International Journal of Applied Science and Mathematics
Volume 9, Issue 3, ISSN (Online): 2394-2894

Spatial Analysis of Soil Radon Gas Concentration in Southwestern Nigeria: A Bayesian Approach

Olawale Basheer Akanbi
Department of Statistics, University of Ibadan, Ibadan Nigeria.
Corresponding author email id: muhdbashola@yahoo.com

Date of publication (dd/mm/yyyy): 17/05/2022

Abstract **– Radon gas is a leading cause of lung cancer after smoking. According to Environmental Protection Agency (EPA) and Surgeon General, about 20,000 lung cancer deaths are recorded each year and are caused by radon gas Radon is readily available in all types of soil and moves freely into homes via cracks on the floor, hence can contaminate our drinking water. Previous studies have explored the danger of soil radon gas using various classical methods such as ordinary kriging to capture spatial attributes of the data sets to make useful predictions. In this study, two Bayesian approaches are considered namely, A Bayesian linear regression (Independent) Model which ignores the spatial dependency as the baseline model, and Bayesian Geostatistical Model with an additional spatial dependency parameter to quantify the relationships between Radon gas concentration and the covariate structures are considered. The effects of the environmental variables were investigated and the radon potential map of the study sites was produced using four picocuries (200 Bqm-3) as the threshold. The radon potential map indicated that 22 locations already have radon concentrations above the four picocuries threshold. The two approaches produced similar posterior means and standard deviations. However, the WAIC and DIC values indicated that the Bayesian Geostatistical Model provided a better fit for the data, hence justifying the inclusion of the spatial parameter.**

Keywords **– Bayesian Geostatistical Model, Spatial Parameter, Deviance Information Criteria, Radon Gas Concentration, Geogenic Radon Potential.**

I. INTRODUCTION

Radon gas is a common public health risk in many parts of the world including; America and the UK and its adjudged to be the leading cause of lung cancer after smoking. Most people are exposed to radon in their homes, schools, and workplaces. Radon is an invisible, odorless naturally occurring gas found in soil, rocks, and water, hence can move freely into the environment. Radon is a decay product of radium with a half-life of 3.8 days hence, the inhalation of the short-lived product has been linked to an increase in the risk level of lung cancer. In recent times, there has been a global effort to improve on the environment through the implementation of environment-friendly policies, and the development of new remediation techniques by different countries to safeguard their environment in line with WHO standard and other relevant environmental agencies. Radon exposure guidelines exist to guide personal and regulatory decision-making and are based on the estimated risk to public health based on a specific exposure level. In Canada, health Canada published a radon exposure guideline in 2007 of 200 Bqm-3 above which, they recommend action be taken to reduce radon exposure levels. In the United States, the Environmental Protection Agency set a radon exposure guideline of 148 Bqm-3 and the world health organization sets its radon guideline for action at 100 Bqm-3. In Malaysia, soil gas radon and permeability assessment in mapping radon risk areas of Perak State in the country were carried out, Nuhu (2021). Similarly, in South Korea, Spatial modeling of radon potential mapping using deep learning algorithms was considered, Panahi (2022). In Nigeria, a Bayesian study has been done on the determinant of different gases, like a greenhouse, Akanbi (2018) but radon gas was not a well-known radioactive gas then, owing to this ignorant, the danger of this gas is not a subject of interest among Nigerians until when its risk level was determined Over the Lithological Units of a Southwestern Nigerian University residential soil, Deborah (2020). But unfortunately, the government did not take any serious action to sanitize the people of its danger and prevention on the environment in line with WHO existing standards. Hence, the need to create awareness and consciousness of the underlying risk of radon gas by exploring 150 locations in southwestern Nigeria using the Bayesian statistics approach. This study aims to investigate the impact of spatial correlation structures on soil radon gas concentration using the Bayesian approach. Bayesian methods are widely used in live sciences, geology, engineering, and economics to solve their related problems. This was fully discussed by Zellner (1971). Based on the nature of the data, spatial analysis for detecting spatial patterns of geo-features according to their spatial or nonspatial attributes (Anselin 1992, Goodchild et al. 1992) was investigated. Geostatistical modeling is viewed as a hierarchical specification, which naturally led to the adoption of a Bayesian framework for inference and suitable Markov chain Monte Carlo (MCMC) for model fitting. Figure 1 depit the spatial map of the study location. It shows the proximity of the twenty-two locations considered in the Nigeria.

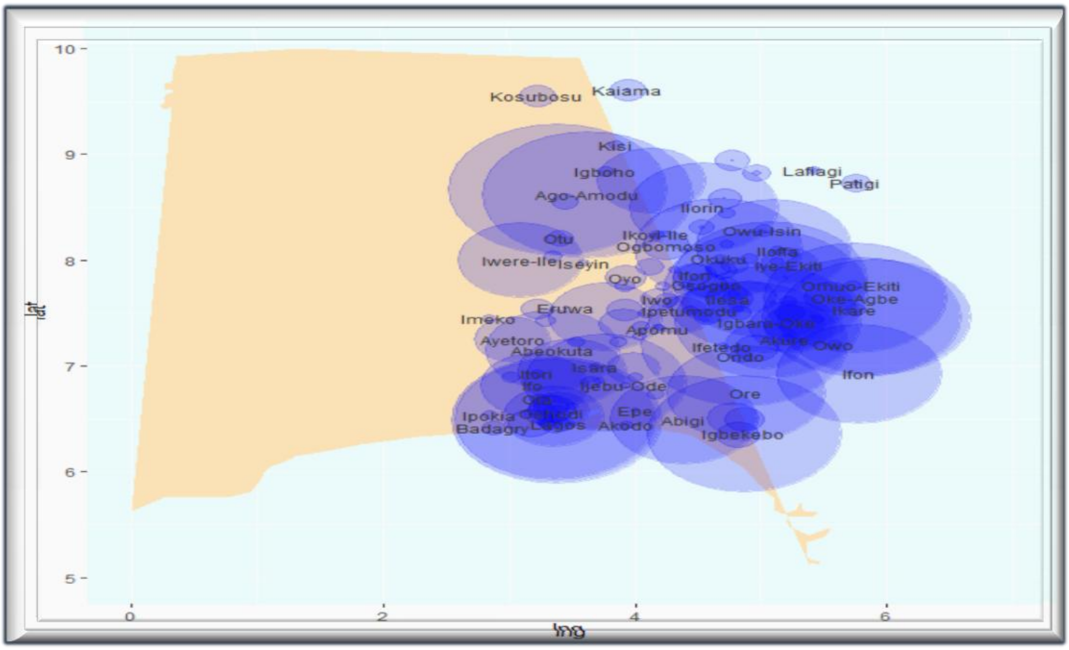

Fig. 1. Spatial map of the study locations.

Source: Google search.

II. Bayesian Regression

In this study, we consider two linear Bayesian regression models: Independent linear model with no spatial correlation structure; and Bayesian geo-statistical model with spatial correlation. In the two models let y_{ir} be the observed Radon gas (R_n) concentration at location i for replicate r. let x_{ir} be a vector of length k = 2, representing the covariates comprising of soil permeability and GRP. Measurement of R_n was initially skewed and was log-transformed to attained normality.

2.1. *Bayesian Linear Regression Model*

Bayesian Linear Regression model (independent structure). In this model we assume that locations were independent and that Radon gas (R_n) emissions were affected by the designated covariates independently at each location so that;

$$y_{ir} \ = \beta_0 + \sum_{k=1}^{2} \beta_k x_{irk} \ + \varepsilon_{ir} \tag{1}$$

Where β_k are the regression coefficients and ε_{ir} is the residual under the independence and normality assumption.

2.1.1. *Likelihood Construction for Bayesian Linear Regression Model*

The response variable y is the log transformation of Radon gas.

$$y_i \sim N(0, \sigma^2)$$

$$L(y) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum y^2}{2\sigma^2}} \tag{2}$$

2.1.2. *Elicitation of Prior Distribution for Bayesian Linear Regression Model*

$$\beta \sim u(0,1) \tag{3}$$

$\sigma^2 \sim IG(\propto, \beta)$ with the probability distribution function given as;

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\propto)} (\sigma)^{-\alpha-1} e^{(-\frac{\beta}{\sigma})} \tag{4}$$

Ignoring the constants, eq. (4) becomes;

$$p(\sigma^2) \propto e^{(-\frac{\beta}{\sigma})} (\sigma)^{-\alpha-1} \tag{5}$$

The combination of eq. (3) and eq. (5) yields the joint prior given as;

$$p(\beta, \sigma^2) \propto e^{(-\frac{\beta}{\sigma})} (\sigma)^{-\alpha-1} \tag{6}$$

2.1.3. *Posterior Distribution for Bayesian Linear Regression Model (BLRM)*

The posterior distribution for the Bayesian linear regression model is formulated by combining eq. (2) and eq. (6). This yields;

$$p(\beta, \sigma^2 / y) \propto (2\pi\sigma^2)^{-\frac{n}{2}} (\sigma)^{-\alpha-1} e^{-\frac{1}{\sigma^2}(\frac{\sum y^2}{2} + \frac{\beta}{1})} \tag{7}$$

2.2. *Bayesian Geo-Statistical Model (Spatial Model) [BGSM]*

The second model is an extension of the first model with additional parameter to account for the spatial correlation between the experimental sites.

The additional term is modelled as a random effect with variance reflecting the spatial correlation, so that eq. (1) becomes;

$$y_{ir} \ = \beta_0 + \sum_{k=1}^{2} \beta_k x_{irk} \ + s_i + \varepsilon_{ir} \tag{8}$$

$$y \sim N(x\beta + s, \sigma^2 I)$$

where I is a $r \times r$ indentity matrix and s is the spatial correlation parameter.

2.2.1. *Likelihoodfor the BGSM (Spatial Model)*

The probability distribution function is given as;

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2 I}(y - x\beta - s)^2} \tag{9}$$

The likelihood function in eq. (9) becomes;

$$L(y) = (2\pi\sigma^2 I)^{-\frac{n}{2}} e^{-\frac{\sum(y_i - x\beta - s_i)^2}{2\sigma^2}} \tag{10}$$

2.2.2. *Elicitation of Prior for the BGSM*

The model parameters are assigned the following probability distributions; $\beta \sim u(0,1)$.

The regression coefficients are assigned a uniform distribution since we do not have any prior information on them, according to Gelman (2008).

$\sigma^2 \sim$ non informative Jeffrey prior and $s \sim N(0, \tau^2\varphi)$ where τ^2, φ represent the spatial correlation and the correlation matrix respectively with $\varphi_{ij} = f(d_{ij}, \omega, \gamma)$ between s_i and s_j.

Here, d_{ij}, ωaredistance between s_i *and* s_j *and* the rate of decrease in spatial correlation per unit distance respectively, with large ω suggesting a rapid decrease in spatial correlation and γ controls the amount of smoothing. Thomas et al (2004) advised a value of $\gamma = 1$. This is common in literature.

In order to estimate the hyper parameters parameters, we assigned unique priors as required in Bayesian statistics. Hence, we assume a uniform and proper inverse gamma distribution for τ^2, ω respectively as follows; $\tau^2 \sim IG(\propto, \beta)$ and $\omega \sim u(0,1)$

The designated priors can be expressed as;

$$p(\beta, \omega) = 1 \tag{11}$$

Since β, ω are uniformly distributed.

$$p(\sigma^2) = \frac{1}{\sigma^2} \tag{12}$$

$$p(\tau^2) = \frac{\beta^\alpha}{\Gamma(\propto)} (\tau)^{-\alpha-1} e^{(-\frac{\beta}{\tau})} \tag{13}$$

Ignoring the constants, eq. (13) becomes;

$$p(\tau^2) \propto e^{(-\frac{\beta}{\tau})} (\tau)^{-\alpha-1} \tag{14}$$

The combination of eq. (11), eq. (12) and eq. (14) yields the joint prior for second model given as;

$$p(y/\beta, \sigma^2, \omega, \tau^2) \propto \frac{(\tau)^{-\alpha-1}}{\sigma^2} e^{(-\frac{\beta}{\tau})} \tag{15}$$

2.2.3. *Posterior Distribution for the BGSM*

The posterior distribution for the Bayesian linear regression model is formulated by combining eq. (11) and eq. (16). This yield; $p(\beta, \sigma^2, \omega, \tau^2/y) \propto (2\pi\sigma^2 I)^{-\frac{n}{2}} e^{-\frac{\sum(y_i - x\beta - s_i)^2}{2\sigma^2}} . \frac{(\tau)^{-\alpha-1}}{\sigma^2} e^{(-\frac{\beta}{\tau})}$

$$p(\beta, \sigma^2, \omega, \tau^2/y) \propto (2\pi\sigma^2 I)^{-\frac{n}{2}} . \frac{(\tau)^{-\alpha-1}}{\sigma^2} e^{-\left[\frac{\sum(y_i - x\beta - s_i)^2}{2\sigma^2 I} + \frac{\beta}{\tau}\right]} \tag{16}$$

2.3. *Model Comparison*

The two Bayesian models using the deviance information criterion (DIC) were compared.The DIC is based on the posterior expectation of the deviance D and the effective number of parameters P_D in themodel, and is expressed as: $DIC = P_D + D$.

The DIC is easily computed from the samples generated through MCMC.

III. Results and Discussion

The following tables and figures discuss the results of the analysis carried out for this study. Table 1 presents the summary statistics of variables considered for the modelling. It shows that the average/mean Radon, soil air perm and GRP values are 10.43, 8.81, 8.51, respectively. See Table 1 for detailed summary statistics.

Table 1. Descriptive Statistics of the Variables.

Descriptive	Radon	Soil Air Perm	GRP
Mean	10.43	8.81	8.511
Median	4.76	4.01	3.965
Maximum	49.59	87.00	39.191
Minimum	0.28	0.59	0.152
Std. Dev.	12.65	17.41	10.370
Skewness	1.72	3.50	1.82
Kurtosis	1.71	13.83	2.17
Jarque-Bera	3.11	3.71	2.48
Probability	0.86	0.31	0.29
Sum	1565.13	1321.21	1276.58
Sum Sq. Dev.	39.83	45.19	39.54
Observations	150	150	150

3.1. *Fitting Distributions*

Table 2 shows the distribution fitting for the dependent variable, Radon concentration. It could be seen from the table that Lognormal distribution best fit the data with minimum AIC of 988.5565. The other distributions like Weibull, t-distribution, gamma, normal, and exponential also produced AIC of 1003.806, 1077.011, 1006.286, 1190.797 and 1005.526 respectively.

Table 2. Fitting distributions for Logarithmic Radon Concentration.

Distributions	Loglik	AIC
Weibull	-499.9031	1003.806
t-distribution	-535.5056	1077.011
Gamma	-501.143	1006.286
Normal	-593.3987	1190.797

Distributions	Loglik	AIC
Exponential	-501.7629	1005.526
Lognormal	-492.2783	988.5565

3.2. *Radon Potential Map*

Figure 2 give a summary plot of the Radon potential for the locations under study. The radon potential map is produced that indicate areas with high, medium or low potential for elevated indoor radon levels. Based on soil measurement, kemski et al. (1992) proposed the following classifications: low (<25 KBqm^{-3}), medium (25-100 KBqm^{-3}), elevated (100 KBqm^{-3}-250 KBqm^{-3}), high (> 250 KBqm^{-3}) and this is used for the classification of the radon concentration. After the classification, we have 128 low concentrations, 22 medium, 0 elevated and 0 high as shown in Figure 2.

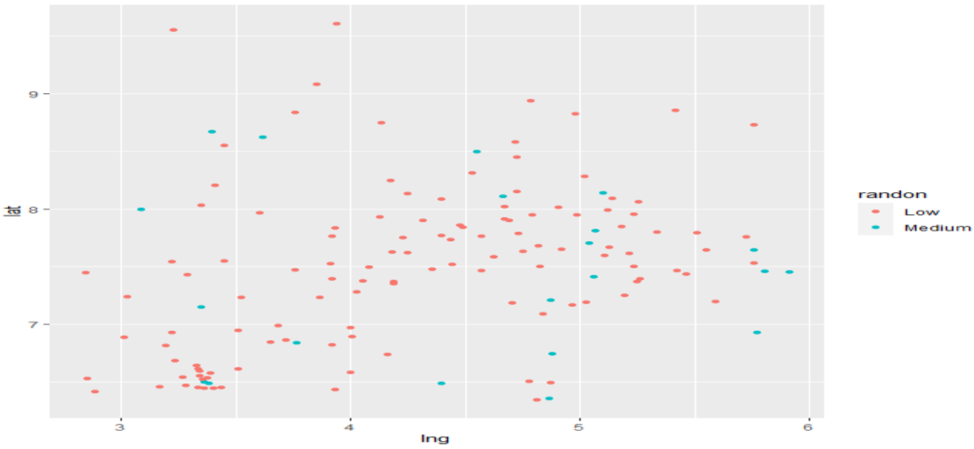

Fig. 2. Radon potential map of the study region.

3.3. *Bayesian Results*

Figure 3 shows the posterior plot of independent linear model using uniform prior. The distribution shown is approximately normal with balanced tails on both sides. Consequently, we consider other conjugate priors and present the outputs in Figure 4.

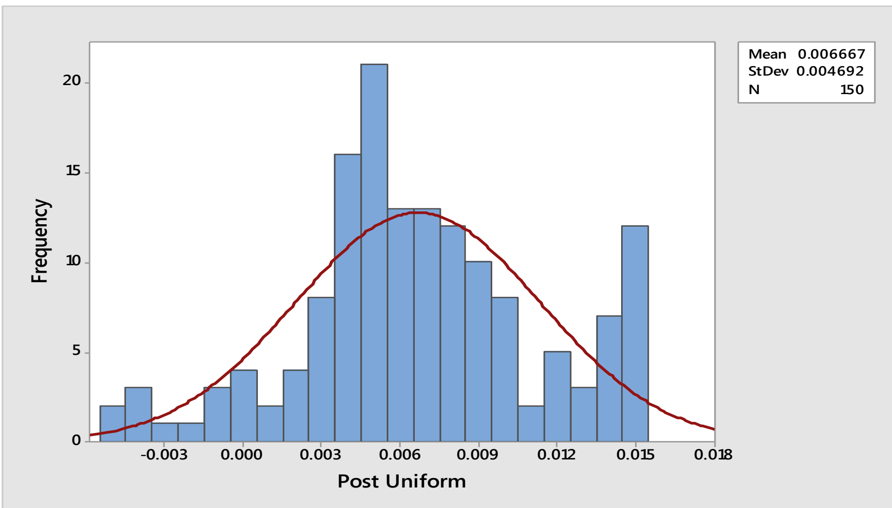

Fig. 3. Histogram and density plot of Posterior estimate using Uniform prior.

Table 3 presents the modelling output when inverse-gamma prior is used. As noted earlier, inverse-gamma prior seems more appropriate than the uniform conjugate prior considered in Figure 3.

Table 3. Posterior summaries using Inverse-Gamma prior for the BLRM.

	Post mean	Post Std.	25%	75%
Intercept	0.451	0.037	0.426	0.473
SAP	-0.652	0.049	-0.684	-0.619
GRP	1.082	0.039	1.055	1.108
Precision	0.217	0.015	0.171	0.275

Figure 4 shows the posterior intercept for the independent linear model (BLRM) using Inverse-Gamma prior. The graph shows that inverse-gamma prior produced a more normally distributed output than when a uniform conjugate prior is used. This implies that the conjugate prior of inverse-gamma is more appropriate compare to when a uniform prior is considered.

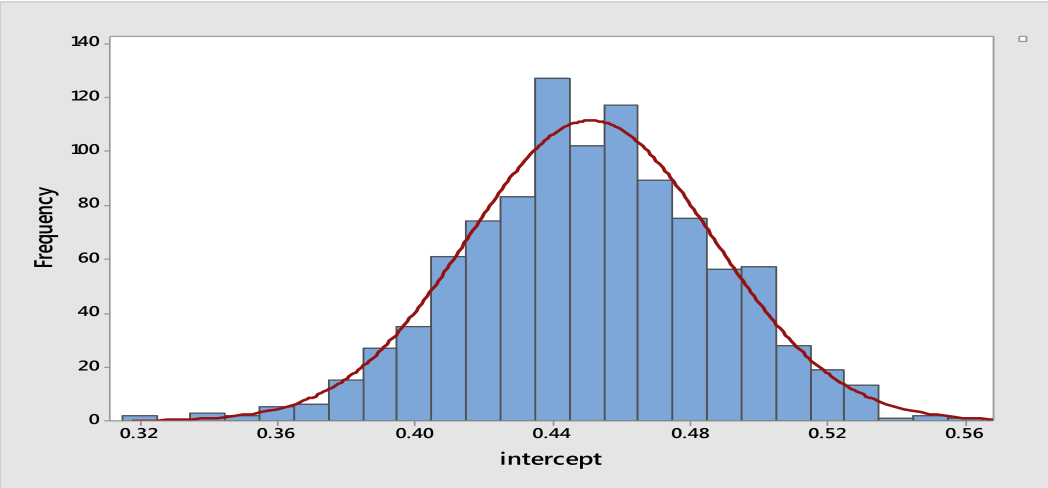

Fig. 4. Histogram and density plot of Posterior estimate of the intercept using inverse-gamma prior and for the BLRM.

Since inverse-gamma conjugate prior produce a more expected result on the intercept over the uniform prior. We adopt the inverse-gamma for the estimation of the parameters. Figure 5 shows the posterior estimate (SAP)for the independent linear model (BLRM) using Inverse-Gamma prior.

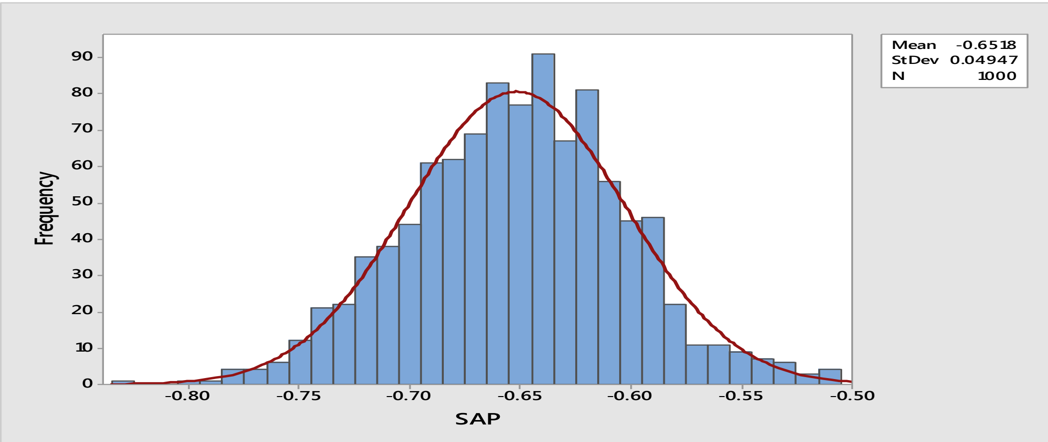

Fig. 5. Histogram and density plot of Posterior estimate of SAP using inverse-gamma prior and for the BLRM.

Figure 6 shows the posterior estimate (GRP)for the independent linear model (BLRM) using Inverse-Gamma prior.

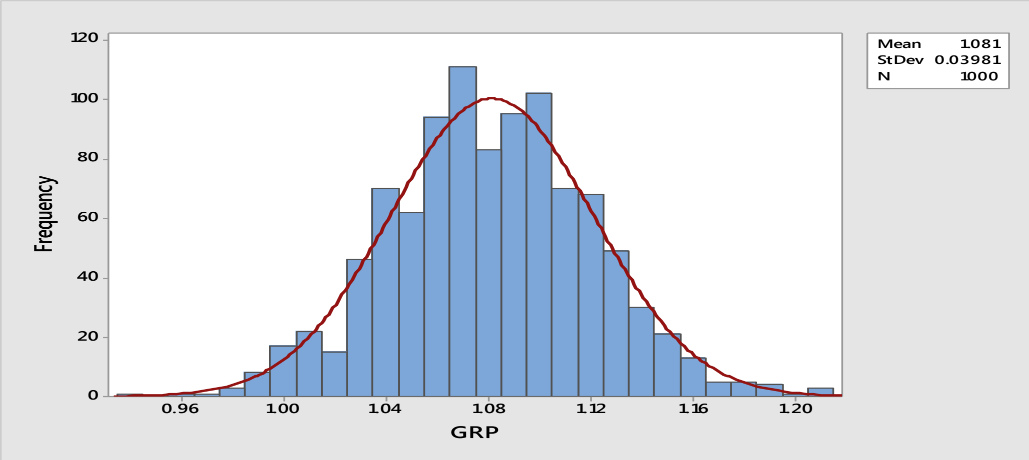

Fig. 6. Histogram and density plot of Posterior estimate of GRP using inverse-gamma prior and for the BLRM.

3.4. *Bayesian Geo-Statistical Model*

Table 4 shows the posterior mean, posterior standard deviation, Naïve standard deviation and the time series standard deviation which takes care of the autocorrelation of the spatial effect.

Table 4. Bayesian Spatial Regression (BGSM).

	Post Mean	Post SD	Naïve SE	Time-Series SE
(Intercept)	0.4004	0.42022	0.148571	0.094960
SAP	-0.6706	0.07907	0.027954	0.027954
GRP	1.0912	0.04061	0.014357	0.014357
Sigma sq	0.1531	0.01574	0.005563	0.005563
Tau sq	0.1020	0.01049	0.003709	0.003709

Table 5 shows the quantiles for the variables under the BGSM. Using the boxplot, colum2 is the least quartile, column 4 represents the median or the second quartile. See Table 5 for detailed quantiles output.

Table 5. Quantiles for each variable under the BGSM.

	2.5%	25%	50%	75%	97.5%
(Intercept)	-0.21684	0.1917	0.3825	0.5777	1.0280
SAP	-0.80707	-0.7028	-0.6575	-0.6115	-0.5877
GRP	1.02459	1.0722	1.0994	1.1134	1.1423
sigma.sq	0.13010	0.1436	0.1548	0.1611	0.1751
tau.sq	0.08673	0.0957	0.1032	0.1074	0.1167

Figure 7 shows the posterior plotsfor the Bayesian Geo-statistical model using the inverse gamma prior. The figure shows the BGSM estimate for the intercept. As shown, it is skewed to the left. Although, it can be approximated to normal since the tail of the distribution is no evident or significantly different from the right tail. The skewness is observed in values not in tail effect.

International Journal of Applied Science and Mathematics
Volume 9, Issue 3, ISSN (Online): 2394-2894

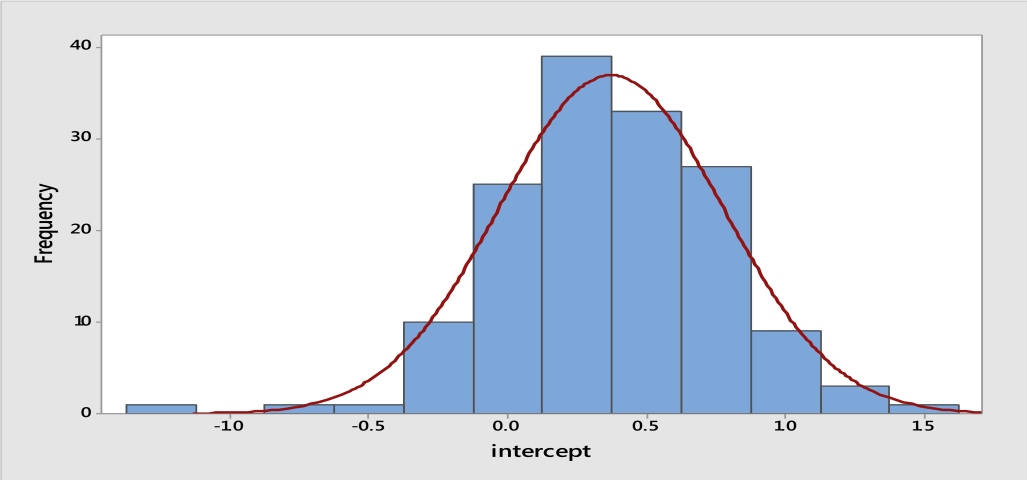

Fig. 7. Histogram and density plot of Posterior estimate of the intercept using inverse- gamma prior and for the BGSM.

Figure 8 presents the graph of the posterior estimate of GRP using inverse-gamma prior and for the BGSM. The distribution is normally distributed.

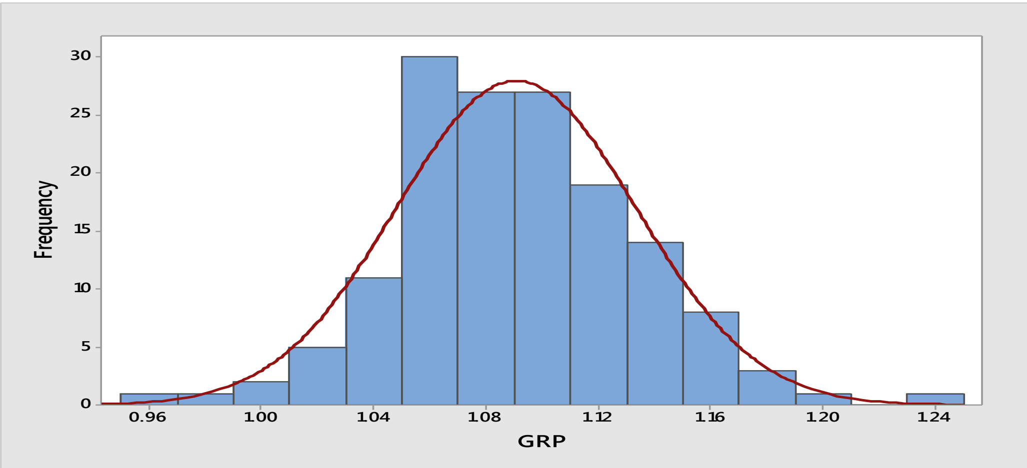

Fig. 8. Histogram and density plot of Posterior estimate of GRP using inverse-gamma prior and for the BGSM.

Figure 9 presents the graph of the posterior estimate of SAP using inverse-gamma prior and for the BGSM. The distribution is approximately normally distributed, however, with some skewness component to the left.

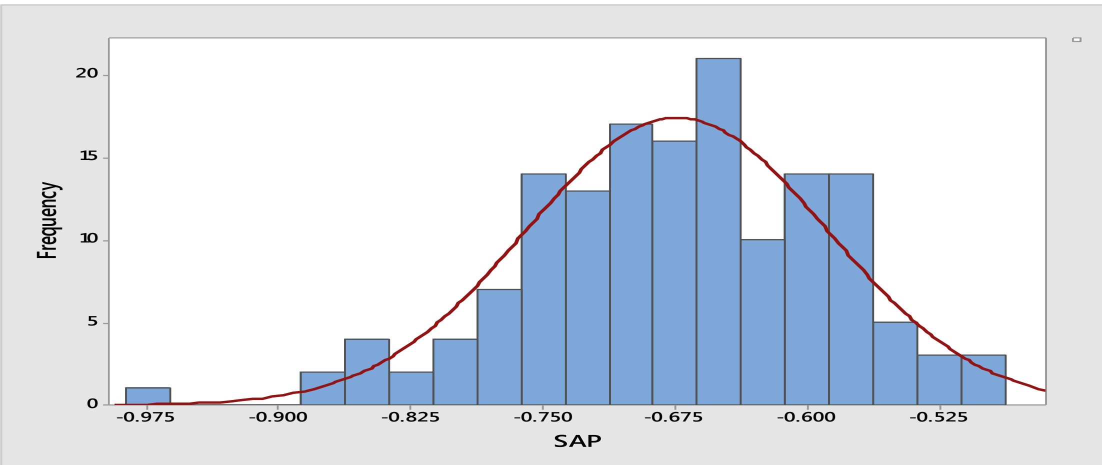

Fig. 9. Histogram and density plot of Posterior estimate of SAP using inverse-gamma prior and for the BGSM.

Figure 10 presents the graph of the posterior estimate of TAU-SQ using inverse-gamma prior and for the BGSM. The distribution is approximately normally distributed, however, with some skewness component to the right.

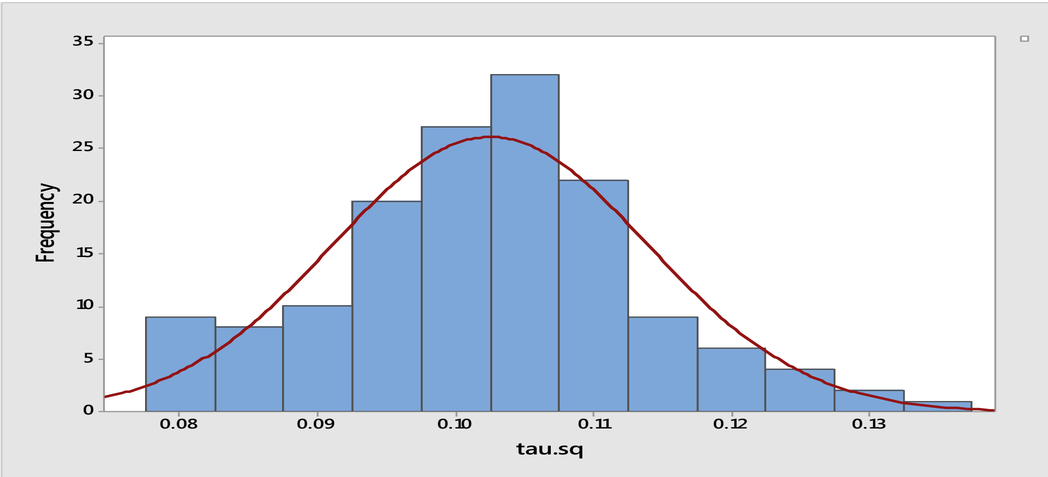

Fig. 10. Histogram and density plot of Posterior estimate of TAU-SQ using inverse-gamma prior and for the BGSM.

3.5. *Numeric Summary of Results Across Model Criteria*

Table 6. Comparism of results across model criteria.

	Independent Bayesian Regression Model		**Bayesian Geo-statistical Model**	
	Post mean	Post SD	Post mean	Post SD
Intercept	0.451	0.037	0.4004	0.42022
SAP	-0.652	0.049	-0.6706	0.07907
GRP	1.082	0.039	1.0912	0.04061
WAIC	8168.775		18.02771	
DIC	2183729		17928.18	

IV. Summary and Discussion of Results

Following the results obtained, it is observed that the radon produced on the potential map of the 150 locations indicated that 22 out of 150 study sites exceeded an action level threshold of 4 picocuries (200Bqm^{-3}). This implies that with the locations spatially correlated with the locations with low radon concentrations presently can increase over the years. Similarly, the Soil Air Permeability (SAP) and Geogenic Radon Potential (GRP) were influential in soil radon gas concentration. Therefore, in Table 6, the posterior means for SAP and GRP were -0.652 and 1.082 under the BLRM and -0.671 and 1.091 for the BGSM respectively. These values quantify the effects of the covariates on soil radon gas in the study locations. Also, the spatial dependency parameter has a mean value of 0.102 which clearly shows that the covariate structures are spatially correlated. Furthermore, the WAIC were (8168.775,18.02771), while the DIC were (2183729, 17928.18) for the BLRM and BGSM respectively.

V. Conclusion

The two models seemed to fit the data well as they produced closed posterior estimates. However, the model which provides a noticeably better fit to the data is the Bayesian geo-statistical spatial regression model as it produces lower WAIC and DIC values compared to the independent Bayesian linear regression model.Thus, the inclusion of spatial dependency in the model helps to quantify the relationship between soil radon gas and the geologic factors.

REFERENCES

[1] O.B. Akanbi (2018). Determinants of Greenhouse gases in Nigeria using a Bayesian Approach. *Research Journali's Journal of Mathematics*5 (3), 1-13.

[2] I. Anselin andA. Getis. (1992): Spatial statistical analysis and geographic information systems. *The Annals of Regional science*. 26(1), 19-33.

[3] T.E. Deborah , K. C. S. Mynepalli, O. Rachel, A. Yinka., A. Olusegun, O. Babakayode and M. O. Olatunde(2020). 'Determination of Residential soil gas radon risk indices over the Lithological units of a Southwestern Nigeria University'. Sci Rep. doi: 10.1038/s41598 -020-64217-8.

[4] A. Gelman. (2008): Objection to Bayesian Statistics. *International Society for Bayesian Analysis*.vol.3, number 3, 445-449

[5] M. Goodchild., R. Haining., and S. Wise (1992). 'Integrating GIS and Spatial Data Analysis—Problems and Possibilities'. *Int. Journal of Geographical Information System*. 6 (5): 407 – 423.

[6] J. Kemski, R. Klingel, H. Schneiders, A. Siehl and J. Wiegand (1992): Geological structure and geo chemistry controlling radon in soil gas. *Radiat. Prot. Dosim*. 1992; 45(1-4): 235-239

[7] M. Panahi., P. Yariyan.,F. Rezaie., S. Won Kim., A. Sharifi., A. A. Alesheikh, J. Lee, S. Kim., J. Yoo andL. Saro(2022). 'Spatial modeling of radon potential mapping using deep learning algorithms'. Geocarto International. DOI: 10.1080/10106049.2021.2022011

[8] H. Nuhu, S. Hashim, M. Aziz Saleh, M. Syazwan Mohd Sanusi, A. Hussein Alomari,and M.H. Jamal (2021).'Soil gas radon and soil permeability assessment: Mapping radon riskareas in Perak State, Malaysia'. PLoS ONE 16(7):e0254099. https://doi.org/10.1371/journal.pone.0254099

[9] C.D. Thomas, A. Cameron, R.E. Green, M. Bakkene, L.J. Beaumont, Y.C. Collingham, E. Marinez, F.D. Siqueira., A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. Jaarsveld, G. F. Midgley,L. Miles, M. A. Ortega-Huerta, A.T. Peterson, O.I. Phillips and S. E.Williams (2004). 'Extinction risk from climate change' Nature Volume 427 pages145–148.

[10] A. Zellner (1971): *An introduction to Bayesian Inference in Econometrics*. Wiley, New York.

AUTHOR'S PROFILE

Dr. O.B. Akanbi, He lives in Ibadan, Nigeria. He is a lecturer in Statistics Department, University of Ibadan, Nigeria. His areas of research are Econometrics, Bayesian Inference & Modelling, Sampling Techniques and Operations research. He has Professional diploma (Statistics), B.Sc. (Statistics), M.Sc. (Statistics) and Ph.D. (Statistics) from the University of Ibadan, Nigeria. Dr. Akanbi has to his credit a number of refereed journal articles and conference papers within and outside Nigeria. He is a reviewer to some local and international journals. He has also written different statitical books on Sample survey and operartions research. He is a member of some statistical societies; International Statistical Institute (ISI); International Society for Bayesian Analysis (ISBA); University of Ibadan Laboratory for Interdisciplinary Statistic--al Analysis (UI-LISA); International Biometric Society (IBS) Group Nigeria (GNi); Statistics & Probability African society (SPAS); Nigerian Statistical Association (NSA); and Nigerian Young Statisticians Group (NYSG).