f_{ann} and $f_{\text{ann},j}$ pair” and from their output y is iteratively improved until the condition $E_{\text{avg}} \leq E_{\text{user}}$ is achieved. The derivative information from the eq. (4.6) and eq. (4.7) play an important role in determining both step-size and update direction which result in much accurately predicted output y compared to the conventional ANN approach. The pseudo code and the flow chart (refer to Fig 4-3) for the proposed approach are shown below.

Pseudo Code of the Correction model based Approach

Step 1: Feed $(x_1, x_2, x_3, \dots, x_n)$, to f_{ann} to which gives an output y ;

Step 2: Feed $(x_1, x_2, x_3, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n)$ to the candidate correction model $f_{\text{ann},j}$ leading to $x_{j,c}$, an approximation of x_j ;

Step 3: Calculate the objective function,
$$E_{\text{obj}} = \frac{x_j - x_{j,c}}{|x_j|} * 100$$
 ;

Step 4: If $E_{\text{obj}} \ll E_{\text{user}}$, then return y and **STOP**, else, calculate residue $R = x_j - x_{j,c}$ and sensitivity of the correction model (eq. (4.6));

Step 5: Calculate δ which is the ratio of residue and sensitivity. Update $y = y - \alpha * \delta$ (To accelerate convergence, δ is multiplied by a constant $\alpha > 0$);

Step 6: Go to **Step 2** and updated y ;

Step 7: Return y and **STOP**.

Fig 4-3: Flow-chart illustrating the Correction model based ANN correction approach.

4.2 Ranked statistical performance measures for IAQ models

In this section a brief introduction of different performance measures used to assess the performance of different interpolation schemes is discussed. The rank (or weightage) of the performance measures change depending on whether the assessed air quality models

estimate extreme-end (peak-end, low-end), and mid-range contaminant concentrations. For example while assessing the performance of the air quality models when estimating extreme-end (peak-end, low-end) contaminant concentrations the rank of the performance parameter FV is four whereas the same parameter is ranked sixth when the assessed air quality models estimate mid-range contaminant concentrations. Since our study is focused on the indoor radon concentrations in the Ohio homes, we will follow the performance measures suggested for IAQ models discussed in [71]. Table 4.1 shows the ranked performance measures for IAQ models along with the acceptable limits for each measure based on extensive literature review [71].

Table 4.1: Ranked statistical performance measures for IAQ models.

Rank	IAQ Models		
	Peak-End Estimates	Low-End Estimates	Mid-Range Estimates
1	$0.9 \leq \rho \leq 1.0$	$0.9 \leq \rho \leq 1.0$	$0.9 \leq r \leq 1.0$
2	$0.75 \leq m \leq 1.25$	$0.75 \leq m \leq 1.25$	$0.75 \leq m \leq 1.25$
3	$-25 \leq (b/C0) \% \leq 25$	$-25 \leq (b/C0) \% \leq 25$	$-25 \leq (b/C0) \% \leq 25$
4	$-0.5 \leq FV \leq 0.5$	$-0.5 \leq FV \leq 0.5$	$0 \leq NMSE \leq 0.25$
5	$0.8 \leq Fa2 \leq 1.2$	$0.8 \leq Fa2 \leq 1.2$	$-0.25 \leq FB \leq 0.25$
6	$0.8 \leq RHC \leq 1.2$	$0 \leq MCM_2 \leq 1.2$	$-0.5 \leq FV \leq 0.5$
7	$0 \leq MCM_2 \leq 1.2$	$0.8 \leq MG \leq 1.2$	Bootstrap CI over r , $NMSE$, FB
8	$0.8 \leq MG \leq 1.2$	$0.8 \leq VG \leq 1.2$	$0.8 \leq Fa2 \leq 1.2$
9	$0.8 \leq VG \leq 1.2$	Bootstrap CI over MG , VG	$0 \leq MCM_2 \leq 1.2$
10	Bootstrap CI over MG , VG	Scatter plots	$0.8 \leq MG \leq 1.2$
11	Scatter plots	QQ plots	$0.8 \leq VG \leq 1.2$
12	QQ plots	-	Bootstrap CI over MG , VG
13	-	-	$0.8 \leq IOA_r \leq 1.0$
14	-	-	Scatter plots
15	-	-	QQ plots
16	$0.8 \leq IOA_r \leq 1.0$	$0.8 \leq IOA_r \leq 1.0$	-
17	$-15 \leq A_p \leq 15$	-	-

The following representation is used in the performance measures equations: y_k is the output of the correction model based ANN; d_k is the corresponding desired output of the k^{th} sample in the validation data, and N_v represents the number of validation samples.

4.2.1 Mean of training and validation data (M_t , M_v)

The mean statistics of training and validation data are expressed as

$$M_t = \frac{1}{N_v} \sum_{k=1}^{N_v} d_k, \quad (4.8)$$

and

$$M_v = \frac{1}{N_v} \sum_{k=1}^{N_v} y_k. \quad (4.9)$$

4.2.2 Standard deviation of training and validation data (SD_t , SD_v)

SD_t and SD_v statistics measure the variability from the mean statistics in training and validation data, respectively. They are expressed as

$$SD_t = \sqrt{\frac{1}{N_v} \sum_{k=1}^{N_v} |d_k - M_t|}, \quad (4.10)$$

And

$$SD_v = \sqrt{\frac{1}{N_v} \sum_{k=1}^{N_v} |y_k - M_v|}. \quad (4.11)$$

4.2.3 Least square slope (m) and intercept (b) regression statistics

The least square slope ' m ' and intercept ' b ' for a line " $y = b + mx$ " is expressed as

$$m = \left[\frac{N_v \sum_{k=1}^{N_v} d_k y_k - \left(\sum_{k=1}^{N_v} d_k \right) \left(\sum_{k=1}^{N_v} y_k \right)}{\sum_{k=1}^{N_v} d_k^2 - \left(\sum_{k=1}^{N_v} d_k \right)^2} \right], \quad (4.12)$$

and

$$b = \frac{\left[\sum_{k=1}^{N_v} y_k - m \left(\sum_{k=1}^{N_v} d_k \right) \right]}{N_v}. \quad (4.13)$$

4.2.4 Paired peak accuracy (A_p)

Paired peak accuracy (A_p) measure serve as a supplementary test of model performance and is given by

$$A_p = \frac{y_r - d_r}{d_r} * 100\%, \quad (4.14)$$

where d_r is the peak value of the validation data and y_r is the corresponding output from the correction model based ANN.

4.2.5 Geometric mean bias (MG)

Geometric mean bias (MG) is defined as the exponential function of the difference between means of the logarithmic function of observed values and logarithmic function of predicted values and is expressed as

$$MG = \exp \left[\overline{\ln d_k} - \overline{\ln y_k} \right]. \quad (4.15)$$

4.2.6 Geometric mean variance (VG)

Geometric mean variance (VG) is defined as the exponential function of mean of square of the difference between logarithmic function of observed values and logarithmic function of observed values and is expressed as

$$VG = \exp\left(\overline{(\ln d_k - \ln y_k)^2}\right). \quad (4.16)$$

4.2.7 Fractional bias (*FB*)

Fractional bias (*FB*) is defined as the mean error that measures the residual of the observed and predicted concentrations. Typically, *FB* varies between -2 and +2 with an ideal value of zero. It is expressed as

$$FB = 2 \left[\frac{\overline{d_k} - \overline{y_k}}{\overline{d_k} + \overline{y_k}} \right]. \quad (4.17)$$

4.2.8 Fractional variance (*FV*)

The fractional variance (*FV*) is defined as the normalization of the mean bias of the variances of the observed and predicted values. The ideal value of *FV* is zero.

$$FV = 2 \left[\frac{\sigma_{d_k}^2 - \sigma_{y_k}^2}{\sigma_{d_k}^2 + \sigma_{y_k}^2} \right], \quad (4.18)$$

where σ_{y_i} is the variance of the observed values *i.e.*, the input and σ_y is the variance of the corresponding outputs from correction model ANN.

4.2.9 Normalized Mean Square Error (*NMSE*)

Normalized mean square error (*NMSE*) emphasizes the scatter of the whole data and estimates the total deviation between observed and predicted values. To ensure that the *NMSE* is not biased towards over prediction or under prediction normalization is done by the denominator. Smaller values of *NMSE* indicate better model performance, the ideal value being zero.

$$NMSE = \frac{\overline{(d_k - y_k)^2}}{d_k * y_k}. \quad (4.19)$$

4.2.10 Factor of Two (*Fa2*)

Factor of two (*Fa2*) is defined as the percentage of predictions of the observed values within a factor of two. The ideal value of *Fa2* is 1 (or 100%).

$$Fa2 = \text{Fraction of data satisfying } 0.5 \leq \frac{y_k}{d_k} \leq 2.0. \quad (4.20)$$

4.2.11 Coefficient of correlation (*r*)

Coefficient of correlation (*r*) describes the proportional change of the means of two quantities. The ideal value of *r* is 1.0 which implies good agreement between the observed and predicted values.

$$r = \frac{\left(\overline{(d_k - \bar{d}_k)} * \overline{(y_k - \bar{y}_k)} \right)}{\sigma_{y_k} * \sigma_{d_k}}. \quad (4.21)$$

4.2.12 Spearman correlation coefficient (ρ)

Spearman correlation coefficient (ρ) is defined as the correlation coefficient (*r*) between the ranked variables. It's ideal value is 1.0 and is expressed as

$$\rho = \frac{\sum_{k=1}^{N_v} (d_k - \bar{d}_k) * (y_k - \bar{y}_k)}{\sqrt{\sum_{k=1}^{N_v} (d_k - \bar{d}_k)^2 * \sum_{k=1}^{N_v} (y_k - \bar{y}_k)^2}}. \quad (4.22)$$

4.2.13 Model comparison measure (*MCM₂*)

Model comparison measure (*MCM₂*) is expressed as

$$MCM_2 = Avg(afb) + 2 * Std(afb), \quad (4.23)$$

and

$$AFB = 2 \left[\frac{|\overline{d_k} - \overline{y_k}|}{|\overline{d_k} + \overline{y_k}|} \right], \quad (4.24)$$

where Avg (AFB) and Std (AFB) are the average and standard deviation of the absolute fractional bias (AFB) which is calculated from eq. (4.24). The smaller value of MCM_2 indicates that the model on average compares best with observations.

4.2.14 Robust highest concentration (RHC)

Robust highest concentration (RHC) performance measure is used to study the peak concentrations that are highly variable. It is expressed as

$$RHC = C(R) + \Theta \times \ln \left(\frac{3R-1}{2} \right), \quad (4.25)$$

Where $C(R)$ is the R^{th} largest value; Θ is the average of the $R-1$ largest values; and R is the minimum of the number of values used to distinguish the “upper end” of the distribution and number of values exceeding a threshold value. R value is “26” for frequency distributions involving a year’s worth of values (averaging times of 24 hr or less).

4.2.15 Revised Index of Agreement (IOA_r)

Revised Index of Agreement (IOA_r) indicates the sum of the magnitudes of the differences between the predicted and observed deviations about the observed mean of the sum of the magnitudes of the ideal model and observed deviations about the observed mean.

IOA_r ranges from -1 to +1 and is expressed as

$$IOA_r = 1 - \frac{\sum |y_p - y_{ip}|}{2 \sum |y_{ip} - M_o|}, \quad (4.26)$$

where M_o is the mean of the observed values.

4.2.16 Scatter plots

Scatter plots are used to determine the bias and error in the observed and predicted values. They are better suited for correlation measures.

4.2.17 Quantile-Quantile ($Q-Q$) plots

Quantile-Quantile ($Q-Q$) plots are cumulative frequency distributions of observed and modeled values over their entire ranges. They provide a visual characterization of the scattering of predicted and estimated values with respect to the 1:1 line. Over the 1:1 line signifies a model over prediction and under the 1:1 line signifies a model under prediction.

Quality of the proposed correction based ANN model against the existing techniques is evaluated based on these performance measures and is discussed in next section.

4.3 Results and Comparison

For the correction modeling approach, out of the two candidate correction models the model with the least validation error is selected for the implementation of root finding algorithm along with the less-accurate original ANN model. Table 4.2 shows the training and validation errors of both the candidate correction models and it is clear that the first correction model with latitude as output has the least validation error and is preferred over the latter. Since the correction model and the original less accurate conventional ANN model form a mutually supportive pair while implementing the algorithm, it is logical to keep the structure of both the models same. For this purpose, initially the number of hidden neurons in the correction model is fixed at 10.

Table 4.2: Training and Validation Errors of candidate correction models.

Model(s)	Training Error (%)	Validation Error	
		E_{avg} (%)	E_{worst} (%)
Candidate Correction Model 1	15.88	15.79	49.82
Candidate Correction Model 2	21.19	23.15	59.73

As the correction model is determined, the proposed pseudo code is applied to the mutually supportive pair of the conventional ANN model and the correction model, each with 10 hidden neurons and the radon output is estimated. The validation error E_{avg} shown in eq. (3.6) is calculated. It is observed that the validation error E for the correction model ANN approach is well below the E_{user} of 5% compared to the conventional ANN approach as shown in Table 4.3.

Table 4.3: Comparison of Validation Errors of the Conventional ANN and Correction Model Based ANN Models.

Model(s)	Validation Error	
	E_{avg} (%)	E_{worst} (%)
Conventional ANN	12.13	72.19
Correction model based ANN	3.40	35.61

Further, the performance of the proposed method is compared with the traditional interpolation techniques and the conventional ANN method based on various model performance measures discussed in Section 4.2. ArcGIS geostatistical analyst is used for estimating radon data utilizing traditional interpolation techniques such as kriging, RBF, IDW, GPI, and LPI. The same set of training and validation data that were used to model

radon concentrations in the ANN based approaches are used for modeling radon concentrations from the interpolation techniques.

Table 4.4 and Table 4.5 present a summary of the performance measures of the traditional interpolation techniques; the conventional and proposed ANN based models for evaluation purposes. BOOT v2.0 is used to compute the performance measures of M , SD , r , $NMSE$, FB , $Fa2$, MG , VG and FV . The remaining performance measures of b/C_0 , m , IOA , A_p , RHC , MCM_2 , ϱ can be computed using Microsoft® Excel and the equations presented in Section 4.2.

The ranked performance measures discussed in Section 4.2 are used to determine the best model in estimating the extreme-end (peak-end, low-end), and mid-range radon concentrations. This is achieved by eliminating all the models whose performance measures are not within the limit mentioned in Table 4.1. Elimination of the models is a step by step procedure where in each step a different performance measure is used to determine which models are to be considered for further evaluation. Initially, all the models are evaluated based on rank one performance measure and those models that does not satisfy the rank one performance measure criterion are eliminated. In the next step, evaluation of the remaining models is continued based on the rank two performance measure and so on. This evaluation procedure is continued until one is left with only one model which is considered the best compared to other models in predicting radon concentrations.

From Table 4.4, one can observe that only the correction model based ANN was within the acceptance criterion of $0.9 \leq \varrho \leq 1.0$ (ranked one for extreme-end estimates) and $0.9 \leq r \leq 1.0$ (ranked one for mid-range estimates). Apart from ϱ and r , the traditional interpolation techniques and the conventional ANN model failed to satisfy the $0.75 \leq m \leq 1.25$ criterion (ranked two for the extreme-end and mid-range estimates) and $-25 \leq b/C_0 (\%) \leq 25$ criterion

(ranked three for the extreme-end and mid-range estimates). The correction model based ANN satisfied the rank two and rank three criteria as well. Therefore, at this step, the traditional interpolation techniques and the conventional ANN model should not be considered for further evaluation. Alternatively, incorporating more variables such as gravel deposits and the presence of glacial sand that are correlated to radon concentrations helps in reconsidering these models [36]. However, even after reconsideration only the correction model based ANN meet the criterion of $-0.5 \leq FV \leq 0.5$ (ranked four for the extreme-end estimates) while the remaining interpolation techniques were out of range. At this point, we can conclude that only the correction model based ANN is considered best compared to the remaining models in estimating the extreme-end values of radon concentrations.

Let us continue the evaluation of the remaining models after reconsideration in estimating the mid-range values. From Table 4.4 and Table 4.5, it is clear that kriging, LPI, conventional ANN approach and the correction model based ANN approach meet the criteria of $0 \leq NMSE \leq 0.25$ (ranked four) and $-0.25 \leq FB \leq 0.25$ (ranked five). At this stage, all the above interpolation techniques are considered for further evaluation. However, only the correction model based ANN satisfied the $-0.5 \leq FV \leq 0.5$ criterion (ranked six) and the remaining interpolation techniques are eliminated. Eventually, even in the case of estimating the mid-range values of radon concentration the correction model based ANN proved to be the best model.

As mentioned in Chapter 2, the training data pattern greatly affects the accuracy of traditional interpolation schemes. Since the training data in this case is randomly selected from the total radon dataset no traditional interpolation technique works universally best which leads to poor interpolation and eventually these techniques fail to satisfy the above criterion for the ranked performance measures. From Table 4.4 and Table 4.5, it is clear that

the performance measures of conventional ANN model is better than all the traditional interpolation techniques because its performance does not entirely depend on the training data pattern and the network weights are adjusted accordingly during training. However, the model accuracy of the conventional ANN model is not enough to satisfy the ranked performance parameters criterion. Table 4.4 and Table 4.5 shows that the correction model based ANN model satisfies the criterion for all the ranked performance measures. Therefore, it can be concluded that the proposed approach is flexible and it interpolates data accurately for any randomly generated dataset even with a lesser degree of correlation between them.

Table 4.4: Performance measures computed for traditional interpolation techniques,
Conventional ANN and Correction model based ANN model.

Model(s)	M	SD	E_{avg} (%)	$NMSE$	r	ρ	m	(b/C_0) %
Observed	4.22	2.05	0.00	0.00	1.000	1.000	1.000	0.00
Kriging	3.71	1.03	13.96	0.25	0.369	0.417	0.186	69.35
GPI	3.38	0.21	18.98	0.34	0.151	0.166	0.015	78.48
LPI	3.69	1.00	15.96	0.25	0.382	0.441	0.185	68.96
IDW	3.76	1.29	17.98	0.30	0.261	0.337	0.164	72.77
RBF	3.72	1.15	15.52	0.29	0.274	0.335	0.153	72.93
Conventional ANN	3.85	1.05	12.13	0.18	0.583	0.597	0.297	61.44
Correction model based ANN	4.51	1.83	3.40	0.03	0.947	0.909	0.845	22.33

Table 4.5: Performance measures computed for traditional interpolation techniques,
Conventional ANN and Correction model based ANN model (Cont.).

Model(s)	<i>FB</i>	<i>FV</i>	<i>MCM₂</i>	<i>Fa2</i>	<i>IOA_t</i>	<i>A_p</i>	<i>VG</i>	<i>MG</i>	<i>RHC</i>
Observed	0.000	0.000	0.000	1.00	1.000	0.00	1.00	1.000	1.000
Kriging	0.128	0.661	0.846	0.69	0.124	-64.89	1.19	1.076	0.369
GPI	0.222	1.636	0.952	0.62	0.043	-70.86	1.21	1.137	0.151
LPI	0.133	0.693	0.874	0.58	0.124	-66.21	1.18	1.079	0.382
IDW	0.114	0.553	0.967	0.67	0.042	-64.45	1.26	1.084	0.261
RBF	0.125	0.563	0.952	0.71	0.079	-65.18	1.24	1.084	0.274
Conventional ANN	0.092	0.649	0.786	0.78	0.203	-49.46	1.14	1.038	0.583
Correction model based ANN	-0.067	0.113	0.549	1.00	0.728	0.42	1.04	0.909	0.947

To complete the performance analysis of all the models, the graphical tools *i.e.*, the scatter plots and $Q-Q$ plots for all the interpolation schemes are shown in Fig 4-4 to Fig 4-10 and Fig 4-11 to 4-17 respectively. The interpolation scheme that has the observed and predicted value points plotted close to the 1:1 line (identity line) for both the scatter plots and the $Q-Q$ plots is considered as the best model.

Fig 4-4: Scatter plot of kriging technique.

Fig 4-5: Scatter plot of GPI technique.

Fig 4-6: Scatter plot of LPI technique.

Fig 4-7: Scatter plot of IDW technique.

Fig 4-8: Scatter plot of RBF technique.

Fig 4-9: Scatter plot of Conventional ANN model.

Fig 4-10: Scatter plot of Correction Model Based ANN model.

Fig 4-11: $Q-Q$ plot of kriging technique.

Fig 4-12: $Q-Q$ plot of GPI technique.

Fig 4-13: Q - Q plot of LPI technique.

Fig 4-14: Q - Q plot of IDW technique.

Fig 4-15: $Q-Q$ plot of RBF technique.

Fig 4-16: $Q-Q$ plot of Conventional ANN technique.

Fig 4-17: $Q-Q$ plot of Correction model based ANN technique.

Considering the closeness of predicted values to the given values of each performance measure and graphical representations, one can infer that the correction model based ANN approach proposed in this paper is the best model for modeling and estimating the extreme-end (peak-end, low-end), and mid-range radon concentrations.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This thesis demonstrates the use of ANNs for modeling radon concentrations in Ohio. The available radon concentration data was reformatted to suit ANN training by converting the zip code information into latitude and longitude data. From Table 3.2, the best neural model, *i.e.* ANN model with 10 hidden neurons, trained using quasi-Newton is chosen as the best MLP3 approach for the purpose of estimating radon concentrations in zip codes with no available data. From Table 3.1, it can be seen that more training data results in a lesser validation error. Based on this observation, it is recommended that 90% of the available data be used for training to accurately predict the radon concentrations in the remaining zip codes with no available data. However, from Table 3.2 it is clear that the condition $E_{\text{avg}} \leq E_{\text{user}}$ is not achieved by the conventional ANN approach.

Therefore, in the second phase of the thesis, a new correction based ANN modeling approach for estimating radon concentrations is proposed. Simple MLP3 networks are used to develop both the original less-accurate model and the correction model. The proposed approach is compared with the conventional ANN modeling approach and traditional interpolation techniques. From Table 4.4 and Table 4.5, it can be confirmed that the

proposed correction model based ANN modeling approach delivers the best model compared to other models in predicting both the extreme-end (peak-end, low-end), and mid-range radon concentrations. The proposed approach also eliminates the need for additional knowledge input *i.e.*, uranium which tends to increase the structural complexity as in the case of KBNNs.

5.2 Future Work

Exploring further advanced techniques based on ANN modeling could lead to better prediction results. Furthermore, several other techniques exist in literature that have not yet been explored for application to this problem, *e.g.* Support Vector Machines, which is being used extensively in the field of image processing for reconstruction of images. Alternatively, a different root finding method can be applied in the correction based ANN approach. This requires mathematical knowledge in assessing the pros and cons of different root finding methods necessitating additional research.

References

1. Price, P.N., and Gelman A., *Statistics: A Guide to the Unknown*, 4th ed., Florence, KY: Cengage Learning, 2006.
2. “Radon Information,” [online document], 2010 June 10th, available: <http://www.epa.gov/radon/pubs/citguide.html>.
3. Harrell, J., Belsito, M.E., and Kumar, A., “Radon hazards associated with outcrops of Ohio shale in Ohio,” *Environmental Geology and Water Science*, Vol. 18, pp. 17-26 1991.
4. Harrell, J., McKenna, J.P., and Kumar, A., *Geological controls on indoor Radon in Ohio, Investigation no. 144*, Division of Geological Survey, Dept. of Natural Resources, State of Ohio, Vol. 36, 1993.
5. Heydinger, A., Kumar, A., and Harrell, J., *An indoor Radon information system. Environmental Software*, Vol. 6, pp. 194-201, 1991.
6. Kumar, A., Harrell, J., and Heydinger, A., “Ohio goes online to combat indoor Radon,” *EM (by Air and Waste Management Association)*, pp. 10-12, 2001.

7. Kumar, A. and Varadarajan, C., “Development of Ohio radon information system. Proceeding of Environmental Data Analysis–Assessing Health and Environmental Impacts,” *Developing Policy, and Achieving Regulatory Compliance Conference, Air and Waste Management Association*, Oak Brook, IL 2005.
8. Kumar, A., Heydinger A., and Harrell, J., *Development of an indoor information system for Ohio and its application in the study of the geology of radon in Ohio, Prepared for Ohio Air quality Development Authority*, pp. 192, 1990.
9. Kumar, A., Sud, A., and Heydinger, A., “Application of ORACLE 7.3 database management system for the development of an environmental database,” *Environmental Progress*, Vol. 17, pp. F11-F14, 1998.
10. Ojha, S., Thomas, S.J., and Kumar, A., “Experience in integrating Geographical Information Systems (GIS) to an indoor Radon database,” *Environmental Progress*, Vol. 20, pp. O7-O10, 2001.
11. Joshi, A., Manne, G.K., and Kumar, A., “Management of Ohio’s Radon data with MS access/SQL server 7.0,” *Environmental Progress*, Vol. 21, pp. D8-D12 2002.
12. Kumar, A., Tandale, A., Kalapati, R.S., and Ghose, S., “Management of radon mitigation data in the state of Ohio,” *Environmental Progress*, Vol. 22, pp. O19-O24 2003.
13. Kumar, A., Varadarajan, C., and Kadiyala, A., “Management of Radon Data in the State of Ohio, U.S.A,” *The Open Environmental and Biological Monitoring Journal*, Vol. 4, pp. 57-71, 2003.

14. Kumar, A., Velagapudi, S., Nimmatoori, P., Kadiyala, A., and Manthena, D., *Analysis of Indoor Radon Mitigation Systems in Ohio, USA, in Radon: Properties, Applications and Health Hazards*, New York: Nova Science Publishers, 2012.
15. Akkala, A., Devabhaktuni, V., Kumar, A., and Bhatt, D., “Development of an ANN interpolation scheme for estimating missing radon concentrations in Ohio,” *The Open Environmental and Biological Monitoring Journal*, Vol. 4, pp. 21-31, 2011.
16. Wang, F., Devabhaktuni, V., and Zhang, Q.J., “Neural network structures and training algorithms for RF and microwave applications,” *International Journal of RF and Microwave Computer Aided Engineering*, Vol. 9, pp. 216–240, 1999.
17. Devabhaktuni, V., Xi, C., Wang, F., and Zhang, Q.J., “Robust training of microwave neural models,” *International Journal of RF and Microwave Computer Aided Engineering*, Vol. 12, pp. 109-124, 2002.
18. Zhang, Q.J., Gupta, K.C., and Devabhaktuni, V., “Artificial neural networks for RF and microwave design-From theory to practice,” *IEEE Transactions of Microwave Theory*, Vol. 51, pp. 1339-1350, 2003.
19. Malik, N., “Artificial neural networks and their applications,” *National Conference on Unearthing Technological Developments and their Transfer for Serving Masses GLA ITM*, Mathura, India, 2005.
20. Wang, F., Devabhaktuni, V., Xi, C., and Zhang, Q.J., “Neural network structures and training algorithms for RF and microwave applications,” *International Journal of RF and Microwave Computer Aided Engineering*, Vol. 9, pp. 216-40, 1999.

21. Wang, F. and Zhang, Q.J., "Knowledge-based neural models for microwave design," *IEEE Transactions on Microwave Theory*, Vol. pp. 45, 2333-2343, 1997.
22. Akkala, A., Bhatt, D., Devabhaktuni, V., and Kumar, A., "Knowledge-Based Neural Network Approaches for Modeling and Estimating Radon Concentrations," *Environmental Progress and Sustainable Energy*, doi:10.1002/ep.11617, 2012.
23. Li J., and Heap A.D., "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecological Informatics*, Vol 6, no. 3-4, pp. 228-241, 2011.
24. Akkala, A., *Development of artificial neural networks based interpolation techniques for modeling and estimation of radon concentrations in ohio*, M.S Thesis, Department of Electrical Electronics and Computer Sciences, University of Toledo, 2010.
25. C.O. Ward., "Better global polynomial approximation for image rectification," *International Journal of Modeling & Simulation*, Vol. 28, no. 3, pp. 299-308, 2008.
26. De Carvalho, L.G., Alves, M.D.C., De Oliveira, M.S., Vianello, R.L., Sedyama, G.C., and De Carvalho, L.M.T., "Multivariate geostatistical application for climate characterization of Minas Gerais State, Brazil." *Theoretical and Applied Climatology*," Vol. 102, no. 3-4, pp. 417-428, Nov 2010.
27. Hwang, Y., Clark, M., Rajagopalan, B., and Leavesley, G., "Spatial interpolation schemes of daily precipitation for hydrologic modeling," *Stochastic Environmental Research and Risk Assessment*, Vol. 26, no. 2, pp. 295-320, Feb 2012.

28. Kavian, A., Fathollah, N.Y., Habibnejad, M., and Soleimani, K., "Modeling Seasonal Rainfall Erosivity on a Regional scale: A case Study from Northeastern Iran," *International Journal of Environmental Research*, Vol. 5, no. 4, pp. 939-950, Fall 2011.
29. Eldrandaly, K.A., and Abu Zaid, M.S., "Comparison of Six GIS-Based Spatial Interpolation Methods for Estimating Air Temperature in Western Saudi Arabia," *Journal of Environmental Informatics*, Vol. 18, no. 1, pp. 38-45, Sep 2011.
30. Chen, Tong-bin., Lei, Mei., Yang, Jun., Guo, Qing-jun., Song, Bo., and Zhou, Xiao-yong., "Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis," *Chemosphere*, Vol. 82, no. 3, pp. 468-476, Jan 2011.
31. Uygur, V., Irvem, A., Karanlik, S., and Akis, R., "Mapping of total nitrogen, available phosphorous and potassium in Amik Plain, Turkey," *Environmental Earth Sciences*, Vol. 59, no. 5, pp. 1129-1138, Jan 2010.
32. Chahouki, M.A.Z., Chahouki, A.Z., and Ahvazi, L.K., "Comparing Geostatistical Approaches for Mapping Soil Properties in Poshtkouh Rangelands of Yazd Province, Iran," *Vegetos*, Vol. 24, no. 1, pp. 77-88, Jun 2011.
33. Abichou, T., Chanton, J., Powelson, D., Fleiger, J., Escoriaza, S., Lei, Y., and Stern, J., "Methane flux and oxidation at two types of intermediate landfill covers," *Waste Management*, Vol. 26, no. 11, pp. 1305-1312, 2006.
34. Li, W., Liu, W.L., Zhang, Y., Yin, W., and Liu, Y., "On the comparison of spatial interpolation methods of marine temperature and salinity based on Arcgis software: a

- case study of Tianjin coastal waters in the Bohai bay,” *Sixth International Symposium on Digital Earth: Models, Algorithms, and Virtual Reality*, Vol. 7840, 2010.
35. Burrough P.A, McDonnell R.A., *Principles of geographical information systems*, Clarendon Press: Oxford, 1998.
 36. “ArcGis 9.2 Desktop Help,” [Online document], 2010 June 10th, available: [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How_Radial_Basis_Functions_\(RBF\)_work](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How_Radial_Basis_Functions_(RBF)_work).
 37. Li, Y.J., Fang, Y.J., and Li, Q., “Application of A Neural Network Predictive Control Based on GGAP-RBF for the Supercritical Main Steam,” *Fourth International Conference on Machine Vision (ICMV 2011): Machine Vision, Image Processing, and Pattern Analysis*, Vol. 8349, 2012.
 38. Li, X., Li, Y.Y., Seem, J.E., and Lei, P., “Maximum Power Point Tracking for Photovoltaic Systems Using Adaptive Extremum Seeking Control,” *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pp. 1503-1508, 2011.
 39. Lin, Wen-Bin., Chen, Chien-An., and Chiang, Huann-Keng., “Design and Implementation of a Sliding Mode Controller Using a Gaussian Radial Basis Function Neural Network Estimator for a Synchronous Reluctance Motor Speed Drive,” *Electric Power Components and Systems*, Vol. 39, no. 6, pp. 548-562, 2011.
 40. Purwanto, Eswaran, C., Logeswaran, R., and Rahman, A.R.A., “Prediction Models for Early Risk Detection of Cardiovascular Event” *Journal of Medical Systems*, Vol. 36, no. 2, pp. 521-531, Apr 2012.

41. Sarevska, M., and Mastorakis, N., "Regular Antenna Array Synthesis Using Neural Network," *Recent Researches in Telecommunications, Informatics, Electronics and Signal Processing*, pp. 37-40, May 2011.
42. Yu, Hwa-Lung., Chan, Chang-Chuan., Cheng, Tsun-Jen., and Chen, Chu-Chih., "Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter PM 2.5 concentrations," *Atmospheric Environment*, Vol. 54, pp. 1-8, 2012.
43. Palmer, David J., "Use of regression kriging to develop a Carbon: Nitrogen ratio surface for New Zealand," *Geoderma*, Vol. 183-184, pp. 49-57, Aug 2012.
44. "Remote-sensing image analysis and geostatistics," *International Journal of Remote Sensing*, Vol. 33, no. 18, pp. 5644-5676, Sep 2012.
45. Goyal, Manish Kumar., and Adeloye, A.J., "Downscaling of Precipitation for Lake Catchment in Arid Region in India using Linear Multiple Regression and Neural Networks," *Open Hydrology Journal*, pp. 122-136, 2010.
46. Fooladmand, Hamid Reza¹., "Evaluation of some equations for estimating evapotranspiration in the south of Iran," *Archives of Agronomy & Soil Science*, Vol. 57, no. 7, pp. 741-752, Oct 2011.
47. Kumar, A.R.S., Ojha, C.S.P., Goyal, M.K., Singh, R.D., and Swamee, P.K., "Modeling of Suspended Sediment Concentration at Kasol in India Using ANN, Fuzzy Logic, and Decision Tree Algorithms," *Journal of Hydrologic Engineering*, Vol. 17, no. 3, pp. 394-404, Mar 2012.

48. Popa, M., Ileana, I., Vosniakos, F., Popa, D., and Golumbeanu, M., "Study on the role of neural networks in environmental pollution forecasts," *Journal of Environmental Protection and Ecology*, Vol. 12, no.4, pp. 1948-1953, 2011.
49. Carnevale, C., Finzi, G., Guariso, G., Pisoni, E., and Volta, M., "Surrogate models to compute optimal air quality planning policies at a regional scale," *Environmental Modeling and Software*, Vol. 34, pp. 44-50, Jun 2012.
50. Chaki, S., and Ghosal, S., "Application of an optimized SA-ANN hybrid model for parametric modeling and optimization of LASOX cutting of mild steel," *Production Engineering*, Vol. 5, no. 3, pp. 251-262, 2011.
51. Taskin, Mustafa., Caligulu, Ugur., and Gur, Ali., "Modeling adhesive wear resistance of Al-Si-Mg-/SiC PM compacts fabricated by hot pressing process, by means of ANN," *International Journal of Advanced Manufacturing Technology*, Vol. 37, no. 7-8, pp. 715-721, 2008.
52. Balcilar, M., Dalkilic, A.S., Suriyawong, A ., Yiamsawas, T., and Wongvvises, S., "Investigation of pool boiling of nanofluids using artificial neural networks and correlation development techniques," *International Communications in Heat and Mass Transfer*, Vol. 39, no. 3, pp. 424-431, Mar 2012.
53. Ghanbary, F., Modirshahla, N., Khosravi, M ., and Behnajady, M.A., "Synthesis of TiO₂ nanoparticles in different thermal conditions and modeling its photo catalytic activity with artificial neural network," *Journal of Environmental Sciences-China*, Vol. 24, no. 4, pp. 750-756, 2012.

54. Oloke, J.K., Lateef, A., and Adesiyun, M.O., "Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm," *Renewable Energy: An International Journal*, Vol. 46, pp. 276-281, Oct 2012.
55. "Artificial intelligence methods applied for quantitative analysis of natural radioactive sources," *Annals of Nuclear Energy*, Vol. 45, pp. 73-79, Jul 2012.
56. Zarei-Hanzaki, A., Abedi, H.R., and Haghdadi, N., "Artificial neural network modeling to predict the high temperature flow behavior of an AZ81 magnesium alloy," *Materials & Design*, Vol. 39, pp. 390-396, Aug 2012.
57. Atashi, Hossein., and Hassanpourfard, Mahtab., "Studying the abrasion behavior of rubbery materials with combined design of experiment-artificial neural network," *Chinese Journal of Polymer Science (Springer Science & Business Media B.V.)*, Vol. 30, no. 4, pp. 520-529, Jul 2012.
58. Mosavi, M.R., Sam, A., and Vaghei, Y., "Recovery and grade accurate prediction of pilot plant flotation column concentrate: Neural network and statistical techniques," *International Journal of Mineral Processing*, Vol. 110-111, pp. 140-154, Jul 2012.
59. Paik, Byeong-Cheon., "Inference model derivation with a pattern analysis for predicting the risk of microbial pollution in a sewer system," *Stochastic Environmental Research & Risk Assessment*, Vol. 26, no. 5, pp. 695-707, Jul 2012.
60. Geng, Z.Q., Wang, Y.Q., Zhang, Y.Y., and Zhu, Q.X., "Integrated Input Training Neural Network PCA and RBF for Chemical Process Modeling," *Materials, Mechatronics and Automation*, pp. 467-469, 2011.

61. Coutinho, J., Kumar, A., and Kadiyala, A., *Development of artificial neural network models for predicting maximum daily PM10 concentration near Cincinnati, Ohio*, In: Benjamin V, Jozsi S (9) Horizons in Earth Science Research 2, Nova Science Publishers Inc. ISBN: 978-1-61668-709-0, 2010.
62. Diaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G., and Moncada-Herrera, J.A., “A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile,” *J Atmos Environ*, Vol. 42, pp. 8331–8340, 2008.
63. Caselli, M., Trizio, L., de Gennaro, G., and Ielpo, P., “A simple feedforward neural network for the PM10 forecasting: Comparison with a radial basis function network and a multivariate linear regression model,” *J Water Air Soil Poll*, Vol. 201, pp. 365–377, 2009.
64. Papanastasiou, D.K., Melas, D., and Kioutsoukis, I., “Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized mediterranean city,” *J Water Air Soil Poll*, Vol. 182, pp. 325–334, 2007.
65. Haykin, S., *Neural Networks – A comprehensive foundation*, New Jersey: Pearson Education, 1999.
66. Bishop, C.M., *Neural networks for pattern recognition*, Oxford: Oxford Press, 1995.
67. Karlik, B., and Olgac, A.V., “Performance analysis of various activation function in generalized MLP architectures of neural networks,” *Int. J. Artificial Intelligence and Expert Systems*, Vol. 1, pp. 111–122, 2011.

68. Mendhurwar, K., Sundani, H., Khan, Z., Bhattacharya, P., and Devabhaktuni, V.,
Device modeling using correction model based ANN and support vector regression,
International Symposium on Microwave and Optical Technology, New Delhi, India, 2009.
69. Zhang, Q.J., *NeuroModeler*, Department of Electronics, Carleton University, 1125
Colonel By Drive, Ottawa K1S5B6, Ontario, Canada, 1999.
70. Peressini, A.L., Sullivan, F.E., and Uhl Jr., J.J., *The mathematics of nonlinear programming*,
New York: Springer, 1988.
71. Kadiyala, A., and Kumar, A., *Guidelines for operational evaluation of air quality models:
Outdoor and Indoor Air Quality Models*, Saarbrücken: LAP LAMBERT Academic
Publishing, 2012.

Appendix A

Source Code for Correction Model Based ANN Modeling approach

```
m = 1; c = 0;
m=1;
for k=1:110

    fid1 = fopen('data.dat');
    inp = fscanf(fid1, '%g',330);
    inp1 = inp(m);
    inp2 = inp(m+1);
    outp = inp(m+2);
    inpv = [inp1 inp2];
    fclose(fid1);

    ybad = orig(inpv);
    fid2 = fopen('trad.txt','a');
    fprintf(fid2, '%f\n',ybad);
    fclose(fid2);

    inp1c = ybad;
    inp2c = inp2;
    inpc = [inp1c inp2c];
    outpc = correction1(inpc);
    fid3 = fopen('mod.txt','a');
    fprintf(fid3, '%f\n',outpc);
    fclose(fid3);

    R = outpc-inp1;
    fid4 = fopen('error.txt','a');
    fprintf(fid4, '%f\n',R);
    fclose(fid4);
```

```

x(1)=-1.0+(2.0)*(inp1c-(1.0)) / ((12.24989796) - (1.0));
x(2)=-1.0+(2.0)*(inp2c-(-84.79165)) / ((-80.53175) - (-84.79165));
z(1) = 1.0 / ( 1.0 + exp(-1.0 * (2.8966+x(1)*(1.84256)+x(2)*(4.24076))));
z(2) = 1.0 / ( 1.0 + exp(-1.0 * (2.55345+x(1)*(-7.48095)+x(2)*(18.9876))));
z(3) = 1.0 / ( 1.0 + exp(-1.0 * (10.3315+x(1)*(8.56266)+x(2)*(-17.4902))));
z(4) = 1.0 / ( 1.0 + exp(-1.0 * (3.9155+x(1)*(-10.1916)+x(2)*(25.4966))));
z(5) = 1.0 / ( 1.0 + exp(-1.0 * (4.87492+x(1)*(3.81473)+x(2)*(6.86083))));
z(6) = 1.0 / ( 1.0 + exp(-1.0 * (7.82337+x(1)*(8.1067)+x(2)*(-12.013))));
z(7) = 1.0 / ( 1.0 + exp(-1.0 * (-2.16329+x(1)*(12.3752)+x(2)*(-12.6631))));
z(8) = 1.0 / ( 1.0 + exp(-1.0 * (5.63282+x(1)*(-0.774374)+x(2)*(21.7944))));
z(9) = 1.0 / ( 1.0 + exp(-1.0 * (0.0256262+x(1)*(-5.97796)+x(2)*(18.1607))));
z(10) = 1.0 / ( 1.0 + exp(-1.0 * (1.42298+x(1)*(0.396632)+x(2)*(-0.100451))));

y(1) = 3.17523+z(1)*(1.12634)+z(2)*(-1.54783)+z(3)*(-0.715785)+z(4)*(1.50854)+z(5)*(-
0.514922)+z(6)*(0.877342)+z(7)*(0.2171)+z(8)*(-0.300982)+z(9)*(0.269883)+z(10)*(-
4.0892);

output(1) = 38.438617+(y(1)-(0.0))*((41.934916) - (38.438617))/((1.0)-(0.0));

bw = [1.84256 -7.48095 8.56266 -10.1916 3.81473 8.1067 12.3752 -0.774374 -5.97796
0.396632];
dw = [1.12634 -1.54783 -0.715785 1.50854 -0.514922 0.877342 0.2171 -0.300982 0.269883 -
4.0892];
ew = 3.17523;
sensitivity1 = 0.0;
for i = 1:10
sensitivity1 = sensitivity1 + dw(i)*z(i)*(1-z(i))*bw(i);
end
sensitivity1;
sensitivity2 = sensitivity1 * ((12.24989796) - (0.1)/2.0);
sensitivity = sensitivity2 * ((41.934916) - (38.448563));

J = sensitivity;
R;
alpha = 0.1;
Euser = 0.1;iter = 0; ymod = ybad;
Eobj = ((inp1-outpc)/inp1)*100;
while(abs(Eobj)>Euser)
deltay = abs(-R/J);
if(ybad<outp)
ybad = ymod + (alpha * deltay);
elseif(ybad>outp)
ybad = ymod - (alpha * deltay);
end
ymod = ybad;
if(ybad>outp)

```



```

break;
end
inp = [ybad inp2];
outpc2 = correction1(inp);
Eobj = ((inp1-outpc2)/inp1)*100;
R = outpc2 -inp1;
iter = iter + 1;
if (iter > 1000)
    c= c+1;
    break;
end
end
ybad;
fid5 = fopen('revmod.txt','a');
fprintf(fid5, '%f\n',ybad);
fclose(fid5);
m = m+3;
k = k+1;
end

```

Appendix B

Screenshots of the *Neuromodeler* Software

B.1 *Neuromodeler* Software Main Window

Fig B-1: Screenshot of the main window of the *Neuromodeler* software.

B.2 New Model Creation Window in the *Neuromodeler* Software.

Fig B-2: Screenshot of the new neural model creation window in the *Neuromodeler* software.

B.3 Training Window of the *Neuromodeler* Software.

Fig B-3: Screenshot of the training window of the *Neuromodeler* software, while training the MLP3 of the Conventional ANN technique.

B.4 Testing Window of the *Neuromodeler* Software.

Fig B-4: Screenshot of the testing window of the *Neuromodeler* software, while validating the MLP3 of the Conventional ANN technique.

B.5 Screenshot of the *Neuromodeler* Software comparison window.

Fig B-5: Screenshot of the *Neuromodeler* software window, showing the validation data (magenta line) and the predicted data (yellow line).

B.6 Screenshot of interpolated radon concentration surface by ArcGis spatial analyst.

Fig B-6: Estimated radon concentrations obtained for Ohio by using four traditional interpolation techniques.