Journal of Remote Sensing & GIS
ISSN: 2469-4134

Alao et al., J Remote Sensing & GIS 2016, 5:4
DOI: 10.4182/2469-4134.1000185

Research Article Open Access

Differentiating Non-Homoscedasticity and Geospatially Extreme Outliers for Urban and Rural Landscape Dataset Using Pearson's Product Moment Correlation Coefficients for Quantitating Clustering Tendencies in Non-Vaccinated Measles Populations in Nigeria

Samuel Alao, Komi Mati and Benjamin Jacob*

Department of Global Health, College of Public Health, University of South Florida, Tampa, FL, USA

Abstract

Linearized Models on measles vaccination related centroids in literature cannot provide pertinent data for local government measles managers. Spatial analysis is a cost cutting epidemiological tool for large scale immunization programs. A multivariate regression model was constructed to determine anthropogenic related covariates. In addition, we quantitated the clustering tendencies in the auto- correlated dataset using orthogonal eigenvectors and also illustrated problem hot spots for effective vaccine coverage. Data was retrieved from Demographic Health survey 2013 for Nigeria (N=28,337). Poverty, illiteracy level, and no vitamin A supplements were strong determinants of measles non-vaccination at a statistically significant level of (P<0.0001). The first order autocorrelation statistics (DW=0.1647, P<0.0001), (DW=0.2406, P<0.0001); and second order correlation (Moran's I=0.456, Z score=1208), (Moran's I=0.442, Z score=608) demonstrated a positive spatial autocorrelation for rural and urban geo-locations respectively. Land cover land use (LCLU) maps from Google earth and Diva-GIS were uploaded into ArcMap to visually represent the hot spot areas. Significant Mapped data showed that children not vaccinated against measles are clustered in the rural areas of Muslim dominated northern parts of Nigeria.

Keywords: Autocorrelation; Eigenvector; Clustering; Spatial filtering

Introduction

Nigeria is one of the most populous countries in the world with an estimated population of 162 million people at 2013. The country is divided into 36 states with the federal capital territory (Abuja) as the seat of government. However, despite the success recorded in measles vaccination globally and associated reduction in under-five children mortality, Nigeria still lags behind by 42% coverage. 90%-95% coverage is needed to achieve herd immunity required for regional elimination [1]. Nigeria's under-five mortality is 143 per 1000 at 2010 and makes it the highest in Africa sub region. 1 million children die before the age of five and measles accounts for 4% of these deaths within this age group [2,3]. Despite, 85% proven measles vaccine effectiveness when given to infants at 9 months: the challenges with cold chain, logistics, and program failures has significantly diminished the impact of immunization effectiveness [4].

Similarly, measles deaths occur in countries like Nigeria where vital registration systems cannot provide reliable information on cause-specific mortality. Because of these challenges, WHO has relied on mathematical models to estimate the global burden of measles. For example, multi-variate logistic regression model has been used to study the prevalence rates of measles, rubella, scarlet fever, and the differences among measles and other rash and febrile illness(RFIs) [5]. However, logistic regression was inadequate in addressing the consistency of measles diagnosis from other RFIs based on clinical and laboratory results. Likewise, Ref. [6] applied logistic regression analysis on 2007 demographic and Health Survey for Bangladesh (BDHS) and discovered the positive influence of maternal education on the uptake of measles vaccination.

However, dichotomous Linear models cannot heuristically optimize datasets due to inconspicuous non-quantitated heteroscedasticity. Hence, due to non-quantitated non-homoscedastic measles-related predictors: linearized residual forecasts of geo-referenced measles related primary sampling units/neighborhoods geospatially over lied on urban and non-urban geo-classified LULC polygons currently in literature cannot provide pertinent data for local government measles managers. Subsequently, patient population proportional ratio cannot be measured on urban diffusing land cover nor on rural geo-classified, undisturbed pasture grasslands. Previous measles related models have not been objectively incorporated into a measles landscape surveillance data system and instead relied on vaccination coverage data as the primary indicator of local disease burden.

However, Poisson probability paradigms can render elucidative explanatory count-variable responses and the residuals from the model output e.g., (pseudo R^2 values at a 95% confidence interval in PROC REG). Consequently, these models could either consistently capture the effects of large measles mortality where low vaccination coverage is reported, or show periods of low mortality between outbreaks when high vaccination coverage is reported in specific neighborhoods.

Similarly, these frequentist models have revealed over-dispersion due to extreme observations (outliers) and missing interactions in the independent variables [7]. Over- Poissonian variation is commonly due to violation of the assumption that the variance is equal to the mean [8,9]. In statistics, over dispersion is the presence of greater

***Corresponding author:** Benjamin Jacob, Department of Global Health, College of Public Health, University of South Florida, Tampa, FL, USA, Tel: 8139742011; E-mail: bjacob1@health.usf.edu

Received November 04, 2016; **Accepted** December 08, 2016; **Published** December 10, 2016

Citation: Alao S, Mati K, Jacob B (2016) Differentiating Non-Homoscedasticity and Geospatially Extreme Outliers for Urban and Rural Landscape Dataset Using Pearson's Product Moment Correlation Coefficients for Quantitating Clustering Tendencies in Non-Vaccinated Measles Populations in Nigeria. J Remote Sensing & GIS 5: 185. doi: 10.4182/2469-4134.1000185

variability in a data set than would be expected based on a given statistical model.

A negative binomial regression model with a non-homogenous gamma distributed mean can compensate for over dispersion in PROC GENMOD for a geo-sampled dataset of unvaccinated measles count in LULC forecasting vulnerability model. The regression residual rendered from the compensatory model cannot quantitate clustering tendencies in urban and non -urban geo-classified supervised or unsupervised LULC dataset [8].

Furthermore, a first-order autocorrelation cluster models in SAS can provide a probabilistic framework to predict the unobserved elements of a dynamic landscape process, such as true measles incidence in an urban commercial environment, when provided with observed elements of the dynamic process, such as reported measles cases and vaccination coverage.

Traditional Measles data analyses often study the form, direction, and strength of the relationship exhibited by two quantitative variables measured on a single set of n observations. A scatterplot in SAS can visualize this relationship, with a conventional correlation coefficient describing the direction and strength of a straight-line relationship of the overall landscape measles analyses pattern. Positive spatial autocorrelation means that geographically nearby values of a variable tend to be similar on a measles land covariate map: high values tend to be located near high values, medium values near medium values, and low values near low values. According to Griffiths [10], families with similar economics, tend to organize themselves in a way that concentrates similar household attributes on a map-creating positive spatial autocorrelation amongst many variables.

The Durbin Watson (DW) test is a well-known formal method of testing if intra-cluster serial error correlation; and this can constitute a serious obstacle undermining a model's inferential suitability. The DW tests can detect serial error autocorrelation in the geo-referenced measles unvaccinated related cluster model by assuming that error is normally distributed independently of predictor variables with zero mean and constant variance [11].

Recent quantitative geographical analysis methods have supplemented risk mapping, geo-referenced explanatory data by decomposing the Moran's I into synthetic variates, whose linear combinations constitute a spatial filter logistic model, with a generalized linear model specification to determine residual auto correlation [12]. The eigenvector filtering approach is a non-parametric technique that removes the inherent autocorrelation from generalized linear regression models by treating it as a missing variable [10].

However, the purpose of a non-parametric spatial filtering technique is to control for autocorrelation with a set of proxy variables rather than to identify a global autocorrelation parameter for a spatial process [10]. The basis for this procedure is the decomposition of the Moran's *I* into orthogonal and uncorrelated measles unvaccinated risk related geo-referenced map-able pattern components in SAS (e.g., AUTOREG, PROC VARIOGRAM). This orthogonal decomposition would make the latent spatial correlation represented by the geographic configuration of important unsupervised local government area stratified clusters ("hot spot") as described by a given spatial weights matrix.

These corresponding fractionalized, measles unvaccinated related eigenvectors can then be optimally employed as predictor variables in a robust predictive regression equation for determining geo-spatiotemporal covariates associated with specific measles vaccination or unvaccinated cluster-oriented parameters. However, unexplained geo-referenceable data clustering may be only artefactual as a result of differential case reporting, unknown demographic changes or duplication of case data [13].

Spatial analysis is a cost effective epidemiologic tool for detecting and predicting trends of disease outbreaks especially for large scale interventions in low resource countries [14]. Similarly, spatial autocorrelation serves as a diagnostic tool for model mispecifications common in frequentist models, spatial non-homoscedasticity and extreme observations [15].

Research on the application of GIS and remote sensing for measles vaccination surveillance is sparse in Nigeria. Additionally, no study had quantitated clustering tendencies for measles non-vaccination in Nigeria. However, Bharti et al. employed spatial techniques in determining the role of epidemiologic connectivity and seasonally variability for measles incidence in Nigeria. Therefore, we intend to apply spatial autocorrelation techniques to qualitatively quantitated the clustering tendencies in our experimental dataset to help local measles managers focus supplementary immunization activities to high risk communities. Our research objectives were:

1. To construct Poisson regression model to determine anthropogenic related covariates associated with each diva-GIS land cover map.
2. To quantitate measles unvaccinated clustering tendencies in auto-correlated dataset using orthogonal synthetic fractionalized eigenvectors.
3. Remotely summarizing the distribution of urban and non -urban measles unvaccinated centroids and their representative clinical population in every district.

Methodology

Methods

Variables of interest were derived from a systematic review of the important literature and analyzed through a generalized linear model to determine the suitability of geo-sampled normalized predictors to our geo-referenced measles unvaccinated outcome variable. Model selection for our response variable (measles non-vaccination) was based on Akaike information criterion (AIC) and Pseudo R square value (R^2) in SAS 9.4. Similarly, multicollinearity in the independent variables was excluded to preserve homoscedascity using variance inflation factor (<10) for our model in PROC REG.

We utilized Poisson regression model for the Nigerian DHS 2013 dataset, where the non-vaccinated measles count was the dependent variable on the Left-hand side of the equation and the right-hand side of the equation contained the predictor variables (literacy, wealth index, religion). The Poisson model follows the regression equation illustrated in [16]:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

Which was further rewritten as SST=SSM+SSE, where SS was notation for sum of squares and *T*, *M*, and *E* in PROC GEN MOD were the descriptions for total quantized model error estimates. The squares of the sample correlation (r^2) given as the ratio of SSM/SST was the proportion of variability for the geo-sampled, normalized measles non-vaccinated dataset [16]. Likewise, the r value was the strength of

association between measles non-vaccinated count i.e., (y_i) to the fitted values i.e., $\hat{y}_l$

In addition, we applied a negative binomial regression due to the presence of extra-Poissonial variations observed within in the Poisson regression model. The Poisson distribution is a special case of the negative binomial distribution where the mean approximates the standard deviation [9]. The negative binomial distribution reduced the extra-Poissonial variation by collapsing the outliers to the mean: in so doing, the variance was equal to the mean [17].

In the linear regression analyses, of the geo-sampled, measles unvaccinated related, normalized dataset *k*; the null hypothesis was H_0: k=0 and the alternative hypothesis was H_a: k>0. The log-likelihood (i.e., LL) for the models were noted. We employed the likelihood ratio (LR) test to compute the LR statistic using -2(LL) (Poisson) and the LL (i.e., negative binomial) as reported in (http://www.ats.ucla.edu/).

Furthermore, we applied PROC AUTOREG as a form of spatial error (SE) model in evaluating all probabilistic residual estimates from the measles unvaccinated model. The spatial error, measles unvaccinated epidemiologic model treats spatial dependency primarily as a nuisance; in the same way, statistical approaches often treat temporal serial correlation as a noise to be eliminated [10,18]. The auto-regressive process involves a geo-sampled geo-referenced, variable Y as a function of nearby neighborhood *Y* values and the error residuals of *Y* as a function of nearby *Y* residuals [10].

The Yule-Walker estimate is one of the fastest methods of quantitating autoregressive error model computationally and is the default in PROC AUTOREG. Harvey refers Yule-Walker method to as two-step transform method in 1981. Similarly, other methods of quantitating autoregressive errors are the maximum likelihood (ML), unconditional least squares (ULS) and iterated Yule-Walker method [19].

The vector (φ) for the geo-sampled asymptotic normalized measles unvaccinated related autoregressive parameters is given by the mathematical expression $\varphi=(\varphi_1, \varphi_2, \ldots. \varphi m)'$ and subsequently portrayed the variance matrix of the error vector $v=(v_1,\ldots, v_n)'$ to be Σ, $E\, v\, v'=\Sigma=\sigma^2 v$. The matrix υ was calculated with an unbiased/consistent variance (σ^2) when the autoregressive parameters φ are known using the generalized least squares (GLS). Likewise, the Yule-Walker method allows for alternation of β estimation with autoregressive parameter (φ) estimation utilizing the generalized least squares for the given autocorrelation function [20]. However, the basic building blocks for quantitating autoregressive errors using YW methods is to generate OLS of β estimation and subsequently derive autoregressive parameter φ from autocorrelation function of OLS residuals. Similarly, v is derived from estimates of φ, and Σ is generated from estimate of v and OLS estimate of φ^2. The autocorrelation corrected estimates of the regression parameters β are computed by GLS, utilizing the values derived from Σ estimated matrix. These are the Yule-Walker estimates [19].

We utilized the mathematical concept of Kalman filter algorithm utilized by Jacob et al. [17] in transforming less precise measurements or random variations to more accurate estimates of unknown variables instead of those based on single measurement. Kalman filters algorithm is a set of mathematical equations that allows for data transformation through the inverse Cholesky root of υ and subsequently compute the GLS output through an iterative process.

Additionally, we also applied PROC AUTOREG for Durbin Watson statistics to detect inherent residual error coefficients in the regression analysis for measles unvaccinated related data. The DWPROB option in PROC AUTOREG allows for the determination of significance level within our predictor variables. The DW statistics tests the null hypothesis H_o: $\varphi_1=0$ against H_1: $\varphi_1>0$. Likewise, the residuals associated the observation at time *t* is given by e_t in this study. Then, the Durbin Watson test is given by the equation below $d=\frac{\sum_{t=2}^{T}(e_t-e_t)^2}{\sum_{t=1}^{T}e_t^2}$

where T was the number of geo-referenced measles unvaccinated related count observations [21].

The value of d in the DW test for the measles unvaccinated related model is approximately 2 (1-r), where r is the sample autocorrelation of the residual. Similarly, the distribution of d falls within the values 0 and 4 and is symmetrical around 2; while, d=2 shows no serial correlation in errors [22]. A value of d<1 implies that the errors within the model are serially correlated to one another, and may violate the assumption of error independence in a given regression model. Durbin Watson d Values <2 indicates a positive serial correlation and d>2 signifies successive error terms are on average are much different from one another i.e. negatively correlated [20].

Pearson product moment correlation coefficient is a form of spatial autocorrelation method, employed in this research to quantitate geo-referenced unsupervised predicative LULC urban and non- urban geo-classified explanatories in AUTOREG. We constructed a PROC VARIOGRAM in SAS 9.4 to generate a scatter plot of standardized versus summed nearby standardized geo-referenced measles unvaccinated related neighborhoods on geo-classified LULC maps. In so doing, the geographical distribution of geo-referenced measles unvaccinated related cluster units and their associated co-variates was cartographically delineated on a geo-classified LULC maps in ArcGIS.

A global indicator of spatial association (GISA) compares variations in means of all features to differences between means for each neighbor in an entire study location. Indicators utilized in global spatial statistics to estimate spatial autocorrelation include Moran's I and Geary C. The values ranges from +1 (Positive spatial autocorrelation) to -1(Negative spatial autocorrelation) and the SAS/ArcGIS output transforms these values into z scores for statistical hypothesis testing (i.e., statistically significant or not). Large z scores signify an intense clustering of high values (hotspots) and a low z score demonstrates more intense clustering of low values (cold spots).

Moran's index is given as: $I_i=\frac{x_i-\bar{X}}{s_i^2}\sum_{J=1,J\neq 1}^{n}w_{i,j}(x_j-\bar{X})$ (2.1)

Where x_i is a distinct attribute for feature *i*, $\bar{X}$ is the mean of the subsequent attribute, $w_{i,j}$ is the spatial weight between neighborhood points *i* and *j* as described in Ref. [23].

Likewise, $s_i^2=\frac{\sum_{J=1,J\neq 1}^{n}(x_j-\bar{X})}{n-1}$ where n is the total number of features (2.2)

As such, a geo-referenced dataset of measles related primary sampling unit in Nigeria was digitally overlaid onto sub-meter geo-classified google earth data. Initially, we generated an unsupervised LULC classification in ArcGIS using urban and non- urban polygon data. We considered a linearized model framework for a regressively determining statistically significant landscape co-variates associated with each polygon using ArcGIS geo-classification and then constructed a Moran's I spatial autocorrelation measles unvaccinated primary sampling unit model.

Results

Data description

Demographic Health survey 2013 for Nigeria was drawn from primary sampling units referred to as clusters chosen from enumeration areas for the survey. A total of 40,680 households were chosen from 904 clusters within urban and non-urban zones of the country. This primary sampling frame/cluster or neighborhood is based on the population and housing census conducted in Nigeria in 2006 [24]. However, the geographic coordinates of sampled households were displaced by 0-2 km of positional error in the urban and 0-5 km in the rural zones, with 1% of the rural clusters further displaced randomly by a maximum of 10 km to protect the confidentiality of the respondents. Urban zones are often taken as population estimate of more than 20,000 individuals whose economic activity were non-agrarian [24]. A sample size (N=17,522) represent the number of under-five not vaccinated in urban and rural zones based on their vaccination cards or reported by their mother.

Table 1 was and output we generated in PROC FREQ for each of the independent variables to illustrate the frequencies of measles non-vaccination among its covariates. The results demonstrated that measles non-vaccination was higher within poor households (79.9%) when compared to the wealthy households (46.3%) in the wealth index class. Similarly, the level of maternal literacy revealed an impact of measles vaccination coverage on the Nigerian population. (81.98%) of children from non-literate mothers, were not vaccinated against measles when compared to (43.3%) of children from literate mothers that were not vaccinated. Children that received vitamin A supplement had a lower proportion of measles non-vaccination (34.9%) vis à vis the population of children without vitamin A supplement (79.1%). Likewise, among the religious groups in Nigeria, Christianity had a lower proportion of measles related non-vaccination (41.6%) in relation to Islam (74.9%) and Traditionalist (79.0%). Additionally, higher proportion of non-vaccination was noticed among the Fulani (85.7%) and Hausa (79.7%) ethnic populations of northern Nigeria when compared to Igbo (36.4%), Yoruba (37%) in the and others (55.6%) (Table 1).

PROC GENMOD with a Poisson distribution in SAS 9.4 was utilized in analyzing suitable predictors of interest based on literature reviews to our outcome variable and the LINK function was log (Table 2). We indicated the SCALE OPTION; although, it's a default in SAS 9.4 and enables the researcher to appreciate the effect of scaled Pearson chi-square on over dispersion (Table 3). The results obtained, satisfied the significant level of <0.05 for most of our regressors; except for distance (P-value=0.51) in rural, literacy in urban zone (P-value=0.782) and Hausa within the ethnicity group of the rural zone (P-value=0.713) as reported in Table 2 below.

Predictors	Not Vaccinated	Percent (%) not vaccinated
Wealth Index		
Wealthy (n=15706)	7270	46.29
Poor (n=12631)	10093	79.91
Literacy		
Literate (n=15367)	6730	43.80
Not Literate (n=12970)	10633	81.98
Residence		
Urban (n=9591)	4340	45.25
Rural (n=18746)	13023	69.47
Vitamin A supplement		
Vitamin A (+) (n=11413)	3978	34.85
Vitamin A (-) (n=16924)	13385	79.09
Religion		
Christianity (n=11649)	4851	41.64
Islam (n=16405)	12286	74.89
Traditionalist (n=272)	215	79.04
Ethnicity		
Yoruba (n=3222)	1199	37.21
Hausa (n=8749)	6974	79.71
Igbo (n=3088)	1123	36.37
Fulani (n=2257)	1935	85.73
Others (n=11010)	6121	55.59

Table 1: Frequency Distribution of Unvaccinated Measles Count with Predictor Variables.

The goodness of fit assessment comes as part of diagnostic output generated in PROC GENMOD. The deviance value divided by the degree of freedom is used as an indicator of model dispersion in this study. A value <1 is taken as under-dispersed while values >1 is taken as an over-dispersed dataset. The goodness of fit assessment of our Poisson model showed an over-dispersion of the experimental dataset by a factor of 5.3 and 3.4 respectively for urban and rural zones within the Nigerian study site (Table 3).

Furthermore, the generalized autocorrelation test in PROC AUTOREG revealed serial errors that are dependent and positively

Urban	Variable	Estimate	P-Value	Rural	Variable	Estimate	P-value
	Poverty	0.0007	<0.0001		Poverty	0.0009	<0.0001
	Wealthy	0.0000			Wealthy	0.0000	
	Vit. A (-)	0.0000			Vit. A (-)	0.0000	
	Vit. A (+)	-0.0063	<0.0001		Vit. A (+)	-0.0071	<0.0001
	Illiteracy	0.0072	<0.0001		Illiteracy	0.0047	<0.0001
	Literacy	-0.0016	0.7818		Literacy	-0.0078	0.0418
	Christianity	-0.0128	<0.0001		Christianity	-0.0226	<0.0001
	Islam	-0.0106	<0.0001		Islam	-0.0224	<0.0001
	Fulani	0.0019	<0.0001		Fulani	0.0002	<0.0022
	Igbo	0.0005	<0.0001		Igbo	-0.0009	<0.0001
	Hausa	0.0009	<0.0001		Hausa	0.0000	0.7132
	Yoruba	0.0015	<0.0001		Yoruba	0.0005	<0.0001

Table 2: Poisson Regression Model Comparing Measles Unvaccinated Counts for Urban and Rural Zones.

	Poisson		Negative Binomial	
Statistics	Rural	Urban	Rural	Urban
Deviance	3.3748	5.3508	1.1328	1.1583
Scaled Pearson χ^2	3.2012	4.7921	1.0173	0.8394
Dispersion			0.0411	0.1250

Table 3: SAS output illustrating the goodness of fit assessment criteria.

Durbin Watson Spatial Autocorrelation Test				
Zones	**DW**	**Durbin's t**	**Prob**	**Pseudo R^2**
Rural	0.1657	1.0831	0.1394	0.7506
Urban	0.2733	1.3225	0.0930	0.6468

Table 4: Generalized Autocorrelation Test for Measles Unvaccinated Count for rural and urban zones.

Parameter Estimates						
Rural	**Variable**	**Estimate**	**SE**	**t Value**	**Sig. Level**	
	Intercept	1.4516	0.1470	9.88	<0.0001	
	Y lag	0.9791	0.001981	494.31	<0.0001	
	Pseudo R^2					0.9587
Urban						
	Intercept	1.5879	0.18050	8.80	<0.0001	
	Y lag	0.94647	0.00357	270.01	<0.0001	
	Pseudo R^2					0.9307

Table 5: The first-order measles unvaccinated related parameter estimates with a lag dependent variable.

correlation within the models for rural (0.166) and urban (0.273) zones despite the linear relationship between the predictor variables and the measles unvaccinated related outcome for both zones (Table 4) as noticeable by pseudo R square (Table 5).

We specified the LAGDEP OPTION for the lagged dependent measles unvaccinated variable(Y_{lag}) in the DW test statistics to quantitate distance at lag1. The first order autocorrelation test result showed that our predictor variables were positively correlated for (rural=1.0831) and (urban=1.3225). Therefore, we reject the H_0 that the residual error is independent of one another for rural and urban zones in the georeferenced measles unvaccinated related epidemiological model. Furthermore, the d statistics compares our independent variables in the auto-regressive model seperated by lags higher than one. Our research analysis demonstrated that the first order correlation tests confirms the assumption of positive spatial autocorrelation in the geo-referenced measles vaccination related dataset.

Furthermore, we applied PROC VARIOGRAM procedure with The LAGDISTANCE OPTION for the computation of autocorrelation index. The LAGDISTANCE in the autocorrelation model signifies the neighborhood size. Likewise, the binary row-averaged weights created for the normalized, asymptotic, epidemiologic measles unvaccinated measles related dataset within the spatial dependence procedure generates a Moran's coefficient equivalent to the regression slope of Moran's scatter plot [15] (Table 6).

The Moran's index and the Geary's c ratio derived from PROC VARIOGRAM revealed a positive spatial autocorrelation for rural (MI=0.456, Gc=0.550) and urban zones (0.442, Gc=0.556) as seen in Table 6. The ratio of the Moran's index for rural and urban zones revealed a 103% clustering tendencies for measles non-vaccination to occur in rural communities compared to urban neighborhoods within Nigerian study site. Similarly, the displayed z-score demonstrates more clustering intensity for mealses non-vaccination in non-urban geo-locations (z=1208) compared to urban geo-classified landscapes (z=608).

Discussion

Measles vaccination coverage in Nigeria is 42% based on the latest data from Demographic Health Survey (DHS) 2013. Which is a clear departure from attaining the 90-95% herd immunity needed for regional elimination? The challenge in the past had been where to find the 58% unvaccinated population that would benefit from supplementary immunization activities? Since different neighborhoods within states have different coverage rates. Delineating and targeting hotspots for optimizing immunization interventions has been suggested as a potential mechanism for efficiently using limited vaccines in low income countries like Nigeria [25].

Covariates generated in our frequentist model mirrored similar findings in the literature of factors associated with measles non-vaccination [26]. Glatman-Freedman et al. [27] explored the effects of social determinants on immunization programs in low income countries while Ataguba et al. [28] researched on socio-economic inequalities in immunization coverage in Nigeria. However, our findings do not significantly demonstrate that Euclidean distances to geo-referenced health center constitute a barrier to measles vaccination; which contrast some findings in some literature [2]. Vitamin A supplements are given concurrently with the MMR vaccine at 9 and 15 months according to the Expanded Program on Immunization (EPI) schedule and potentially explains the high correlation noted within the regression model for measles unvaccinated outcome [3].

The northern states of Nigeria are confronted by the challenges of illiteracy, poverty, and religious unrest in the form of Boko haram and vaccines coverage is about half of that in southern states [2]. Boko Haram is an anti-state Islamic group that forbids western education and have engaged in wanton destruction of religious, health and educational infrastructures in the North. Hotspot analysis clearly demonstrated that communities like Bama, Chibok, Baga, Konduga in Bornu state and Gujba, Damaturu, Potiskum in Yobe state (Figure 2) ravaged by the Islamic insurgents were hotspots for measles non-vaccination. These communities in rural northern Nigeria continues to lag behind their southern communities in term of health indices such as infant mortality and vaccination coverage. Similarly, the high number of internally displaced persons (IDPs) from insurgencies, natural disasters and communal clashes between herdsmen and farmers in these northern geo-locations with inadequate access to health care could partly explain our hotspot findings (Figure 2).

The national bureau of statistics (2012) reported health facilities population proportional ratio that is low (e.g., Kano state=11/100,000) and unevenly distributed with rural communities in northern Nigeria severely affected. Government policies over the years of developing urban centers with basic infrastructures to the detriment of rural communities have facilitated the migration of health professionals and educated individuals towards urban areas in search of better opportunities. Nevertheless, a significant number of people residing in the rural communities are poor farmers and do not have access to health care. Traditional healers serve as the primary health care provider to

Autocorrelation Statistics							
Zones	**Assumption**	**Coefficient**	**Observed**	**Expected**	**Std Dev**	**Z**	**Pr>|Z|**
Rural	Normality	Moran's I	0.456	-0.000095	0.000378	1208	<0.0001
		Geary's c	0.550	1.0000000	0.001129	-399	<0.0001
Urban	Normality	Moran's I	0.442	-0.000184	0.000727	608	<0.0001
		Geary's c	0.556	1.0000000	0.002043	-213	<0.0001

Table 6: Spatial statistics derived from PROC VARIOGRAM SAS output.

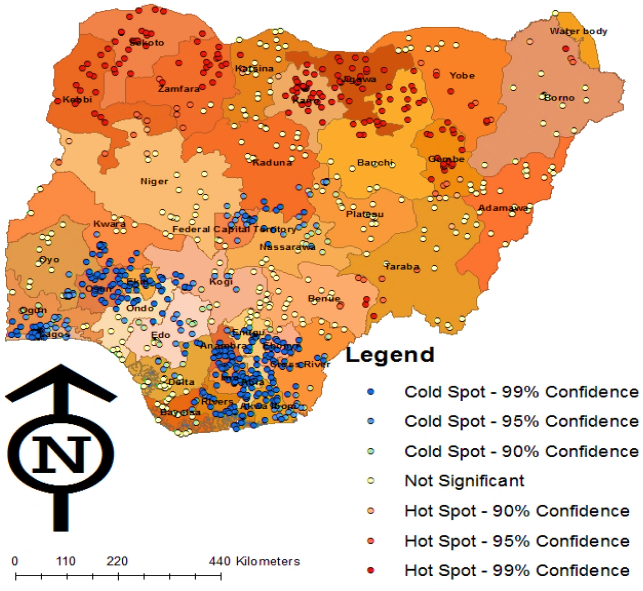

Figure 1: Hotspot Analysis of Measles Non-vaccination in Nigeria at State Level.

non-urban residents and offer herbal, spiritual, and religious practices to them [27]. Although, measles outbreaks occur in crowded populations such as the urban zones. This study is directed towards measles vaccines coverage in Nigeria and our findings showed stochastic explanative geo-referenced clusters, revealing measles unvaccinated trends in the significant proportion of neighborhoods in rural North: which are not as crowded as urban communities but contain a significant population of extended families living in enclaves.

It is not unusual to notice over-dispersion in Poisson probability paradigms and the standard approach for adjusting over-dispersion within the Poissonial regression model involves scaling the deviance and Pearson Chi Square criteria for urban (deviance/DF=4.6255) and non-urban geo-referenced geo-locations (deviance/DF=3.1610). However, the negative binomial regression model allowed for adequate compensation of over-dispersion by collapsing the outliers to the mean when scaled deviance and Pearson chi-square are insufficient (Table 3). The dispersion parameter values for rural (0.041) and urban (0.125) in the statistical output demonstrated that the negative binomial model was justifiable in correcting for over-dispersion.

Durbin Watson test allowed us to identify the inherent structural flaw in the model in urban zones (PSC=0.166) and rural agro-pasture lands (PSC=0.273) as seen in other seasonal, time series dataset. However, a Durbin Watson test <1 may signify that the values for successive error terms are close in value to one another and dependent. In this case, we have serially correlated positive errors and the DW test statistics was applied to elucidatively quantitate the distance lag between centroids points within neighborhoods at the Nigerian study site. Similarly, the Durbin t statistics at the first lag (lag1) demonstrated a positive serial autocorrelation with our autoregressive model for urban (Durbin's t=1.3225 lag=0.9465) and non-urban landscapes (Durbin's t=1.083 lag=0.9791).

We had similar attributes of measles unvaccinated related clusters aggregating in communities spread across northern parts of the country; and also, noticed dissimilar attributes of measles vaccinated related cluster clustering at the southern parts of the Nigerian study site. The dichotomized measles non-vaccinated related data revealed more clustering tendencies in the non-urban zones (Moran's I statistics=0.456, p=0.001) compared to the urban zones (Moran's I statistics=0.442, p=0.001). Likewise, it is of note that the northern and southern parts of the Nigerian Study site are demographically distinct (Figure 1). The rural communities in the north are densely populated and widely separated from one another by a large expanse of semi-arid lands. Similarly, public health care facilities are poorly distributed to serve these communities.

Significant outliers (HL and LH hotspots) within the North-central region of the country (Figure 1) could be informative in terms of socio-demographic characteristics and epidemiologic connectivity. The seasonal migratory nature of the Hausa/Fulani ethnic groups toward the south in search of pastures for their livestock, the influence of Abuja located in the north central as the capital of the country has attracted people from different regions of the country into the north central for trade and job opportunities and the influence of the sophisticated southern Yoruba and Igbo ethnic nationalities on the Hausa/Fulani population in north central could have explained these extreme values (Figure 3).

As measles immunization coverage increases, it becomes important to identify risk factors for measles non-vaccination and implement supplementary immunization activities within these high-risk groups. Additionally, it is also pertinent to provide high-quality immunization services and treatment of complications arising from nosocomial measles infection during epidemics [4].

Conclusion

Our findings demonstrate that clustering tendencies for measles non-vaccination tend to occur in rural communities. Hopefully, this finding from this study could help increase coverage especially in northern Nigeria by targeting high risk geo-locations, taking into consideration socio-economic factors, religious and ethnic differences.

Limitations

Based on the merits of this research, future quantification of

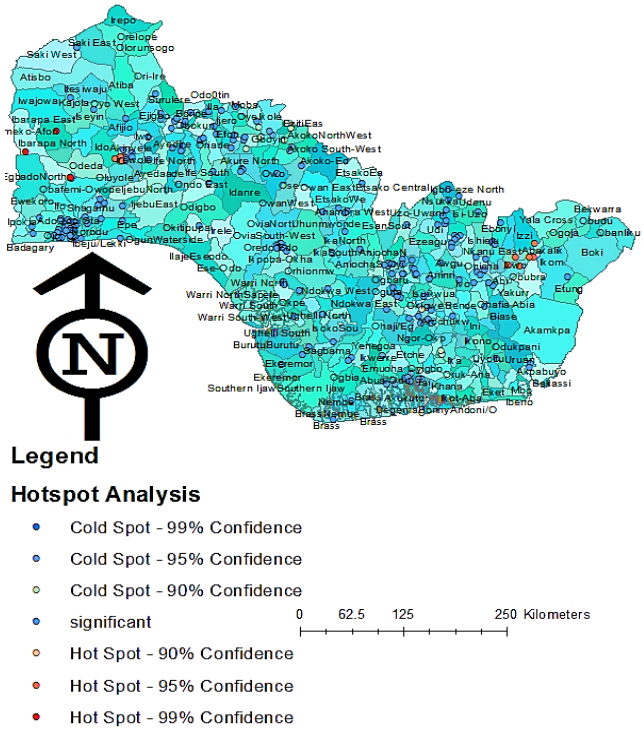

Figure 2: Demonstrating Hotspot Analysis for Measles Non-vaccination within rural communities in Northern Nigeria.

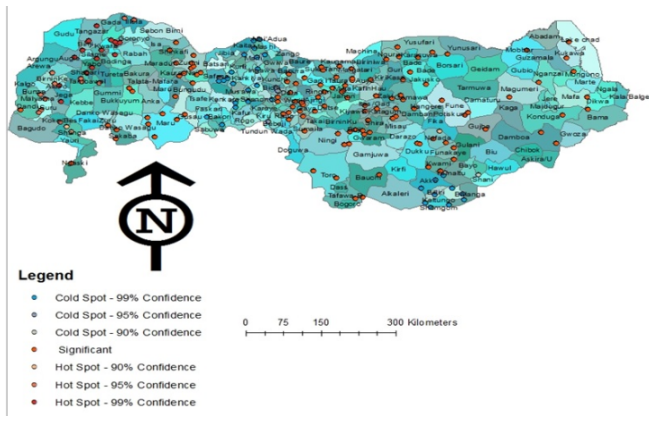

Figure 3: Hotspot Analysis of Measles Non-vaccination coverage Within Urban Communities in Southern Nigeria.

measles primary sampling units on urban and non- urban geo-referenced Bayesian generalized theoretical model may spatially adjust misspecification due to improperly geo- referenced measured covariates. Non-quantitated Bayesian inconspicuous uncertainties can create propagating non-linear erroneous variability in remotely sensed predictors [17]. However, by utilizing conditional predictive prior probability and a subjective maximum likelihood estimation in a Bayesian weighted matrix, all residual forecast from autoregressive equation can be rectified. This would include misspecification such as spatial heteroscedasticity, excessively negative skewed leptokurtic distribution, extreme observation (outliers), and geo-coordinates with poor positional accuracy and other.

References

1. Goodson JL, Masresha BG, Wannemuehler K, Uzicanin A, Cochi S (2011) Changing epidemiology of measles in Africa. J Infect Dis 204: S205-S214.
2. Metcalf CJE, Tatem A, Bjornstad ON, Lessler J, O'Reilly K, et al. (2015) Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. Epidemiol Infect 143: 1457-1466.
3. Ophori EA, Tula MY, Azih AV, Okojie R, Ikpo PE (2014) Current trends of immunization in Nigeria: prospect and challenges. Trop Med Health 42: 67-75.
4. Akramuzzaman SM, Cutts FT, Hossain MJ, Wahedi OK, Nahar N, et al. (2002) Measles vaccine effectiveness and risk factors for measles in Dhaka, Bangladesh. Bull World Health Organ 80: 776-782.
5. Li H, Ao R, Peng L, Li D, Zhang J (2014) Case classification in measles surveillance system under the Two-level Logistic Model. Zhonghua liuxingbingxue zazhi 35: 57-60.
6. Haque SR, Bari W (2013) Positive role of maternal education on measles vaccination coverage in Bangladesh. Int J Psychol Behav Sci 3: 11-17.
7. McCue T, Carruthers E, Dawe J, Liu S, Robar A, et al. (2008) Evaluation of generalized linear model assumptions using randomization.
8. Griffith DA (2006) Assessing Spatial Dependence in Count Data: Winsorized and Spatial Filter Specification Alternatives to the Auto-Poisson Model. Geogr Anal 38: 160-179.
9. Haight FA (1967) Handbook of the Poisson distribution. Wiley.
10. Griffith DA (2013) Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.
11. Durbin J, Watson GS (1950) Testing for serial correlation in least squares regression: I. Biometrika 37: 409-428.
12. Chun Y, Griffith DA, Lee M, Sinha P (2016) Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. J Geogr Syst 18: 67-85.
13. Godoy P, Domínguez A, Alcaide J, Camps N, Jansà JM, et al. (2004) Characteristics of tuberculosis patients with positive sputum smear in Catalonia, Spain. Eur J Public Health 14: 71-75.
14. Sartorius B, Cohen C, Chirwa T, Ntshoe G, Puren A, et al. (2013) Identifying high-risk areas for sporadic measles outbreaks: lessons from South Africa. Bull World Health Organ 91: 174-183.
15. Griffith DA (2009) Methods: spatial autocorrelation. International Encyclopaedia of Human Geography. pp: 396-402.
16. Weiss NA (1999) Introductory Statistics. Addison Wesley.
17. Jacob BG, Novak RJ, Toe L, Sanfo MS, Afriyie AN, et al. (2012) Quasi-likelihood techniques in a logistic regression equation for identifying *Simulium damnosum* sl. larval habitats intra-cluster covariates in Togo. Geo Spat Inf Sci 15: 117-133.
18. Ward MD, Gleditsch KS (2008) Spatial Regression Models. SAGE Publications.
19. Institute S (2012) SAS/ETS 12.1 User's Guide: SAS Inst.
20. Jacob BG, Mendoza D, Ponce M, Caliskan S, Moradi A, et al. (2014) Pseudo R^2 Probablity Measures, Durbin Watson Diagnostic Statistics and Einstein Summations for Deriving Unbiased Frequentistic Inferences and Geoparameterizing Non-Zero First-Order Lag Autocorvariate Error in Regressed Multi-Drug Resistant Tuberculosis Time Series Estimators. Am J Appl Math Stat 2: 252-301.
21. Draper NR, Smith H (2014) Serial Correlation in the Residuals and the Durbin-Watson Test. Applied Regression Analysis. 3rd edn. pp: 179-203.
22. Chen Y (2016) Spatial autocorrelation approaches to testing residuals from least squares regression. PloS One 11: e0146865.
23. Anselin L (1995) Local indicators of spatial association-LISA. Geographical analysis 27: 93-115.
24. Alegana VA, Atkinson PM, Pezzulo C, Sorichetta A, Weiss D, et al. (2015) Fine resolution mapping of population age-structures for health and development applications. J R Soc Interface 12: 20150073.
25. Azman AS, Lessler J (2015) Reactive vaccination in the presence of disease hotspots. Proc Royal Soc London B: Biol Sci 282: 20141341.
26. Favin M, Steinglass R, Fields R, Banerjee K, Sawhney M (2012) Why children are not vaccinated: a review of the grey literature. International health 4: 229-238.
27. Glatman-Freedman A, Nichols K (2012) The effect of social determinants on immunization programs. Hum Vaccin Immunother 8: 293-301.
28. Ataguba JE, Ojo KO, Ichoku HE (2016) Explaining socio-economic inequalities in immunization coverage in Nigeria. Health Policy Plan 2016: czw053.