PRINT ISSN 1119-8362
Electronic ISSN 1119-8362

Full-text Available Online at
https://www.ajol.info/index.php/jasem
https://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 26 (2) 211-218 February 2022

# Geospatial Evaluation of Open Defecation Indicators from Across the 36 States and Federal Capital Territory in Abuja, Nigeria

**[*1]AJAO, IO; [2]ODEYEMI, CA; [1]OBAFEMI, OS**

*[*1]Department of Mathematics and Statistics, [2]Department of Surveying and Geoinformatics, Federal Polytechnic, Ado-Ekiti, Nigeria*

**Corresponding Author Email: isaacoluwaseyiajao@gmail.com, Tel: 08035252017*

**ABSTRACT:** This paper analyzes the geoSpatial serial correlations of open defecation indicators such as water availability, unimproved sanitation, literacy level and Gini coefficient from across the 36 states and FCT Abuja in Nigeria for 2018. It examines the relationship that may exist between open defecation and each of the indicators considered. The various spatial model selection diagnostic tests conducted revealed that the Spatial Lag Model is the most appropriate predicting model having the minimum values from some information criteria, results show that the global Moran's I values of open defecation is below 0.2, and spatial lag coefficient is -0.0243, this shows that open defecation in Nigeria decreases by 0.0243% for each additional 1% of water availability and other factors in the neighboring states. The results also show that unimproved sanitation is the only significant predictor for open defecation challenge in Nigeria, based on the available data. The states with highest proportion of open defecation are Kogi, Plateau, and Bayelsa, while states with the lowest include Katsina, Abia, and Akwa-Ibom. This paper provides beneficial policy recommendations for reducing open defecation in Nigeria and areas of focus towards achieving the national vision of making Nigeria open defecation Free by 2025.

**DOI**: https://dx.doi.org/10.4314/jasem.v26i2.6



**Impact factor**: http://sjifactor.com/passport.php?id=21082

**Google Analytics: https://www.ajol.info/stats/bdf07303d34706088ffffbc8a92c9c1491b12470**



**Keywords**: Spatial serial correlation, Open defecation, unimproved sanitation, Spatial lag model

Open defecation refers to the habit of using open spaces and public places for excreting. The habit is dangerous for human health, especially the youths and the elderly. This is obviously a major factor drawing the Nigeria back from achieving the Millenium Development Goals (MDG) because it is practiced in some parts of the country.

Spatial serial correlation records measure the spatial dependence between assessments of a comparable variable in different spots in space. The more the perception esteems are impacted are influenced by perception esteems that are geologically close to them, the more conspicuous the spatial relationship. It is moreover described as the positive or negative relationship of a variable with itself in view of the spatial region of the perceptions. On the other hand Spatial insights is the quantifiable examination of land associations, including spatial appropriation, Spatial serial correlation, and spatial alliance (Mustafa, Rompaey, Cools, and Saadi, 2018). The theoretical models were used to calculate the change coefficient. Connection examination results connoted "spatially packed or grouped." By assessment, non-related or erratic examination results signified "spatially sporadic or arbitrary. The pointers for figuring Spatial serial correlation can be arranged into two social occasions, to be explicit, worldwide Spatial serial correlation markers and neighborhood Spatial serial correlation pointers. The most generally perceived worldwide Spatial serial correlation pointer is the Moran's I measure (Cliff and Ord, 2013). Spatial models have been applied in a collection of orders, for instance, criminal science, demography, monetary issue, the investigation of illness transmission, political hypothesis, and general prosperity. Cressie (1993), Darmofal (2015), LeSage and Pace (2009), and Waller and Gotway (2004) give course understanding introductions. Darmofal (2015) gives a preface to spatial weighting organizations. LeSage and Pace (2009) portray outright, direct, and indirect impacts. Anselin (1988) gives a model preamble to the subject. The objective of this paper is to evaluate open defecation indicators from across the 36 states and Federal Capital Territory in Abuja, Nigeria.

## MATERIALS AND METHODS

Using the regression equation matrix

$$y = \mathrm{X}\beta + \mathrm{u} \qquad (1)$$

**y** is $n \times 1$ vector of output-variable, while **X** is the $n \times K$ covariate matrix, $\boldsymbol{\beta}$ is the $K \times 1$ regression parameter vector, then **u** is the $n \times 1$ vector of errors. It is hypothesized that $u_i$ are identically distributed with

$$E(u_i) = 0 \quad E(u_i^{\,2}) = \sigma^2 \qquad (2)$$

The aim is to test the hypothesis that $u_i$ are not correlated,

$$H_0 : E(\mathrm{uu}') = \sigma^2 \mathrm{I} \qquad (3)$$

*Using Moran's Test of nearby residual correlation:* Most researchers accept that the matrix of spatial weighting denoted by $\mathbf{W}_1$ will produce a better output of spatial links directly for the white noise **u**. As a result, the $H_0$ could be tested by the researcher using the standard Moran $I$ test statistic (Moran 1950)

$$I = \frac{\hat{\mathrm{u}}'\mathrm{W}_1\hat{\mathrm{u}}}{\hat{\sigma}^2\left[\mathrm{tr}\{(\mathrm{W}_1' + \mathrm{W}_1)\mathrm{W}_1\}\right]^{1/2}} \qquad (4)$$

Note that $\hat{\mathrm{u}} = y - \mathrm{X}\hat{\beta}$ are the summarized error values and $\hat{\sigma}^2 = \hat{\mathrm{u}}'\hat{\mathrm{u}}/n$ is the following estimator for $\sigma^2$. From assumption from literatures, from Kelejian and Prucha (2001) it should follow that $I \sim N(0; 1)$ and

$$I^2 = \chi^2(1) \qquad (5)$$

The other case that should be considered is when the person carrying out the research does not have enough information on the weighting matrices if $\mathbf{W}_1, \mathbf{W}_2 \ldots \mathbf{W}q$ sufficiently model the spatial relationship among $u_i$. If this happens, the person can test $H_0$ by making use of the $I(q)^2$ test statistic:

$$I(q)^2 = \begin{bmatrix} \hat{\mathrm{u}}'\mathrm{W}_1\mathrm{u}/\hat{\sigma}^2 \\ \vdots \\ \hat{\mathrm{u}}'\mathrm{W}_q\mathrm{u}/\hat{\sigma}^2 \end{bmatrix}' \Phi^{-1} \begin{bmatrix} \hat{\mathrm{u}}'\mathrm{W}_1\mathrm{u}/\hat{\sigma}^2 \\ \vdots \\ \hat{\mathrm{u}}'\mathrm{W}_q\mathrm{u}/\hat{\sigma}^2 \end{bmatrix} \qquad (6)$$

Where $\Phi = (\phi_{rs})$ and $r, s = 1, \ldots, q$ :

$$\phi_{rs} = \frac{1}{2}\mathrm{tr}\{(\mathrm{W}_r + \mathrm{W}'_r)(\mathrm{W}_s + \mathrm{W}'_s)\} \qquad (7)$$

It follows from Kelejian and Prucha (2001) and Drukker and Prucha (2013) that $I(q)^2 = \chi^2(q)$ under $H_0$

The term Spatial serial correlation can likewise be alluded to as breaking down the relationship of similar variable in various spatial areas. The strategy for investigation are worldwide Spatial serial correlation and nearby Spatial serial correlations. Three potential outcomes are: positive Spatial serial correlation, negative Spatial serial correlation, and no Spatial serial correlation (Wu Hong, 2015).

*Method of worldwide Spatial serial correlation:* Worldwide Spatial serial correlation identifies the spatial example of the entire examination zone and uses a solitary incentive to reflect the autocorrelation of the whole locale. In this paper, the worldwide Spatial serial correlation list is embraced to test the worldwide Spatial serial correlation (Li, Tang, Kong, Liu, and Yang, 2016). The worldwide Spatial serial correlation record is meant as Global Moran's I. The figuring interaction is given in the more elaborated equation beneath

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\sum_{i=1}^{n}(x_i - \bar{x})} \qquad (8)$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{and} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Note that $I$ is the worldwide Spatial serial correlation index; the number of spatial unit data is $n$; the attribute values of the spatial divisions $i$ and $j$ are $xi$ and $xj$, accordingly; while the spatial weight coefficient matrix is $w_{ij}$ to explain the adjacency relation of each spatial divisions.

*Method of nearby Spatial serial correlation:* The autocorrelation of a nearby states by index of nearby autocorrelation, is represented as Local Moran's I. This can be evaluated as (9).

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{s^2} \qquad (9)$$

Note that $I$ is the worldwide Spatial serial correlation index; the number of spatial unit data is $n$; the attribute values of the spatial divisions *i* and *j* are *xi* and *xj*, accordingly; while the spatial weight coefficient matrix is $w_{ij}$ to explain the adjacency relation of each spatial divisions. (Yang, Niu, Tang, and Zhu, 2018).

*Stating the matrix of the Spatial Weight:* What explains the spatial arrangement of the indicators among the states in the country is the matrix of the spatial weight (Liao and Xia, 2016). The binary symmetric spatial weight matrix (*W*) explains the relationships which is adjacent in nature among many states in the country. This can be expressed using the matrix below:

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdot & \cdot & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & \cdot & \cdot & w_{2n} \\ & & & & & \\ w_{n1} & c & \cdot & \cdot & \cdot & w_{nn} \end{bmatrix}$$

*Spatial Lag Model (SLM):* SLMs are the incorporation of spatial lag markers to clarify spatial reliance brought about by externalities and overflow impacts. The expression "lag" depicts indicators in a given space, or target spatial data, that affect the spatial data of neighboring states. The spatial information of neighboring states additionally impact the objective spatial data (Anselin, 1988).

$$Y = \alpha + \rho WY + \beta X + \varepsilon$$
$$\varepsilon \sim \text{iid N}(0, \sigma^2 \text{I}) \qquad (11)$$

The SLM created by merging the discoveries of various findings is shown in Equation (11) (Anselin, 1999) the intercept is α, ρ is the spatial autoregressive coefficient, WY is the spatial lag variable, β is the regression coefficient, X is the predictor variable and $\varepsilon$ is the vector of the error term. The SLM takes out the impedance brought about by Spatial serial correlation and tests the impacts of spatial connections. Not the same as customary regression models, for example, the OLS, SLMs contain an extra spatial lag variable. This involves including a relationship lattice of the examination and neighboring examples into the regression model, where ρ is the spatial autoregressive coefficient. Along these lines, regardless of whether the variable equivalents 0 (ρ , 0) is assessed to decide whether Spatial serial correlation exists in the SLM (Can, 1992).

*Spatial Error Model (SEM):* SEMs surmise that Spatial serial correlation is accessible in the residual terms. This thought hopes to address model errors, particularly the proximity of Spatial serial correlation. In SEMs, terms are normally controlled by expanding the coefficient of spatial residuals λ with the spatial weight lattice. By then, whether or not the coefficient of spatial blunder λ has quantifiable factual importance and counterparts 0 (λ, 0) are evaluated to choose whether Spatial serial correlation exists in the SEM. The model is imparted in Equation (12) (Kelejian and Prucha, 1998).

$$Y = \alpha + \beta X + \varepsilon \text{ with, } \varepsilon = \lambda W_{\varepsilon} + \varepsilon \quad \varepsilon \sim \text{iid N}(0, \sigma^2) \qquad (12)$$

The intercept term is α , the regression coefficient β , the predictor vector is *X*, $\varepsilon$ is the vector of the residual term, the spatial error coefficient is λ, the spatial weight matrix is *W* and the modified error term is ξ. The SEM can be utilized to wipe out the impedance of Spatial serial correlation and get precise assessment results and measurable speculations

*Spatial Durbin Model (SDM):* Includes spatially lagged dependent variable and spatially lagged explanatory indicators:

$$Y = \alpha + \rho WY + \beta X + WX\theta + \varepsilon \qquad (13)$$

where $Y$, $X$, and $\beta$ are defined as above, $\rho$ is the spatial autoregressive parameter, $W$ is the spatial weight matrix (nxn), $\rho WY$ is the endogenous interaction effect, $\theta$ (kx1) vector of unknown parameters, while $WX\theta$ is the exogenous interaction effect (Wang, Chang and Wang, 2019).

*Spatial serial correlation Tests:* The Global Moran (MI), Global Geary (GC), Global Getis-Ords (GO), Moran MI Error Test, LM Error (Burridge), LM Error (Robust), LM Lag (Anselin), LM Lag (Robust) to conform the validity of spatial autoregressive models. The MLE is used to estimate the fit of the spatial autoregressive model. The Log Likelihood Function (LLF), Akaike Information Criterion (AIC), Schwarz Criterion (SC), Amemiya Prediction Criterion (FPE), Hannan-Quinn Criterion (HQ), Rice Criterion (Rice), Shibata Criterion (Shibata), Craven-Wahba Generalized Cross Validation (GCV) are employed to test overall goodness-of-fit. A smaller criterion value denotes a stronger goodness-of-fit (Anselin, 1988).

## RESULTS AND DISCUSSIONS

The data used for this research was obtained from National Bureau of Statistics (NBS) bulletin 2018 edition. It contains indicators on open defecation across the 36 states a in Nigeria. All data analyses were done using STATA14.0 , R 3.6.2, ArcGIS 10.7, and GeoDa 1.14.0

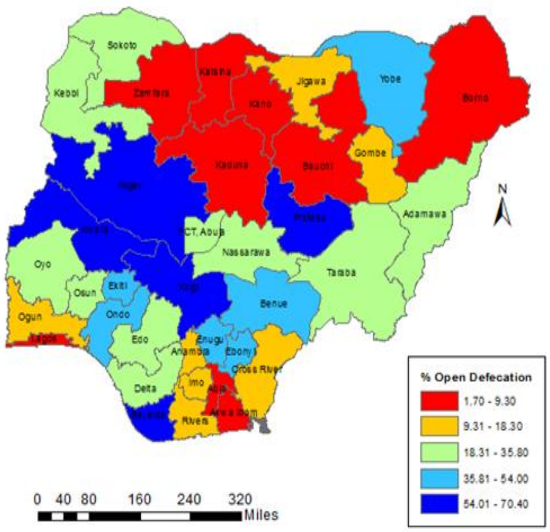


**Fig. 1:** Map showing percentage distribution of open defecation occurrences by states in Nigeria

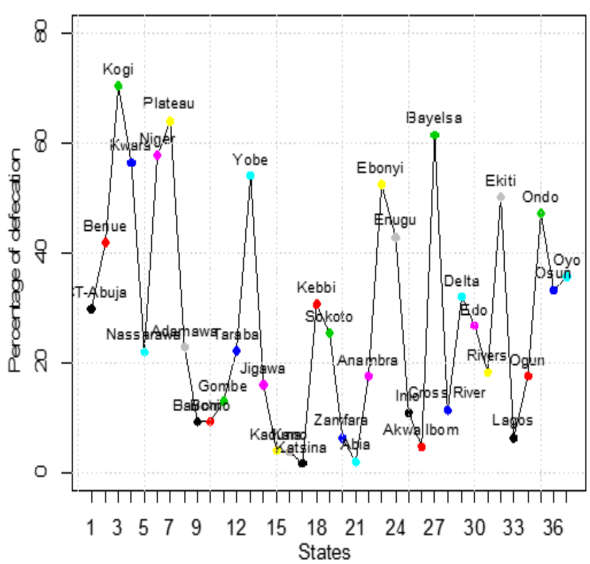


**Fig. 2:** Chart showing percentage distribution of open defecation occurrences by states in Nigeria

It can be observed from the percentage distribution chart in fig. 1 and fig. 2 that the states with the highest proportion of open defecation are Kogi, Plateau, and Bayelsa having 70.4%, 64%, and 61.5% respectively, while states with the lowest proportion include Katsina, Abia, and Akwa-Ibom have 1.7%, 2%, and 4.8% respectively.

*Spatial model selection:* Before analyzing the spatial correlation among open defecation and other indicators, a reliable and adequate model selection has to be made. In this paper, four spatial models which are, OLS, SLM, SEM, and SDM were tested on the data and various regression models and estimates were obtained. Table 1 shows the criteria employed and their respective values from the spatial models:

**Table 1:** Spatial model selection diagnostic criteria

| Criterion | OLS | SLM | SEM | SDM |
|---|---|---|---|---|
| LLF | 153.61200 | 153.5000 | 154.2800 | 152.2300 |
| AIC | 313.5900 | 298.3960 | 360.4190 | 320.1640 |
| Log AIC | 5.7481 | 5.6984 | 5.8873 | 5.7688 |
| SC | 389.8560 | 370.9660 | 428.9440 | 434.2410 |
| Log SC | 5.9658 | 5.9161 | 6.0614 | 6.0736 |
| FPE | 314.1120 | 298.8920 | 360.7250 | 321.6440 |
| HQ | 338.6040 | 322.1980 | 383.2410 | 356.4800 |
| Rice | 327.9610 | 312.0710 | 370.4330 | 352.7920 |
| Shibata | 304.0050 | 289.2750 | 353.1160 | 302.2820 |
| GCV | 319.9550 | 304.4520 | 364.9910 | 333.5840 |

As a rule of thumb, the information criterion with the least value gives the best model. It can be observed from table 1 that SLM give the minimum values of AIC, Log AIC, SC, Log SC, FPE, HQ, Rice, Shibata and GCV. For instance, the AIC value of the SLM is 298.3960, which is lower than those of OLS (153.61200), SEM (360.4190), and SDM (320.1640). This applies to other criteria except the LLF criterion. This revealed that the SLM is suitable for estimating and predicting the open defecation percentage of Nigeria going by the NDHS data of 2018 and will adequately correct the estimation error which be caused by geoSpatial serial correlation.

**Table 2:** Modeling Open defecation indicators using Spatial Lag Model (SLM)

| | | | | | |
|---|---|---|---|---|---|
| Sample Size | = | 37 | | | |
| Wald Test | = | 19.541 | P-Value > Chi2(4) | = | 0.0006 |
| F-Test | = | 4.8853 | P-Value > F(4 , 32) | = | 0.0034 |
| R2 (R-Squared) | = | 0.3791 | Raw Moments R2 | = | 0.7907 |
| R2a (Adjusted R2) | = | 0.3015 | Raw Moments R2 Adj | = | 0.7646 |
| Root MSE (Sigma) | = | 16.8459 | Log Likelihood Function | = | -153.5049 |

| **Defecation** | **Coef.** | **Std. Err.** | **z** | **P>\|z\|** | **[95% Conf.** | **Interval]** |
|---|---|---|---|---|---|---|
| Gini | 55.5829 | 47.6151 | 1.17 | 0.243 | -37.741 | 148.9069 |
| Water | -0.1261 | 0.0996 | -1.27 | 0.206 | -0.3214 | 0.0692 |
| Sanitation | -0.5579 | 0.1488 | -3.75 | 0.001 | -0.8496 | -0.2661 |
| Literacy | -0.4454 | 0.2909 | -1.53 | 0.126 | -1.0157 | 0.1247 |
| _cons | 69.2446 | 28.1433 | 2.46 | 0.014 | 14.0853 | 124.4039 |
| /Rho | -0.0244 | 0.0386 | -0.63 | 0.529 | -0.1000 | 0.0514 |
| /Sigma | 15.3097 | 1.7809 | 8.6 | 0.000 | 11.8192 | 18.8003 |

There is an overall significance (p-value = 0.0034) of the spatial effect of the indicators on the percentage of defecation. The R-square is low (0.3791), but the raw moment R-square is high (0.7907). This shows that the fit is good enough for prediction. It can be observed that the most significant variable is unimproved sanitation (p-value = 0.001), and since the contribution is negative, this means an increase in the percentage of sanitation will have a unit decrease in the percentage of defecation across the states. This paper calculates values of global Moran I, global Geary, global Getis-Ords among others for open defecation percentage of 37 states (FCT Abuja inclusive) in 2018. The results in table 3 shows that the hypothesis of no autocorrelation is not rejected in all the models except in global Getis-Ords test under SEM. For instance, all the Moran I values of defecation percentage in the 37 states fall below 0.2, with p-values greater than significance level 0.05. This indicates an overall non significant spatial positive correlation of open defecation in Nigeria. A non-statistically significant positive global Moran's *I* suggests non existence of Spatial serial correlation across the states boundaries. From table 4, it can be discovered that the p-values exhibited by all the various tests of heteroscedasticity are not lower than the 0.05 significance level. This implies non rejection of null hypothesis of homoscedasticity existence. As a result there is no violation of non-constant variance common in regression models. It can be seen from table 5 that the normality assumption is not rejected using the various tests except for White Lagrange multiplier test under the OLS and SLM having p-values 0.0313 and 0.0196. Others confirmed the normality of the errors

**Table 3:** Spatial serial correlation tests

| Test | OLS | | SLM | | SEM | | SDM | |
|---|---|---|---|---|---|---|---|---|
| | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** |
| Global Moran (MI) | 0.1479 | 0.0692 | 0.1477 | 0.0692 | 0.1259 | 0.1131 | 0.0157 | 0.6535 |
| Global Geary (GC) | 0.8121 | 0.1993 | 0.8121 | 0.1993 | 0.8557 | 0.3074 | 0.9172 | 0.5652 |
| Global Getis-Ords (GO) | -0.6946 | 0.0700 | -0.6946 | 0.0689 | -0.5918 | 0.0000 | -0.0737 | 0.6536 |
| Moran MI Error Test | 0.7951 | 0.4265 | 0.7951 | 0.4265 | 0.7356 | 0.462 | 0.8002 | 0.4236 |
| LM Error (Burridge) | 1.8981 | 0.1683 | 1.8981 | 0.1683 | 1.3779 | 0.2405 | 0.0214 | 0.8838 |
| LM Error (Robust) | 4.4765 | 0.0344 | 4.4765 | 0.0344 | 3.7831 | 0.0518 | 0.2611 | 0.6093 |
| LM Lag (Anselin) | 0.2731 | 0.6013 | 0.2731 | 0.6013 | 0.3134 | 0.5756 | 0.1274 | 0.7211 |
| LM Lag (Robust) | 2.8514 | 0.0913 | 2.8514 | 0.0913 | 2.7185 | 0.0992 | 0.3672 | 0.5445 |

**Table 4:** Spatial Heteroscedasticity tests

| Test | OLS | | SLM | | SEM | | SDM | |
|---|---|---|---|---|---|---|---|---|
| | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** |
| Hall-Pagan LM Test | 0.1130 | 0.7368 | 0.1131 | 0.7366 | 0.5254 | 0.4685 | 0.0064 | 0.9362 |
| Harvey LM Test | 0.9463 | 0.6230 | 0.5860 | 0.7460 | 0.6297 | 0.7299 | 3.0031 | 0.2228 |
| Wald LM Test | 2.3349 | 0.1265 | 1.4460 | 0.2292 | 1.5538 | 0.2126 | 7.4098 | 0.0065 |
| Glejser LM Test | 1.4169 | 0.4924 | 1.0705 | 0.5855 | 1.2497 | 0.5353 | 1.5814 | 0.4535 |
| Machado-Santos-Silver Test | 4.3125 | 0.1158 | 6.1323 | 0.0466 | 4.2310 | 0.1206 | 1.9707 | 0.3733 |
| White-Test Koenker ($R^2$) | 1.3175 | 0.8584 | 1.6266 | 0.8040 | 1.5111 | 0.6797 | 1.6443 | 0.9493 |
| White-Test B-P-G (SSR) | 1.7344 | 0.7845 | 1.8068 | 0.7712 | 1.1281 | 0.7703 | 2.1789 | 0.9025 |
| Cook-Weishberg LM Test | 0.1487 | 0.6998 | 0.1257 | 0.7230 | 0.3922 | 0.5311 | 0.0085 | 0.9266 |

**Table 5:** Spatial Non-normality tests

| Test | OLS | | SLM | | SEM | | SDM | |
|---|---|---|---|---|---|---|---|---|
| | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** |
| Jarqua-Bera LM Test | 2.2087 | 0.3314 | 3.2915 | 0.1929 | 1.2921 | 0.5241 | 2.5950 | 0.2732 |
| White LM Test | 6.9306 | 0.0313 | 7.8641 | 0.0196 | 5.6065 | 0.0606 | 4.7684 | 0.0922 |
| Doornik-Hansen LM Test | 2.5931 | 0.2735 | 3.3811 | 0.1844 | 1.6446 | 0.4394 | 3.1509 | 0.2069 |
| Geary LM Test | -1.1342 | 0.5672 | -1.1342 | 0.5672 | -0.0826 | 0.9595 | -1.1342 | 0.5672 |
| Anderson-Darling Z Test | 0.4516 | 0.7221 | 0.5242 | 0.8157 | 0.3090 | 0.5879 | 0.3936 | 0.6180 |
| D'Agostino-Pearson LM Test | 3.0790 | 0.2145 | 4.6598 | 0.0973 | 1.8740 | 0.3918 | 3.4401 | 0.1791 |

**Table 6:** Spatial Regression Specification Error Test (RESET)

| Test | OLS | | SLM | | SEM | | SDM | |
|---|---|---|---|---|---|---|---|---|
| | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** | **Coef.** | **Signif.** |
| Ramsey RESETF1 Test | 0.888 | 0.3534 | 3.306 | 0.0787 | 5.142 | 0.0302 | 1.707 | 0.2016 |
| Ramsey RESETF2 Test | 0.546 | 0.5850 | 1.852 | 0.1745 | 2.553 | 0.0941 | 1.249 | 0.3024 |
| Ramsey RESETF3 Test | 0.363 | 0.7803 | 2.209 | 0.1083 | 1.709 | 0.1862 | 0.848 | 0.4799 |
| Debenedicts-Giles Reset1 Test | 0.123 | 0.8846 | 0.081 | 0.9226 | 0.091 | 0.9128 | 0.747 | 0.483 |
| Debenedicts-Giles Reset3 Test | 0.603 | 0.7254 | 2.516 | 0.0470 | 0.769 | 0.6006 | 0.597 | 0.7301 |

**Table 7:** Spatial Multicollinearity Diagnostic Tests

| Test | OLS | | SLM | | SEM | | SDM | |
|---|---|---|---|---|---|---|---|---|
| | **Chi2** | **P-value** | **Chi2** | **P-value** | **Chi2** | **P-value** | **Chi2** | **P-value** |
| Farrar Glauber | 28.7104 | 0.0001 | 15.4098 | 0.0173 | 16.5855 | 0.0009 | 48.2808 | 0.0000 |

The results of Spatial Regression Specification Error Test (RESET) as presented in table 6 showed that the null hypothesis that says the model is correctly specified is not rejected using the various model mis-specification tests considering the four models.
There is evidence of multicollinearity among the predictors using the various models as indicated in table 7. This is a common phenomenon in data sets such as this

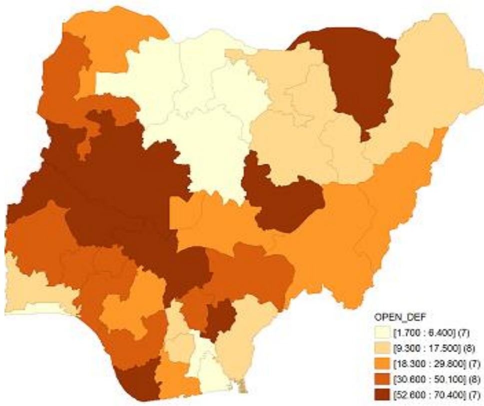


**Fig. 3**: Spatial distribution of open defecation by states

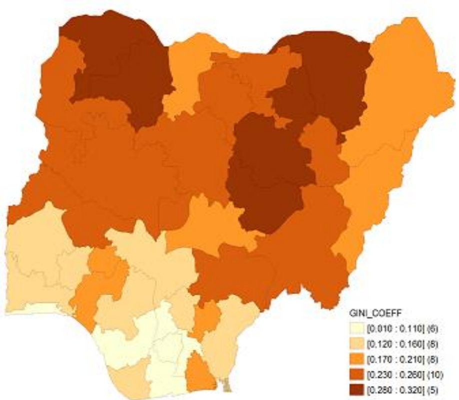


**Fig. 4**: Spatial distribution of Gini coefficients by states

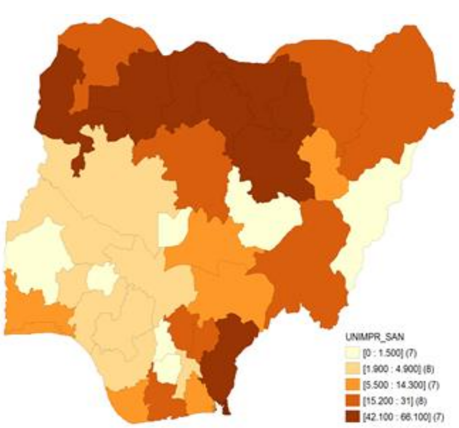


**Fig. 5:** Spatial distribution of unimproved of sanitation by states

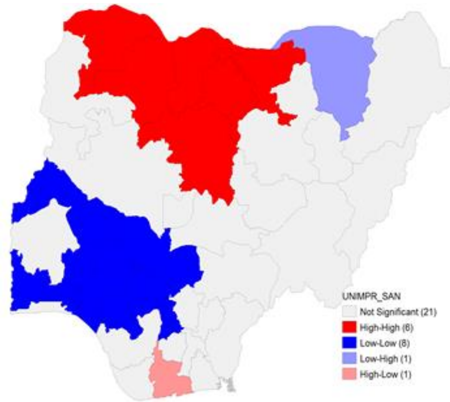


**Fig. 6**: LISA cluster map of unimproved sanitation

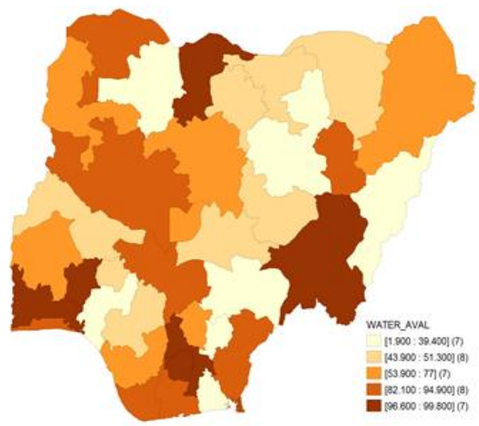


**Fig. 7:** Spatial distribution of water availability by states

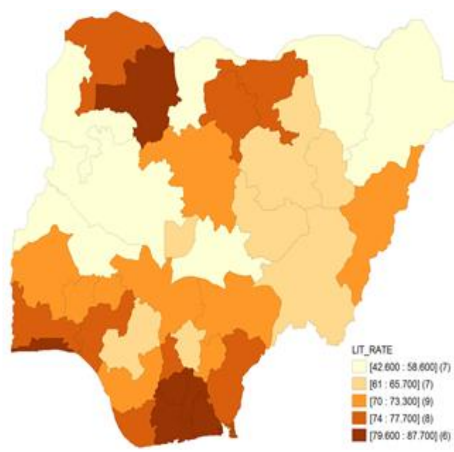


**Fig. 8:** Spatial distribution of literacy rate by states

*Local Indicators of Spatial Association (LISA) analysis:* Anselin (1999) divided the LISA values into four quadrants based on the degree of spatial clustering specifically, High-High, Low-High, Low-Low and High-Low, as illustrated in figure 8

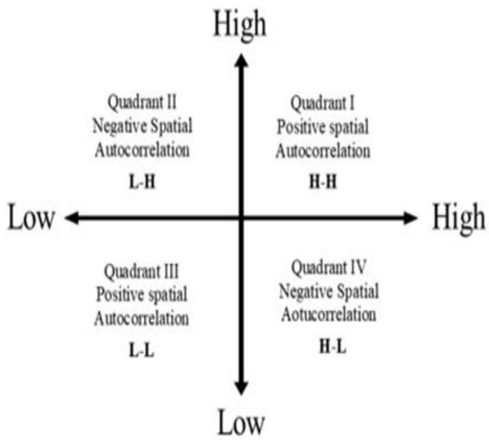


**Fig. 9**: Spatial serial correlation of the LISA analysis

It can be observed from the percentage distribution charts in figs. 1 and 2 that the states with the highest proportion of open defecation are Kogi, Plateau, and Bayelsa, while states with the lowest proportion include Katsina, Abia, and Akwa-Ibom. It is evident from the results displayed in table 1 that the Spatial Lag Model (SLM) is the most appropriate predicting model having the minimum value of the AIC from the selected information criteria. Results also show that the global Moran's I values of open defecation from across the nation is below 0.2, and spatial lag coefficient ρ stands at -0.0243. The results of regression analysis show that unimproved sanitation (p-value = 0.0001) at 5% significance level is the only and most significant predictor for open defecation challenge in Nigeria, based on the available data. The null hypotheses were not rejected in all the essential tests carried out: The Spatial serial correlation test, spatial heteroscedasticity test, spatial non-normality test, and spatial regression specification error test as presented in tables 3, 4, 5, and 6. According to the analysis of LISA cluster maps of open defecation in Nigeria in 2018 using fig. 6, the unimproved sanitation high-high cluster covers 6 states majorly in the North-west, which include Sokoto, Zamfara, Kaduna, Kano, Katsina, and Jigawa. Low-low cluster areas are found in 8 states in the South west, south-East and South-West, which are Ogun, Ekiti, Ondo, Edo, Osun, Kwara, Kogi, and Anambra. High-low cluster is found only in Rivers, while low-high cluster around Yobe. The cluster maps also reveal that 21 states are not significant in term of cluster analysis. The hotspots and coldspots (high-high and low-low clusters respectively) tend to be grouped together in several states, while the high-low and low-high clusters, which indicate comparative outliers, are usually a single area. The high-high quadrant represents high-percentage of unimproved sanitation surrounded by other high- percentage of unimproved sanitation and the low-low quadrant represented low-percentage of unimproved sanitation surrounded by other low-percentage of unimproved sanitation.

*Conclusion*: It can be concluded based on the above results that the percentage of open defecation is still very high in Nigeria. We therefore recommend that the government through the relevant agencies should provide improved and adequate sanitation facilities throughout the country, if the vision of making Nigeria open defecation Free by 2025 will be realistic.

## REFERENCES

Anselin, L. (1995). Local indicators of spatial association: LISA. *Geogr. Anal. 93*, 93–115.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*; Springer Science & Business Media: Dordrecht, The Netherlands; ISBN 978-94-015-7799-1.

Anselin, L. (1999). Spatial Econometrics, Available online: http://www.csiss.org/aboutus/presentations/ files/baltchap.pdf (accessed on 1 December 2018).

Can, A. (1992). Specification and estimation of hedonic housing price models. *Reg. Sci. Urban Econ. 22*, 453–474.

Cliff, A.D; Ord, J.K. (1973). *Spatial Autocorrelation*; Classics in Human Geography Revisited; Pion: London, UK.

Cliff, A.D; Ord, J.K. (1981). *Spatial Processes: Models and Applications*; Pion: London, UK; Volume 266, ISBN 08-85086-081-4.

Cressie, N. (1993). *Statistics for Spatial Data*. Rev. ed. New York: Wiley.

Darmofal, D. (2015). *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press.

Drukker, D. M; I. R. Prucha. (2013). On the $I2(q)$ test statistic for spatial dependence: Finite sample standardization and properties. *Spatial Economic Analysis* 8: 271–292.

Federal Ministry of Water Resources, Nigeria (2013). National Water-Sanitation Policy Implementation Guidelines, Third Draft

Kelejian, H. H.; I. R. Prucha. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104: 219–257.

LeSage, J.; R. K. Pace. (2009). Introduction to Spatial Econometrics. Boca Raton, FL: Chapman & Hall/CRC.

Liao Xianchun,; Xia Enlong (2015). Why China attracts FDI inflows? a perspective of environmental stringency and the degree of corruptibility. *World Economy Study.* 1 (01), 112.

Li Li, Tang Dengli; Kong Ying; Liu Dongjun,; Yang Yuanhua (2016). A study on the impact of FDI on urban Haze pollution: a case study of the Pearl River. Management Comment. 28 (6), 11.

Kelejian, H; Prucha, I. (1998). A generalized spatial two-stage least-squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Financ. Econ.*, *17*, 99–121.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23.

Mustafa, A; Rompaey, A.V; Cools, M.; Saadi, I; Teller, J. (2018). Addressing the determinants of built-up expansion and densification processes at the regional scale. *Urban Stud. 55*, 3279–3298.

National Population Commission (NPC) [Nigeria]; ICF. (2019). Nigeria Demographic and Health Survey 2018. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF. 25-29

UNICEF (2014). Final Report for Study on WASH Challenges in Per-Urban Areas of Nigeria, Africa Infrastructure Foundation.

Waller, L. A; Gotway, C. A. (2004). Applied Spatial Statistics for Public Health Data. Hoboken, NJ: Wiley.

Wen-Ching Wang; Yu-Ju Chang; Hsueh-Ching Wang (2019). An Application of the Spatial Autocorrelation Method on the Change of Real Estate Prices in Taitung City. *ISPRS Int. J. Geo-Inf. 8*, 249

WHO and UNICEF, JMP on Water and Sanitation, 2010, 2011, 2012, 2013 & 2014

Wu Hong (2015). Carbon emission reduction in China: Spatial and Temporal Pattern, Evolution Mechanism and Policy suggestions-theory and method based on Spatial Econometrics. Management World. (11) 3.

Yuanhua Yang; Guohua Niu; Dengli Tang; Mengjue Zhu (2018). Spatial Econometric Analysis of the Effect of Government Governance on Regional Emission Reduction: Evidence from China. *Pol. J. Environ. Stud. 27. 6, 2833-2842*