Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering
Elektronik/Online ISSN: 2547-958X

RESEARCH ARTICLE / ARAŞTIRMA MAKALESI

Investigating the Relationship between Soil Gas Radon and Soil Permeability by Using Artificial Neural Networks

Yapay Sinir Ağları Kullanılarak Toprak Gazı Radonu ve Toprak Geçirgenliği arasındaki İlişkinin Araştırılması

Selin Erzin

Dokuz Eylul University, Science Faculty, Physics Department, 35390, İzmir, TÜRKİYE

Corresponding Author /Sorumlu Yazar : selin.erzin@deu.edu.tr

Abstract

This study aims to explore the relationship between soil gas radon concentration (C_{Rn}) and soil permeability (k). To accomplish this, a single linear regression analysis (SLRA) model and an artificial neural network (ANN) model were built from 142 soil gas C_{Rn} and k measurements collected from the literature. When soil gas C_{Rn} values predicted by both models were compared with those measured, the ANN model outperformed the SLRA model. Furthermore, several performance metrics, including correlation coefficient, root mean square error, relative absolute error, and mean absolute error were determined to examine the prediction capabilities of SLRA and ANN models. The metrics obtained demonstrated that the ANN model exhibited superior performance to the SLRA model, thereby showing the accuracy and applicability of the ANN model for forecasting soil gas C_{Rn} values. The study's findings indicated that the developed ANN model may be utilized to forecast soil gas C_{Rn} values based on soil k values.

Keywords: *Soil gas radon, Soil permeability, Artificial neural networks*

Öz

Bu çalışmanın amacı, toprak gazı radon konsantrasyonu (C_{Rn}) ile toprak geçirgenliği (k) arasındaki ilişkiyi araştırmaktır. Bunu gerçekleştirmek için, literatürden toplanan 142 toprak gazı C_{Rn} ve k ölçümünden bir tek doğrusal regresyon analizi (SLRA) modeli ve bir yapay sinir ağı (YSA) modeli oluşturulmuştur. Her iki model tarafından tahmin edilen toprak gazı C_{Rn} değerleri ölçülenlerle karşılaştırıldığında, YSA modeli SLRA modelinden daha iyi performans göstermiştir. Ayrıca, SLRA ve YSA modellerinin tahmin yeteneklerini incelemek için korelasyon katsayısı, kök ortalama kare hatası, bağıl mutlak hata ve ortalama mutlak hata dahil olmak üzere çeşitli performans ölçütleri belirlenmiştir. Elde edilen ölçütler, YSA modelinin SLRA modelinden daha üstün performans sergilediğini ve böylece toprak gazı C_{Rn} değerlerinin tahmininde YSA modelinin doğruluğunu ve uygulanabilirliğini göstermiştir. Çalışmanın bulguları, geliştirilen YSA modelinin toprak k değerlerine dayalı olarak toprak gazı C_{Rn} değerlerini tahmin etmek için kullanılabileceğini göstermiştir.

Anahtar Kelimeler: *Toprak gazı radonu, Toprak geçirgenliği, Yapay sinir ağları*

1. Introduction

Radon, a radioactive gas prevalent in the environment, is a potential health concern for humans. The natural radionuclides ^{235}U, ^{232}Th, and ^{238}U release radon through their decay chains [1]. There are three radon isotopes: ^{222}Rn, ^{220}Rn, and ^{219}Rn. The most common isotope of radon is ^{222}Rn, which is formed when radium (^{226}Ra) decays in the ^{238}U chain. Consequently, the term "radon" refers to ^{222}Rn in the vast majority of cases, including this study.

^{222}Rn is present in all soils and rocks in the Earth surface. The disintegration of ^{222}Rn creates alpha particles and other radionuclides, which can be breathed by humans. The ionizing radiation's adverse effects on human health are widely known. ^{222}Rn is a significant natural source of ionizing radiation, accounting for approximately fifty percent of the natural radiation dosage because of ionizing radiation [2]. Recent research has connected long-term exposure to relatively low levels of ^{222}Rn concentration (C_{Rn}) to an increased risk of cancer [3-6]. To prevent human exposure, it is vital to anticipate regions with elevated natural radiation levels. Radon potential indicators include soil permeability (k) and soil gas C_{Rn}, which are especially essential in urbanized areas.

Artificial neural networks (ANNs) are intended to mimic the decision-making processes seen in the human brain by simulating the functions of biological neurons. ANNs are composed of linked neurons, which evaluate information and learning patterns from data to make predictions or decisions [7]. The intrinsic complexity and nonlinearity of the ANN structure facilitate the resolution of complex issues, including those involving uncertainties, especially when the fundamental relationship of data is very complicated [8]. Therefore, ANNs are widely employed and regarded as intelligent tools for solving complex issues [9].

In this work, the use of both a single linear regression analysis (SLRA) model and an ANN model to estimate soil gas C_{Rn} values using the soil k values was investigated. To accomplish this, the data of 142 soil gas C_{Rn} and k measurements acquired from the

DOI: https://doi.org/10.21205/deufmd.2025277919

Geliş Tarihi / Received: 31.05.2024
Kabul Tarihi / Accepted: 02.07.2024

literature [1, 10-12] were used. The soil gas C_{Rn} values measured were compared to those predicted by the ANN and SLRA models to assess the efficiency of both models in predicting soil gas C_{Rn} values from the soil *k* values.

2. Materials and Method

2.1. Artificial Neural Networks

An ANN is a sort of machine learning model based on the human brain's structure [13]. ANNs are made up of linked neurons that collaborate for processing information. Neurons are separated into three layers: input, hidden layer(s) and output. The input layer receives input data from the outside world and transmits it into the network [14]. The hidden layer(s) do different computations on the data provided through the input layer, and the results are sent to the output layer [14]. The output layer transmits the data learned by the network to the outside world [15]. Each neuron in a layer is connected to all the other neurons in the layer above it via weighted connections. This form of ANN is also known as a multi-layer feed-forward perceptron (MLP) [16].

ANN performance is highly dependent on the number of hidden layers [17]. In particular, for complex problems where accuracy and the time complexity are the primary limitations, determining the hidden layers' number is an important challenge in the construction of ANNs [18]. ANN models with fewer than three hidden layers exhibited lower accuracy, whereas those with more than three hidden layers were demonstrated to be suboptimal in terms of time complexity [18]. Additionally, the hidden neurons' number is also crucial while developing ANN model. This number is selected based on the intricacy of the input-output relationship [17]. As the relationship develops more complicated, more hidden neurons should be used [17]. As observed by Choobasti et al. [14], using an excessively large number of hidden neurons result in the ANN model's performance in predicting being reduced due to overfitting.

MLP learning is an unconstrained optimization problem that attempts to lower overall error values based on the synaptic weights of the ANN [16]. Using input-output vectors as training data, a learning algorithm iteratively adjusts the synaptic weight values in an MLP to approach the desired behavior [19]. This technique is generally completed in two phases using the backpropagation learning method [17]. To create outputs, data is sent into the ANN through the input layer in the first step [17]. In the second phase, any discrepancies in the expected and actual outputs are communicated from the output layer to the previous layers, with the connection weights modified to lower the error value [20]. After training, ANNs keep learning weights of each neuron in their memory. During the testing phase, a new and previously untested dataset is fed into the ANN to provide forecasts based on the saved learning weights [20]. Finally, the actual values are compared to the predicted values by the ANN to evaluate its prediction ability [20].

2.2. Development of Artificial Neural Network Model

In this study, the use of ANNs to determine the relationship between the soil gas C_{Rn} and soil *k* values was investigated. To accomplish this, the data of 142 soil gas C_{Rn} and *k* measurements acquired from the literature [1, 10-12] were used. In creating the ANN model, the measured *k* value was used as an input parameter, while the determined soil gas C_{Rn} value was used as an output parameter. The details of input and output parameters used in this study are listed in Table 1.

Random samples were selected from the dataset for training and testing. While training dataset was used to build an ANN model and identify its learning weights, testing dataset was used to select the best ANN architecture based on the identified learning weights. To do this, 142 data sets were divided into 20% and 80% sets for testing and training purposes. It has been demonstrated that data preparation before network training can enhance the ANN performance [17, 21]. Thus, the data in this study was normalized from -0.9 to 0.9 using Equation (1) with the ranges given in Table 1.

$$x_{norm} = 1.8 \times \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) - 0.9 \tag{1}$$

where x_{norm} and x denote normalized and actual values, respectively, while x_{max} and x_{min} denote the experimental data's highest and lowest values, respectively.

Table 1. The details of the input and output parameters used in the ANN model developed.

Parameters used	Minimum	Maximum	Mean	Range	Std. Deviation	Kurtosis	Skewness
Input parameter							
Soil *k*	5.2×10^{-14}	2.4×10^{-11}	3.16×10^{-12}	2.40×10^{-11}	5×10^{-12}	3.48	6.92
Output parameter							
Soil gas C_{Rn}	0.37	319.40	38.96	319.03	58.73	2.10	2.51

The purpose of an ANN design is to find the optimal architecture of ANNs for a given task. The trial-and-error approach was employed to examine the optimum design of ANNs. Three primary hyperparameters to consider while selecting the best architecture for ANNs are the number of hidden layers, the number of hidden neurons per hidden layer, and the transfer function of each hidden and output layers [22]. These hyperparameters have a vital role in improving the network's prediction accuracy [22]. The hidden layers' number varies with the problem's nature and complexity [23]. When hidden layers' number rises, the ANN model tends to overfit [24]. As a result, in this work, ANN models were created with one to three hidden layers. As mentioned before, determining the number of hidden neurons in the hidden layer(s) is challenging. In this work, the number of hidden neurons in each hidden layer(s) varied from 1 to 5 by one. To obtain optimal ANN performance throughout the stages of training and testing, most often utilized transfer functions (hyperbolic tangent sigmoid (tansig) and logistic sigmoid (logsig) functions) were employed. The Levenberg-Marquardt back-propagation algorithm was employed during the phase of training. The performance of ANN models was then investigated to determine the best ANN structure. Mean absolute error (MAE) was used to evaluate each generated network size's performance until no appreciable improvement was discovered. The best ANN model consists of three hidden layers, each with five hidden neurons, a tansig transfer function in the hidden layers' neurons and output layer neuron and 78 epochs.

2.3. Simple linear regression analysis

Simple linear regression analysis (SLRA) is the most basic form of RA and is used to examine the relationship between two variables. SLRA was employed in this study to assess the relationship between soil gas C_{Rn} and soil k values, with the SPSS 16.0 software program. The 142 data sets utilized in creating the best ANN model were additionally employed for developing the SLRA model, which resulted in the equation below.

$$\text{Soil gas } C_{Rn} = 25.369 + 4\times10^{12} \times \text{soil } k \qquad R^2=0.134 \qquad (2)$$

In Eq. (2), soil k is in m^2 and soil gas C_{Rn} is in kBq m^{-3}.

3. Results and Discussion

The ANN model's predicted soil gas C_{Rn} values were compared to those obtained experimentally in Figs. 1 and 2 for training and testing samples. Almost all of the forecasts in Figs 1 and 2 are close to the perfect prediction line, which is shown as a solid diagonal line. The coefficient of correlation (r) measures the strength of a linear relationship between two variables.

The following guideline was proposed by Smith [25] for |⍰| values:

$|r| \leq 0.2$ There is a weak correlation between the variables.

$0.2 < |r| < 0.8$ There is a correlation between the variables.

$|r| \geq 0.8$ There is a strong correlation between the variables.

The r values of 0.84 and 0.88 obtained for the training and testing samples in Figs. 1 and 2 demonstrate a significant correlation between the predicted and measured soil gas C_{Rn} values based on Smith [25]. In other words, the predicted and measured C_{Rn} values are not significantly disparate. The results also demonstrate that the soil gas C_{Rn} value can be accurately predicted using the constructed ANN model, provided that the soil k value is known.

The SLRA model (Eq. (2)) has an R^2 value of 0.134, based on the SLRA results. The predicted soil gas C_{Rn} values from Eq. (2) were compared to the experimentally obtained soil gas C_{Rn} values in Fig. 3 for all samples to evaluate the SLRA model's prediction performance. The r value of 0.37 in Fig. 3 illustrates that there is a correlation between the predicted and measured values based on Smith [25]. Furthermore, Fig. 3 shows that the SLRA model cannot accurately estimate soil gas C_{Rn} values from soil k values.

In reality, the r value between predicted and measured soil gas C_{Rn} values is a useful indication of the ANN and SLRA models' prediction ability. In this study, two performance metrics (relative absolute error and root mean square error, denoted as *RAE* and *RMSE*) were calculated using the formulae described below to assess the prediction accuracy of the ANN and SLRA models.

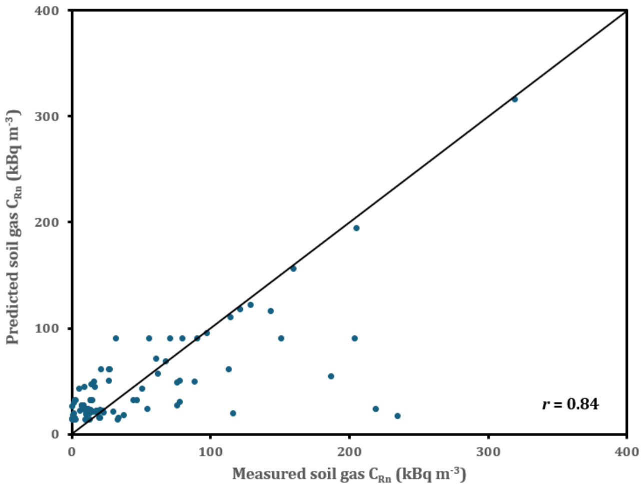

Figure 1. Comparison of the measured soil gas C_{Rn} values with the predicted soil gas C_{Rn} values from the ANN model for training samples.

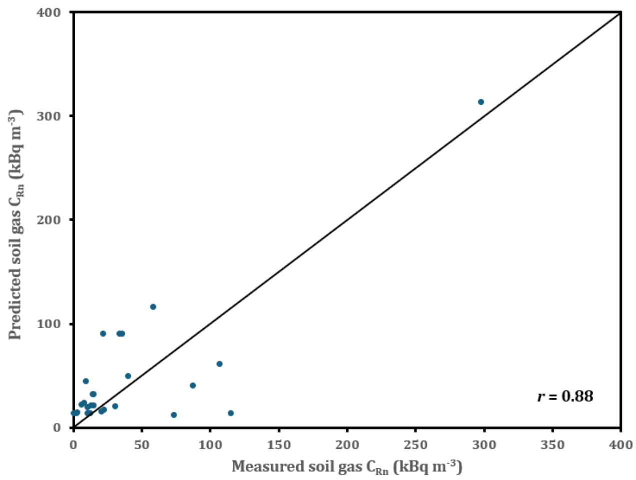

Figure 2. Comparison of the measured soil gas C_{Rn} values with the predicted soil gas C_{Rn} values from the ANN model for testing samples.

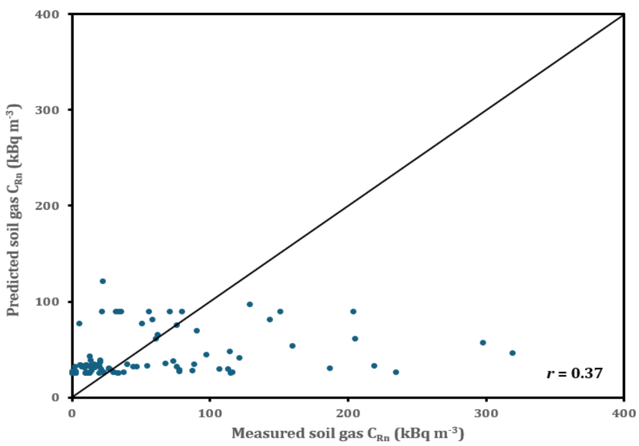

Figure 3. Comparison of the measured soil gas C_{Rn} values with the predicted soil gas C_{Rn} values from the SLRA model for all samples.

$$RAE = \frac{\sum_{i=1}^{N}\left|\left((C_{Rn})_{Pre} - (C_{Rn})_{Exp}\right)\right|}{\sum_{i=1}^{N}\left|\left((C_{Rn})_{Exp} - (C_{Rn})_{Exp\,Mean}\right)\right|} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((C_{Rn})_{Exp} - (C_{Rn})_{Pre}\right)^2} \quad (4)$$

where $(C_{Rn})_{Exp}$ is the measured soil gas C_{Rn}value; $(C_{Rn})_{Pre}$ is predicted soil gas C_{Rn}value; $(C_{Rn})_{Exp\,Mean}$ is the measured soil gas C_{Rn}values' mean value; and N is the sample number. In a perfect forecast, the *RAE* value is 0; the *RAE* value increases as the model prediction error increases. The closer the *RMSE* is to zero, the smaller the difference between forecasts and observations.

Table 2 illustrates the developed ANN and SLRA models' performance metrics. The ANN model performed better in terms of prediction accuracy than the SLRA model based on the computed metrics in Table 2. This result also shows that the developed ANN model is effective and beneficial in predicting soil gas C_{Rn} values. As a result, the soil gas C_{Rn} values can be accurately predicted using the trained ANN structures from the soil *k* values.

Table 2. Performance metrics of the ANN and SLRA models developed in this study.

Model	Data	*r*	*MAE* (kBq m^{-3})	*RMSE* (kBq m^{-3})	*RAE* (kBq m^{-3})
ANN	Training set	0.84	21.66	38.76	0.52
	Testing set	0.88	25.61	35.48	5.45
SLRA	All set	0.37	35.00	54.50	13.30

4. Conclusion

This research investigates the use of SLRA and ANN models to estimate soil gas C_{Rn} values from soil *k* values. To do this, data including 142 soil gas C_{Rn} and soil *k* measurements from the available literature were used. The soil gas C_{Rn} values predicted from ANN and SLRA models were compared to the measured soil gas C_{Rn} values to assess the models' predictive ability. The comparison results demonstrated a better performance of the ANN model than the SLRA model in predicting soil gas C_{Rn} values. In accordance with the study's findings, the developed ANN model may be applied to forecast soil gas C_{Rn} values using the soil *k* values. This study demonstrates the effectiveness of the ANNs to acquire and illustrate complex relationships among the parameters.

Ethics committee approval and conflict of interest statement

This article does not require ethics committee approval. This article has no conflicts of interest with any individual or institution.

References

[1] Nuhu, H., Hashim, S., Aziz Saleh, M., Syazwan Mohd Sanusi, M., Hussein Alomari, A., Jamal, M.H., Abdullah, R.A., Hassan, S.A. 2021. Soil Gas Radon and Soil Permeability Assessment: Mapping Radon Risk Areas in Perak State, Malaysia, PLoS One, Vol. 16(7), e0254099.

[2] WHO. 2011. Guidelines for drinking water quality. World Health Organization, Copenhagen. 2011 Vol 1. 3rd edition).

[3] Alonso, H., Rubiano, J.G., Guerra, J.G., Arnedo, M.A., Tejera, A., Martel, P. 2019. Assessment of Radon Risk Areas in the Eastern Canary Islands using Soil Radon Gas Concentration and Gas Permeability of Soils, Science of The Total Environment, Vol. 664, p. 449–460.

[4] Giustini, F., Ciotoli, G., Rinaldini, A., Ruggiero, L., Voltaggio, M. 2019. Mapping the Geogenic Radon Potential and Radon Risk by Using Empirical Bayesian Kriging Regression: A Case Study from A Volcanic Area of Central Italy. Science of The Total Environment, Vol. 661, p. 449–464.

[5] ICRP. 2010. Lung Cancer Risk from Radon and Progeny and Statement on Radon. ICRP Publication 115. Ann. ICRP, Vol. 40(1), p. 1–64.

[6] UNSCEAR. 2006. Effects of Ionizing Radiation-Volume II: Scientific Annexes C, D and E. United Nations, New York.

[7] Sharshir, S.S., Abd Elaziz, M., Elsheikh, A. 2023. Augmentation and Prediction of Wick Solar Still Productivity Using Artificial Neural Network Integrated with Tree–Seed Algorithm, International Journal of Environmental Science and Technology, Vol. 20, p. 7237–7252.

[8] Araei, A.A. 2013. Artificial Neural Networks for Modeling Drained Monotonic Behavior of Rockfill Materials. International Journal of Geomechanics, Vol. 14(3), 04014005.

[9] Khandelwal, M., Singh, T.N. 2009. Prediction of Blast-Induced Ground Vibration Using Artificial Neural Network. International Journal of Rock Mechanics and Mining Sciences, Vol. 46(7), p. 1214-1222.

[10] Neznal, M., Neznal, M., Matolin, M., Barnet, I., Miksova, J. 2006. The New Method for Assessing the Radon Risk of Building Sites. Project report. State Office for Nuclear Safety, Prague

[11] Cosma, C., Cucoş-Dinu, A., Papp, B., Begy, R., Sainz, C. 2013. Soil and Building Material as Main Sources of Indoor Radon in Băiţa-Ştei Radon Prone Area (Romania), Journal of Environmental Radioactivity, Vol. 116, p. 174-179.

[12] Beltrán-Torres, S., Szabó, K.Z., Tóth, G., Tóth-Bodrogi, E., Kovács, T., Szabó, C. 2023. Estimated Versus Field Measured Soil Gas Radon Concentration and Soil Gas Permeability, Journal of Environmental Radioactivity, Vol. 265, 107224.

[13] Ünal, H.T., Başçiftçi, F. 2022. Evolutionary Design of Neural Network Architectures: A Review of Three Decades of Research, Artificial Intelligence Review, Vol. 55, p. 1723–1802.

[14] Choobbasti AJ, Farrokhzad F, Barari A (2009) Prediction of Slope Stability Using Artificial Neural Network (A Case Study: Noabad, Mazandaran, Iran). Arabian Journal of Science and Engineering, Vol. 2, p. 311–319

[15] Guo, Z., Uhrig, R.E. 1992. Use of Artificial Neural Networks to Analyze Nuclear Power Plant Performance, Nuclear Technology, Vol. 99, p. 36–42

[16] Goktepe B, Agar E, Lav AH (2004) Comparison of Multilayer Perceptron and Adaptive Neuro-Fuzzy System on Backcalculating the Mechanical Properties of Flexible Pavements ARI: The Bulletin of Istanbul Technical University, Vol. 54, p. 65–77.

[17] Erzin, S. 2024. Prediction of the Radon Concentration in Thermal Waters Using Artificial Neural Networks, International Journal of Environmental Science and Technology, Vol. 21(3), p. 7321–732.

[18] Uzair, M., Jamil, N. 2020. Effects of Hidden Layers on the Efficiency of Neural networks. 2020 IEEE 23rd International Multitopic Conference (INMIC), 1-6.

[19] Gopalakrishnan, K. 2010. Effect of Training Algorithms on Neural Networks Aided Pavement Diagnosis, International Journal of Engineering Science and Technology, Vol. (2), p. 83-92.

[20] Erzin, Y., Rao, B.H., Singh, D.N. 2008. Artificial Neural Network Models for Predicting Soil Thermal Resistivity, International Journal of Thermal Sciences, Vol. 47(10), p. 1347–1358.

[21] Motahar, S., Sadri, S. 2021. Applying Artificial Neural Networks to Predict the Enhanced Thermal Conductivity of a Phase Change Material with Dispersed Oxide Nanoparticles, International Journal of Energy Resources, Vol. 45, p. 15092–15109.

[22] Mahmoud, B., Pete, B., Majid, H., Sahar, H. 2021. Ten-Year Estimation of Oriental Beech (Fagus Orientalis Lipsky) Volume Increment in Natural Forests: A Comparison of An Artificial Neural Networks Model, Multiple Linear Regression and Actual Increment, Forestry: An International Journal of Forest Research, Vol. 94(4), p. 598–609.

[23] Muhammad Waseem, A., Monjur, M., Yacine, R. 2017. Trees vs Neurons: Comparison Between Random Forest and ANN for High-Resolution Prediction of Building Energy Consumption, Energy and Buildings, Vol. 147, p. 77-89, ISSN 0378-7788.

[24] Kaastra, I., Boyd, M. 1996. Designing a Neural Network for Forecasting Financial and Economic Time Series. Neurocomputing, Vol. 10(3), p. 215-236.
[25] Smith, G.N. 1986. Probability and Statistics in Civil Engineering: An Introduction, Collins, London