International Journal of Geographic Information System
Vol. 5, No. 1, January 2018, pp. 1-36, E-ISSN: 2576-375X
Available Online at http://acascipub.com/Journals.php

Research article

Introducing an extra binomial variation scheme in a linear-logistic model for fitting a sub-county-level, forecast vulnerability grid-stratified zip code polygon to optimally regressively quantitate measles endemicity in Florida

Ashley S. Curry[a], Toni Panaou[b], Sam Alao[a], Benjamin Jacob[a]*

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

[b]Department of Integrative Biology, Civil and Environmental Engineering, University of South Florida, Tampa, FL

*Corresponding author: E-mail: bjacob1@health.usf.edu

Abstract

The commonly employed regression modeling approach was compared against a remote sensing interpolation technique to determine which model was more robust at identifying vulnerable populations at-risk for endemic measles transmission at the sub-county–level in Florida. Initially, univariate and other descriptive statistics were derived from multiple sociodemographic covariates in Indian River County, the county with the highest reported number of measles cases in the state of Florida in 2015. Subsequently, the zip code with the greatest contribution to these cases was assessed by land use land cover (LULC) type. The dominant LULC was determined to be Urban

Built-Up, and Residential at medium density. The corresponding LULC geoclassificactions and regression estimates were subsequently identified in Hillsborough County, FL. The remote sensing model was determined to be more advantageous for optimally forecasting georeferenceable geolocations of potential, hyperendemic measles,capture point foci as it allowed for at-risk populations to be identified at the zip code level, while addressing geographic heterogeneity of sampled populations.

Keywords: Measles, regression modeling, land use land cover, remote sensing, Florida.

1. Introduction

Determining county-level geographic locations (henceforth geolocations) of unvaccinated vulnerable measles populations has been hampered due to lack of literature contributions. Regression models [Poisson, Logistic, Negative binomial] have revealed hyperendemic foci explanators of other infectious disease processes such as Malaria (Jacob et al. 2009), West Nile Virus (Griffith 2005), Tuberculosis (Jacob et al. 2013),and HIV rates (Kalipeni and Zulu 2008).Unfortunately, exploitation and quantitative interpolation of frequentist and non-frequentist iterative regression algorithms for optimally delineating epidemiological, capture point, and hyperendemic foci from georeferenceable, sub-county level, empirical measles-related, sociodemographic datasets have not been reviewed. Hence, presently it is unknown where vulnerable populations of measles endemicity occur at the county level in the United States.

Commonly, quantitating sub-grid county-level, measles–related,georeferenceable data requires summing up regressed covariates(Alao et al. 2016).Data retrieved from these epidemiological, vulnerability, endemic, regression models can be either the prediction of the outcomes (e.g., geolocation of a sub-county, hyperendemic, epidemiological, capture point), or the testing of hypotheses, on the basis of other related information (e.g.prevalence,case distribution). In so doing, elucidative sub-county, measles, descriptors (landscape geoclassified data) may be revealed which may provide a measure of how well observed clinical outcomes are replicated by the model, based on the proportion of total variation of regression outcomes explained by the model.

In statistics, simple linear regression is a linear regression model with a single explanatory variable(Kenney & Keeping 1962).For example , a simple regression sub-county, forecast, vulnerability, measles model would have ideally two-dimensional sample points with one independent variable and one dependent variable (conventionally, the *x* and *y* coordinates would be part of a Cartesian coordinate system). In such an endemic model, a linear function (i.e., a non-vertical straight line) may predict the dependent variable values (e.g., number of sub-county measles cases within a grid-stratifiable, georeferenceable, remotely sensed polygon) as a function of a dataset of sociodemographic explanative, independent variables. The adjective simple refers to the fact that the outcome variable is related to a single predictor(Hosmer &Lemeshow, 1980).

The ordinary least squares(OLS) method may optimally minimize and regressively quantitate an empirical dataset of georeferenceable, sociodemographic, parameterizable measles-related covariates (e.g., vertical distances between sub-county, grid-stratifiable, capture points and the fitted line) in a forecast, vulnerability, model framework. Under a specific testing research hypothesis, the accuracy of the line through the sample capture points could be measured by the sum of squared residuals.

In statistics, the residual sum of squares (RSS), also known as the sum of squared residuals (SSR), or the sum of squared errors of prediction (SSE), is the sum of the squares of residuals (deviations predicted from actual empirical values of data) (Rao 1973). It is a measure of the discrepancy between the data and an estimation model. A small RSS in a sub-county, forecast, vulnerability measles model indicates a tight fit of the model to the data (Alao et al. 2017). It may be usable as an optimality criterion in parameter selection and model selection. In general, total sum of squares = explained sum of squares + residual sum of squares(Rao 1973).

A multivariate, OLS, specified, sub-county, grid-stratified, measles modelpartitioning may reveal generalizable, unbiased, iteratively interpolative, quantitative, model estimators. In particular, the explained sum of squares measures how much variation there is in the modelled values and this is compared to the total sum of squares, which measures how much variation there is in the observed data, and to the residual sum of squares, which measures the variation in the modelling errors(Hosmer &Lemeshow, 1980).

The explained sum of squares (ESS) would be the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model (Rao 1973). For example, an

empirical, optimizable dataset of time series, measles, forecast, vulnerability explanators may be regressively optiumally quantitated employing $y_i = a + b_1x_{1i} + b_2x_{2i} + \ ... \ + \varepsilon_i$, where y_i is the i^{th}, sub-county, sampled, georeferenceable observation of the response variable (i.e., zip code, prevalence), x_{ji} is the i^{th} observation of the j^{th} explanatory, geoclassified, sociodemographic variable, (e.g., mean annual income) a and b_i arecoefficients, i indexes the landscape geoclassified, spatial trend observations from 1 to n, and ε_i is the i^{th} value of the error term. The goal would be to make this sum as small as possible in the vulnerability forecasts (e.g. sub-county, hyperendemic foci, geolocations). In general, the greater the ESS, the better an estimated infectious disease model performs (Griffith 2005, Jacob et al. 2009).

An OLS, measles-related, regression model may encompass the slope of the fitted line devised from an empirical dataset of sampled, grid-stratified, discrete integer, capture point, count data, (sub-county, georeferenced, measles, prevalence, urban, case distribution,). In statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the squares of the differences between the observed responses (values of the variable being predicted) in the given dataset (Freedman 2005). In the case of a sub-county, forecast, vulnerability, measles model, the slope of the fitted line would equal the correlation between y and x corrected by the ratio of standard deviations of the sampled socio-demographic and other census variables. The intercept of the fitted line in the model would be such that it would pass through the center of mass (x, y) of the sub-county, sampled, measles –realted, capture points.

When an intercept is included in a forecast, vulnerability, sub-county, time series, measlesmodel, then r^2 would simply be the square of the sample correlation coefficient (i.e., r) between the observed outcomes and the observed, explanatory, predictor values. A correlation coefficient is a number that quantifies a type of correlation and dependence, between two or more values (Freedman 2005). A mean pseudo R^2 value or a total pseudo R^2 value can account for the variation at the sub-county (i.e., zip code polygon), differentially stratified, grid-stratifiable, measles, predictive, risk model. In statistics, the coefficient of determination, denoted R^2 or r^2 is the proportion of the variance in the dependent variable that is predictable from the independent variable(s) (Glantz&Slinker 1990). There are several definitions of R^2 that are only sometimes equivalent in a forecast, vulnerability, sub-county, generalizable, measles model (Alao et al. 2017). One class of such cases includes that of a simple linear regression

where r^2 is used instead of R^2. If additional, sub-county-level, explanatoryregressors are included in the time series measles model, R^2 would be the square of the coefficient of multiple correlation.The Sample Multiple Correlation Coefficient, R, is a measure of the strength of the association between the independent explanatory variables and the one dependent prediction variable (Hosmer and Lemeshow1980). In both such cases, the coefficient of determination in a forecast vulnerability, measles, spatiotemporal model would range from 0 to 1.

Important cases where the computational definition of R^2 can yield robust values in an epidemiological, forecast, vulnerability, sub-county, grid-stratifiable, georeferenceable measles model may depend on the diagnostic, model-fitting procedure (e.g., linear regression is conducted without including an intercept). Additionally, negative values of R^2 may occur when fitting non-linear georeferenceable functions to sampled, sub-county-level, time series, measles data. In cases where negative values arise in a predictive, sub-county, epidemiological, measles model, the mean of the data may provide a better fit to the outcomes than the fitted function values, according to a particular criterion (e.g., interpolated regression estimate synthesized from a neighboring county/district/province).

Other regression methods that may be usable in place of OLS for optimally regressively quantitating, clinical, field-operationizable and remote-sampled, sub-county,empirical, parameter estimator datasets of sociodemographic and other specified, temporally dependent, measles covariates including least absolute deviations for minimizing the sum of absolute values of residuals and the Theil–Sen estimator (a line whose slope is the median of the slopes determined by pairs of sample points) (Hosmer and Lemeshow1980).Least absolute deviations (LAD), also known as least absolute errors (LAE), least absolute value (LAV),least absolute residual (LAR), sum of absolute deviations, or the L_1 norm condition, is a statistical optimality criterion and the statistical optimization technique that relies on (Rao 1973).Total Least Squares regression of an empirical sampled, measles, sub-county ,dataset may also find a regression line that fits a set of two-dimensional (2-D) capture points, but unlike OLS, least absolute deviations, and median slope regression, the model may not really be definable as simple linear regression.Hence the vulnerability, measles explanatory forecasts (e.g., geographic locations of hyper-endemic capture points) may not separate the sub-county-level, coordinates of the capture point, epidemiological foci into one dependent and one independent variable which could potentially return a vertical line as its fit.

The intercept (often labeled the constant) is the expected mean value of Y when all X=0 (Hosmer an Lemeshow, 1980). An epidemiologist or measles experimenter may begin with a regression equation with one predictor, X. If X sometimes = 0, the intercept is simply the expected mean value of Y at that value in the model. If X never = 0, then the intercept has no intrinsic meaning. In scientific research, the purpose of a regression model is to understand the relationship between predictors and the response (Freedman 2005). In a linear regression, measles, sub-county, forecast, vulnerability model if X never = 0, there is no interest in the intercept (Alao et al. 2017). As such the model output would not be able to reveal any relationship between X and Y (sampled sub-county covariate and measles prevalence). Regardless, the intercept in the regression model would be required to calculate predicted values (e.g., geolocations of hyperendemic, measles –related unvaccinated, georeferenceable, sub-county capture points). Further, when X never =0 in a time series, sub-county, forecast, vulnerability. measles model this may be a reason for centering X. If an epidemiologist or researcher rescales X so that the mean or some other meaningful value = 0 (e.g., subtracting a constant from X) in the measles model, then the intercept would have meaning.The intercept (often labeled the constant) is the expected mean value of Y when all X=0 (Rao 1973). It would be the mean value of Y at the chosen value of X.

If there are dummy variables in a time series, measles, sub-county, forecast, vulnerability model, the intercept would have more meaning. Dummy coded clinical and remote sampled, temporally dependent,measlesvariables have values of 0 for the reference group and 1 for the comparison group (Alao et al. 2016). Since the intercept would be the expected mean value when X=0 in the model, it would be the mean value only for the reference group when X=0. This is especially important to consider when the dummy coded, explanatory predictor in a measles model is included in an interaction term. For example, if X1 is a continuous variable centered at its mean in a measles sub-county, forecast, vulnerability model, X2 would be a dummy coded predictor, and the model would contain an interaction term for X1*X2. The B value for the intercept is the mean value of X1 (Freedman 2005). Regardless, if regression assumptions are violated in a sub-county, forecast-oriented, vulnerability, measles, temproal model, the renderings (e.g., geolocation of a georeferenceable,hyperendemic,capture point) would be mis-specified.

According to Jacob et al. (2012, 2009), classical assumptions for conducting a robust infectious disease regression analysis in an, epidemiological, regression, chronic infection model (e.g., non-frequentistic, forecast, vulnerability, measles, sub-county, iterative interpolator) include:1) the sample is representative of the sampled population for the inference prediction; 2) the error is a random sampled clinical variable (e.g., prevalence statistic or zip code, case distribution) or a grid-stratified, urban residential land use land cover (LULC) with a mean of zero conditional on the other sampled explanatory variables (e.g., orthogonally decomposable measles sample dataset); 3)the independent variables are measured with no error (note: If this is not so, the prognosticative measles modeling may be done instead employing errors-in-variables model techniques); 4) the predictors are linearly independent, (i.e. it is not possible to express any sampled sub-county, measles predictor as a linear combination of the others); 5) the forecast errors are uncorrelated (Gaussian), that is, the variance–covariance matrix of the errors is diagonal and each non-zero element is the variance of the error; and finally 6) the variance of the error is constant across the observations (e.g. homoscedastic, clinically sampled, parameterizable, measles estimators).

These assumptions imply that the parameter estimates will be unbiased, consistent, and efficient only if the regression assumptions in a forecast, vulnerability, sub-county, measles, georeferenceable model are not violated. It is important to note that actual georeferenceable, measles-related, time series data rarely satisfies regression assumptions (Alao et al. 2016). Variation from the assumptions can sometimes be used as a measure of how far the model is from being useful (Griffith 2005). Many of these assumptions may be relaxed in a sub-county, predictive, measles, capture point, epidemiological, vulnerabilty model employing more advanced treatments (e.g., iterative Bayesian). Reports of statistical analyses usually include analyses of tests on the sample data and methodology for the fit and usefulness of the model(Freedman 2005).

Independent and dependent variables from chronic infection, vulnerability, epidemiological paradigms (e.g., georeferenceable, grid-stratifiable, sub-county, measles, forecast model) may refer to regressed values measured at capture point geographic locations (georeferenced, sub-county, zip-code, levelhouseholds). There may be spatial trends and spatial autocorrelation in sampled, sub-county-level, vulnerability variables that violate statistical assumptions of regression.

Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics (Griffith 2005). Spatial autocorrelation exists because real world phenomena are typified by orderliness, pattern, and systematic concentration, rather than randomness. In other words, spatial autocorrelation in a sub-county, epidemiological, forecast, vulnerability, georeferenceable, measles model means a dependency exists between values of a variable (e.g., race) in neighboring or proximal, sub-county geolocation, or a systematic pattern in values of a measles-related variable across the geographic locations on a LULC map due to underlying common factors.

Spatial autocorrelation would have many interpretations in a georeferenceable, forecast, vulnerability, measles, epidemiological, sub-county model such as a nuisance parameter, self-correlation, map pattern, a diagnostic tool, a missing variables surrogate, redundant information, a spatial process mechanism, a spatial spillover, or the outcome of areal unit demarcation. The most dismissive for modeling sub-county measles sampled, time series datasetsis as a nuisance parameter. Spatial autocorrelation may be captured by a sub-county, measles-related, parameter estimator, error model specification as its presence is necessary for a good description(Alao et al. 2017), but it may not be truly of interest. The measles model may interfere with the estimation of other sub-county sampled measles covariates that are of true interest. This interference may tend to be for measles sampled, temproal parameters such as variances, rather than means in the model.

Interpreting spatial autocorrelation in a forecast, vulnerability, measles model may reveal self-correlation capabilities. Correlation arises from the geographic context within which attribute values occur, and as such may be mathematically expressiblein terms of the Pearson product moment correlation coefficient (r) formula, but with neighboring values of variable Y replacing those of variable X (Griffith 2005). In statistics, the Pearson correlation coefficient (PCC), also referred to as thePearson's r, Pearson product-moment correlation coefficient (PPMCC) or bivariate correlation, is a measure of the linear correlation between two variables X and Y (VerHoef and Cressie 1993). We assumed that by employing autocorrelation common test procedures such as the score test and the likelihood ratio test may be more precisely definable in a georeferenceable, measles forecast, vulnerability, model output. In addition, we assumed statistics and computing formulae may be parsimoniously obtainable for various

small sample procedures as proposed in Skovgaard (2001). The properties of the tests for and their hypotheses may be hence comparable, employing spatial autocorrelation simulations.

Geographic weighted regression is one technique to deal with latent autocorrelation in sub-county, measles, time series datasets. The variables may include values aggregated by sub-county areas (e.g., with aggregated data the modifiable areal unit problem can cause extreme variation in regression parameters. When analyzing data aggregated by political boundaries, postal codes or census areas, results may be very distinct with a different choice of units (Anselin 1995).

Alternatively, a remote sensing, gridded, county level, georeferenceablepolygon may be scrutinized in a geographic information system (GIS) environment. In so doing, cartographicillustrations of a time series, measles sub-county, grid-stratified model, explanatory regressor may be visualized in GIS. Utilizing feature attribute tables and zonal statistics, an epidemiologist or measles researcher may be able to optimally quantitate levels of variation within each georeferenceable, sub county, grid-stratifiable polygon parsimoniously. Because this formulation requires no regression, there would be less uncertainties and/or noisy variables (heteroskedastic) employing the remote sensing sub-grid level stratification in ArcGIS. In regards to GIS, it has been used to map confirmed cases, and local transmission patterns (Wolfson et al.2007), and although studies have demonstrated the link between MMR vaccination and Measles outbreaks, none have yet developed maps utilizing sociodemographic covariates to forecast areas at high risk for potential outbreaks at the county or zip code level.

Therefore, our objectives were a) to construct a regression model to determine covariates associated to hyperendemic, capture point, explanatory foci, 2) to geographically classify sub-county, sampled landscapes, and 3) to derive autocorrelation statistics for quantitating clustering propensities in empirical datasets of sociodemographicparameterizablegeoreferenceable, measles-related, time series covariates.

2. Materials and Methods

2.1Study Site

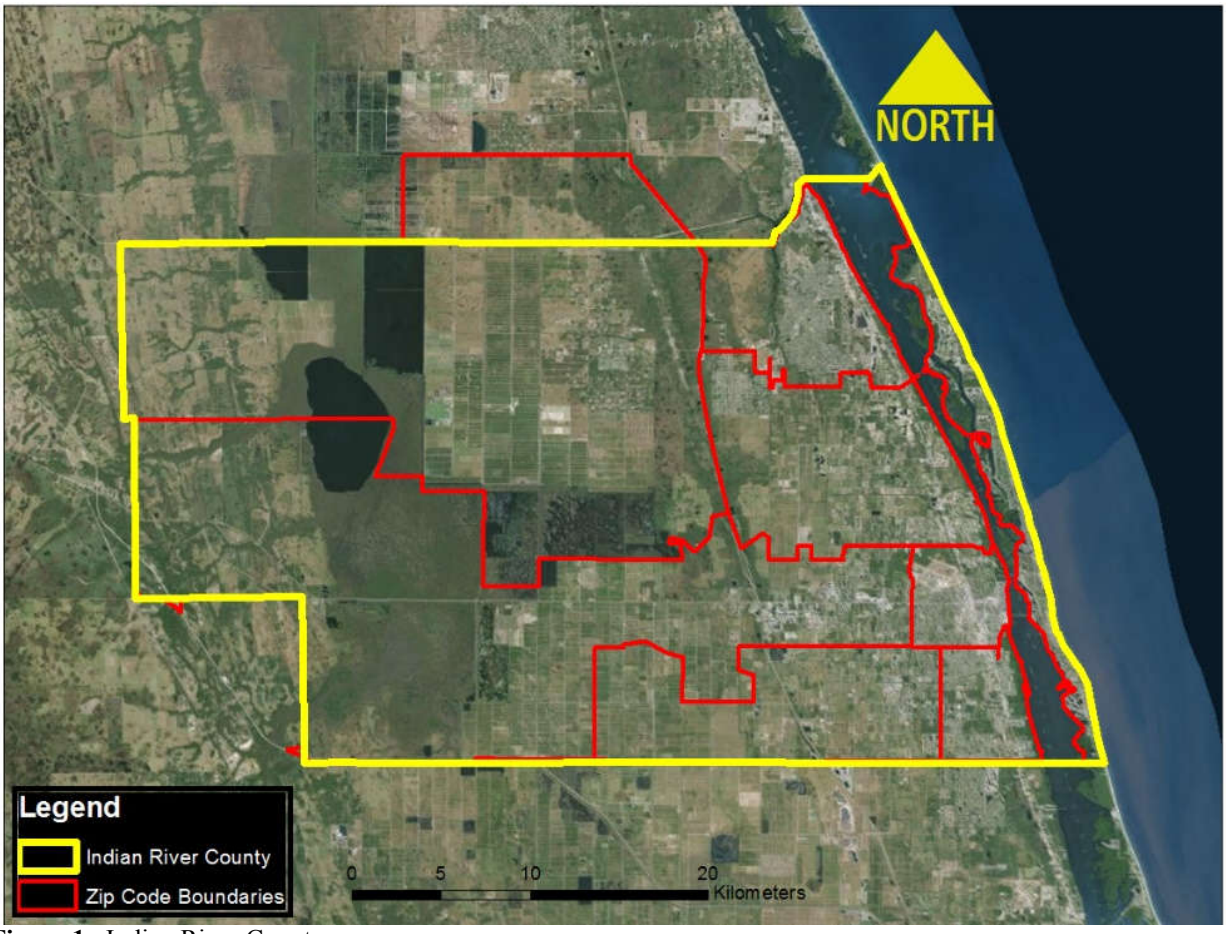

Figure 1. Indian River County

Florida is located in the southeast portion of the U.S. and is the 4th most populous state in the country with a population hovering at just above 20 million people. The state receives upwards of 80 million tourists per year, and is consequently one of the top travel destinations worldwide (Florida Quick Facts, n.d.).

Indian River County demonstrated the highest Measles prevalence in Florida, and served as the study site for this model (Figure 1). Indian River County has a population of 151,000 people, with approximately 19% of them under the age of 18 (U.S. Census Bureau). 26% of those aged 25 years or older have acquired a bachelor's degree (U.S. Census Bureau).

The median household income stands at $46,000, with the poverty rate at 13% (U.S. Census Bureau).

Measles cases in Indian River County were used to model vulnerable populations to Measles in Hillsborough County (Figure 2). This county has a population of 1.2 million people, making it the 4th most populated county in Florida, and is sociodemographically more diverse than Indian River County (Hillsborough County Community Atlas, n.d.). Measles data by county was obtained from the Florida Department of Health (2016).

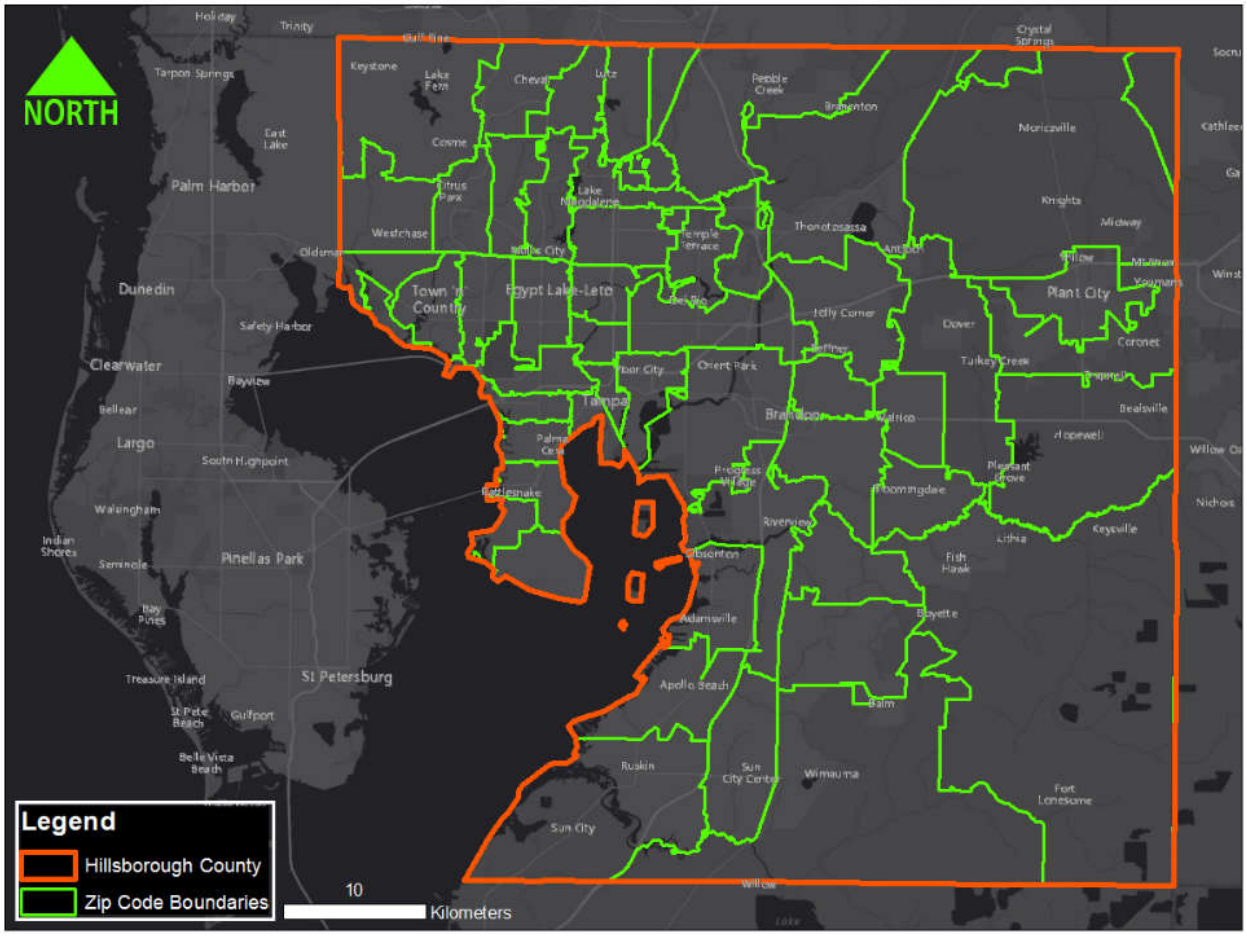

Figure 2. Hillsborough County.

2.2 Land use Land cover classification

Sociodemographic data was obtained from the U.S. Census Bureau (2017). Land use land cover (LULC) data was obtained from the National Land Cover Database (Homer et al., 2015).

A georeferenced epidemiological, time series dataset of measles-related, georeferenced, primary, sampling unit was digitally overlaid onto sub-meter,geoclassified,Google Earth™ data. Initially, we generated an

unsupervised LULC classification in ArcGIS employing urban and non-urban,polygonized, topographic, feature attributes. We considered a linearized model framework for regressively determining statistically significant landscape covariates associated with each polygon using ArcGIS geo-classification.

2.3 Regression Analyses

We employed a binomial coefficient, rearranged to achieve the conditionally convergent series ina sub-county, measles, explanatory, regression model. The plus and minus terms were first grouped in pairs of the sampled sociodemographic and LULC covariate coefficient estimates employing the resulting series based on the actual observational indicator values. The double series was thereby equivalent to Catalan's integral: $\gamma = \int_0^1 \frac{1}{1+x} \sum_{n=1}^{\infty} x^{2^n - 1} dx$. Catalan's integrals are a special case of generalizable formulas due to $J_0\left(\sqrt{z^2 - y^2}\right) = \frac{1}{\pi} \int_0^{\pi} e^{y \cos\theta} \cos(z \sin\theta) d\theta$ where $J_0(z)$ is a Bessel function of the first kind (Rao 1973). The Bessel function is a function $Z_n(x)$ defined in a regression model by using the recurrence relations $Z_{n+1} + Z_{n-1} = \frac{2n}{x} Z_n$ and $Z_{n+1} - Z_{n-1} = -2 \frac{dZ_n}{dx}$ (Freedman 2005), which more recently has been defined as solutions in linear models using the differential equation $x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + \left(x^2 - n^2\right) y = 0$ [6]. by Jacob et al. (2013).Here the Bessel function $J_n(z)$ was defined by the contour integral $J_n(z) = \frac{1}{2\pi i} \oint e^{(z/2)(t - 1/t)} t^{-n-1} dt$ where the contour enclosed the origin and was traversed in a counter-clockwise direction. This function generated: $J_0\left(2i\sqrt{z}\right) = \frac{1}{\pi} \int_0^{\pi} e^{(1+z)\cos\theta} \cos\left[(1-z)\sin\theta\right] d\theta$ $z \equiv 1 - z'$ and $y \equiv 1 + z'$. In mathematics, Bessel

functions are canonical solutions y(x) of Bessel's differential equation: $x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+(x^2-\alpha^2)y=0$ for an arbitrary real or complex number α (i.e., the order of the Bessel function); the most common and important cases are for an α integer or half-integer (Hosmer &Lemeshow, 1980).

Thereafter, to quantify the equivalence in the measles, regression-based, parameter estimators, we expanded $1/(1+x)$ in a geometric series and multiplied the sub-county, sampled, data, feature attributes by x^{2^n-1}, and integrated the term wise as in Jacob et al. (2012). Other series for γ then included $\gamma=\frac{3}{2}-\ln 2-\sum_{m=2}^{\infty}(-1)^m\frac{m-1}{m}\left[\zeta(m)-1\right]$ and $\gamma=\frac{2^n}{e^{2^n}}\sum_{m=0}^{\infty}\frac{2^{mn}}{(m+1)!}\sum_{t=0}^{m}\frac{1}{t+1}-n\ln 2+o\left(\frac{1}{2^n e^{2^n}}\right)$ A rapidly converging limit for γ was then provided for the sub-county, endemic, measles model by $\gamma=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\left(\frac{1}{k}-\frac{\zeta(1-k)}{n^k}\right)\right]=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\frac{1}{k}\left(1+\frac{B_k}{n^k}\right)\right]$ and $\gamma=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\left(\frac{1}{k}-\frac{\zeta(1-k)}{n^k}\right)\right]=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\frac{1}{k}\left(1+\frac{B_k}{n^k}\right)\right]$ where B_k was a Bernoulli number. Another limit formula was then provided by the equation $\gamma=-\lim_{n\to\infty}\left[\frac{\Gamma\left(\frac{1}{n}\right)\Gamma(n+1)n^{1+1/n}}{\Gamma\left(2+n+\frac{1}{n}\right)}-\frac{n^2}{n+1}\right]$. In mathematics, the Bernoulli numbers Bn are a sequence of rational numbers with deep connections to number theory, whereby, values of the first few Bernoulli numbers are B0 = 1, B1 = ±1⁄2, B2 = 1⁄6, B3 = 0, B4 = −1⁄30, B5 = 0, B6 = 1⁄42, B7 = 0, B8 = −1⁄30 (Hosmer &Lemeshow, 1980). Jacob et al. (2009) found if m and n are sampled values and f(x) is a smooth sufficiently differentiable function in a seasonal yellow fever –related, infectious, regression model which is defined for all the values of x in the

interval $[m, n]$, then the integral $I = \int_m^n f(x)\,dx$ can be approximated by the sum (or vice versa)

$$S = \frac{1}{2}f(m) + f(m+1) + \cdots + f(n-1) + \frac{1}{2}f(n)$$

.

The Euler–Maclaurin formula then provided expressions for the difference between the sum and the integral in terms of the higher derivatives f(k) at the end points of the interval m and n. The Euler–Maclaurin formula provides a powerful connection between integrals and sums which can be used to approximate integrals by finite sums, or conversely to evaluate finite sums and infinite series using integrals and the machinery of calculus (Freedman 2005). Hence, for the sub-county, time series, measles, sampled sociodemographic and LULC, geographically classified, regression values, p, we had

$$S - I = \sum_{k=2}^{p} \frac{B_k}{k!}\left(f^{(k-1)'}(n) - f^{(k-1)'}(m)\right) + R$$

where B1 = −1/2, B2 = 1/6, B3 = 0, B4 = −1/30, B5 = 0, B6 = 1/42, B7 = 0, B8 = −1/30, and R was an error term. Note in this research

$$-B_1\,(f(n) + f(m)) = \frac{1}{2}\,(f(n) + f(m))$$

. As such, we re-wrote the regression-based, optimizable,predictive, vulnerability formula as follows:

$$\sum_{i=m}^{n} f(i) = \int_m^n f(x)\,dx - B_1\,(f(n) + f(m)) + \sum_{k=1}^{p} \frac{B_{2k}}{(2k)!}\left(f^{(2k-1)'}(n) - f^{(2k-1)'}(m)\right) + R$$

We then rewrote the mathematical equation more elegantly as

$$\sum_{i=m}^{n} f(i) = \sum_{k=0}^{p} \frac{1}{k!}\left(B_k f^{(k-1)'}(n) - B_k^* f^{(k-1)'}(m)\right) + R$$

with the convention of

$$f^{(-1)'}(x) = \int f(x)\,dx$$

(i.e. the -1th derivation of f is the integral of the function).

Limits to the sub-county, measles, regression model was then rendered by

$$\gamma = \lim_{x\to\infty} \zeta\left(\zeta(z)\right) - 2^x + \left(\frac{4}{3}\right)^x + 1$$

where $\zeta(z)$ was the Riemann zeta function. For the sampled, sub-

county, measles parameter estimator numerical values from 1 to n in the empirical dataset was found to be asymptotic to $\frac{\sum_{k=1}^{n} d(k)}{n} \sim \ln n + 2\gamma - 1$.

In this research, the Bessel functions of the second kind, were denoted by Yα(x), and by Nα(x), which were actually solutions of the Bessel differential equation employing a singularity at the origin (x = 0). This provided an efficient iterative algorithm for g by computing $B_k = \frac{B_{k-1} n^2}{k^2} = A_k = \frac{1}{k}\left(\frac{A_{k-1} n^2}{k} + B_k\right) = U_k V_k = U_{k-1} + A_k$ and $V_k = V_{k-1} + B_k$ $A_0 = -\ln n B_0 = 1 U_0 = A_0$ and $V_0 = 1$ Reformulating this identity rendered the limit $\lim_{n\to\infty}\left[\sum_{k=0}^{\infty} \frac{\left(\frac{n^k}{k!}\right)^2 H_k}{\sum_{k=0}^{\infty}\left(\frac{n^k}{k!}\right)^2} - \ln n\right] = \gamma$ in the measles forecast, model output.

Infinite products involving g also arose from the Barnes G-function using the positive integer n. In mathematics, the Barnes G-function G(z) is a function that is an extension of super factorials to the complex numbers which is related to the Gamma function(Hazewinkel 2001). In this research, this function provided $\prod_{n=1}^{\infty} e^{-1+1/(2n)}\left(1+\frac{1}{n}\right)^n = \frac{e^{1+\gamma/2}}{\sqrt{2\pi}}$ and also the equation $\prod_{n=1}^{\infty} e^{-2+2/n}\left(1+\frac{2}{n}\right)^n = \frac{e^{3+2\gamma}}{2\pi}$. The Barnes G-function was then linearly defined in the time-series dependent, sub-county, measles regression-based, forecast, vulnerability model which then generated $G(z+1) = (2\pi)^{z/2} \exp\left(-\left(z(z+1)+\gamma z^2\right)/2\right) \times \prod_{n=1}^{\infty}\left[\left(1+\frac{z}{n}\right)^n \exp\left(-z+z^2/(2n)\right)\right]$ where γ was the Euler–Mascheroni constant, exp(x) = ex, and ∏ was capital pi notation.

The Euler-Mascheroni constant was then rendered by the expressions $\gamma = -\Gamma'(1) = -\psi_0(1)$ where $\psi_0(x)$ was the digamma function $\gamma = \lim_{s \to 1}\left[\zeta(s) - \frac{1}{s-1}\right]$ and the asymmetric limit form of $\gamma = \lim_{s \to 1^+}\sum_{n=1}^{\infty}\left(\frac{1}{n^s} - \frac{1}{s^n}\right)$ and $\gamma = \lim_{x \to \infty}\left[x - \Gamma\left(\frac{1}{x}\right)\right]$. In mathematics, the digamma function is defined as the logarithmic derivative of the gamma function: $\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ where it is the first of the polygamma functions. In our model the digamma function, ψ0(x) was then related to the harmonic numbers in that $\psi(n) = H_{n-1} - \gamma$ where Hn was the nth harmonic number, and γ was the Euler-Mascheroniconstant. In mathematics, the n[th] harmonic number is the sum of the reciprocals of the first n natural numbers (Freedman 2005). The difference between the n[th] convergent in equation (◇) and γ in our sub-county, regression-based, measles model was then calculable by $\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x - \lfloor x \rfloor}{x^2}dx$ where $\lfloor x \rfloor$ was the floor function which satisfied the inequality $\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x - \lfloor x \rfloor}{x^2}dx$. The symbol g was then $\gamma' \equiv e^{\gamma} \approx 1.781072$. This led to the radical representation of the sampled, sub-county, covariate coefficients as $e^{\gamma} = \left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/3}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/4}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/5}$ which was related to the double series $\gamma = \sum_{n=1}^{\infty}\frac{1}{n}\sum_{k=0}^{n-1}(-1)^{k+1}\binom{n-1}{k}\ln(k+1)$ and $\binom{n}{k}$ which in this reaserch delineated a binomial coefficient.

Thereafter, another proof of product in the measles, forecast, vulnerability regression model was provided by the equation $\frac{\pi}{2}=\left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/4}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/8}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/16}$. The solution was then made even clearer by changing $n\to n+1$. In this research, both these regression-based formulas were also analogous to the product for e which was then rendered by calculating $e=\left(\frac{2}{1}\right)^{1/1}\left(\frac{2^2}{1\cdot 3}\right)^{1/2}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/3}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/4}$.

Unfortunately, extra variation was detected in the estimates in the measles model. A modification of the iterated re-weighted least square scheme and/or a negative binomial non-homogenous regression-based framework conveniently accommodates extra-Poisson variation when constructing seasonal log-linear models employing frequencies or prevalence rates as dependent response variables (Haight 1967).Operationally, these models consist of making iterated weighted least square fit to approximately normally distributed dependent measles l-related explanatory predictor covariate coefficients based on observed rates or their logarithm. Unfortunately, the variance of the time series measles-related observations in log-linear equations are commonly assumed to be constant (Alao et al. 2016).Subsequently, we assumed that introducing an extra-binomial variation scheme into a sub-county,linear-logistic model can be fitted for a Poisson procedure. The probabilities describing the possible outcome of a single trial were then subsequently modeled, as a function of the explanatory predictor variables, using a logistic function.

Thereafter, we constructed a robust negative binomial regression model in SAS with a non-homogenous mean and a gamma distribution by incorporating $\alpha=\frac{1}{\theta}(\alpha>0)$ in equation (2.1). We let $g(\tau_i)$ be the probability density function of τ_i in the model. Then, the distribution $f(y_i|\mathbf{x}_i)$ was no longer conditional on τ_i. Instead it was obtained by integrating $f(y_i|\mathbf{x}_i,\tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i)=\int_0^{\infty} f(y_i|\mathbf{x}_i,\tau_i)g(\tau_i)d\tau_i$. The distribution in the measles, sub-county, forecast, vulnerability, endemic, regression model was then $f(y_i|x_i)=\frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}, y_i=0,1,2..$ The negative binomial distribution was

thus derived as a gamma mixture of random variables. The conditional mean in the model was $E(y_i \mid x_i) = \mu_i = e^{x_i'\beta}$ and the variance in the residual estimates was $V(y_i \mid x_i) = \mu_i\left[1+\frac{1}{\theta}\mu_i\right] = \mu_i[1+\alpha\mu_i] > E(y_i \mid x_i)$.

To further estimate the district-level models, we specified DIST=NEGBIN (p=1) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1 was set p=1, which revealed the variance function $V(y_i \mid x_i) = \mu_i + \alpha\mu_i$ was linear in the mean of the model. The log-likelihood function of the NEGBIN1 model was then provided by $L=\left\{\sum_{j=0}^{y_i-1}\ln\left(j+\alpha^{-1}\exp\left(x_i'\beta\right)\right)\right\}$ Additionally, the equation $-\ln(y_i!)-\left(y_i+\alpha^{-1}\exp\left(x_i'\beta\right)\right)\ln(1+\alpha)+y_i\ln(\alpha)$ was generated. The gradient for the measles model was then quantified employing the equations $\frac{\delta L}{\delta\beta}=\sum_{i=1}^{N}\left\{\left(\sum_{j=0}^{y_i-1}\frac{\mu_i}{(j\alpha+\mu_i)}\right)x_i-\alpha^{-1}\ln(1+\alpha)\mu_i x_i\right\}$ and $\frac{\delta L}{\delta\alpha}=\sum_{i=1}^{N}\left\{-\left(\sum_{j=0}^{y_i-1}\frac{\alpha^{-1}\mu_i}{(j\alpha+\mu_i)}\right)-\alpha^{-2}\mu_i\ln(1+\alpha)-\frac{\left(y_i+\alpha^{-1}\mu_i\right)}{1+\alpha}+\frac{y_i}{\alpha}\right\}$

The negative binomial regression model with variance function $V(y_i \mid x_i) = \mu_i + \alpha\mu_i^2$, was then referred to as the NEGBIN2 model. To estimate this regression-based, temporally dependent measles, sub-county, frequencymodel, we specified DIST=NEGBIN (p=2) in the MODEL statements. A test of the distribution was then performed by examining the hypothesis that $\alpha = \frac{1}{\theta_i} = 0$. A Wald test of this hypothesis was also provided which were the reported t statistics for the estimates in the model. Under the Wald statistical test, the maximum likelihood estimate $\hat{\theta}$ of the sampled measles parameter(s) of interest θ was compared with the proposed value

θ_0, with the assumption that the difference between the two will be approximately normally distributed. The log-likelihood function of the regression models (i.e., NEGBIN2) was then rendered by the equation: $L=\sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j+\alpha^{-1}\right)-\ln\left(y_i!\right)=-\left(y_i+\alpha^{-1}\right)\ln\left(1+\alpha\exp\left(x_i'\beta\right)\right)+y_i\ln\left(\alpha\right)+y_i x_i'\beta\right\}$ whose gradient was $\frac{\delta L}{\delta\beta}=\sum_{i=1}^{N}\frac{y_i-\mu_i}{1+\alpha\mu_i}x_i$. The variance in the measlesmodel was then assessed by $\frac{\delta L}{\delta\alpha}=\sum_{i=1}^{N}\left\{-\alpha^{-2}\sum_{j=0}^{y_i-1}\frac{1}{\left(j+\alpha^{-1}\right)}\alpha^{-2}\ln\left(1+\alpha\mu_i\right)+\frac{y_i-\mu_i}{\alpha\left(1+\alpha\mu_i\right)}\right\}$. The final mean in the model was calculated as: $\frac{pr}{1-p}$, the mode as $\begin{cases}\left\lfloor\frac{p(r-1)}{1-p}\right\rfloor & \text{if } r>1\\ 0 & \text{if } r\leq 1\end{cases}$, the variance as $\frac{pr}{(1-p)^2}$, the skewness as $\frac{1+p}{\sqrt{pr}}$, the kurtosis as $\frac{6}{r}+\frac{(1-p)^2}{pr}$, the moment generating function as $\left(\frac{1-p}{1-pe^t}\right)^r \text{ for } t<-\log p$, the characteristic function as $\left(\frac{1-p}{1-pe^{it}}\right)^r \text{ with } t\in\mathbb{R}$, and the probability generating function as $\left(\frac{1-p}{1-pz}\right)^r \text{ for } |z|<\frac{1}{p}$.

We generated a stepwise backward regression model to tease out any probabilistic uncertainties in the county-level, epidemiological, forecast, vulnerability, endemic, measles model. In statistics, stepwise regression is a method of fitting regression models in which the choice of predictive variables is carried out by an automatic procedure (Rao, 1972). In each step, a county-level, zip code-level, explanatory prognosticator was considered for addition or subtraction from the set of diagnostic variables based on some prespecified criterion. A spatial autocorrelation measles unvaccinated primary sampling unit model was then constructed.

2.4. Autocorrelation Model

A spatial autoregressive measles model was then generated that employed a explanatory variable **Y**, as a function of nearby sampled, georeferenced, sub-county, sociodemographic or geoclassified, LULC covariate. In this research, **Y** had an indicator value 1 (i.e., an autoregressive response) and/or the residuals of **Y** which were values of nearby sampled **Y** residuals (i.e., a SAR or spatial error specification). For time series-dependent modelling parameterizable estimators, the SAR model furnishes an alternative specification that frequently is written in terms of matrix **W (**Griffith 2003). A misspecification perspective was employed for performing a spatial autocorrelation uncertainty estimation analyses using the sampled measlescovariates. The model was built using the $y = X\beta + \varepsilon^*$ (i.e. regression equation) assuming the sampled data had autocorrelated disturbances. The model also assumed that the sampled data could be decomposed into a white-noise component, ε, and a set of unspecified sub-county, level regression models that had the structure $y = X\beta + \underbrace{E\gamma + \varepsilon}_{=\varepsilon^*}$. Jacob et al. (2009) found that white noise in a seasonal, malaria-based regression infection model is a univariate or multivariate discrete-time stochastic process whose terms are independent and independent (i.i.d) with a zero mean. In this research, the misspecification term was E_γ.

A Moran's I spatial autocorrelation measles unvaccinated, primary sampling unit model was constructed where:

y=[(logx1y1) + (logx2y2) + (logx3y3)]/ 1 + [(logx1y1) + (logx2y2) + (logx3y3)]

y = Measles contribution, aka the proportion of contribution to total cases. We then interpolated the regression estimate devised from a georeferenceable, measles, cluster model for the Hillsborough County study site. In so doing, we were able to determine a vulnerability endemicity map for measles transmission.

We generated univariate statistics and regression models by employing zip code level (TIGER/Line with Selected Demographic and Economic Data 2015) data for geospatially regressively summarizing the county-level, socio-demographic epidemiological, covariate coefficients. We generated a misspecification term

for constructing an explanative model in PROC REG. Multiple data layers were created using different coded values for the various known county-level, georeferenceable, asymptotically, normalized, county-level, sociodemographic feature attributes.

The land cover type most associated with the polygon with the highest contribution to measles cases in Hillsborough County was identified within that polygon, and again within georeferenced, zip code polygons of a differing county, along with sub-level grid-stratified polygons at the zip code and neighborhood level.

3. Results

A stepwise backward regression analysis was employed in selecting our variable of interest set at 0.1 (Figures 3a and 3b). Three predictors passed the selection test criteria (male age 10-14 years, Latinos, and occupied housing units). Additionally, the variance inflation factor of <5 was utilized in checking for multicollinearity within the regression model.

The statistical regression equation utilized in the linear frequentist regressive model was given by:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_1 - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

The left-hand side of the equation is the independent outcome(i.e., proportion of measles cases in Hillsborough County), while the right-hand side are the predictor variables (Alao et al. 2016). This was further rewritten as SST=SSM+SSE, where SS was notation for sum of squares and T, M, and E in PROC REG were the descriptions for total quantized model error estimates. The adjusted R^2=0.998 demonstrated a linearized, normalized geosampled measles related dataset.

The result also showed that males aged between 10-14 years and occupied housing units (p=<0.001) were significant at $\alpha = 0.05$.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: Measles Measles

Number of Observations Read	8
Number of Observations Used	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.11386	0.03795	1190.14	<.0001
Error	4	0.00012756	0.00003189		
Corrected Total	7	0.11399			

Root MSE	0.00565	R-Square	0.9989
Dependent Mean	0.37375	Adj R-Sq	0.9980
Coeff Var	1.51093		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-0.00829	0.00861	-0.96	0.3899	0
M10_14yrs	M10_14yrs	1	0.00023165	0.00001609	14.40	0.0001	1.68299
Latino	Latino	1	0.00000485	0.00000204	2.38	0.0760	1.51512
OHU	OHU	1	0.00003492	0.00000104	33.54	<.0001	2.10399

Figure 3a. SAS output results.

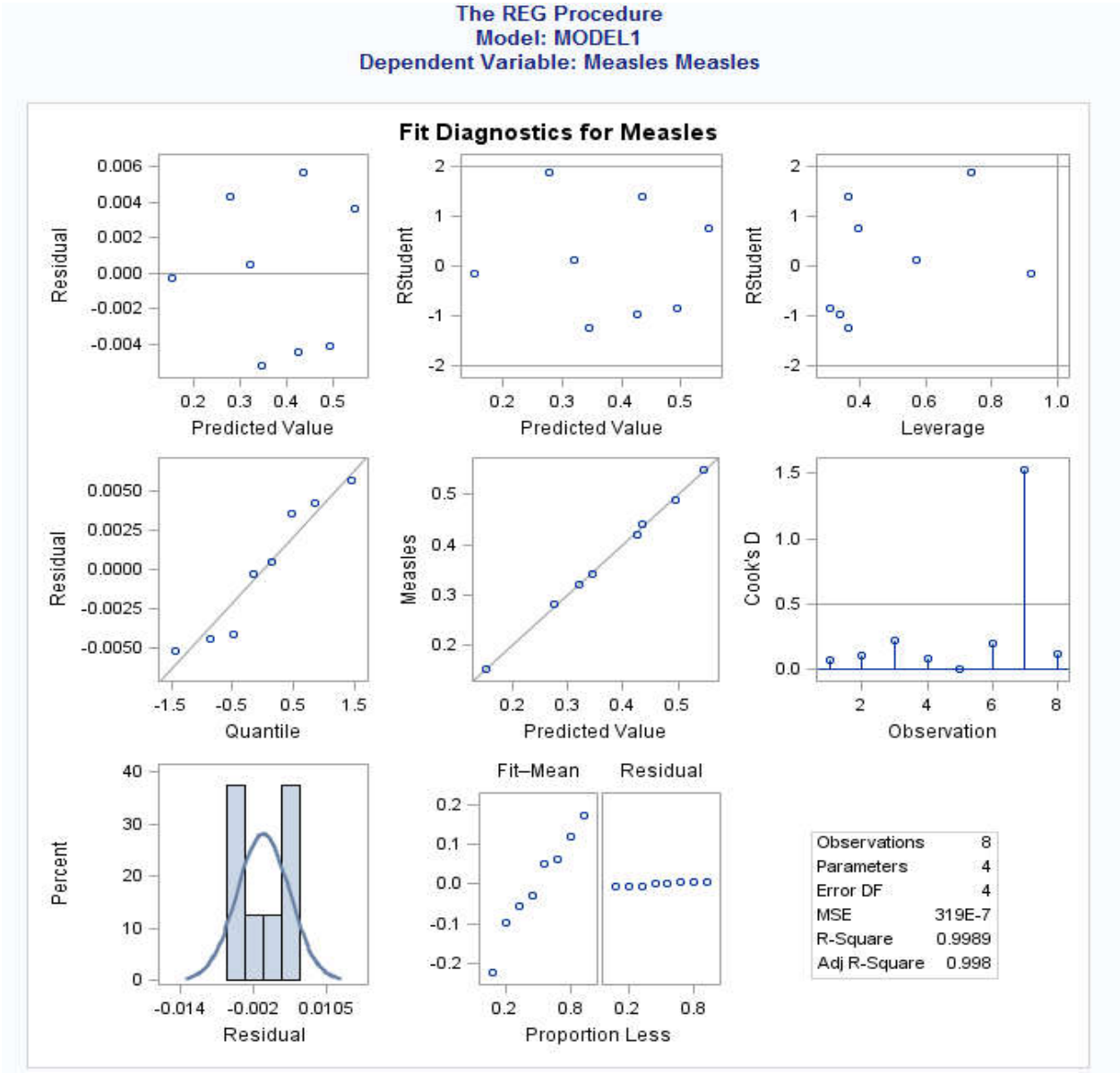

Figure 3b. SAS output results.

Based on the significant covariates associated with Measles in Indian River County, the dominant LULC predictor types were determined to be Agriculture and Urban Built-Up (Figure 4). Zip Code 32958 demonstrated the greatest contribution to Measles cases in Indian River County (18.4%), as determined by the population of the zip code weighted to define its contribution to the total number of cases. Urban Built-Up, specifically at medium density, presented as the dominant LULC associated with Measles cases in this county (Figures 4 and 5).

International Journal of Geographic Information System
Vol. 5, No. 1, January 2018, pp. 1-36, E-ISSN: 2576-375X
Available Online at http://acascipub.com/Journals.php

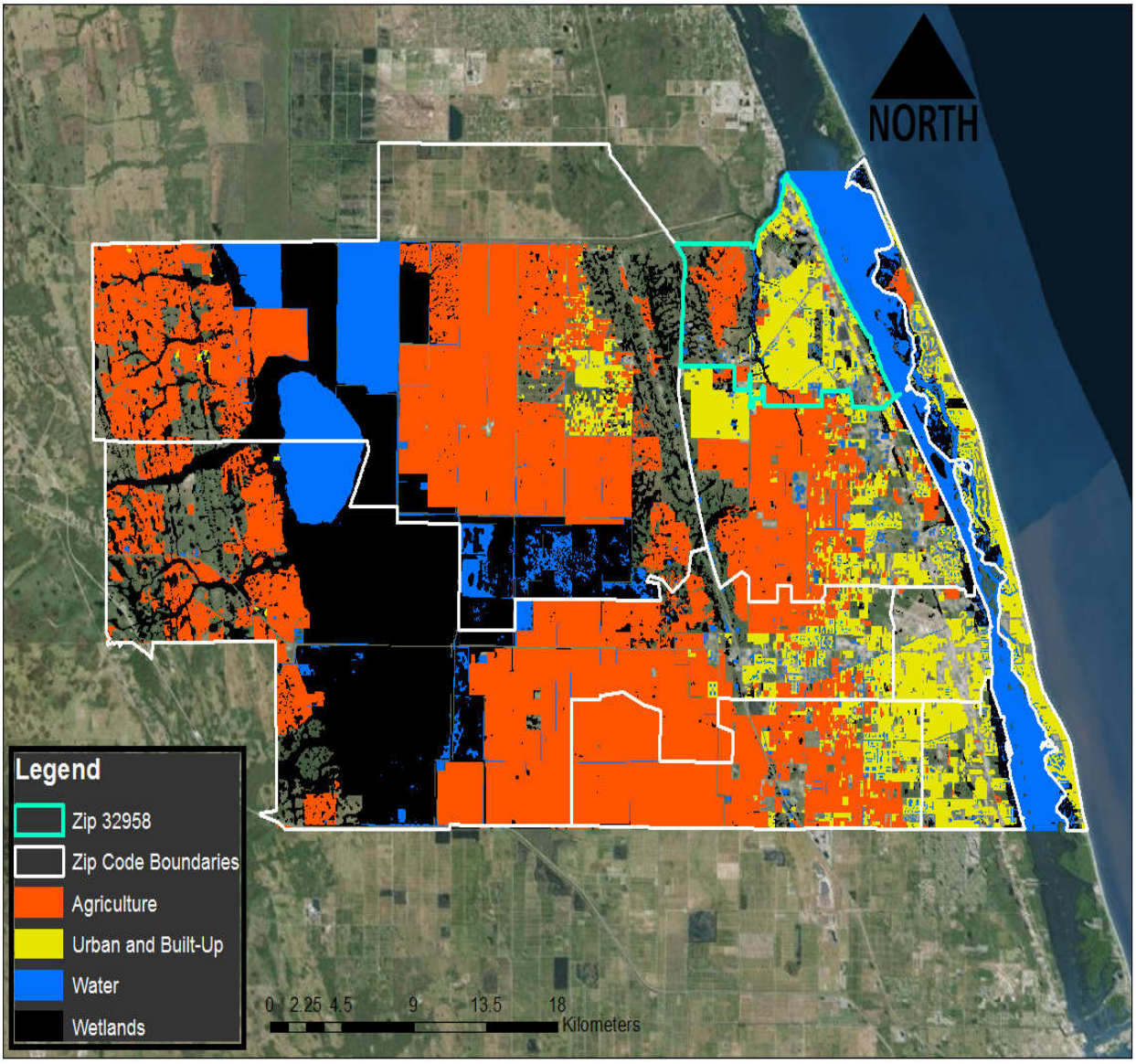

Figure 4. Land use land cover display of Indian River County divided by zip code. Zip code 32958 is highlighted as the greatest contributor to Measles cases.

International Journal of Geographic Information System
Vol. 5, No. 1, January 2018, pp. 1-36, E-ISSN: 2576-375X
Available Online at http://acascipub.com/Journals.php

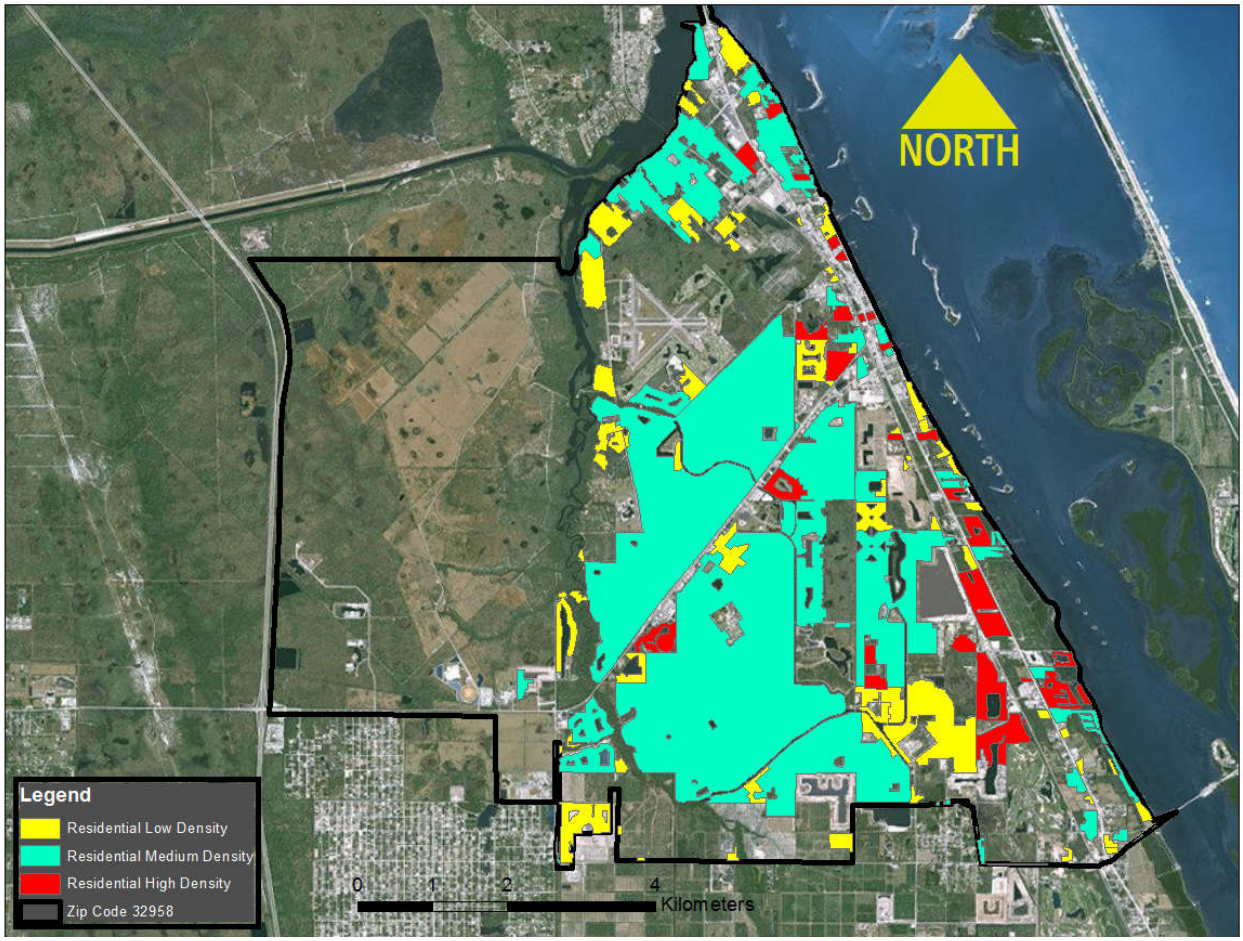

Figure 5. Dominant land use land cover type (residential), and the sub classifications of low, medium, and high. Medium density urban land cover greatest indicator for Measles cases.

Like many locations, Hillsborough County is not homogeneous in population demographics, or in respect to land use and land cover (Figure 6). Thus, two zip codes were examined to demonstrate how these differences can influence models intended to target at-risk communities. Examining two zip code polygons in Hillsborough County – one heterogeneous (Figure 7), one homogeneous (Figure 8) by LULC type – at-risk populations for Measles cases were identified by grid cell, depictingspecific geolocations where these populations are clustered. In regression modeling, both polygons would appear as significant locations for at-risk populations for Measles. The sub-grid level polygons demonstrate heterogeneous locations have significant variation. Identifying at-risk populations through regression analysis at the zip code level results in compounding the error intrinsic to each model.

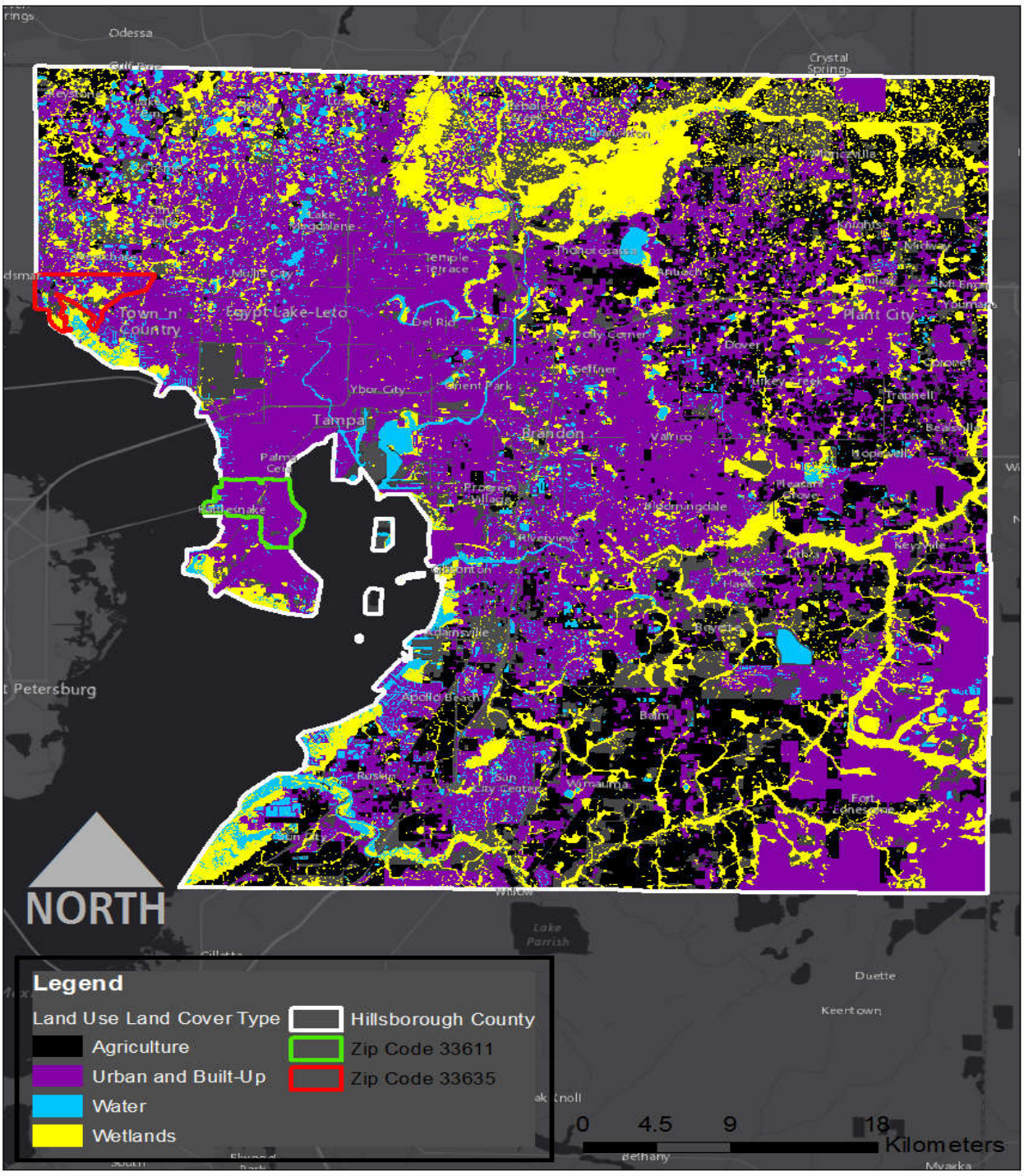

Figure 6. Land use land cover map of Hillsborough County.

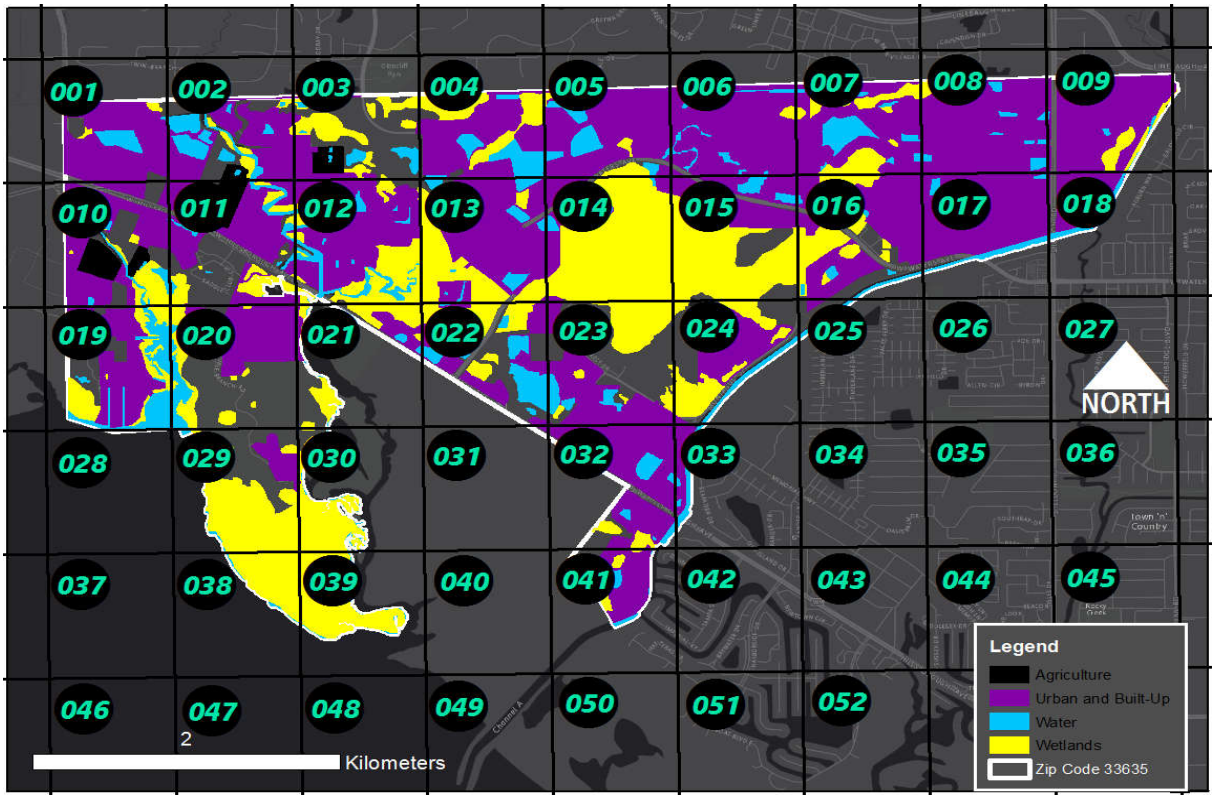

Figure 7. Zip code 33635 LULC, a heterogeneous polygon.

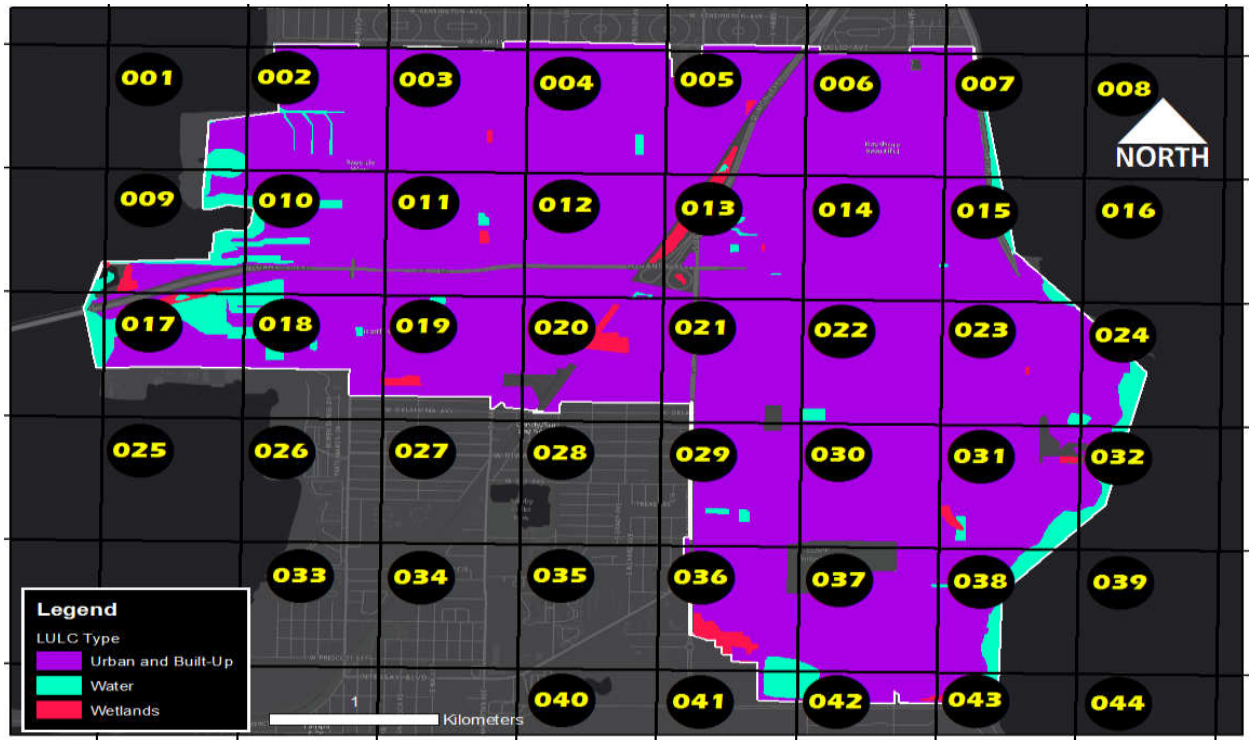

Figure 8. Zip code 33611, a homogeneous polygon.

Additionally, due to heterogeneity of LULC and sociodemographics at the county and zip code level, regression modeling is unable to provide specific at-risk locations, and can consequently overlook at-risk populations in statistically non-significant polygons. By simply quantifying proportions of significant covariates, there is no specificity engrained, thus no ability to show where said covariates can be located within the polygon (Figures 9, and 10).

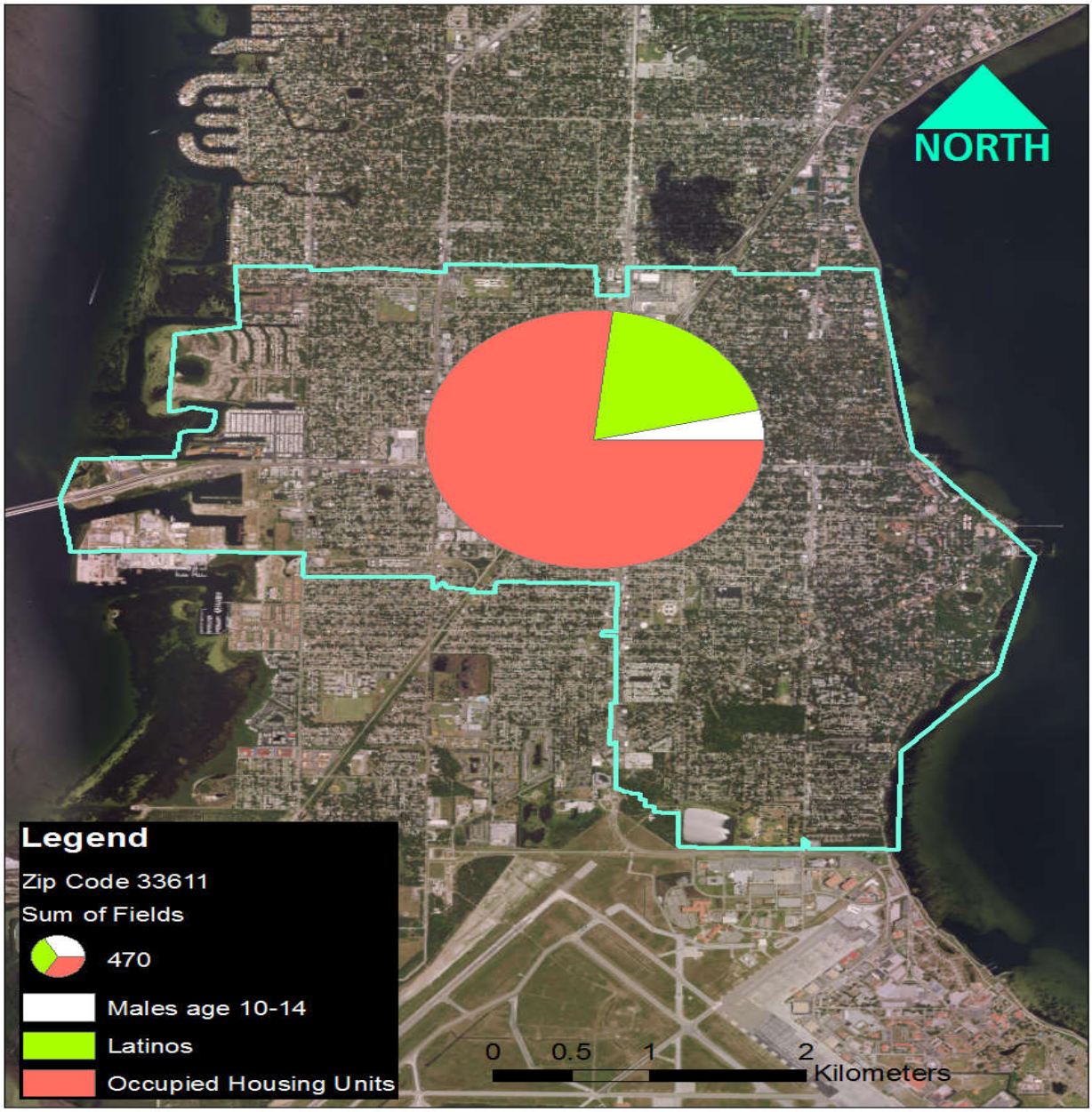

Figure 9. Proportion of statistically significant covariates within zip code 33611.

International Journal of Geographic Information System
Vol. 5, No. 1, January 2018, pp. 1-36, E-ISSN: 2576-375X
Available Online at http://acascipub.com/Journals.php

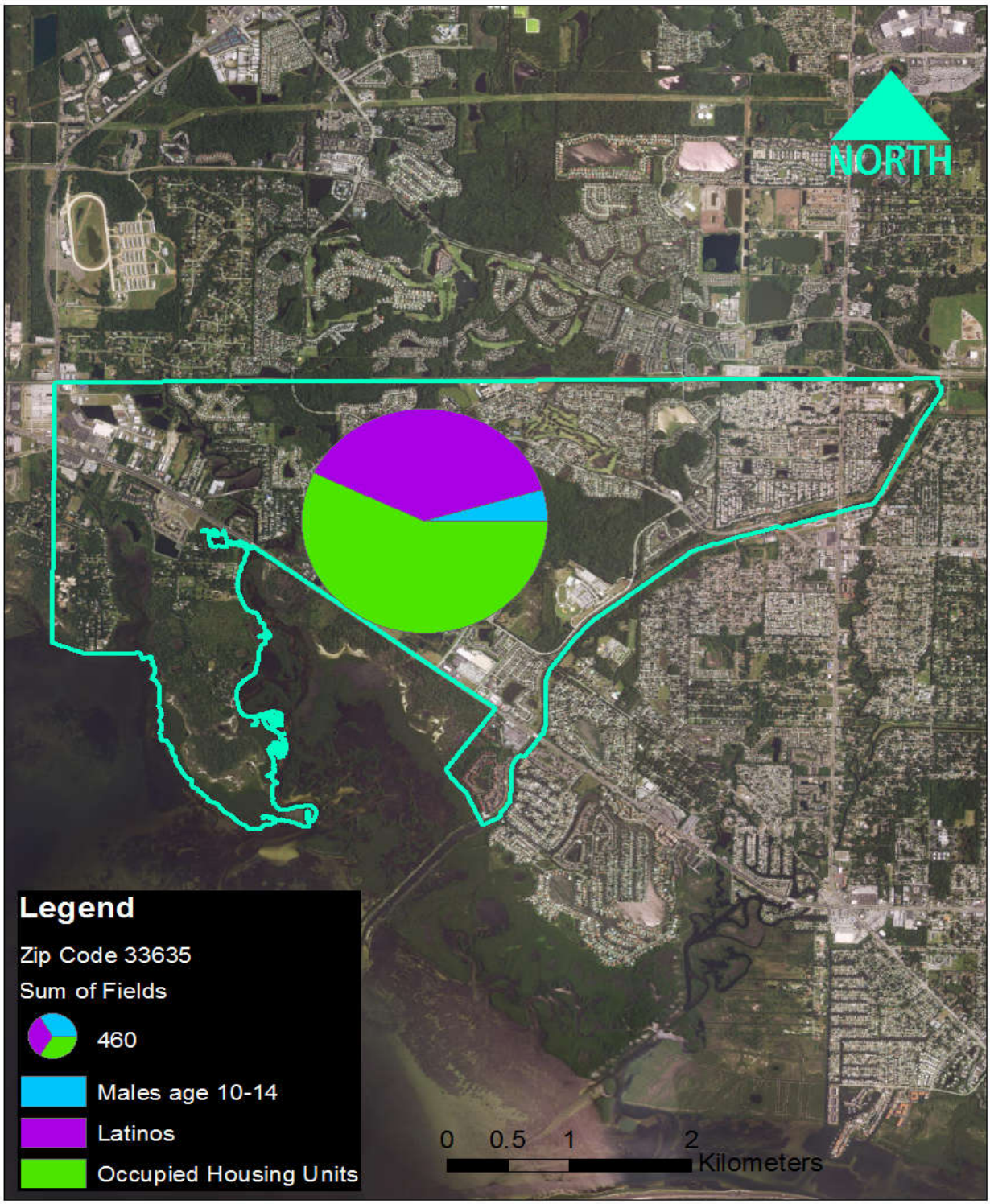

Figure 10. Proportion of statistically significant covariates within zip code 33635.

Grid cell 012 of zip code 33635 (see Figure 12), and grid cell 011 of zip code 33611 (see Figure 11) were examined for differences in vulnerability at the neighborhood level. Both grid cells, the seemingly homogeneous zip code polygon, along with the heterogeneous zip code polygon show variation at the neighborhood level. Grid cell 012 of zip code 33635 identifies a highly vulnerable population end-to-end with a neighborhood with less risk (Figure 13). Grid cell 011 of zip code 33611 may be more homogeneous than 33635, however, there are locations within the grid cell that are not areas of vulnerability for Measles cases (Figure 14). Thus, the remote sensing methodology results in reduced error, and identification of more specific and accurate geolocations.

Figure 11. Residential LULC in zip code 33611.

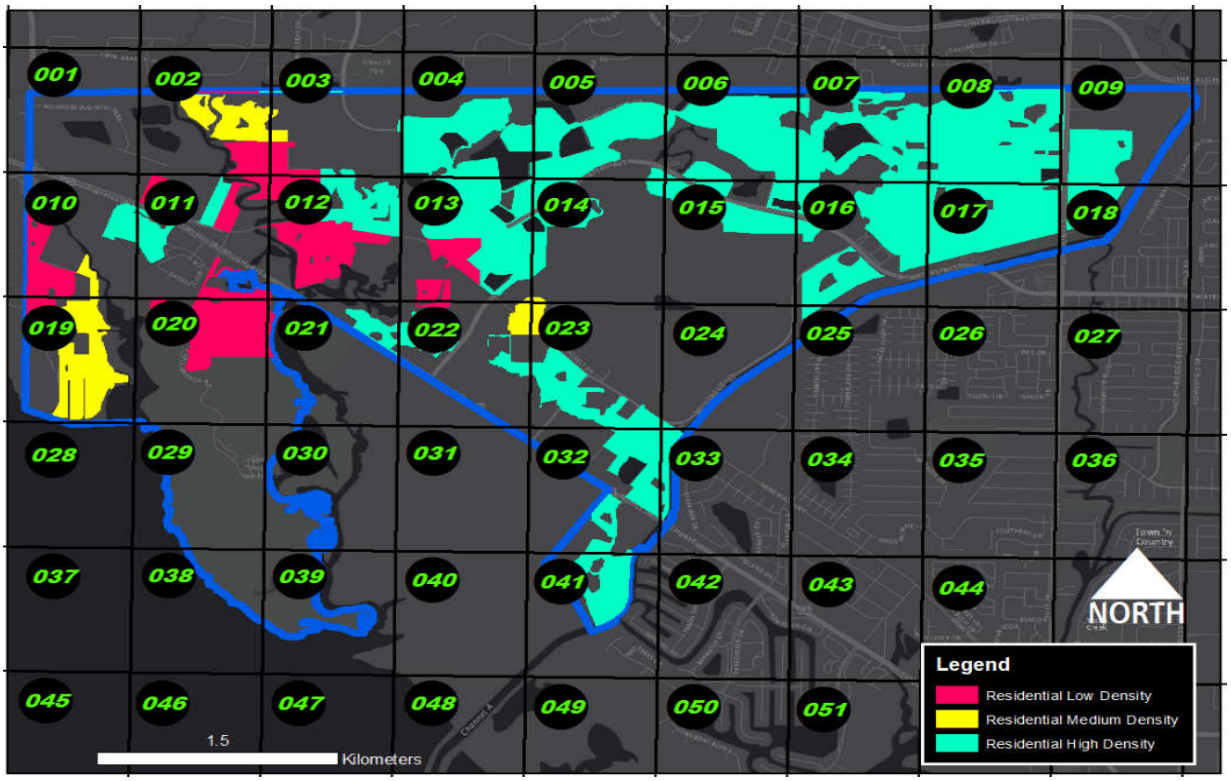

Figure 12. Residential LULC within zip code 33635.

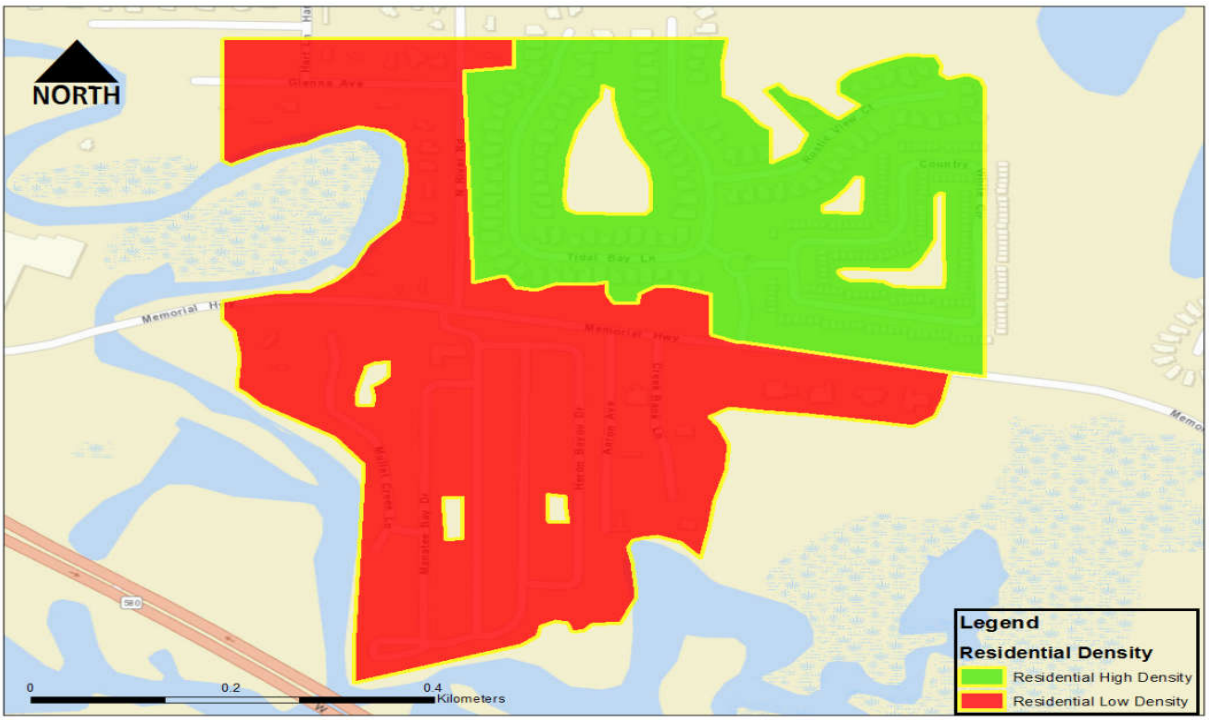

Figure 13. Heterogeneous LULC at the neighborhood level of zip code 33635.

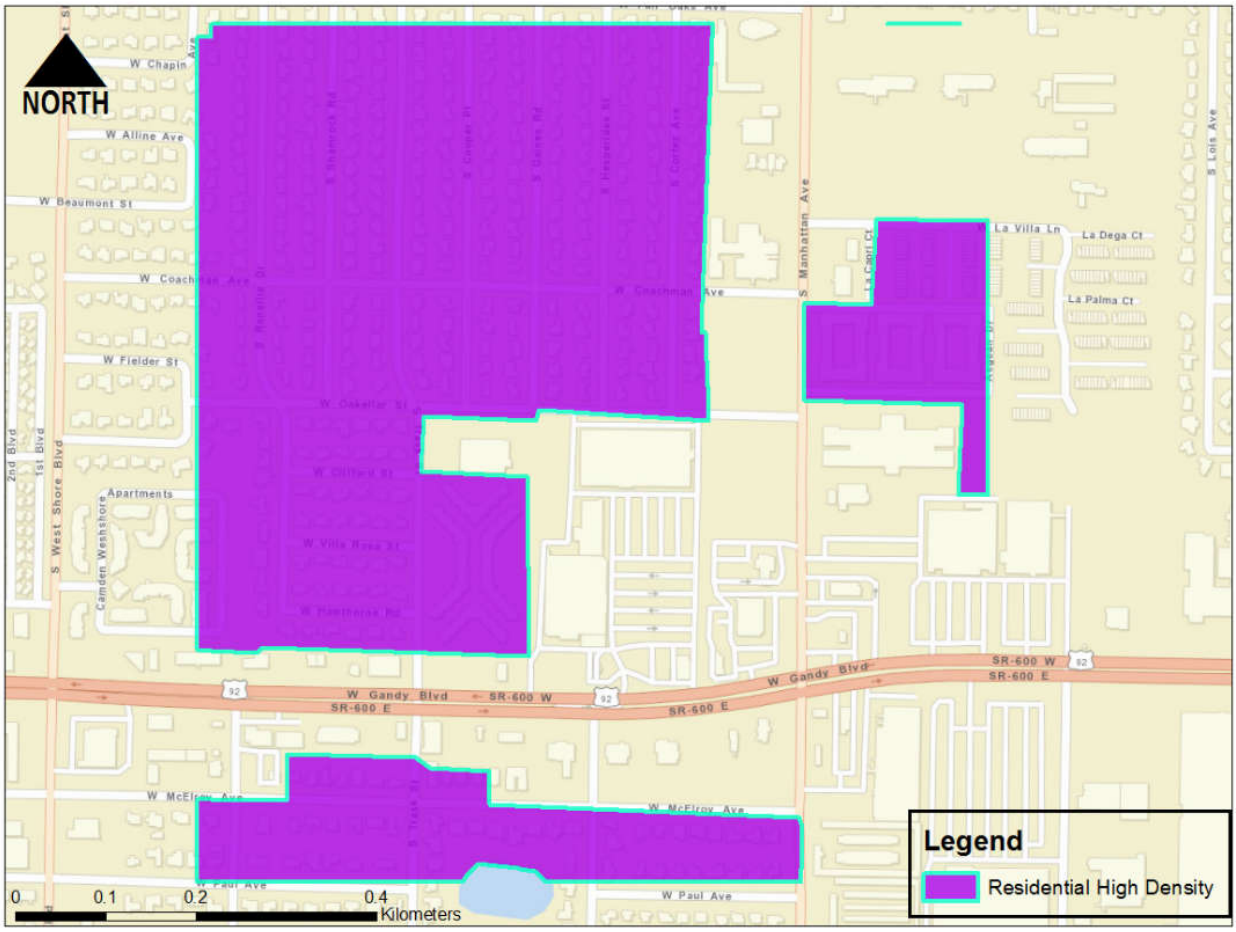

Figure 14.LULC of zip code 33611 at the neighborhood level.

4. Discussion

We constructed a regression model using multiple sociodemographic and LULC covariates. Overdispersion is often encountered when fitting very simple infectious disease parametric models, such as those based on the non-Gaussian distribution (Jacob et al. 2009, Griffith 2005). The distribution had one free parameter which did not allow for the variance to be adjusted independently of the mean. If over-dispersion is a feature, an alternative model with additional free parameters may provide a better fit (Haight 1967). In the case of count data, a mixture model like the negative binomial distributioncan be proposed instead in a forecast, vulnerability model in which the mean of the a distribution can itself be thought of as a random variable drawn – in this case – from the gamma distribution thereby introducing an additional free parameter.Note the resulting negative binomial distribution for rectifying an over-dispersed measles model variation.

This analysis determined that males aged 10-14, Latinos, and the number of occupied housing units were statistically significant indicators of Measles cases. Using the dominant Land Use Land Cover type associated with the polygon of greatest contribution to Measles cases in the sample population, it was determined that Urban Built-Up, specifically medium density, LULC type represented geographical locations that may be targeted for MMR vaccination outreach programs.

We compared two sampling strategies for quantitating sociodemographic and three dimensional heterogeneous covariates in Indian River County. In literature, the most commonly employed methodology applied for sub grid classification of a non-homogeneous land cover covariant is to tabulate regression statistics (pseudo R^2) within each grid cell. We compared this regression tactic with a remote sensing endmember irradiance frequency approximation. We found that the addition of heteroscedastic sub-grid level models increased the uncertainty, as all the models violated the assumption of independent covariates (multicollinearity). Hence, adding the pseudo R^2 values would only increase the levels of error variance. Conversely, the remote sensing partitioning of the grid by spectral reflectance allowed us to determine the amount of heterogeneity in the landscape covariates and the other explanatory sociodemographic independent variables. Our recommendation is that for stratifying Measles sociodemographic and landscape data that a grid stratifiable remote sensing algorithm be employed for quantitating levels of variability in regression covariates at the county level. The remote sensing methodology of identifying vulnerable measles populations shows variability at the neighborhood level that may be important when assessing at-risk communities and allocating resources (Delmelle et al., 2014).

Interpreting spatial autocorrelation as a sub-county geolocation map pattern can emphasize conspicuous trends, gradients, swaths or mosaics across a measles, forecast, vulnerability model.Consider a constant, in a measles model which is the degenerate case (i.e., a constant has no variance) of perfect positive spatial autocorrelation: once the value of a constant is known at a single sub-county location, it would be known at all locations. Next, consider a variable that portrays a north-south (or east-west) linear trend across a measles predictive map. If this variable has a mean of zero, then it would be geometrically orthogonal to and uncorrelated with the constant in the autocorrelation model. These north-south and east-west oriented, explanative, linear trend, forecastable measles variables also maybe orthogonal and uncorrelated. A variable with mean zero whose values' magnitudes form a 3-dimensional

symmetric mound in the center of a map constitutes yet another mutually orthogonal and uncorrelated map pattern (Griffith 2003). Sub-county, sampled, sociodemographicand LULC explanative measles regressors can display maximum levels of positive spatial autocorrelation when geographic variance is present in the covariates, and may be described as global geographic patterns

Alternating sequences of moderately large mounds and basins with either an east-west or a north-south orientation sub-county, measles, grid-stratified, zip code polygons can portray moderate positive spatial autocorrelation, and constitute regional map patterns. Alternating sequences of small mounds and basins with either an east-west or a north-south orientation can portray weak positive spatial autocorrelation which may also constitute local map patterns. This fragmentation continues through randomness (zero spatial autocorrelation) to arrangements of increasingly alternating values (i.e., single value measles-oriented, mounds and basins), which may portray increasing negative spatial autocorrelation. Most substantive variables have geographic distributions that can be described by linear combinations of some subset of these mutually orthogonal and uncorrelated varying-sized mound-basin map patterns (Griffth 2003).

References

Alao, S., Mati, K., and 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- vaccinate. *Journal of Remote Sensing & GIS*, *5*(4). https://doi.org/10.4182/2469-4134.1000185.

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Curtale, F. et al. (2010). Description of two measles outbreaks in the Lazio Region, Italy (2006-2007). Importance of pockets of low vaccine coverage in sustaining the infection. *BMC Infect Dis*, *10*, 62. https://doi.org/10.1186/1471-2334-10-62.

Delmelle, E. et al. (2014). Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, *28*(5), 1107–1127. https://doi.org/10.1080/13658816.2013.871285.

Elam-Evans, L. D., Yankey, D., Singleton, J. A., & Kolasa, M. (2014). National, state, and selected local area vaccination coverage among children aged 19-35 months - United States, 2013. *MMWR: Morbidity & Mortality Weekly Report*, *63*(33), 741–748. https://doi.org/10.15585/mmwr.mm6433a1.

Florida Department of Health. (2016). FLHealthCHARTS Data Viewer. Retrievedfrom http://www.flhealthcharts.com/charts/OtherIndicators/NonVitalIndNoGrpCountsDataViewer.aspx?cid=0129.

Freedman, D. A. (2005). *Statistical Models: Theory and Practice*, Cambridge University Press.

Glantz, Stanton A. &Slinker, B. K. (1990). Primer of Applied Regression and Analysis of Variance. McGraw-Hill.

Griffith, D. A. (2005). A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. *International Journal of Health Geographic*, *4*(18). https://doi.org/10.1186/1476-072X-4-18.

Griffith, D. A. (2003). Spatial filtering (pp. 91–130). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24806-4_4.

Hazewinkel, M. (2001). The algebra of quasi-symmetric functions is free over the integers. *Advances in Mathematics*, *164*(2), 283–300. https://doi.org/10.1006/aima.2001.2017.

Hillsborough County Community Atlas. (n.d.). Retrieved from http://www.hillsborough.communityatlas.usf.edu/demographics/default.asp?ID=12057.

Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, *9*(10), 1043–1069. https://doi.org/10.1080/03610928008827941

Jacob, B. G. et al. (2009). A heteroskedastic error covariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asympotical efficient estimates from ecological sampled Anopheles arabiensis aquatic habitat covariates. *Malaria Journal*, *8*(1), 216. https://doi.org/10.1186/1475-2875-8-216.

Jacob, B. G. et al. (2013). A Bayesian Poisson specification with a conditionally autoregressive prior and a residual Moran's coefficient minimization criterion for quantitating leptokurtic distributions in regression-based multi-drug resistant tuberculosis treatment protocols. *Journal of Public Health and Epidemiology*, *5*(3), 122–143. https://doi.org/10.5897/JPHE12.076.

Jacob, B. G. et al. (2012). Quasi-likelihood techniques in a logistic regression equation for identifying Simulium damnosum s.l, larval habitats intra-cluster covariates in Togo. *Geo-Spatial Information Science*, *15*(2), 117–133. https://doi.org/10.1080/10095020.2012.714663.

Kalipeni, E., & Zulu, L. (2008). Using GIS to model and forecast HIV/AIDS rates in Africa, 1986–2010. *The Professional Geographer*, *60*(1), 33–53. https://doi.org/10.1080/00330120701724061.

Kenney, J. F. and Keeping, E. S. (1962). Linear regression and correlation. Ch. 15 in *Mathematics of Statistics*, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 252–285.

Kouadio, I. K., Kamigaki, T., & Oshitani, H. (2010). Measles outbreaks in displaced populations: a review of transmission, morbidity and mortality associated factors. *BMC International Health and Human Rights*, *10*(5). https://doi.org/10.1186/1472-698X-10-5.

Rao, C.R. (1973). Linear statistical inference and its applications (2nd ed.). New York: John Wiley & Sons.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, *67*(337), 112–115. https://doi.org/10.1080/01621459.1972.10481212.

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, *28*(1), 3–32. https://doi.org/10.1111/1467-9469.00223

State of Florida.com - Florida Quick Facts. (n.d.). Retrievedfrom http://www.stateofflorida.com/facts.aspx.

US Census Bureau. (2017). TIGER/Line® with Data. Retrievedfrom https://www.census.gov/geo/maps-data/data/tiger-data.html.

Ver Hoef, J. M., & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, *25*(2), 219–240. https://doi.org/10.1007/BF00893273.

WHO. (2017). Measles. *WHO*. Retrieved from http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles/en/.

Wolfson, L. J. et al. (2007). Has the 2005 measles mortality reduction goal been achieved? A natural history modelling study. *The Lancet*, *369*(9557), 191–200. https://doi.org/10.1016/S0140-6736(07)60107-X.