Pseudo R² 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

Benjamin G. Jacob[1,*], Daniel Mendoza[2], Mario Ponce[3], Semiha Caliskan[1], Ali Moradi[4], Eduardo Gotuzzo[3], Daniel A. Griffith[5], Robert J. Novak[1]

[1]Department of Global Health, College of Public Health, University of South Florida, Tampa Fl
[2]Department of Environmental and Occupational Health, College of Public Health, University of South Florida, Tampa Fl
[3]Instituto de Medicina Tropical Alexander Von Humboldt-Universidad Peruana Cayetano Heredia, Lima, Peru
[4]LECOM School of Medicine 5000 Lakewood Ranch Blvd Bradenton, FL
[5]School of Economic, Political and Policy Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX
*Corresponding author: bjacob1@health.usf.edu

Abstract In randomized clinical trials using clustered multi-drug resistant tuberculosis (MDR-TB) data, groups of human population are routinely assigned to treatments; whereas, observations are taken on the individual subjects using clinically-oriented explanatory covariate coefficient estimates for identifying sites of hyperendemic transmission. Further, standard methods for data analyses of clinical MDR-TB data postulate models relating observational parameters to the response variables without accurately quantitating varying observational intra-cluster error coefficient effects. Implicit in this assumption is that the effect of these error coefficient estimates are identical. However, non-differentiation of varying and constant residual within-cluster covariate coefficient uncertainty effects in a time-series clinical MDR-TB endemic transmission model can lead to misspecified forecasted predictors of endemic transmission zones (e.g., mesoendemic). In this research we constructed multiple georeferenced autoregressive hierarchical models accompanied by non-generalized predictive residual uncertainty non-normal diagnostic tests employing multiple covariate coefficient estimates clinically-sampled in San Juan de Lurigancho Lima, Peru. Initially, a SAS-based hierarchical agglomerative polythetic clustering algorithm was employed to determine high and low MDR-TB clusters stratified by prevalence data. Univariate statistics and Poisson regression models were then generated in R and PROC NL MIXED, respectively. Durbin-Watson statistics were derived. A Bayesian probabilistic estimation matrix was then constructed employing normal priors for each of the error coefficient estimates which revealed both spatially structured (SSRE) and spatially unstructured effects (SURE). The residuals in the high MDR-TB explanatory prevalent cluster revealed two major uncertainty estimate interactions: 1) as the number of bedrooms in a house in which infected persons resided increased and the percentage of isoniazid-sensitive infected persons increased, the standardized rate of tuberculosis tended to decrease; and, (2) as the average working time and the percentage of streptomycin-sensitive persons increased, the standardized rate of MDR-TB tended to increase. In the low MDR-TB explanatory time series cluster single marital status and building material used for house construction were important predictors. Latent explanatory non-normal error probabilities in empirically regressed MDR-TB clinical-sampled covariate estimates can be robustly spatiotemporally quantitated employing a first-order autoregressive resdiualized model and a Bayesian diagnostic uncertainty estimation matrix.

Keywords: Multi-Drug Resistant Tuberculosis (MDR-TB), Durbin-Watson statistics, Bayesian, San Juan de Lurigancho

Cite This Article: Benjamin G. Jacob, Daniel Mendoza, Mario Ponce, Semiha Caliskan, Ali Moradi, Eduardo Gotuzzo, Daniel A. Griffith, and Robert J. Novak, "Pseudo R² 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." *Journal Name*, vol. 2, no. 5 (2014): 252-301. doi: 10.12691/ajams-2-5-1.

1. Introduction

Prevalence of Multidrug-resistant tuberculosis, (MDR-TB), defined as a disease caused by strains of *Mycobacterium tuberculosis* that are resistant to at least isoniazid and rifampin, the two most important first-line anti-TB drugs, appeared after the introduction of rifampicin in 1966. Until 1990, most MDR-TB cases occurred in patients receiving prolonged, inappropriate therapy; while sporadic outbreaks of primary transmission occurred, the magnitude and impact was relatively limited (Cegielski et al., 2006). In the early 1990's, several large outbreaks of MDR-TB unfolded in hospitals and institutions in the United States implicating MDR-TB as a major public health threat (Frieden et al., 1993; Pearson et al., 1992). High rates of nosocomial transmission to health care workers and human immunodeficiency virus (HIV) positive patients in particular, were documented. From 1999 through 2002, the median prevalence of MDR-TB in new case-patients was at critical levels (>6.5%) in specific regions of the world, including the Baltic states and other eastern European countries (Zignol et al., 2006). Subsequent nosocomial and institutional outbreaks in Italy, Spain, Russia, and Chile made it clear that MDR-TB ranked among the most serious public health issues facing the world.

Recently, public health awareness about MDR-TB has been re-enforced by the occurrence of extensively drug-resistant (XDR)-TB outbreaks defined as a form of MDR-TB with additional resistance to fluoroquinolones and at least one of the second-line injectable drugs used in tuberculosis treatment: amikacin, kanamycin and capreomycin. MDR-TB raises concerns of future TB epidemic treatment options, and jeopardizes the major gains made in TB control and progress on reducing TB deaths, especially among people living with HIV/AIDS.

In implementing an MDR-TB control program, newer predictive spatial statistical algorithms may provide a powerful robust tool for identifying high prevalence areas for understanding where at-risk populations are geographically located for resource targeting and cost-effective control. Traditionally, cluster-based algorithms have been useful for estimating the force of morbidity of MDR-TB.for generating statistics in a sampled population for comparing the disease burden across various environmental settings, with different underlying incidence rates. For example, Becerra et al., (2000) sought to refine the definition of MDR-TB transmission 'hot spots' first described in 1994, by the World Health Organization (WHO), the International Union against Tuberculosis and Lung Disease (IUATLD) and other partners who launched the Global Project on Anti-TB drug resistance surveillance. They obtained estimates of two global MDR-TB indicators, MDR-TB incidence per 100,000 population per year and expected numbers of new patients with MDR-TB per year using a hierarchical explanatory residual within-cluster based time-series analyses for various global regions where data were available. They concluded that it was useful to include georeferenced covariate coefficient estimates of underlying TB –related time series incidence rates and of absolute numbers of MDR-TB cases for seasonally defining indicators of MDR-TB transmission 'hot spots'.

Furthermore, according to their findings, estimating the absolute number of MDR-TB patients was critical for planning the delivery of directly observed MDR-TB therapy and the rational procurement of second-line drugs. In another hierarchical residual explanatory time series MDR-TB regression model analysis, Shafer et al (1995) ascertained the role of the HIV and Mycobacterium tuberculosis transmission on MDR-TB emergence in New York City using drug susceptibilities and restriction-fragment-length-polymorphisms of TB cases at a city hospital between two nine-month periods (1987/1988 and 1990/1991). The proportion of TB patients with MDR increased from 10% (27/267) to 17% (38/222; P =.03). Among MDR-TB patients of known HIV status, the proportion with HIV increased from 16% (3/19) to 58% (22/38; P =.006). HIV-infected MDR-TB patients were more likely than HIV-seronegative MDR-TB patients to have initial MDR (88% vs 56%; P = .03). Among 56 MDR-TB cases, 12 had unique patterns while 44 belonged to only one of six groups. The hierarchical, residual ,explanatory, endemic transmission-oriented, risk-analyses revealed that 75% (27/36) of MDR-TB patients during the 1990/1991 period were infected with strains cultured from HIV-seronegative patients during the 1987/1988 period.

Mapping, and analyzing prevalence or incidence MDR-TB data with conventional hierarchical residual explanatory statistical approaches; however, can be problematic as spatiotemporal clinical residualized model outputs can be affected by random variation due to population variability leading to a loss of statistical power when cases are assigned to subgroups. Spatiotemporal explanatory classification of MDR-TB endemic transmission-oriented data is common as the distribution of clinical and environmental sampled georeferenced predictor covariate coefficient estimates usually encompasses several geographic areas (see Gandhi et al. 2006).

To reduce this within-subject standard deviation, previous research in other medical disciplines have employed a compound Poisson approach for detection of residual clustering of varying and constant georeferenced explanatory time series covariate coefficient estimates by testing individual areas that may be combined with their neighbors. For example, Besag and Newell (1991) proposed a hierachical within residualized explanatory cluster-based model to screen for collections of childhood leukemia cases in northern England; whereby each georefernced classified sub-location was based on the number of neighbors that had to be combined in order to contain a minimum number of cases (i.e., cluster sampled size). This method scanned the data for collections of cases that appeared to be unusual clusters. To do so, the hierarchical explanatory time series intra-cluster diagnostic error detection algorithm centered a circular window on each sub-region. This window was then expanded to include neighboring regions until the total number of cases in the window reached a user-specified threshold, *k*. Then, the population size inside the window was compared to that expected under an average or expected frequency rate. They found no evidence for clustering of leukemia cases in the years surveyed (1975-85). Waller et al. (1994) used the same method to quantitate prevalence survey patterns in leukemia in

upstate New York. They did not find strong evidence for clustering, although there was a suggestion of some clustering in one county. They did however recommend employing the technique to prioritize areas for further study. Le, Petkau, and Rosychuk (1996) used a modification of the Besag and Newell hierarchical residual algorithimic error detection method to examine whether time series cancer clusters appeared near pulp and paper mills in British Columbia, Canada. This format relied on a pre-determined cluster size for each test, which was provided by Le et al. (1996) who generated a testing algorithm for the automatic selection of infectious disease–oriented time series cluster sizes. The method successfully re-identified several knownexplanatory clusters of different types of cancers.

Commonly, Besag and Newell's method calculates two statistics: *l* which is the local statistic, the number of regions required for the window centered over an individual region to contain *k* cases. To evaluate whether the *k* cases form a cluster, the method looks to see whether the number of cases in the window are unlikely for the window's population at risk. The null hypothesis employing the classified intra-cluster time series explanatory error detection algorithm is that there is no clustering, (e.g., a Poisson disease rate does not exist across the epidemiological study area). Thus, the case count inside the window would be proportional to the population at risk, otherwise the null hypothesis is rejected. Following the null spatial model as defined by Besag and Newell (1991), cases that are distributed among the areas in an epidemiological interventional study site may be tabulated to be proportional to a sampled population size employing a common disease rate. The method can calculate probability for *l* under the null spatial model employing $P(L \leq 1) = 1 - \sum_{x=0}^{k-1} \frac{e^{-\lambda} \lambda^x}{x!}$. This expression may also calculate probability that *l* has reached or exceeded the value predicted by the null hypothesis (*L*) in a robust, explanatory, spatiotemporal, endemic, transmission-oriented, MDR-TB-related, forecasting, risk model. In general in MDR-TB time series data analyses it is 1 minus the probability that *l* is less than *L*, (i.e., the probability that there are fewer than *k* cases in the area) (see Ghandhi et al. 2006). The probability of 0 through *k*-1 MDR-TB cases may then be found by summing the Poisson term from $x = 0$ to $x = k - 1$. Lambda, (λ), is the average or expected case count, the average or expected disease frequency multiplied by the population-at-risk. The term *e* will then indicate the exponential function in the explanatory, clinical, MDR-TB, endemic, transmission-oriented, explanatory, risk-related, model residual forecasts.

When performing a Besag and Newell analysis however, there is a necessity to calculate *l* and its significance for all the georeferencable explanatory clusters identified. By so doing, the algorithm will list all clusters that have a probability less than the significance level specified, (e.g., alpha). The default alpha normally in the MDR-TB explanatory residual forecasts thereafter would then be calculated using 0.05. Unfortunately, it is difficult to determine the appropriate cluster size *k* of a prior and thus this test clearly will face multiple testing problems for time series MDR-TB risk modeling

spatiotemporal, clinical, uncertainity-based covariate coefficients by repeating the test with a number of plausible values of *k*.

Spatial statistics may help quantitate local and global clustering in explanatory, spatiotemporal-sampled, clinical and/or environmental MDR-TB endemic transmission-oriented data. For example, the standard spatial scan statistic is a maximum likelihood (ML) ratio test statistic based on a circular window of variable size scanning an epidemiological geographical area under surveillance to determine hotspots. In recent years, there has been much effort invested in designing efficient algorithms for geographically locating "high discrepancy" regions in statistical software (e.g., SAS/GIS) with methods ranging from fast heuristics for special cases, to generalized orthogonalizable digitized grid-based matrices (see Jacob et al. 2013, Jacob et al. 2010b, Ghandhi 2006). Employing spatial scan statistics and explanatory time series approximation cluster-based error-detection algorithms, the number of contributions to the computational studies of spatiotemporal MDR-TB endemic transmission may robustly be constructed based on sampled clinical and/or environmental georeferenced explanatory observational covariate coefficient estimates. By so doing, spatial statistics may elucidate the mechanics of MDR-TB transmission by prioritizing seasonal-sampled, georeferenced ,district-level, explanatory covariate coefficients. Thus, covariates for identifying spatial distribution of high-risk populations and random time series heterogeneity in resistant strains may be efficiently quantitated employing explanatory, residual-based, hierarchical ,cluster-based, error diagnostics for determining multivariate heteroscedastic parameters, for example, from hierarchical explanatory intra-cluster-based regression model residuals. This is vital since inconspicuous latent uncertainty coefficients and error probabilities found in an empirical-sampled dataset of MDR-TB regression model residual forecasts have revealed that errors in variance estimation can substantially alter numerical predictions of the model by inflating the value of test statistic thereby, increasing the chance of a Type I error - incorrect rejection of the null hypothesis (Jacob et al. 2010b).

An approximation, explanatory, georeferencable, hierarchical ,residualized, intra-cluster, error- detection algorithm for a large class of discrepancy-oriented MDR-TB endemic transmission-oriented risk model derived functions may also improve the operationalizable approximation of prior methods employing the Kulldorff scan statistic. Kulldorff's framework assumes that counts c_i are Poisson distributed with $c_i \sim Po(qb_i)$, where *bi* represents the known census population of cell *si* and *q* which then in a predictive, epidemiological, explanatory, spatiotemporal,endemic, transmission-oriented,predictive, risk model would represent the unknown underlying MDR-TB infection rate (see Cressie 1993). Since extensions of these simple approximation algorithms can be seasonally generated using explanatory, clinical-based, endemic, transmission-oriented, MDR-TB, covarite, coefficient, measurement values within sub-meter resolution orthogonalizable grid-based matrices (see Jacob et al. 2013, Jacob et al. 2010b, Ghandhi 2006), extant methods may be statistically customized for efficiently

regressing georeferenced, clinically-oriented, explanatory, time series MDR-TB datasets.

For large, spatiotemporal, empiricial, MDR-TB–related, explanatory datasets, an intra-cluster-based, diagnostic, error-detection algorithm can also be constructed using spatial scans to examine whether small-space, spatiotemporal-dependent, streaming, algorithms yield accurate residualized, interpolated ,forecasts targeting explanatory endemic regions based on clinical density count data (e.g., resistant strain data). Streaming algorithms may also provide optimal answers to the discrepancy maximization problem commonly found in predictive, spatiotemporal, clinical, endemic, transmission-oriented, MDR-TB, epidemiological, risk models using space –time, explanatory covariate coefficients as the input. Streams can be denoted as an ordered sequence of points (or "updates") that must be accessed in order and can be read only once or a small number of times.Much of the streaming literature is concerned with computing statistics on frequency distributions that are too large to be stored. For this class of problems, there is a vector $a = (a_1, \ldots, a_n)$ (initialized to the zero vector 0) that has updates presented to it in a stream. The goal of these algorithms is to compute functions employing considerably less space than it would take to represent a precisely. A notable special case is when c=1 (only unit insertions are permitted).Besides the above frequency-based problems, some other types of problems have also been studied for MDR-TB predictive modeling. Many MDR-TB time series graph problems are solved in the setting where the adjacency matrix or the adjacency list of the graph is streamed in some unknown order. There are also some problems that are very dependent on the order of the stream (i.e., asymmetric MDR-TB model functions), such as counting the number of inversions in a stream and finding the longest increasing subsequence.

Minimizing the discrepancy of a set system is a fundamental problem in combinatorics associated to explanatory infecious diease models such as MDR-TB predictive epidemiological risk models. Combinatorics is a branch of mathematics concerning the study of finite or countable discrete structures. Aspects of combinatorics include counting the structures of a given kind and size (enumerative MDR-TB-related combinatorics), for deciding when certain criteria can be met, and constructing and analyzing objects meeting the criteria (as in combinatorial designs and matroid theory), finding "largest", "smallest", or "optimal" objects (extremal combinatorics and combinatorial optimization), and studying combinatorial structures arising in an algebraic context, or applying algebraic techniques to combinatorial problems (algebraic combinatorics). Combinatorial problems arise in many areas of pure mathematics, notably in algebra, probability theory, topology, and geometry, and combinatorics also has many applications in mathematical optimization, computer science, ergodic theory and statistical physics. Many combinatorial may be considered in isolation, giving an *ad hoc* solution to a problem arising in a dataset of explanatory time series MDR-TB predictive risk model residually forecasted derivatives in some mathematical contex.

Many combinatorial problems can be efficiently solved for MDR-TB related series-parallel graphs or partial k-trees. In graph theory, a *k*-tree is a chordal graph all of whose maximal cliques are the same size $k + 1$ and all of whose minimal clique separators are also all the same size *k*. Optimal infectious disease –related epidemiological MDR-TB risk-related *k*-trees would be exactly the maximal graphs with a given treewidth, graphs to which no more edges can be added without increasing their treewidth. The graphs that have treewidth at *k* would then be exactly the subgraphs of MDR-TB risk-related *k*-trees, and for this reason they would be considered partial *k*-trees. Every explanatory, spatiotemporal, predictor, *k*-tree may be formed by starting with a $(k + 1)$-vertex complete graph and then repeatedly adding vertices in such a way that each added vertex has exactly *k* neighbors that form a clique(Cressie 1993).

Certain explanatory, spatiotemporal, MDR-TB *k*-trees (e.g., with $k \geq 3$) may be also the graphs formed by the edges and vertices of stacked polytopes.These polytopes are formed by starting from a simplex and then repeatedly gluing simplices onto the faces of the polytope; this gluing process will mimicic the construction of *k*-trees by adding vertices to a clique.Every stacked MDR-TB-related polytope would then form a *k*-tree. Unfortunately, not every *k*-tree comes from a stacked polytope since MDR-TB *k*-tree is the graph of a stacked polytope if and only if no three $(k + 1)$-vertex cliques have *k* vertices in common (see Spencer 1985).

Many combinatorial problems can be efficiently solved for partial *k*-trees. However,the edge-coloring problem is one of a few combinatorial problems for which no linear-time algorithm has been obtained for partial *k*-trees. The best known algorithm solves the problem for partial *k*-trees *G* in time $O\left(n\Delta 22(k+1)\right)$ where *n* is the number of vertices and Δ is the maximum degree of *G*. A linear algorithm which optimally edge-colors may be provied for given partial MDR-tb related *k*-tree for fixed *k*. The edge-coloring problem is one of a few combinatorial problems for which no efficient algorithms have been obtained for series algorithms. - In any system of *n* sets in a universe of size n, there always exists a coloring which achieves discrepancy 6\sqrt{n} (Spencer 1985). The original proof of Spencer was existential in nature, and did not give an efficient algorithm to find such a coloring in a spatiotemporal, predictive, epidemiological MDR-TB endemic, transmission-oriented, explanatory, georeferencable risk model . Recently, a breakthrough work of Bansal (2010) gave an efficient algorithm which finds such coloring. His algorithm was based on an Semidefinite programming SDP relaxation of the discrepancy problem and a clever rounding procedure in the risk model residually forecasted derivatives.

Semidefinite programming (SDP) is a subfield of convex optimization concerned with the optimization of a linearized objective function (that is, a function to be maximized or minimized) over the intersection of the cone of positive semidefinite matrices with an affine space, (i.e., a spectrahedron). Semidefinite programming is a relatively new field of optimization which is of growing interest for several reasons. Many practical time serie endemic trasnmission-oriented explanatory MDR-TB problems in clinical operations research can be modeled or approximated using semidefinite programming. In automatic control theory, MDR-TB –related SDP's may be

employed in the context of linear matrix inequalities. SDPs are in fact a special case of cone programming and can be efficiently solved by interior point methods (Cresssie 1993). All linear programs for predictive MDR-TB epidemiological risk modeling can be expressed as SDPs, and via hierarchies of SDPs the solutions of polynomial optimization problems can be approximated. Semidefinite programming has been used in the optimization of complex systems. In recent years, some quantum query complexity problems have been formulated in term of semidefinite programs which may be employed also in epidemiological MDR-TB risk forecast modeling.

Linearized programming problem is one in which we wish to maximize or minimize a linear objective function of clinical-sampled MDR-TBendemic trasnmission-oriented predicot variables over a polytope. In semidefinite programming,an experimenter may use real-valued vectors and to quantitate the dot product of vectors; nonnegativity constraints on real variables in LP are replaced by semidefiniteness constraints on matrix variables in SDP (see Hazewinkle 2002). Specifically, a general semidefinite programming problem can be defined as any mathematical programming problem of the form

$$\min_{x^1,\ldots x^n \in \mathbb{R}^n} \sum_{i,j \in [n]} c_{i,j}\left(x^i \bullet x^j\right)$$

$$subject\ to \sum_{i,j \in [n]} a_{i,j,k}\left(x^i \bullet x^j\right) \le b_k \ \forall k.$$

A time series, forecasting, regression-based, endemic, trasnmission-oriented, MDR-TB epidemiological $n \times n$ matrix M is positive semidefinite if it is the gramian matrix of some vectors (i.e. if there exist vectors $x^1,\ldots x^n$ such that $m_{i,j} = x^i \bullet x^j$ for all i, j). In linear algebra, the Gramian matrix (or Gram matrix or Gramian) of a set of vectors $v_1,\ldots,v_n$ in an inner product space is the Hermitian matrix of inner products, whose entries are given by $G_{ij} = \left\langle v_j, v_i \right\rangle$. For finite-dimensional real vectors with the usual Euclidean dot product, the Gram matrix is simply $G = V^T V$ (or $G = V^* V$ for complex vectors using the conjugate transpose), where V is a matrix whose columns are the vectors v_k (Griffith 2003)An important application for MDR-TB time series risk modeling would be to compute linear independence: a set of vectors is linearly independent if and only if the Gram determinant (the determinant of the Gram matrix) is non-zero (see Cressie 1993). If this is the case, an MDR-TB experimeneter may denote this as $M \succeq 0$. Note that there are several other equivalent definitions of being positive semidefinite in MDR-TB risk modeling. For example, positive semidefinite endemic, transmission-oriented, explanatory matrices generally would have only non-negative eigenvalues and have a positive definite square root. In order to quantitate parameter estimator significance levels in a endemic trasnmission-oriented explanatory MDR-TB risk model. Furthermore, there would be a necessity to denote the $\mathbb{S}^n$ of all $n \times n$ clinically-related, time-series, symmetric matrices. The space would have to equipped with the inner product where *tr* denotes the trace where

$$\left\langle A, B \right\rangle_{\mathbb{S}^n} = tr\left(A^T B\right) = \sum_{i=1, j=1}^{n} A_{ij} B_{ij}.$$ in the risk model. An MDR-TB experimeneter could rewrite the time series mathematical program as

$$\min_{X \in \mathbb{S}^n} \left\langle C, X \right\rangle_{\mathbb{S}^n}$$

$$subject\ to \left\langle A_k, X \right\rangle_{\mathbb{S}^n} \le b_k, k = 1,\ldots,m$$
$$X \succeq 0$$

where entry i, j in C which may be then given by $c_{i,j}$ and A_k in an $n \times n$ matrix having i, j th entry $a_{i,j,k}$. Note if an an the experimenter adds slack variables appropriately to the risk model, the SDP can be converted to one of the form

$$\min_{X \in \mathbb{S}^n} \left\langle C, X \right\rangle_{\mathbb{S}^n}$$

$$subject\ to \left\langle A_i, X \right\rangle_{\mathbb{S}^n} \le b_i, i = 1,\ldots,m$$
$$X \succeq 0$$.

In an optimization problem, a slack variable is a variable that is added to an inequality constraint to transform it to an equality (Cressie 1993). Introducing a slack variable into an endemic trasnmission-oriented explanatory MDR-TB risk-related forecasting model would replace an inequality constraint with an equality constraint and a nonnegativity constraint in the forecasted deriavtives for effectively targeting the statsitically significant clinical variables.In linear programming for constructing a robust, spatiotemrporal, endemic, trasnmission-oriented, explanatory, MDR-TB risk model it may be required to turn an inequality into an equality where a linear combination of the clinical variables is less than or equal to a given constant in the former. As with the sampled clinical-sampled explanatory variables in the augmented constraints, the slack –oriented explanatory MDR-TB predictor variables may not take on negative values, as the Simplex algorithm requires them to be positive or zero.

The simplex algorithm operates on linear programs in standard form, that is linear programming problems of the form, $c^T \bullet X$ subject to $A_X = b, x_i \ge 0$ with $x = (x_1,\ldots,x_n)$. The clincal-sampled MDR-TB variables of the problem, $c = (c_1,\ldots,c_n)$ would be the coefficients of the objective function, A a $p \times n$ matrix, and $b = (b_1,\ldots,b_p)$ constants with $b_j \ge 0$. There is a straightforward process to convert any time series, explanatory, endemic, trasnmission-oriented, MDR-TB risk-related linear program into one in standard form so this results in no loss of generality in the forecasted derivatives..In geometric terms, the feasible region $A_X = b, x_i \ge 0$ may be a possibly unbounded convex polytope. Convex polytope is a special case of a polytope, having the additional property that it is also a convex set of points in the n-dimensional space $\mathbf{R}^n$ (Cressie 1993). By so doing, a simple characterization of the extreme sampled clinical points or vertices namely $x = (x_1,\ldots,x_n)$ in the epidemiological MDR-TB risk model may be identified where an extreme point may be quantiated parsimoniously, if and only if the subset of column vectors A_i corresponding to the nonzero entries of x $(x_i \ne 0)$ are linearly independent in the risk model.

For convenience, an SDP for a robust, endemic, trasnmission-oriented , explanatory, MDR-TB risk model may be specified in a slightly different equivalent forms. For example, linear expressions involving nonnegative spatiotemporal -sampled MDR-TB-related clinically oriented scalar variables may be added to the program specification. This remains an SDP because each clinical sampled variable can be incorporated into the time series matrix X as a diagonal entry (X_{ii} for some i). To ensure that $X_{ii} \geq 0$, constraints $X_{ij} = 0$ can be added for all $j \neq i$ in the model residuals. By introducing the slack variable $y \geq 0$, the inequality $A_X \leq b$ can be converted to the equation $A_X + y = b$. Slack variables may give an embedding of a polytope $P \to (R_{\geq 0})^f$ into the standard *f*-orthant, where *f* is the number of constraints (facets of the polytope). This MDR-TB risk map may then be expressed in terms of the *constraints* (e.g.,linear functionals, covectors).Slack variables are dual to generalized barycentric coordinates, and, dually to generalized barycentric coordinates (which are not unique but can all be realized), are uniquely determined, but cannot all be realized. Dually, generalized barycentric coordinates express a polytope with *n* vertices (dual to facets), regardless of dimension, as the *image* of the standard $(n-1)$-simplex, which has *n* vertices – the map is onto:

$\Delta^{n-1} \to P$ and expresses points in terms of the *vertices* (points, vectors). The endemic trasnmission-oriented risk map would be classified one-to-one if and only if the polytope is a simplex, in which case the map would be an isomorphism. As such, the model would correspond to any clincal-sampled point not having a unique generalized barycentric coordinate.

As another example, note is that for any positive semidefinite spatiotemporal MDR-TB epidemiological matrix X, there would exist a set of vectors $\{v_i\}$ such that the i, j entry of X is $X_{i,j} = (v_i, v_j)$ the scalar product of v_i and v_j. Therefore, a MDR-TB-related explanatory SDPs may be formulated in terms of linear expressions on scalar products of vectors. Given the solution to the SDP in the standard form, the vectors $\{v_i\}$ can be recovered in $O(n^3)$ time (e.g., by using an incomplete Cholesky decomposition of X).

The Cholesky decomposition of a Hermitian positive-definite matrix $\mathbf{A}$ is a decomposition of the form $\mathbf{A} = \mathbf{L}\mathbf{L}^*$ where $\mathbf{L}$ is a lower triangular matrix with real and positive diagonal entries, and $\mathbf{L}*$ denotes the conjugate transpose of $\mathbf{L}$. Every Hermitian MDR-TB related positive-definite matrix (and thus also every real-valued symmetric positive-definite matrix) would have a unique Cholesky decomposition. If the matrix $\mathbf{A}$ is Hermitian and positive semi-definite, then it still has a decomposition of the form $\mathbf{A} = \mathbf{L}\mathbf{L}*$ if the diagonal entries of $\mathbf{L}$ are allowed to be zero. When $\mathbf{A}$ has real entries, $\mathbf{L}$ has real entries as well and the factorization may be written $\mathbf{A} = \mathbf{LL}$ (Griffith 2003).

Analogously to linear programming, given a generalized MDR-TB related SDP of the form

$$\min_{X \in \mathbb{S}^n} \langle C, X \rangle_{\mathbb{S}^n}$$
$$\text{subject to} \langle A_i, X \rangle_{\mathbb{S}^n} \leq b_i, i = 1, \ldots, m$$
$$X \succeq 0$$

(the primal problem or P-SDP), may define the *dual* semidefinite program (D-SDP) as

$$\max_{y \in \mathbb{R}^m} \langle b, y \rangle_{\mathbb{S}^n}$$
$$\text{subject to} \sum_{i=1}^m y_i A_i \preceq C$$

where for any two matrices P and Q, $P \succeq Q$ would signify $P - Q \succeq 0$. The weak duality theorem states that the value of the primal SDP is at least the value of the dual SDP. Therefore, any feasible solution to the dual SDP lower-bounds in a MDR-TB risk model would involve the primal SDP value, and conversely, any feasible solution to the primal SDP upper-bounds would then involve the dual SDP value in the forecasted derivatives. This is because

$$\langle C, X \rangle - \langle b, y \rangle = \langle C, X \rangle - \sum_{i=1}^m y_i b_i$$

where the last

$$= \langle C, X \rangle - \sum_{i=1}^m y_i \langle A_i, X \rangle = \left\langle C - \sum_{i=1}^m y_i A_i, X \right\rangle \geq 0,$$

inequality is because both matrices are positive semidefinite, where the result of this function is sometimes referred to as duality gap.

In optimization problems in applied mathematics, the duality gap is the difference between the primal and dual solutions. If d^* is the optimal dual value and p^* is the optimal primal value then the duality gap is equal to $p^* - d^*$. This value in a endemic trasnmission-oriented risk model would always be greater than or equal to 0. The duality gap is zero if and only if strong duality holds (Hazewinkle 2002). Otherwise the gap is strictly positive and weak duality holds. In general given two dual pairs of clinical-sampled MDR-TB predictor variables separated locally convex spaces (X, X^*) and (Y, Y^*) then given the function $f : X \to \mathbb{R} \cup \{+\infty\}$, an experimenter may define the primal problem by $\inf_{x \in X} f(x)$. If there are constraint conditions in the risk model, these can be built into the function f by letting $f = f = I_{constraints}$ where I is the indicator function. Then, if the expermnter lets $F : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a perturbation function in the risk model such that $F(x, 0) = f(x)$ the duality gap may be quantitated by the difference given by $\inf_{x \in X} [F(x, 0)] - \sup_{y^* \in Y^*} [-F^*(0, y^*)]$ where F^* is the convex conjugate in the sampled time series MDR-TB risk-related clinical variables.

Interestingly, im computational optimization, another "duality gap" is often reported, which is the difference in value between any dual solution and the value of a feasible but suboptimal iterate for the primal problem. This alternative "duality gap" quantifies the discrepancy between the value (e.g., sampled MDR-TB endemic trasnmission-oriented explanatory variable) while deriving suboptimal iterates for the primal problem and the value of the dual problem. The value of the dual problem is,

under regularity conditions, equal to the value of the convex relaxation of the primal problem: The convex relaxation is the problem arising replacing a non-convex feasible set with its closed convex hull and with replacing a non-convex function with its convex closure, that is the function that has the epigraph that is the closed convex hull of the original primal objective function(Hazewinkle 2002).

Under a condition known as Slater's condition, the value of the primal and dual SDPs may be equal in a time-series, explanatory, endemic, transmission-oriented, epidemiological, risk model. This is known as strong duality. Unlike for linear programs, however, not every MDR-TB-related SDP would satisfy strong duality; in general, the value of the dual SDP may lie strictly below the value of the primal. Suppose the primal problem (P-SDP) is bounded below and strictly feasible in a spatiotemporal MDR-TB risk forecasting model(i.e., there exists $X_0 \in \mathbb{S}^n, X_0 \succ 0$ such that $\langle A_i, X_0 \rangle_{\mathbb{S}^n} = b_i$, $i = 1,\dots,m$)[equation 1.1]. Then there is an optimal solution y^* to (D-SDP) and $\langle C, X^* \rangle_{\mathbb{S}^n} = \langle b, y^* \rangle_{\mathbb{R}^m}$. Suppose the dual problem (D-SDP) is bounded above and strictly feasible in the model (i.e., $\sum_{i=1}^{m} (y_0)_i\, A_i \prec C$ for some $y_0 \in \mathbb{R}^m$). Then there is an optimal solution X^* to (P-SDP) and the equality from (1.1) holds.

Consider three random variables A, B, and C in a robust MDR-TB risk model. By definition, their correlation coefficients $\rho_{AB}, \rho_{AC}, \rho_{BC}$ are valid if and

only if $\begin{pmatrix} 1 & \rho_{AB} & \rho_{AC} \\ \rho_{AB} & 1 & \rho_{BC} \\ \rho_{AC} & \rho_{BC} & 1 \end{pmatrix} \succeq 0$ Suppose that an

experimenter knows from some prior knowledge (empirical results of a MDR-TB clinical experiment, for example) that $-0.2 \le \rho_{AB} \le -0.1$ and $0.4 \le \rho_{BC} \le 0.5$. The problem of determining the smallest and largest clincial sampled values that ρ_{AC} can take may be then given by: minimize/maximize x_{13} subject to $-0.2 \le x_{12} \le -0.1$, $0.4 \le x_{23} \le 0.5$ $x_{11} = x_{23} = x_{33} = 1$

$\begin{pmatrix} 1 & x_{12} & x_{13} \\ x_{12} & 1 & x_{23} \\ x_{13} & x_{23} & 1 \end{pmatrix} \succeq 0$ where an exprimenter could set

$\rho_{AB} = x_{12}, \rho_{AC} = x_{13}, \rho_{BC} = x_{23}$ to obtain paramter estimator significance levels. This can be formulated by an SDP. The experimenter may then handle the inequality constraints by augmenting the variable matrix and introducing slack variables, for example

$$tr\left(\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{12} & x_{13} & 0 & 0 & 0 \\ x_{12} & 1 & x_{23} & 0 & 0 & 0 \\ x_{13} & x_{23} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \right)$$

$= x_{12} + s_1 = -0.1$

Solving this SDP would then render the minimum and maximum values of $\rho_{AC} = x_{13}$ as -0.978 and 0.872 respectively

Semidefinite programs are important tools for developing approximation algorithms for NP-hard maximization problems. The first approximation algorithm based on an SDP is due to Goemans and Williamson (JACM, 1995). They studied the MAX CUT problem: Given a graph $G = (V, E)$, output a partition of the vertices V so as to maximize the number of edges crossing from one side to the other. This problem can be expressed as an integer MDR-TB related quadratic program by maximizing $\sum_{(i,j) \in E} \frac{1 - v_i v_j}{2}$, such that each $v_i \in \{1, -1\}$. Unless $P = NP$, however the program would not be able to solve this maximization problem in a spatiotemrpaol MDR-TB epidemiological forecasting risk model efficiently. However, Goemans and Williamson observed a general three-step procedure for attacking this sort of problem by relaxing the integer quadratic program into an SDP and solving the SDP to within an arbitrarily small additive of error. Thereafter, by rounding the SDP solution an approximate solution to the original integer quadratic program in the risk model may be obtained. For MAX CUT, the most natural relaxation is $\max \sum_{(i,j) \in E} \frac{1 - v_i v_j}{2}$, such that $\|v_i\|^2 = 1$, where the maximization is over vectors $\{v_i\}$ instead of integer scalars. (Griffith 2003).This relaxation in an MDR-TB related SDP in the model would be associated to the objective function and constraints which are all linear functions of vector inner products. Solving the MDR-TB - related time series SDP would then render a set of unit vectors in R^n . Importantly, since the vectors in a MDR-TB forecasting endemic transmission-oriented explanatory model cannot be collinear (see Jacob et al. 2013c, Jacob et al. 2010b), the value of this relaxed program can only be higher than the value of the original quadratic integer program. Finally, a rounding procedure would be needed in the dataset of the residually forecasted derivatives targeting the statistically significant MDR-TB related clinical predictors to obtain a partition. Goemans and Williamson simply chose a uniformly random hyperplane through the origin and divided the vertices according to which side of the hyperplane the corresponding vectors lay. Straightforward analysis shows that this procedure achieves an expected approximation ratio (i.e., performance guarantee) of $0.87856 - \varepsilon$. The expected value of the cut would then be the sum over edges of the probability in the risk model outputs where the edge is cut, which may be proportional to the angle $\cos^{-1}\langle v_i, v_j \rangle$ between the vectors at the endpoints of the edge over π. Comparing this probability to $(1 - \langle v_i, v_j \rangle)/2$, in expectation the ratio is always at least 0.87856. Assuming the Unique Games Conjecture, it can be shown that this approximation ratio is essentially optimal.

In computational complexity theory, the Unique Games Conjecture is a conjecture made by Subhash Khot in 2002.

The conjecture postulates that the problem of determining the approximate value of a certain type of game, known as a *unique game*, has NP-hard algorithmic complexity. It has broad applications in the theory of hardness of approximation. If it is true, then for many important problems in MDR-TB epidemiological risk modeling it is not only too hard to get an exact solution (as postulated by the P versus NP problem), but also too hard to get a good approximation for proper clinical and/or environmental parameter estimator significance testing.. There are however important implications for constraint satisfaction problems in endemic tranmission-oriented MDR-Tb forecasting risk models. The formulation of the unique games conjecture is often used in hardness of approximation (Hazewinkle 2002).The conjecture postulates the NP-hardness of the following promise problem known as label cover with unique constraints. For each edge, the colors on the two vertices in a robust MDR-TB risk model would then be restricted to some particular ordered pairs. In particular, unique constraints means that for each edge none of the ordered pairs have the same color for the same node.This means that an instance of label cover in the risk model with unique constraints over an alphabet of size k can be represented as a graph together with a collection of permutations $\pi_e : [k] \rightarrow [k]$, one for each edge e of the endemic trasnmission-oriented graph. An assignment to a label cover instance gives to each vertex of G a value in the set $[k]$, often called "colours."
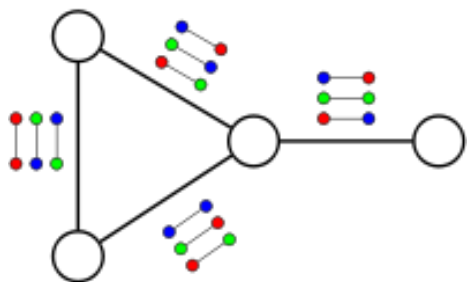

Figure 1.1. The 4 vertices in a predictive MDR-TB epidemiological risk model that have been assigned colors while satisfying the constraints at each edge
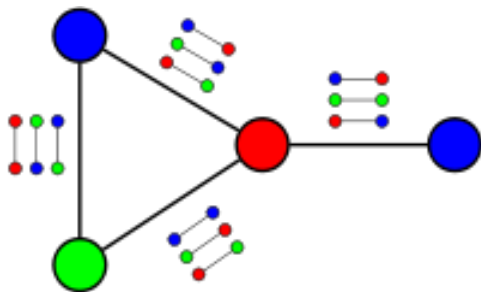

Figure 1.2. A solution to the unique label cover instance for a predictive MDR-TB epidemiological risk model

Unfortunately, such instances in a time-series, explanatory, endemic, trasnmission-oriented, MDR-TB risk model may be strongly constrained in the sense that the colour of a vertex uniquely would define the colours of its neighbours, and hence for its entire connected component. Thus, if the input instance admits a valid assignment in the risk model, then such an assignment can be found efficiently by iterating over all colours of a single node. In particular, the problem of deciding if a given instance admits a satisfying assignment can be solved in polynomial time.
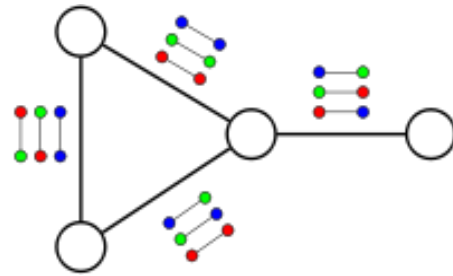

Figure 1.3. An instance of unique label cover that does not allow a satisfying assignment in a MDR-TB risk model
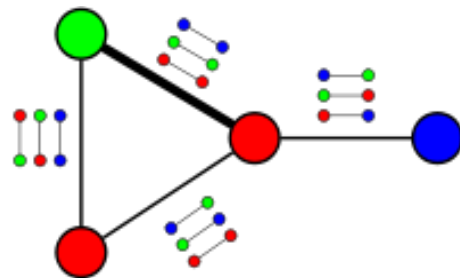

Figure 1.4. An assignment that satisfies all edges except the thick edge in a MDR-TB risk model

The sampled MDR-TB endemic trasnmission-oriented explanatory time-series valu**e** of a unique label cover instance would then be the fraction of constraints that can be satisfied by any assignment. For satisfiable instances, this value would be 1. On the other hand, it may be very difficult for an experimenter to determine the value of an unsatisfiable MDR-TB predictive risk model , even approximatively. The unique games conjecture would formalize this difficulty. More formally, the (c, s) gap label cover problem with unique constraints would be (L_{yes}, L_{no}): $L_{\text{yes}} = \{G$ where, some assignment satisfies at least a c-fraction of constraints in $G\}$ and $L_{\text{no}} = \{G$: Every assignment satisfies at most an s-fraction of constraints in $G\}$where G is an instance of the label cover problem with unique constraints in the model output (Cressie 2003).The unique games conjecture states that for every sufficiently small pair of constants ε, $\delta > 0$, there exists a constant k such that the $(1 - \delta, \varepsilon)$ gap label cover problem with unique constraints over alphabet of size k is (**N**on-deterministic Polynomial-time hard), NP-hard (Hazewinkle 2002).

Non-deterministic Polynomial-time hard in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP". A problem H is NP-hard if and only if there is an NP-complete problem L that is polynomial time Turing-reducible to H (i.e., L $\leq_{\text{T}}$ H). In other words, L can be solved in polynomial time by an oracle machine with an oracle for *H*. Informally, an MDR-TB experimenter could think of an algorithm that can call such an oracle machine as a subroutine for solving *H*, and solves *L* in polynomial time, if the subroutine call takes only one step to compute. NP-hard-related forecasting MDR-TB endemic trasnmission-oriented problems may be of any type: decision problems, search problems, or optimization problems. Instead of time series MDR-TB

graphs, the label cover problem can be formulated in terms of linear equations. For example, suppose that an MDR-TB experimater has a system of linear equations over the integers modulo 7: $x_1 \equiv 2 \cdot x_2 \pmod 7$, $x_2 \equiv 4 \cdot x_5 \pmod 7$ $x_1 \equiv 2 \cdot x_7 \pmod 7$ This would an instance of the label cover problem in a time-series, robust MDR-TB epidemiological risk model with unique constraints. For example, the first equation corresponds to the permutation $\pi_{(1, 2)}$ where $\pi_{(1, 2)}(x_1) = 2x_2$ modulo 7 .A P-problem in a MDR-TB risk model whose solution time is bounded by a polynomia is always also NP (Hazewinkle 2000). If a problem is known to be NP, and a solution to the problem is somehow known, then demonstrating the correctness of the solution can always be reduced to a single P (e.g., MDR-TB polynomial time) verification. If P and NP are *not* equivalent in a time series MDR-TB epidemiological risk model, then the solution of NP-problems requires an exhaustive search. Linear programming, long known to be NP and thought *not* to be P, was shown to be P by L. Khachian in 1979. It is an important unsolved problem to determine if all apparently NP problems are actually P in a MDR-TB epidemiological risk model . A problem is said to be NP-hard if an algorithm for solving it can be translated into one for solving any other NP-problem(Haxewinkle 2002). It may be much easier to show that a problem of statistical significance testing in a time series MDR-TB model is NP than to show that it is NP-hard . A problem which is both NP and NP-hard is called an NP-complete problem (Cressie 1993).

In more recent work a new randomized algorithm has been employed to find a coloring based on a restricted random walk (E.G., "Edge-Walk"). The algorithm and its analysis uses basic linear algebra and is truly constructive in that it does not appeal to the existential arguments. The algorithm may then provide significant contributions to literature such as a new proof of Spencer's theorem and the partial coloring lemma employing time- series explanatory MDR-TB predictive risk model parameter estimators. One of the most basic techniques for combinatorial optimization is linear programming relaxation which may be applicable to time series, explanatory, MDR-TB, predictive, endemic, transmission-oriented, risk modeling. Phrasing optimization in a language more suitable for time series MDR-TB epidemiological forecasting risk models would however, require a constraint matrix $A \in \mathbb{R}^{m \times m}$, a target vector $b \in \mathbb{R}^n$ to render $x \in \{0,1\}^n$ so as to minimize $\|Ax - b\|_\infty$. In order to derive robust clinical explanatory MDR-TB residualized forecasts, an experimenter could relax this 'discreteness' in the coefficinet estimates and instead solve the linear program $\min_{y \in [0,1]^n} \|Ay - b\|_\infty$ in the derivatives. This can be done efficiently employing a restricted random walk algorithm and linear algebra. The next step, and often the most challenging one for most epidemiological infectious disease predictive explanatory risk modeling exercises, would be to round the fractional solution $y \in [0,1]^n$ to an integer solution in $\{0,1\}^n$ while maintaining minmal "loss". How well an MDR-TB experimenter

rounds off these intergers would be vital for determining robustness of a constraint time series matrix A , and general vectors b for obtaining robust, endemic, transmission-oriented, explanatory, time-series, estimators. This may be captured efficiently by the notion of linearized discrepancy introduced by Lovasz, Spencer and Veszttergombi (LSV):

$$lindisc(A) = \max_{y \in [0,1]^n} \min_{x \in \{0,1\}^n} \|Ax - Ay\|_\infty .$$

LSV introduced the notion of parameter estimation originally as a generalization of discrepancy which could be formulated theretically in a time-series, explanatory, MDR-TB, endemic, transmission-oriented, explanatory, predictive, risk model as follows: $disc(A) = \min_{\varepsilon \in \{1,-1\}^n} \|A_\varepsilon\|_\infty$.

This equation would correspond exactly to the notion of discrepancy in the risk model outputs. On the other hand, an MDR-TB experimenter could also write $disc(A) = 2 \max_{x \in \{0,1\}^n} \|Ax - Ae/2\|$, where e denotes the vector. Thus, $disc(A)$ would correspond to discrepancy variables in a robust, linearized, explanatory, time-series predictive, epidemiological, MDR-TB model especially when an experimenter is trying to round a particular ($e/2$) fractional solution. The remarkable result of LSV is that the definition seems to be much weaker than *lindisc* which has to round all fractional solutions in order that there is a natural extension, (i.e., discrepancy,) of *disc* but the derivatives are much stronger. As such, for a matrix $A \in \mathbb{R}^{m \times n}$ and $S \subseteq [n]$, for robust, explanatory, time- series, MDR-TB, risk-related modeling, an experimenter could let A_S denote the $\mathbb{R}^{m \times |S|}$ sub-matrix corresponding to the columns of A indexed by S . Then, by defining $(A) = \max_S disc(A_s)$ the time-series, explanatory, clinical-sampled, parameter estimators may be parsimoniously quantiated. Discrepancy is a natural and more "robust" version of discrepancy (Cressie 1993).

In a time series, MDR-TB, epidemiological, endemic, transmission-oriented, risk model 'discrepancy can be small, but it may carry more structural information than other residually forecasted derivatives. For example, if an MDR-TB experimenter lets A be a random $1, -1$ matrix in a time-series, explanatory, endemic, transmission-oriented, epidemiological, risk model with the constraint that each row has sum 0 . Then, $disc(A) = 0$, but $disc(A) = \Omega(\sqrt{n})$ will make intuitive sense in the risk model as random matrices may be expected to have little structure in the predictive residualized covaraite coefficinet estimates. It is also worth noting that several notable results in discrepancy theory have been bound by issues of discrepancy. For example, Spencer's original proof as well as Gluskin's argument revealed that for all matrices $A \in [-1,1]^{n \times n}$, $disc(A) = O(\sqrt{n})$. LSV may then show the following connection between linearized, time-series, explanatory, MDR-TB model discrepancies. For any matrix A , $lindisc(A) \leq herdisc(A)$ (i.e., heterogenous disc A). [eqn 1.2] (Cressie 1993) . In other words, any fractional solution for a linear program of the form

Equation 1.2 can be rounded to an integer solution with an additive error of at most $herdisc(A)$ in a robust explanatory, endemic, transmission-oriented, predictive, epidemiological, MDR-TB risk model.

Suppose a experimeneter has a fractional solution $y \in [0,1]^n$ in the MDR-TB model time -series parameter estimators statiatical sinificance tests. The goal of the experimenter may then be to find an $x \in \{0,1\}^n$ in the epidemiological, risk model such that $\|Ay - Ax\|_\infty$ is small. He or she could then construct the risk model in such a mannerism that x may be parsimoniously quantitated by iteratively making y more integral. The experimenter could then write the binary expansion of each of the sampled clinical/environmental parameter estimator coordinates of y : for each $i \in [n], y_i = 0.y_{i1}y_{i2}\cdots$. To avoid unnecessary technical issues, the experimeneter may further suppose that each coordinate has a finite expansion of length m. An MDR-TB experimenter could then robustly build a sequence of solutions $y^\upsilon = y, y^1, \dots, y^m = x$ in the epidemiological risk model such that the coordinates of y^i when written in binary will have expansions of length at most $m - i$. Thereafter, by deriving y^1 ; the time s-series explanatory clinical MDR-TB parameter estimators will be efficiently quantitated such that all the estimator significance levels will be rendered accurately. Then, if an experimenter lets $S = \{i \in [n] : y_{im} = 1\}$ by defining *herdisc* in the MDR-TB epidemiological risk model, there would exist a vector $\varepsilon \in \{1, -1\}^S$ such that $\|A_S \varepsilon_S\|_\infty \leq herdisc(A)$. Therefater, by letting $y^1 = y + \varepsilon_S / 2^m$ in the model outputs, interpreting ε as a vector in $\{1, -1, 0\}^n$ may be conducted in a natural mannerism. Clearly, the binary expansions of the coordinates of y^1 then would have length at most $m - 1$ in the time series, explanatory MDR-TB-related, risk model residual forecasted derivatives. Further, $\|Ay - Ay^1\|_\infty = \|A_S \varepsilon_S\|_\infty / 2^m \leq herdisc(A) / 2^m$ in the regressed derivatives. Iterating this argument, a MDR-TB experimenter would then derive $y^1, \dots, y^m = x$ such that $\|Ay^i - Ay^{i+1}\|_\infty \leq herdisc(A) / 2^{m-i}$. Therefore,

$$\|Ay - Ax\| \leq \sum_{i=0}^{m-1} \frac{herdisc(A)}{2^{m-i}} \leq herdisc(A). \quad [1.2]$$

As appealing as equation 1.2 for timely quantitating spatiotemporal -sampled MDR-TB clinical paramter estimators is, the method comes with one important caveat. The original motivation for examining linearized programming relaxations was that an MDR-TB experimeneter could solve them efficiently in a time-series, explanatory, predictive, epidemiologica,l risk model. For this to make sense, however, an experimenter would need to perform the rounding procedure efficiently . Optimally, to make the rounding efficient, in a time -series,

explanatory, clinically-related, endemic, trasnmission-oriented, MDR-TB risk model, an experimenter would need to find a small discrepancy solution efficiently (e.g., find ε_S given A_S). Unfortunately, this in general is very difficult as recently shown by Charkiar, Newman and Nikolov. Furthermore, Spencer's original proof as well as Gluskin's proof do not give an efficient algorithm for finding a good coloring in an edge algorithm for efficently quantitating time-series MDR-TB explanatory covariate coefficients in order to derive robust endemic transmission-oriented clinical parameter estimator significance threshold levels.

For vectors $a^1, \dots, a^n \in \{1, -1\}^n$, there exists $\epsilon \in \{1, -1\}^n$ such that for every $j \in [n]$, $\left|\langle a^j, \epsilon \rangle\right| = O(\sqrt{n})$ (Theorem 1.1), Gluskin's proof follows the partial coloring approach with the crucial lemma proved using a volume argument. The partial coloring method was introduced by Beck in 1981 and all proofs of Theorem 1.1 and many other important discrepancy results in fact use this method. Here, instead of looking for a $\epsilon \sim \{1, -1\}^n$ solution as in the theorem, an MDR-TB experimenter would search for a $\{1, 0, -1\}^n$ solution first in the risk model. The main idea is to instead look for a $\{1, 0, -1\}^n$ solution which has $\Omega(n)$ support. The experimenter would then recurse on the set of clincal sampled MDR-TB clinical and/or environmental coordinates which are set to 0. If everything goes optimally therefater, the geometric decrease would reveal an the ambient dimension in the forecasted derivatives. By so doing, geometrically decreasing discrepancy bounds may be tolerated in the derivatives. This is fundamentally inherent based on one important argument. This is the Minkowski's theorem employing the pigeon-hole principle with exponentially has many "holes" which is rather non-algorithmic in nature.

Importantly, For vectors $a^1, \dots, a^n \in \{1, -1\}^n$, there exists $\epsilon \in \{1, 0, -1\}^n$ such that for every $j \in [n]$, $\left|\langle a^j, \epsilon \rangle\right| = O(\sqrt{n})$ and $|Support(\epsilon)| = \Omega(n)$. To prove the above partial-coloring-lemma, a MDRE-TB experimenter would have to rephrase the problem employing geometric risk modeling language. For example, he or she may choose to let $\mathcal{K} \subseteq R^n$ be the symmetric convex set (symmetric meaning $x \in \mathcal{K}$ implies $-x \in \mathcal{K}$) which may be then defined in a robust MDR-TB forecasting risk model as follows for which may then be re-written as : $\mathcal{K} = \{x : \langle a^j, x \rangle \leq \Delta, \forall j \in [n]\}$. Optimally, the experimenter would want to show that $\mathcal{K}$ contains a $\{1, 0, -1\}^n$ lattice point of large support. This derivation may be qunatiated indirectly by proving that $\mathcal{K}$ instead contains a lot of clinical/environmentral sample points from $\{1, 0, -1\}^n$. Gluskin would do this by a clever volume argument for constructing a MDR-TB risk model by first

showing that the volume of $\mathcal{K} \cap [-1,1]^n$ was large and then applying Minkowski's theorem to show that there are many lattice points. To lower bound the first volume, Gluskin actually would work in the Gaussian space. But if a MDR-TB experimenter looks at a predictive, risk-related, forecasting, endemic, transmission-oriented analysis, a clear advantage would be that the projections would behave better in the Gaussian space. For example, if an MDR-TB experimenter employes a set like $S = \left\{ x \in R^n : |x_1| \le 1 \right\}$, then the Lebsgue mesasure of S would be infinite in the residual model outputs. If the experimenter projects along the first sampled clinical coordinates the outputs would become finite.

Lebesgue integration is a mathematical construction that extends the integral to a larger class of functions which may extend the domains on which a series of MDR-TB-related explanatory functions could be defined. Jacob et al. (2013) proved for non-negative MDR-TB functions with a smooth enough output graph such as continuous quantitated process, the Lebesgue integral plays an important role, since the risk model robustness would involve a pivotal portion of the axiomatic theory of probability functions on closed bounded intervals. This according to the authors was the area under the curve on the graph which could be defined as the integral and computed using Euclidean techniques of approximation of a study region by cartographically displaying time-series SAS/GIS constructed polygons.

In Kolmogorov's probability theory, the probability P of some event E, denoted $P(E)$, is usually defined such that P satisfies the Kolmogorov axioms. These assumptions can be summarized in a MDR-TB epidemiological forecasting risk model by letting (Ω, F, P) be a measure space with $P(\Omega) = 1$. Then (Ω, F, P) would be a probability space, with sample space Ω, event space F and probability measure P. However, as the need to consider more irregular, time-series ,MDR-TB functions arise it may become clear that more careful approximation techniques would be needed to define a suitable integral for the risk model. Also, an experimeneter might wish to integrate on spaces more general than the real line; the Lebesgue integral may provide the right abstractions needed to do this in a robust MDR-TB time-series explanatory risk model.In probability theory, the Donsker's theorem (i.e., central limit theorem) identifies and quantitates time-series explanatory stochastic processes. Employing the Laplace distribution, the regressed endemic, transmission-oriented, MDR-TB-related, explanatory, covariate coefficients would then involve the inverse of the kurtosis of the predictors. By the classical central limit theorem for fixed x, however, the random variable $Gn(x)$ (e.g., explanatory-sampled, georeferenced, operationalizable, field and/or environmental specified, MDR-TB, predictor variable) would converge in distribution to Gaussian (i.e., normal) randmomized variable when $G(x)$ with zero mean and variance $F(x)(1-F(x))$ as the sample size n grows

Suppose an MDR-TB experimeneter wants to employ a standardized Gaussian vector $g \sim \mathcal{N}(0,1)^n$. Then, for any unit vector $v \in R^n, \langle v, g \rangle$ has the standard normal distribution. Now, suppose the experimenter has several risk-based, explanatory, endemic, transmission-oriented, unit vectors $v_1, \ldots, v_m \in R^n$. Then, the clinical/environmental random variables $X_i = \langle v_i, g \rangle$ would be individually standard normals, but would be correlated with one another. Sidak's lemma (1967) states that no matter what the sampled correlations of X_i 's are, to bound the probability that none of the 's is too large. X_i The "worst-behaviour" an MDR-TB experimenter could thus expect for the sampled clinical/environmental data is to be independent. Concretely: Let $v_1, \ldots, v_m \in R^n$ and let $g \sim \mathcal{N}(0,1)^n$ be a standard Gaussian vector in a MDR-TB epidemiological risk model. Then, for all $t_1, \ldots, t_m \in R_+$. In the epidemiological risk model a experimenter may derive $C_i = \left\{ x \in R^n : |\langle v_i, x \rangle| \le t_i \right\}$.

Then, Sidak's lemma says that for $g \sim \mathcal{N}(0,1)^n$,

$$\Pr[g \in C_1 \wedge C_2 \wedge \cdots \wedge C_m]$$
$$\ge \Pr[g \in C_1].\Pr[g \in C_2]\ldots\Pr[g \in C_m].$$

The correlation conjecture asserts that this inequality is in fact true for all symmetric convex sets (in fact, an MDR-TB experimenter would only need to look at $m = 2$ in the forecasting model). Sidak's lemma says the conjecture is true for slabs. It is also known to be true for ellipsoids. The statement for ellipsoids also has a discrepancy implication leading to a vector generalization of Spencer's theorem as pointed out by Krzysztof Oleszkiewicz.

The second inequality required for constructing a robust endemic , transmission-oriented explanatory MDR-TB risk model is a comparison inequality due to Kanter. The lemma essentially would let an MDR-TB experimenter quantitate certain relations between any two sampled related distributions p, q to their product distributions p^n, q^n. Thereafter, the experimenter could employ the notion of peakedness of the clincal distributions. For instance, p, q may be two symmetric MDR-TB related distributions on R^m for some m. Thus, p may be less peaked than q (written $p \preceq q$) if for all symmetric MDR-TB related convex sets $\mathcal{K}$, $p(\mathcal{K}) \le q(\mathcal{K})$. Intuitively, this means that p is putting less of its mass near the origin than q (hence the term less peaked). For example, $\mathcal{N}(0,2) \preceq \mathcal{N}(0,1)$. Therefater, according to Kanter's lemma, the peakedness relation tensorises in the MDR-TB data may be parsimoniously quantiated provided unimodality exists. A univariate distribution is unimodal if the corresponding probability density function has a single maximum and no other local maxima (Cressie 1993). Then, by letting p, q be two symmetric MDR-TB distributions on R^n such that $p \preceq q$ and by letting μ be a unimodal distribution on R^m, the product distributions $p \times \mu$, $q \times \mu$ on $R^{n \times m}$ would satisfy $p \times \mu \preceq q \times \mu$. The proof of this paramter estimation technique is not too hard,

but is non-trivial in that it uses the Brunn-Minkowski's inequality. Combining the technique with the not-too-hard fact that the standard Gaussian distribution is less peaked than the uniform MDR-TB endemic transmission-oriented distribution on $[-1,1]$, μ would be the uniform distribution on $[-1,1]$. Then, $\mathcal{N}(0,1)^n \preceq \mu^n$.

In mathematics, Minkowski's theorem is the statement that any convex set in $\mathbb{R}^n$ which is symmetric with respect to the origin and with volume greater than $2^n \, \mathrm{d}(L)$ contains a non-zero lattice point. Suppose that L is a lattice of determinant $\mathrm{d}(L)$ in the n-dimensional real vector space $\mathbb{R}^n$ in a spatiotemporal MDR-TB epidemiological, endemic, transmission-oriented, predictive, risk model and S is a convex subset of $\mathbb{R}^n$ that is symmetric with respect to the origin. Then, if x is in S in the risk model then $-x$ would also be S. Minkowski's theorem states that if the volume of S is strictly greater than $2^n \, \mathrm{d}(L)$, then S must contain at least one lattice point other than the origin(Cressie 1993). The following argument may then prove Minkowski's theorem for the special case of $L = \mathbb{Z}^2$ in a robust, explanatory, MDR-TB, epidemiological, endemic, trasnmission-oriented, risk model. Thereafter, the clinically residually forecasted derivatives could be generalized to arbitrary lattices in arbitrary dimensions.

Consider a time-series, explanatory, time-series, predictive, epidemiological, MDR-TB risk map $f : S \to \mathbb{R}^2, (x,y) \mapsto (x \bmod 2, y \bmod 2)$. Intuitively, this map could cut the plane into 2 by 2 squares. then the squares would be stacked on top of each other. Clearly $f(S)$ would have an area ≤ 4. Suppose f were injective in the MDR-TB epidemiological risk model. Then, the pieces of S cut out by the squares would stack up in a non-overlapping way. Since f is locally area-preserving(Cressie 1993), this non-overlapping property in the risk model would render forecasted derivatives for all of S area-preserving, so the area of $f(S)$ would be the same as that of S, which would then be greater than 4. This not being the optimal case, f would not be injective, and $f(p_1) = f(p_2)$ would be rendered during the analyses of the sampled clinical MDR-TB explanatory clinical/environmental data points p_1, p_2 in S. Moreover, the definition of f that $p_2 = p_1 + (2i, 2j)$ for some sampled integers i and j, where i and j could not be both zero. Then, since S is symmetric about the origin, $-p_1$ would also be a point in S in the risk model. Since S would be convex, the line segment between $-p_1$ and p_2 would lay entirely in S, and in particular the midpoint of that segment would lie in S. In other words, $\frac{1}{2}(-p_1 + p_2) = \frac{1}{2}(-p_1 + p_1 + (2i, 2j)) = (i, j)$ would lay in S. (i,j) which would be a lattice point in the MDR-TB epidemiological risk model and not the origin since i and j could not both be zero.

Nikhil's argument studies a carefully constructed semi-definite programming relaxation of the problem and then gives a new and amazing rounding algorithm for the SDP. For vectors $a^1, \ldots, a^n \in [1,-1]^n$, thyen there would exist $\varepsilon \in \{1, 0, -1\}^n$ such that for every $j \in [n]$, $\left| \langle a^j, \epsilon \rangle \right| = O(\sqrt{n})$ and $\left| Support(\epsilon) \right| = \Omega(n)$. If an MDR-TB experimenter rephrases the problem in geometric language for fitting a robus,t explanatory, spatiotemporal, predictive, endemic, transmission-oriented, risk model and then lets $\mathcal{K} \subseteq \mathbb{R}^n$ be the symmetric convex set, the forecasted derivatives would reveal $\Delta = O(\sqrt{n})$. Additionally, in the MDR-TB risk model symmetrical output $x \in \mathcal{K}$ would then imply $-x \in \mathcal{K}$.

Interestingly, the partial coloring lemma may be equivalent to determining whether $\mathcal{K}$ contains a $\{1, 0, -1\}^n$ lattice point in a robust, explanatory, MDR-TB, predictive, epidemiological, risk model. As it turns out, however, an experimenter would need to find a lattice point in $\mathcal{K}$ in the predictive risk model in order to adequately perform robust parameter estimator significance testing. Any MDR-TB clinical/environmental sampled point with many $\{1, -1\}$ (or close to $\{1, -1\}$) coordinates will serve equally well. Concretely, a experimeneter would attempt to quantitate $\mathcal{K}$ as to determine if there exists $x \in \mathcal{K}$ such that $\left| \{i : |x_i| = 1\} \right| = \Omega(n)$. Thus, derivation of a robust, MDR-TB, endemic, transmission-oriented, explanatory, risk model output would then be equivalent to finding a vertex of $\mathcal{K}$ which would then be tight on $\Omega(n)$ coloring constraints. For intuition, a MDR-TB experimenter could use the distance from the origin as a proxy for efficiently quantitating how many clinical/environmental coordinates are close to 1 in the absolute value in the risk model explanatory output. Thus, the goal for robust MDR-TB predictive risk modeling may be to find a vertex of $\mathcal{K}$ as far away from the origin as possible in order to parsimoniously quantitate the sampled clincial explanatory covariates based on their statistcal significance levels. The starting point optimally would be the all-zeros vector in the empirical sampled dataset. The x in the MDR-TB epidemiological, endemic, transmission-oriented, explanatory, predictive, risk model may be then updated by employing Brownian motion.

The GBM (i.e., exponential Brownian motion) is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion. Brownian motion is the presumably random drifting in a mathematical model used to describe random movements (Cressie 1993). In 1956 Skorokhod and Kolmogorov defined a separable metric d, called the *Skorokhod metric*, on the space of cadlag functions on [0,1], such that convergence for d to a continuous function is equivalent to convergence for the sup norm, and showed that G_n converges in law in $\mathcal{D}[0,1]$ to the Brownian bridge. Let (M, d) be a metric space, and let $E \subseteq \mathbb{R}$. A function $f : E \to M$ is called a càdlàg function if, for every $t \in E$, the left limit $f(t-) := \lim_{s \uparrow t} f(s)$ exists;

andthe right limit $f(t+) := \lim_{s\downarrow t} f(s)$ exists and equals $f(t)$. That is, f is right-continuous with left limit. In mathematics, a càdlàg (French "continue à droite, limite à gauche"), RCLL ("right continuous with left limits"), or corlol ("continuous on (the) right, limit on (the) left") function is a function defined on the real numbers (or a subset of them) that is everywhere right-continuous and has left limits everywhere(Hazwewinkle 2002).. Càdlàg functions are important in the study of stochastic processes that admit (or even require) jumps, unlike Brownian motion, which has continuous sample paths (Cressie 1993). The collection of càdlàg functions on a given domain is known as Skorokhod space.

The set of all càdlàg functions from E to M is often denoted by $D(E;M)$ (or simply D) and is called Skorokhod space. Skorokhod space can be assigned a topology that, intuitively allows us to "wiggle space and time a bit" (whereas the traditional topology of uniform convergence only allows us to "wiggle space a bit"). For simplicity, take $E = [0,T]$ and $M = \mathbb{R}^n$ — see Billingsley 1999 for a more general construction. We must first define an analogue of the modulus of continuity, $\varpi'_f(\delta)$. For any $F \subseteq E$, set $w_f(F) := \sup_{s,t \in F} |f(s) - f(t)|$ and, for $\delta > 0$, define the càdlàg modulus to be

$$\varpi'_f(\delta) := \inf_{\Pi} \max_{1 \leq i \leq k} w_f([t_{i-1}, t_i]),$$

where the infimum runs over all partitions $\Pi = \{0 = t_0 < t_1 < \ldots < t_k = T\}$, $k \in \mathbb{N}$, with $\min_i(t_i - t_{i-1}) > \delta$. This definition makes sense for non-càdlàg f (just as the usual modulus of continuity makes sense for discontinuous functions) in a MDR-TB risk model and thus it can be shown that f is càdlàg if and only if $\varpi'_f(\delta) \to 0$ as $\delta \to 0$. Then, if a experimenter lets Λ denote the set of all strictly increasing, continuous bijections from E to itself (these are "wiggles in time"). Then by letting $\|f\| = \sup_{t \in E} |f(t)|$ in the risk model , robust uniform norms may be denoted on functions on E. The experimenter can then define the Skorokhod metric σ on D by

$$\delta(f,g) := \inf_{\lambda \in \wedge} \max \{\|\lambda - I\|, \|f - g o \lambda\|\},$$

where $I : E \to E$ is the identity function. In terms of the "wiggle" intuition, $\|\lambda - I\|$ would measure the n MDR-TB size of the "wiggle in time", and $\|f - g\bigcirc\lambda\|$ measures the size of the "wiggle in space". It can be shown that the Skorokhod metric is indeed a metric. The topology Σ generated by σ is called the Skorokhod topology on D (Hazewinkle 2002).

By an application of the Arzelà–Ascoli theorem, one can show that a sequence $(\mu_n)_{n=1,2,\ldots}$ of probability measures on Skorokhod space D is tight if and only if both the following conditions are met:

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(\{f \in D \,\|\,\|f\| \geq a\}) = 0, \quad \text{and}$$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(\{f \in D \,|\, \varpi'_f \geq \varepsilon\}) = 0 \ \text{ for all } \varepsilon > 0.$$

sequence $\{f_n\}_{n \in \mathbb{N}}$ of continuous functions on an interval $I = [a,b]$ is uniformly bounded if there is a number M such that $|f_n(x)| \leq M$ for every function f_n belonging to the sequence, and every $x \in [a,b]$. The sequence is *equicontinuous* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|f_n(x) - f_n(Y)| < \varepsilon$ whenever $|x - y| < \delta$ for all functions f_n in the sequence. Succinctly, a sequence is equicontinuous if and only if all of its elements admit *the same* modulus of continuity. In simplest terms, the theorem can be stated as follows: Consider a sequence of real-valued continuous functions $\{f_n\}_{n \in \mathbb{N}}$ defined on a closed and bounded interval [a, b] of the real line. If this sequence is uniformly bounded and equicontinuous, then there exists a subsequence (f_{nk}) that converges uniformly.The converse is also true, in the sense that if every subsequence of $\{f_n\}$ itself has a uniformly convergent subsequence, then $\{f_n\}$ is uniformly bounded and equicontinuous based on the a proof Let $I = [a,b] \subset \mathbb{R}$ be a closed and bounded interval. If $\mathbf{F}$ is an infinite set of functions $f : I \to \mathbb{R}$ which is uniformly bounded and equicontinuous, then there is a sequence f_n of elements of $\mathbf{F}$ such that f_n converges uniformly on I.

Importantly, by fixing an enumeration $\{x_i\}_{i \in \mathbb{N}}$ of rational MDR-TB clinical/environmental explanatory , endemic, trasnmission-oriented, covariate values in I, F would be uniformly bounded, by the set of points $\{f(x_1)\}_{f \in \mathbb{F}}$ which would alsoe be bounded, based on the Bolzano-Weierstrass theorem. In mathematics, specifically in real analysis, the Bolzano–Weierstrass theorem, is a fundamental result about convergence in a finite-dimensional Euclidean space $\mathbb{R}^n$. The theorem states that each bounded sequence in $\mathbb{R}^n$ has a convergent subsequence. An equivalent formulation is that a subset of $\mathbb{R}^n$ is sequentially compact if and only if it is closed and bounded there is a sequence $\{f_{n1}\}$ of distinct functions in F such that $\{f_{n1}(x_1)\}$ converges. Repeating the same argument for the sequence of sampled MDR-TB epidemiological sampled points $\{f_{n1}(x_2)\}$, there would be a subsequence $\{f_{n2}\}$ of $\{f_{n1}\}$ such that $\{f_{n2}(x_2)\}$ converges.By induction this process can be continued forever, and so there is a chain of subsequences $\{f_{n1}\} \supseteq \{f_{n2}\} \supseteq \cdots$ such that, for each $k = 1, 2, 3, \ldots$, the subsequence $\{f_{nk}\}$ would converge at $x_1, \ldots, x_k$. Now if an MDR-TB experimenter forms the diagonal subsequence $\{f\}$ in the risk model whose mth term f_m is the mth term in the mth subsequence $\{f_{nm}\}$ then, f_m would converge at every rational point of I. Therefore,

given any $\varepsilon > 0$ and rational x_k in I, there would be an integer $N = N(\varepsilon, x_k)$ such that $|f_n(x) - f_m(x)| < \dfrac{\varepsilon}{3}, n, m \geq N$. Since the family F would be equicontinuous, for this fixed ε and for every x in I in the MDR-TB risk model, there would be an open interval U_x containing x such that $|f(s) - f(t)| < \dfrac{\varepsilon}{3}$ for all $f \in$ F and all s, t in I such that $s, t \in U_x$. The collection of intervals $U_x, x \in I$ would then form an open cover of I. Since I is compact, this covering would then admit a finite subcover $U_1, ..., U_J$. Then there would exist an integer K such that each open interval U_j, $1 \leq j \leq J$ would contains a rational x_k with $1 \leq k \leq K$. Finally, for any $t \in I$, there are j and k so that t and x_k belong to the same interval U_j. For this choice of k,

$$|f_n(t) - f_m(t)| \leq \begin{bmatrix} |f_n(t) - f_n(x_k)| \\ + |f_n(x_k) - f_m(x_k)| \\ + |f_m(x_k) - f_m(t)| \end{bmatrix} < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \quad \text{for}$$

all $n, m > N = \max\{N(\varepsilon, x_1), ..., N(\varepsilon, x_K)\}$.

Consequently, the MDR-TB sequence $\{f_n\}$ would be uniformly Cauchy, and therefore wouold converge to a continuous function.

The definitions of boundedness and equicontinuity can also be generalized in a MDR-TB epidemiological risk model to the setting of arbitrary compact metric spaces and, more generally still, compact Hausdorff spaces. Let X be a compact Hausdorff space, and let $C(X)$ be the space of real-valued continuous functions on X. A subset $\mathbf{F} \subset C(X)$ is said to be equicontinuous if for every $x \in X$ and every $\varepsilon > 0$, x has a neighborhood U_x such that $\forall y \in U_x, \forall f \in F : |f(y) - f(x)| < \varepsilon$ (Hazewinkle 2002). A set $\mathbf{F} \subset C(X, \mathbf{R})$ is said to be pointwise bounded if for every $x \in X$, $\sup\{|f(x)| : f \in F\} < \infty$. A version of this holds also in the space $C(X)$ of real-valued continuous functions on a compact Hausdorff space X (Dunford and Schwartz 1958):Let X be a compact Hausdorff space. Thus, a subset $\mathbf{F}$ of $C(X)$ in a MDR-TB risk model is relatively compact in the topology induced by the uniform norm if and only if it is equicontinuous and point wise bounded. The Arzelà–Ascoli theorem would thus be a fundamental result in the study of the algebra of continuous functions on a compact Hausdorff space.

Various generalizations may be then constructed employing MDR-TB risk models. For instance, functions can assume clinical covariate coefficient values in a metric space or (Hausdorff) topological vector space with only minimal changes to the statement (Kelley and Namioka 1982) Kelley (1991)]:Let X be a compact Hausdorff space and Y a metric space. Then $\mathbf{F} \subset C(X, Y)$ is compact in the compact-open topology if and only if it is equicontinuous, pointwise relatively compact and closed.Here the pointwise relatively compact means that

for each $x \in X$, the set $\mathbf{F}_x = \{f(x) : f \in \mathbf{F}\}$ is relatively compact in Y.The proof given can be generalized in a way that does not rely on the separability of the domain. On a compact Hausdorff space X, in a MDR-TB risk model, for instance, the equicontinuity could be used to extract, for each $\varepsilon = 1/n$, a finite open covering of X such that the oscillation of any function in the family is less than ε on each open set in the cover. The role of the rationals can then be played by a set of sampled clinical points drawn from each open set in each of the countably many covers obtained in this way, and the main part of the proof proceeds exactly as above.

Whereas most formulations of the Arzelà–Ascoli theorem assert sufficient conditions for a family of functions to be (relatively) compact in some topology, these conditions are typically also necessary. For instance, if a set $\mathbf{F}$ is compact in $C(X)$, the Banach space of real-valued continuous functions on a compact Hausdorff space with respect to its uniform norm, then it is bounded in the uniform norm on $C(X)$ and in particular is pointwise bounded. Let $N(\varepsilon, U)$ be the set of all functions in $\mathbf{F}$ whose oscillation over an open subset $U \subset X$ is less than $\varepsilon : N(\varepsilon, U) = \{f | osc_U f < \varepsilon\}$. For a fixed $x \in X$ and ε, the sets $N(\varepsilon, U)$ form an open covering of $\mathbf{F}$ as U varies over all open neighborhoods of x. Choosing a finite subcover then gives equicontinuity.To every function g that is p-integrable on [0, 1], with $1 < p \leq \infty$, associate the function G defined on [0, 1] by $G(x) = \int_0^x g(x)dt$. Let $\mathbf{F}$ be the set of functions G corresponding to functions g in the unit ball of the space $L^p([0, 1])$. If q is the Hölder conjugate of p, defined by $1/p + 1/q = 1$, then Hölder's inequality implies that all functions in $\mathbf{F}$ satisfy a Hölder condition with $\alpha = 1/q$ and constant $M = 1$.It follows that $\mathbf{F}$ is compact in $C([0, 1])$. This means that the correspondence $g \to G$ defines a compact linear operator T between the Banach spaces $L^p([0, 1])$ and $C([0, 1])$. Composing with the injection of $C([0, 1])$ into $L^p([0, 1])$, one sees that T acts compactly from $L^p([0, 1])$ to itself. The case $p = 2$ can be seen as a simple instance of the fact that the injection from the Sobolev space $H_0^1(\Omega)$ into $L^2(\Omega)$, for Ω a bounded open set in $\mathbf{R}^d$, is compact.When T is a compact linear operator from a Banach space X to a Banach space Y, its transpose T^* is compact from the (continuous) dual Y^* to X^*. This can be checked by the Arzelà–Ascoli theorem.Indeed, the image $T(B)$ of the closed unit ball B of X is contained in a compact subset K of Y. The unit ball B^* of Y^* defines, by restricting from Y to K, a set $\mathbf{F}$ of (linear) continuous functions on K that is bounded and equicontinuous. By Arzelà–Ascoli, for every sequence $\{y*n\}$, in B^*, there is a subsequence that converges uniformly on K, and this implies that the image $T^*(y_{n_k}^*)$ of that subsequence is Cauchy in X^*.

When f is holomorphic in an open disk $D_1 = B(z_0, r)$ in a MDR-TB model with modulus bounded by M, then (for example by Cauchy's formula) its derivative f' would have a modulus bounded by $4M/r$ in the smaller disk $D_2 = B(z_0, r/2)$. If a family of holomorphic functions on D_1 is bounded by M on D_1, it follows that the family F of restrictions to D_2 is equicontinuous on D_2. Therefore, a sequence converging uniformly on D_2 can be extracted. This is a first step in the direction of Montel's theorem. In complex analysis, an area of mathematics, Montel's theorem refers to one of two theorems about families of holomorphic functions which give conditions under which a family of holomorphic functions is normal.Arzelà–Ascoli theorem is a fundamental result of mathematical analysis giving necessary and sufficient conditions to decide whether every sequence of a given family of real-valued continuous functions defined on a closed and bounded interval has a uniformly convergent subsequence. The main condition is the equicontinuity of the family of functions. The theorem is the basis of many proofs in mathematics, including that of the, Montel's theorem in complex analysis.

Kolmogorov (1933) showed that when F is continuous, the supremum $\sup_t G_n(t)$ and supremum of absolute value, $\sup_t |G_n(t)|$ converges in distribution to the laws of the same functionals of the Brownian bridge $B(t)$, s.In 1952 Donsker stated and proved employing a general extension for the Doob-Kolmogorov heuristic approach. Donsker proved that the convergence in law of G_n to the Brownian bridge holds for Uniform [0,1] distributions (e.g. regressively quantiated explanatory clinically oriented MDR-TB variables) with respect to uniform convergence in t over the interval [0,1]. By the classical central limit theorem, for fixed x, the random variable $G_n(x)$ converges in distribution to a Gaussian (normal) random variable $G(x)$ with zero mean and variance $F(x)(1 - F(x))$ as the sample size n grows (Hazewinkle 2002). HDonsker's formulation may not be useful for constructing a robust MDR-TB epidemiological risk model because of the problem of measurability of the functionals of discontinuous processes. However, in probability theory, Donsker's theorem identifies a certain stochastic process as a limit of empirical processes. It is sometimes called the functional central limit theorem. A centered and scaled version of empirical distribution function F_n in a time-series, explanatory, epidemiological, forecasting, risk-related, MDR-TB model would then define an empirical process employing $G_n(x) = \sqrt{n}(F_n(x) - F(x))$ indexed by $x \in \mathbf{R}$. The sequence of $G_n(x)$ in the risk model forecasted derivatives would bethen random elements of the Skorokhod space $\mathcal{D}(-\infty, \infty)$ which could then converge in distribution to a Gaussian process G with zero mean .Brownian motion is among the simplest of the continuous-time stochastic (or probabilistic) processes, and it is a limit of both simpler and more complicated stochastic processes (e.g., random walk and Donsker's theorem).

Random walk may be defined in a time series explanatory MDR-TB epidemiological predictive risk model formally by taking independent sampled clinical sampled randomized variables $Z_1, Z_2, \ldots$, where each variable is either 1 or -1, with a 50% probability for either value, set to $S_0 = 0$ and $S_n = \sum_{j=1}^{n} Z_j$. The MDR-TB related time series $\{S_n\}$ would then be simple random walk on $\mathbb{Z}$. This series (i.e., the sum of the sequence of -1s and 1s) would render the distance measurements in the empirical sampled dataset of clinical regressors, if each part of the walk is of length one in the model. The expectation $E(S_n)$ of S_n then would be zero. That is, as the mean of all the sampled clinical data approaches zero, the number of clinical MDR-TB regression paramter estimated values would increase. This follows by the finite additivity property of expectation $E(S_n) = \sum_{j=1}^{n} E(Z_j) = 0$. A similar calculation, employing the independence of the sampled MDR-TB time series randomized variables and the fact that $E(Z_n^2) = 1$, would then reveal that $E(S_n^2) = \sum_{j=1}^{n} E(Z_j^2) + \sum_{i=1}^{n} \sum_{j=1}^{n} 2E(Z_j Z_i) = n$. This would hint that $E(|S_n|)$ is the expected translation distance after n steps, based on the order of $\sqrt{n}$. In fact the risk model forecasted derivatives may be qunatiated using $\lim_{n \to \infty} \frac{E(|S_n|)}{\sqrt{n}} = \sqrt{\frac{2}{\pi}}$. This result may reveal that statsitcal diffusion is ineffective for mixing MDR-TB sampled time-series explanatory clinical paramter estimators because of the way the square root behaves for large N in the model. Thereafter, if a MDR-TB experimenter tet $\mathcal{D} = \mathcal{D}[0,1]$ be the space of real-valued clincal functions x on $[0,1]$ in the risk model, all right-continuous varaibles that and have left-hand limits may be expressed employing $x(t+) = \lim_{s \downarrow t} x(s)$ exists,

$x(t+) = x(t)$ for all $0 \leq t < 1$,

$x(t-) = \lim_{s \uparrow t} x(s)$ exists for all $0 < t \leq 1$. In probabilistic literature, such a function is also said to be a cadlag function (Cressie 1993). Introducing a norm on $\mathcal{D}$ by setting $\|x\| = \sup_{0 \leq t \leq 1} |x(t)|$, then $\mathcal{D}$ in the MDR-TB epidemiological predictive risk model would become a Banach space.

A Banach space is a vector space with a metric that allows the computation of vector length and distance between vectors and is complete in the sense that a Cauchy sequence of vectors always converges to a well defined limit in the space (Cressie 1993). A sequence $x_1, x_2, x_3, \ldots$ an empirical-sampled dataset of MDR-TB related paramter estimators would be a *Cauchy* sequence, if for every positive sampled clinical environmental sampled explanatory endemic trasnmission-oriented covariate estimate value ε, there is a positive integer N

such that for all the values $m, n > N$, $|x_m - x_n| < \varepsilon$, where the vertical bars denote the absolute value (Griffith 2003). Cauchy formulated such a condition by requiring $x_m - x_n$ to be infinitesimal for every pair of infinite m, n. To define Cauchy sequences in a robust explanatory MDR-TB epidemiological forecasting risk model thenin any metric space X, the absolute value $|x_m - x_n|$ would have to replac the *distance* $d(x_m, x_n)$ (where $d : X \times X \to \mathbf{R}$ with some specific properties, between x_m and x_n. Formally, a metric space (X, d) in an empirical sampled dataset of MDR-TB -related parameter estimator time series sequence $x_1, x_2, x_3, \ldots$ is Cauchy, if for every positive clinical sampled real number $\varepsilon > 0$ there is a positive integer N such that for all positive integers $m, n > N$, the distance $d(x_m, x_n) < \varepsilon$. Roughly speaking, the terms of the sequence then would be getting closer and closer together in a way that suggests that the time series explanatory MDR-TB related sequence ought to have a limit in X. Nonetheless, such a limit may not always exist within X in the model in order to be able to efficiently render robust forecast derivatives.

A metric space X in a time –series, explanatory, predictive, epidemiological, MDR-TB risk model where every Cauchy sequence converges to an element of X would be considered complete.This non-separability in the model paramters causes well-known problems of measurability however in the theory of weak convergence of measures on the space. To overcome this inconvenience, A.V. Skorokhod introduced a metric (and topology) under which the space $\mathcal{D}$ becomes a separable metric space. Although the original metric introduced by Skorokhod would have a drawback for quantiating uncertainty estimates in a predictive MDR-TB epidemiological forecasting risk model, in the sense that the metric space obtained would not be complete, it may be possible to construct an equivalent metric (i.e., giving the same topology) in the model under which the space $\mathcal{D}$ becomes a separable and complete metric space. This metric may be defined as follows: Let Λ denote the class of strictly increasing continuous spatiotemporal MDR-TB endemic transmission-oriented, explanatory, clinically oriented, risk-related mappings of $[0,1]$ onto itself. Then,

for $\lambda \in \Lambda$, and $\|\lambda\| = \sup\limits_{0 \le s < t \le 1} \left| \log\left\{ (t-s)^{-1} \left(\lambda(t) - \lambda(s) \right) \right\} \right|$.

Furthermore, for $x, y \in \mathcal{D}$ an MDR-TB experimenter could

define $\quad d(x, y) = \inf\limits_{\lambda \in \Lambda} \max \left\{ \|\lambda\|, \sup\limits_{0 \le t \le 1} \left| x(t) - y(\lambda(t)) \right| \right\}$.

The topology generated by this metric in the endemic trasnmission-oriented risk model forecasted derivatives would reveal the Skorokhod topology where the complete separable metric space $\mathcal{D}$ would be the Skorokhod space (cf. also Skorokhod topology). This space is very important in the theory of random processes (cf. also Stochastic process).

Fortunately, the general theory of weak convergence of probability measures on metric spaces and, in particular, on the space $\mathcal{D}$ is well developed in literature.For example, Jacob et al. (2014)employed the process $G(x)$

which was written as $B(F(x))$ in the previously mentioned *S. damnosum s.l.* explanatory time series, epidemiological endemic trasnmission oriented risk model to map hyperendemic foci at a riverine study site as defined by an emepirical sampled datast of georeferenced observations of positively autocorrelated productive larval habitats where B was a standard Brownian bridge on the unit interval.In, Jacob et al. (2014) the context of a topological group of sampled georefernced riverine larval habitats was defined as a sequence (x_k) in a topological group G which was a Cauchy sequence since for every open neighbourhood U of the identity in G there existed some sampled explanatory endemic trasnmission-oriented explanatory measurement indicator value N such that whenever $m, n > N$ it followed that $x_n x_m^{-1} \in U$. The authors checked this for the spatial *S.damnosum s.l.* – related neighbourhoods employing the local base of the identity in G. Furthermore, the authors defined a binary relation based on the spatiotemporal regressed field - sampled larval habitat Cauchy sequence which then determined G such that (x_k) and (y_k) were equivalent if for every open neighbourhood U of the identity in G there existed some N such that whenever $m, n > N$;thus it followed that $x_n y_m^{-1} \in U$. This relation was an equivalence relation. It was reflexive since the time- series sampled *S.damnosum s.l.* –related sequences were Cauchy sequences. The risk model was symmetric since $y_n x_m^{-1} = x_m y_n^{-1} \in U^{-1}$ which by continuity of the inverse was parsimoniously quantitated employing another open neighbourhood of the identity. Additionally, the model was transitive since $x_n z_l^{-1} = x_n y_m^{-1} y_m z_l^{-1} \in U'U''$ where U' and U'' were open neighbourhoods of the identity such that $U'U'' \subseteq U$. In the epidemiological risk model the pairs existed by the continuity of the group operation.

Interestingly, there may be a concept of Cauchy sequence which may be defined in a group G in a time-series, explanatory, predictive, epidemiological, endemic, transmission-oriented, MDR-TB risk model. For example, suppose a MDR-TB experimeneter lets $H = (H_r)$ be a decreasing sequence of normal subgroups of G of finite index in a predictive, time-series, explanator, y risk model. Then, a sequence (x_n) in G would be Cauchy (w.r.t. H) if and only if for any r there is N such that $\forall m, n > N, x_n x_m^{-1} \in H_r$. Technically, this would be the same thing as quantitating a topological group MDR-TB-related explanatory Cauchy sequence for a particular choice of topology on G, namely that for which H is a local base. The set C of such Cauchy sequences would then form a group for the componentwise product, which would thereafter parsimoniously reflect a set C_0 of null sequences (s.th. $\forall r, \exists N, \forall n > N, x_n \in H_r$) which would be a normal subgroup of C. The factor group C / C_0 would then be the completion of G with respect to H in the model output targeting the hyperendemic trasnmission-orienetd foci in an epidemiological interventional study site. A MDR-TB experimenter could then show that this completion in the derivatives is

isomorphic to the inverse limit of the sequence (G / H_r). An example of this construction, familiar in number theory and algebraic geometry is the construction of the p-adic completion of the integers with respect to a prime *p*. In this case, *G* would be the integers under addition, and H_r would be the additive subgroup consisting of integer multiples of p^r. Therefater, if *H* is a cofinal sequence (i.e., any normal subgroup of finite MDR-TB epidemiological endemic transmission-oriented explanatory indices that contain some H_r), then this completion would be canonical in the sense that the gerivatives would be isomorphic to the inverse limit of $(G / H)_H$, where *H* varies over all normal subgroups of finite clinically-oriented indices.

Importantly, in constructive mathematics, Cauchy sequences often must be given with a modulus of Cauchy convergence to be useful. Thus, if $(x_1, x_2, x_3, \ldots)$ is a MDR-TB related time-series Cauchy sequence in the set *X*, then a modulus of Cauchy convergence for the sequence would be a function α from the set of explanatory endemic trasnmission, oriented covariate coefficient values to itself, such that $\forall k \forall m, n > \alpha(k), |x_m - x_n| < 1/k$. Clearly, any sequence with a modulus of Cauchy convergence is a Cauchy sequence (Cressie 1993). The converse (that every MDR-TB-related Cauchy sequence has a modulus) follows from the well-ordering property of the natural numbers (let $\alpha(k)$ be the smallest possible *N* in the definition of Cauchy sequence, taking *r* to be $1/k$). However, this well-ordering property may not hold in in th epidemiological risk model On the other hand, the converse in the risk model paramter estimators may follows directly from the principle of dependent choice which presently is accepted by constructive mathematicians. Thus, moduli of Cauchy convergence are needed directly only by MDR-TB experimeneters do not wish to use any form of choice. That said, using a modulus of Cauchy convergence for regressing an empirical sampled dataset of clinical MDR-tb clinical parameter estimaors can simplify both definitions and theorems in constructive analysis. Perhaps even more useful for time series epidemiological MDR-TB risk modelingh are regular Cauchy sequences, which as sequences with a given modulus of Cauchy convergence (e.g., $\alpha(k) = k$ or $\alpha(k) = 2^k$). Any Cauchy sequence with a modulus of Cauchy convergence is equivalent (in the sense used to form the completion of a metric space) to a regular Cauchy sequence; this can be proved without using any form of the axiom of choice(Griffith 2003). Regular Cauchy sequences were used by Errett Bishop in his Foundations of Constructive Analysis, but they have also been used by Douglas Bridges in a non-constructive textbook .However, Bridges also works on mathematical constructivism; the concept has not spread far outside of that milieu. Furthermore, in a hyperreal continuum , a real sequence of MDR-TB epidemiological endmeic trasnmission-orienetd explanatory clinical predictors would be reflected as a real sequence $\langle u_n : n \in \mathbb{N} \rangle$ which

then would have a natural hyperreal extension, defined for hypernatural explanatory endemic trasnmission-orieted clinical time series sampled values *H* of the index *n* in addition to the usual natural *n*. The sequence is Cauchy if and only if for every infinite *H* and *K*, the values u_H and u_K are infinitely close, or adequal, i.e. $st(u_H - u_K) = 0$ where "st" is the standard part function(Cressie 1993).

Recently there have been attempts to capture unobserved error and space-time interactions using streaming algorithms for error detection of clustering clinical and/or environmental MDR-TB georeferenced district-level explanatory predictor covariate coefficient estimates within a spatial scan matrix. For example, Oeltmann et al. (2008) assessed clustering of cases employing differentially corrected global positioning system (DGPS) technology, to seasonally map TB patient households according to drug-susceptibility testing results, for evaluating an MDR-TB outbreak among US-bound Hmong Refugees in Thailand. The empirical-sampled clinical and/or environmental data were analyzed with a spatial scan error statistic that employed a varying-sized cylinder to encapsulate cases within the radius of the cylinders which was then used to tabulate a *p*-value and log-likelihood ratio for determining the statistical significance of the MDR-TB clusters detected. In the context of cluster analyses *p*-values seasonally quantitated may be relative to the null model (see Waller and Jacquez 1995). Prevalence ratios and 95% confidence intervals for each exposure group was thereafter calculated. The cluster-error detection algorithm indicated an outbreak of MDR-TB in the population, specifically in areas in which TB rates were already elevated. The model also quantitated the extra-Poison variation in the sampled empirical datasets using a negative binomial regression with a gamma distributed non-homogenous mean. A common way to deal with overdispersion for cluster-based count data is to use a generalized linear model (GLM) framework, where the most common approach is a "quasi-likelihood," with Poisson-like assumptions (i.e., quasi-Poisson) or a negative binomial model (see Jacob et al. 2010b).

Additionally, the time-series residualized, uncertainty, diagnostic, clinically-oriented, explanatory, georeferencable, cluster-based tests provided evidence for the presence of these so-called secondary clusters, (i.e., MDR-TB georeferenced spatial clusters not overlapping with the most likely cluster but with significantly large likelihood ratio). These secondary clusters had an associated *p*-value but they were calculated ignoring the existence of the most likely cluster-based, hierarchical, explanatory, predictive, error variance in the residualized forecasts. The consequence of this in endemic mapping MDR-TB is that the *p*-values can be overly conservative in the model outputs leading to a loss in power of the explanatory observational predictor covariate error coefficient estimates. Further, the *p*-values testing a second georeferenced MDR-TB cluster would have to alternatively be calculated conditionally based on the presence of the primary cluster–based, explanatory, georeferenced, uncertainty, residualized, parameter estimators as their values could be smaller than those delivered by the predictive regression-oriented equation. This would then lead to an overall misspecified MDR-TB model.

In many instances MDR-TB researchers choose not to quantitate explanatory, residualized, hierarchical, cluster-based,georeferenced, observational, explanatory, predictor, covariate error coefficient estimates. Ignoring within cluster-based error in sampled clinical and environmental MDR-TB explanatory covariate coefficient estimates, however, would lead to sign reversals of the factor covariances, inflation of factor variances, and other misspecifications which could hamper the ability of a model to make accurate predictions of high risk populations. 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 2003). Spatial autocorrelation leads to biased standard errors and/or biased parameter estimates, as well as artificially inflated degrees of freedom in sampled hierarchical residual cluster-based predictor covariate coefficient estimates (Wakefield, 2003).

The Durbin Watson (DW) test is a well-known formal method of testing if intra-cluster serial error correlation is a serious problem undermining a model's inferential suitability (e.g., assessing the confidence in the predicted value of a dependent variable in a hierarchical residual cluster-based model). The Durbin-Watson test can be used to test the hypothesis $H_0 : \rho s = H_1 : \rho s = \rho^{(s)}$ where $\rho^{(s)}$ is any ρs such that $0 < |\rho s| <$ (Cressie, 1993). The DW tests can test for serial error autocorrelation in an endemic transmission-oriented MDR-TB model by assuming that epsilon (ε) is stationary and normally distributed with mean zero. This statistic then tests the null hypothesis H_o that the georeferenced hierarchical residual cluster-based errors are uncorrelated against the alternative hypothesis H_1 that the errors are first order autoregressive [AR(1)]. Therefore, if ρ are the autocorrelation error coefficients in a spatiotemporal hierarchical residual cluster-based endemic transmission-oriented MDR-TB model, then $w : H_o : \rho s = \rho s$ for some non-zero ρ with $|\rho| < 1$ could be used to quantitate the explanatory georeferenced predictor covariate error coefficient estimates in the sampled clinical and environmental parameters.

Recent years have seen a virtual explosion in the application of cross-sectional spatial growth regression models for TB-related predictive, spatiotemporal, epidemiological risk modeling. While undoubtedly considerable progress has been made, most applications ignore risk model uncertainty. Model uncertainty arises from two sources: (i) the spatial weight or connectivity structure assigned to regions that form the observational basis of spatial data samples (e.g., an empirical dataset of MDR-TB related predictive clinical risk model parameter estimators) , and (ii) specific explanatory variables included. The first source of time series MDR-TB epidemiological risk-related model uncertainty is unique to spatial regression predictive modeling since conventional regression models assume independence between sample observations (regions). The hallmark of spatial growth time series explanatory MDR-TB-related regression models is the spatial weight matrix that distinguishes these from non-spatial growth regressions. The specification of this matrix would be typically constructed by means of geographic criteria, such as contiguity (sharing a common border) or distance, including nearest neighbor distance. Uncertainty regarding a time series TB-related spatial weight matrix has long been recognized by experimenters who typically check whether clinical estimates and inferences are similar when alternative spatial weight structures are employed.

The second source of uncertainty in a time series MDR-TB epidemiological explanatory endemic trasmission-oriented risk model arises in conventional as well as spatial growth regression models since growth theories are not sufficiently explicit about which specific factors underlie the data-generating process for growth regressions. Hence, experimenters are faced with a dilemma regarding the large number of potentialtime series explanatory MDR-TB –related regressors.There is a trade-of between arbitrary selection of a small subset of sampled clinical variables which may give rise to omitted variables bias, and the introduction of a large set of variables that may increase the dispersion of the estimated coefficients, making it difficult to identify important factors.Spatial growth MDR-TB time series explanatory regression models produce estimates and inferences that are conditional on both the particular spatial weight matrix used to specify which observational units (regions) are linked and the set of explanatory variables employed. Selection of an appropriate MDR-TB-related spatial weight matrix and explanatory variables are central to a time series risk-related epidemiological analysis (Jacob et al. 2010b).

Spatial quantile regression can provide much more information on spatial data than the conditional mean regression analysis.Thus in a seasonal MDR-tb-related epidemiological risk model an experimeneter may develop a structure of spatial quantile regression allowing functional coefficients, under a robust semiparametric framework Firstly, w treat data as observed over a space of general dimension N. Denote the set of integer lattice points in N-dimensional Euclidean spaceby ZN, where $N \geq 1$ and $Z = \{0, \}1, \}2, \ldots\}$. A point $i = (i1, \ldots, iN)$ in ZN is referred to as a site. Spatial data are modeled as finite realizations of vector stochastic processes indexed by $i \in ZN$, that is, random fields. We will consider strictly stationary $(d + k + 1)$ -dimensional random fields of the form $\{(Yi, Xi, Ui) : i \in ZN\}$, where Yi, with values in R, Xi, with values in Rd, and Ui, with values in Rk, are defined over a probability space (,F,P). Secondly, we treat spatial quantile regression in a general context of robust spatial regression. In a number of applications, a crucial problem consists in describing and analyzing the influence of the covariates (Ui,Xi) on the real-valued response Yi. In spatial context, this study is particularly difficult due to the possibly highly complex spatial dependence among the various sites. The traditional approach to this problem consists in assuming that Yi has finite expectation, so that spatial conditional mean regression function $g : (x, u)7 \rightarrow g(x, u) :=$ E[Yi | Xi = x, Ui = u] may be well defined and clearly carries relevant information on the dependence of Y on X and U (cf., [14,25,26]). Differently, Hallin et al. [15]

proposed spatial conditional quantile regression, defined by $q\tau : (x, u)7 \to q\tau(x, u) := Q[Y_i | X_i = x, U_i = u]$, (1.1) which provides more comprehensive information on the dependence of Y on X and Uthrough different $0 < \tau < 1$ (see [23] and [41]), where $q\tau (x, u)$ satisfies $P[Y_i < q\tau(x, u) | X_i = x, U_i = u] = \tau$; see also the robust spatial conditional regression in [24]. As is well known

in the nonparametric literature, when $d + k > 3$, both spatial regression functions g(x, u) and $q\tau(x, u)$ can not be well estimated nonparametrically with reasonable accuracy owing to the curse of dimensionality. Because of complex spatial interaction, this issue on how to avoid the curse of dimensionality becomes particularly important, which has been addressed by Gao et al. [9] and Lu et al. [27] for spatial conditional mean regression g(x, u) under least squares partially linear and additive approximation structures, respectively.

A MDR-TB experimenter particularly may be concerned with avoiding the curse of dimensionality for spatial quantile regression analysis, and, for generality, consider a general spatial regression that takes conditional quantile regression $Q(Y_i | U_i, X_i)$ as a special case, to be approximated by a popular linear structureallowing for functional coefficients in the form $(U_i, X_i) = X_{i1}\beta_1(U_i) + \cdots + X_{id}\beta_d(U_i)$, (1.2) with the functional coefficients $\beta_j(\cdot)'s$ defined by minimizing $E\rho(Y_i - (U_i, X_i)) = E\rho(Y_i - X_{i1}\beta_1(U_i) - \cdots - X_{id}\beta_d(U_i))$, (1.3)associated with ρ(y) by which we denote hereafter for a general loss function [see Section2], over a class of functional coefficient linear functions of the form (U_i, X_i) in (1.2). In the subsequent, when considering τ th quantile regression, we will denote by $\rho\tau(z) = |z| + (2\tau - 1)z$ with $0 < \tau < 1$, instead of $\rho(\cdot)$, for the loss function, under which the resulting (U_i, X_i) in (1.2) is the spatial quantile regression with functional coefficientsthat we are mainly concerned with in this paper. Let $X_i = (X_{i1}, \ldots, X_{id})T$. As intraditional linear regression when a baseline effect is desired, we set $X_{i1} \equiv 1$. The regimeU$_i$ is a vector ofexplanatory variables, and $\beta_1(u), \ldots, \beta_d(u)$ are unknown smooth functionsof u to be estimated, with the dimension k of U$_i$ usually small, say k = 1 or 2.

Competing specifications are usually non-nested alternatives so that conventional statistical procedures such as likelihood ratio tests are inappropriate (LeSage and Fischer 2008).Model averaging provides a formal approach that can be used to incorporate model uncertainty in spatial MDR-TB related endemic transmission-oriented explanatory regression models which arises from selecting both the spatial weight matrix and the clinic explanatory variables when making inferences about model parameters. Instead of selecting a single model, this approach proposes toaverage estimates across different models. Bayesian model averaging represents one powerful approach to making parameter inference unconditional on model specification issues.

A Bayesian treatment also has the ability to properly account for high variance of estimates in geographic areas and clarify overall spatial trends and patterns, regardless of distribution of data (Hastie and Tibshirani, 1990). There is a great deal of literature on Bayesian model averaging for non-spatial regression models.

For example, work by Fernandez et al. (2001a) considers cases where the number of possible models is sufficiently large so that calculation of posterior probabilities for all models is difficult or infeasible. A Markov chain Monte Carlo model comparison methodology proposed by Madigan and York (1995) has gained popularity in the mathematical statistics and econometrics literature. An extension to spatial autoregressive regression models is provided by LeSage and Parent (2007). LeSage and Fischer (2008) include simultaneous comparison of models based on both alternative explanatory variables and spatial weight matrices, albeit concentrating on the class of k-nearest neighbor spatial weight matrices. From a technical point of view, using numerical integration techniques from these models may help obtain posterior model probabilities for MDR-TB-related specifications with different k-nearest spatial weight matrices which may then be used to obtain Bayesian model averaged estimates. The computational costs of this procedure, however, makes it an impractical choice for a large set of alternative spatial weight MDR-TB-related time series matrices. A methodology may improve on LeSage and Fischer (2008) by adopting Bayesian information criterion (BIC) posterior model weights to overcome such computational costs, and thus allowing for the consideration of a wide range of weight matrices as potential spatial structures underlying the spillovers in the sampled MDR-TB-related data.

Many Bayesian approaches for analyzing spatial disease patterns focus on mapping spatially smoothed disease rates (Clayton and Kaldor, 1987). Mapping parameters in a spatial Bayesian probabilistic regression matrix can produce stable estimates for the cell-specific disease rates by shrinkage to the overall rate or by averaging over neighboring cells. A Bayesainistic description can be provided using a simple exact uncertainty spatially-dependent cluster-based detection algorithm in an ArcGIS cyberenvironment for quantitating large spatial regions in a highly MDR-infected study area. For example, Jacob et al. (2013) employed an eigenfunction decomposition algorithm associated with a Moran's coefficient to investigate district-level non-linearity in an empirical dataset of spatiotemporal-sampled MDR-TB parameter estimators sampled in San Juan de Lurigancho (SJL) Lima, Peru. The non-parametric technique attempted to remove the inherent autocorrelation in the model by introducing appropriate synthetic surrogate variants. The authors then constructed a robust Bayesian probabilistic Poisson model to generate unbiased estimators for qualitatively assessing resistance to four commonly used drugs in TB treatment: isoniazid, rifampin, ethambutol, and streptomycin. Initially, data of residential addresses and of individual patients with smear-positive MDR-TB were geocoded in ArcGIS. Next, the sampled data were matched interactively within the geodatabase. The MDR-TB attributes were then calculated and digitally overlaid onto sub-meter resolution satellite data within a 1 km buffer of 31 georeferenced health centers using a 10 m^2 grid-based algorithm. Global autocorrelation statistics were then generated by decomposing the sampled data into positive and negative

spatial filter eigenvectors using the eigenfunction decomposition algorithm. Bayesian Poisson projections were then derived employing normal priors for each of the sampled resistant strains. A Residual Moran's coefficient (MC) minimization criterion was then applied to the clinical coefficients spatially quantitated from the decomposition algorithm to detect any unaccounted latent autocorrelation error in the estimators. The model accounted for approximately 14% pseudo-replicated information and exhibited positive residual autocorrelation.

Bayesian methods may also provide some shrinkage and spatial smoothing of raw standardized MDR-TB hierarchical explanatory residual intra-cluster-based error estimates, which are strongly influenced by sampled clinical and environmental population size. x, a data point in general. This may in fact be a vector of values. θ, the parameter of the data point's distribution, i.e., $x \sim p(x|\theta)$. This may in fact be a vector of parameters., α, the hyperparameter of the parameter, i.e., $\theta \sim p(\theta|\alpha)$. This may in fact be a vector of hyperparameters. X, a set of n observed data points, i.e., $x_1, \ldots, x_n$. $\tilde{x}$, a new data point whose distribution is to be predicted. Bayesian inference for MDR-TB epidemiological risk model. The prior distribution is the distribution of the parameter(s) before any data is observed, i.e. $p(\theta|\alpha)$. The prior distribution might not be easily determined. In this case, we can use the Jeffreys prior to obtain the posterior distribution before updating them with newer observations. The sampling distribution is the distribution of the observed data conditional on its parameters, i.e. $p(X|\theta)$. This is also termed the likelihood, especially when viewed as a function of the parameter(s), sometimes written $L(\theta; X) = p(X|\theta)$. The marginal likelihood (sometimes also termed the *evidence*) is the distribution of the observed data marginalized over the parameter(s), i.e. $p(X|\alpha) = \int_\theta p(X|\theta) p(\theta|\alpha) d\theta$. The posterior distribution is the distribution of the parameter(s) after taking into account the observed data. This is determined by Bayes' rule, which forms the heart of Bayesian inference:

$$p(\theta|X,\alpha) = \frac{p(X|\theta)p(\theta|\alpha)}{p(X|\alpha)} \propto p(X|\theta)p(\theta|\alpha)$$. Note

that this is expressed in words as "posterior is proportional to likelihood times prior", or sometimes as "posterior = likelihood times prior, over evidence". The posterior predictive distribution is the distribution of a new data point, marginalized over the posterior:

$$p(\tilde{x}|X,\alpha) = \int_\theta p(\tilde{x}|\theta) p(\theta|X,\alpha) d\theta$$. The prior

predictive distribution is the distribution of a new data point, marginalized over the prior:

$$p(\tilde{x}|\alpha) = \int_\theta p(\tilde{x}|\theta) p(\theta|\alpha) d\theta.$$

Bayesian theory calls for the use of the posterior predictive distribution to do predictive inference, i.e., to predict the distribution of a new, unobserved data point. That is, instead of a fixed point as a prediction, a distribution over possible points is returned. Only this way is the entire posterior distribution of the parameter(s) used. By comparison, prediction in frequentist statistics often involves finding an optimum point estimate of the parameter(s)—e.g., by maximum likelihood or maximum a posteriori estimation (MAP)—and then plugging this estimate into the formula for the distribution of a data point. This has the disadvantage that it does not account for any uncertainty in the value of the parameter, and hence will underestimate the variance of the predictive distribution.

(In some instances, frequentist statistics can work around this problem. For example, confidence intervals and prediction intervals in frequentist statistics when constructed from a normal distribution with unknown mean and variance are constructed using a Student's t-distribution. This correctly estimates the variance, due to the fact that (1) the average of normally distributed random variables is also normally distributed; (2) the predictive distribution of a normally distributed data point with unknown mean and variance, using conjugate or uninformative priors, has a student's t-distribution. In Bayesian statistics, however, the posterior predictive distribution can always be determined exactly—or at least, to an arbitrary level of precision, when numerical methods are used.)

Note that both types of predictive distributions have the form of a compound probability distribution (as does the marginal likelihood). In fact, if the prior distribution is a conjugate prior, and hence the prior and posterior distributions come from the same family, it can easily be seen that both prior and posterior predictive distributions also come from the same family of compound distributions. The only difference is that the posterior predictive distribution uses the updated values of the hyperparameters (applying the Bayesian update rules given in the conjugate prior article), while the prior predictive distribution uses the values of the hyperparameters that appear in the prior distribution.

Suppose a process is generating independent and identically distributed events E_n, but the probability distribution is unknown. Let the event space Ω represent the current state of belief for this process. Each model is represented by event M_m. The conditional probabilities $P(E_n|M_m)$ are specified to define the models. $P(M_m)$ is the degree of belief in M_m. Before the first inference step, $\{P(M_m)\}$ is a set of *initial prior probabilities*. These must sum to 1, but are otherwise arbitrary. Suppose that the process is observed to generate $E \in \{E_n\}$. For each $M \in \{M_m\}$, the prior $P(M)$ is updated to the posterior $P(M|E)$. From Bayes' theorem: [4]

$$P(M|E) = \frac{P(E|M)}{\sum_m P(E|M_m) P(M_m)} \bullet P(M)$$ Upon observation

of further evidence, this procedure may be repeated. For a set of independent and identically distributed observations $E = \{e_1, \cdots, e_n\}$, it may be shown that repeated application of the above is equivalent to

$$P(M|E) = \frac{P(E|M)}{\sum_m P(E|M_m) P(M_m)} \bullet P(M)$$ Where

$$P(E|M) = \prod_k P(e_k|M).$$ This may be used to optimize

practical calculations. By parametrizing the space of models, the belief in all models may be updated in a single step. The distribution of belief over the model space may then be thought of as a distribution of belief over the parameter space. The distributions in this section are expressed as continuous, represented by probability densities, as this is the usual situation. The technique is however equally applicable to discrete distributions. Let the vector θ span the parameter space. Let the initial prior distribution over θ be $p(\theta|\alpha)$, where α is a set of parameters to the prior itself, or *hyperparameters*. Let $E = \{e_1, \cdots, e_n\}$ be a set of independent and identically distributed event observations, where all e_i are distributed as $p(e|\theta)$ for some θ. Bayes' theorem is applied to find the posterior distribution over θ:

$$p(\theta|E,\alpha) = \frac{p(E|\theta,\alpha)}{p(E|\alpha)} \cdot p(\theta|\alpha)$$

$$= \frac{p(E|\theta,\alpha)}{\int_\theta p(E|\theta,\alpha)\, p(\theta|\alpha)\, d\theta} \cdot p(\theta|\alpha)$$

where $p(E|\theta,\alpha) = \prod_k p(e_k|\theta)$.

Bayesian approaches would be very useful for capturing gradual, regional changes in the sampled MDR-TB parameters, and may be useful in detecting abrupt, localized changes indicative of 'hot spot' clustering. In recent years, models including both spatially structured random effects (SSRE) and spatially unstructured random effects (SURE) have been very popular in Bayesian infectious disease hierarchical cluster-based regression models. For example, the model proposed by Besag et al., (1991) incorporated both SSRE and SURE in a single Bayesian model which Ghosh et al., (1999) used to analyze leukemia data. Waller and Zelterman, (1997) extended this model to incorporate spatiotemporal effects for county level lung cancer rates in Ohio. Besag et al., (1995) have suggested a prior specification for the SSRE more suitable for detection of spatial clusters. As noted by Ferreira et al. (2002), Bayesian frameworks can easily define a prior for the clusters as well as incorporate predictor covariate error coefficient estimates and extra-Poisson variation. Bayesian methods which may thus model the random and true variation in a spatiotemporal-sampled endemic transmission—oriented MDR-TB clinical dataset of georeferenced clinical explanatory observational covariate coefficients

In this paper we propose a first-order autocorrelation and a Bayesian model in which we compute probabilities of potential georefernced MDR-TB clusters using multiple clinical and environmental parameters sampled in SJL a district in Lima, Peru. A statistical framework is presented for the analysis using the sampled data generated from community-based surveys conducted in the SJL study site. Our models partitioned the sources of infection into those from within the household and those from the community at large. The observational predictor covariate coefficient estimates reflecting these sources of infection were then quantitated as functions of the risk factors. Instead of focusing on populations of carriers and susceptibles,

solely, as in previous cluster-based MDR-TB research, emphasis instead was placed on identifying the residual-based error coefficients and their primary influences, thus directing the research as a spatial autoregressive process. This was done by dividing the study area into small-area units in ArcGIS, and by assigning the risk of obtaining MDR-TB to each sampled unit, based on clinical, socioeconomic and demographic characteristics of the population.

Initially, we generated a dataset of Durbin Watson test statistics. A non-parametric eigenvector filtering technique was then used to remove inherent spatial autocorrelation from a generalized linear regression (GLM) model generated using the sampled clinical and environmental MDR-TB predictor variables by treating it as a missing variable (i.e., first order) effect (Getis and Griffith, 2002; Griffith, 2000). To expand the inferential basis with a random effect, a generalized linear mixed model (GLMM) was then constructed in SAS/GIS to account for latent cluster-based error autocorrelation components in the model. We specified a likelihood function for the sampled data and a prior distribution for the parameter estimates. The response variable prior distribution included the model statement and the error variance prior distribution which in this research was the gamma distribution. Markov models were then used to incorporate the residual within intra- cluster-based explanatory georeferenced predictor covariate coefficient estimates to derive transition probabilities, formulate the likelihood function, and calculate the ML estimates. In probability theory, a Markov model is a stochastic model that assumes the Markov property. Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable. A stochastic process has the Markov property if the conditional probability distribution of future states of the process depends only upon the present state, not on the sequence of events that preceded it Given valid assumptions about the nature of variance autocorrelation uncertainty in Bayesian applications, we assumed the serial correlation consistent standard error estimators generated from a spatially weighted distance function error matrix model, may develop and implement MDR-TB control strategies in the SJL study site by determining residual explanatory georeferenced covariate coefficient estimates associated to prolific clusters based on clinical and environmental-sampled data. Since, the possible existence of non-normal error probabilities in residual forecasts is a major concern in the application of linear and non-linear regression analysis using spatiotemporal-sampled MDR-TB data including the analysis of variance, as the presence of these error coefficients can invalidate residual intra-cluster-based tests of significance that assume the effect and residual error variances are uncorrelated and normally distributed. Therefore, our objectives in this research were to: (1) generate a stepwise regression model using multiple predictor variables (2) filter all latent autocorrelation in residual estimates using a stepwise negative binomial regression with a gamma distributed mean; and, (3) construct Bayesian random-effects hierarchical generalized linear model (HGLM) specifications for adjusting the SSRE and SURE in a cluster-based model to identify high risk populations of MDR-TB for implementing control strategies in SJL.
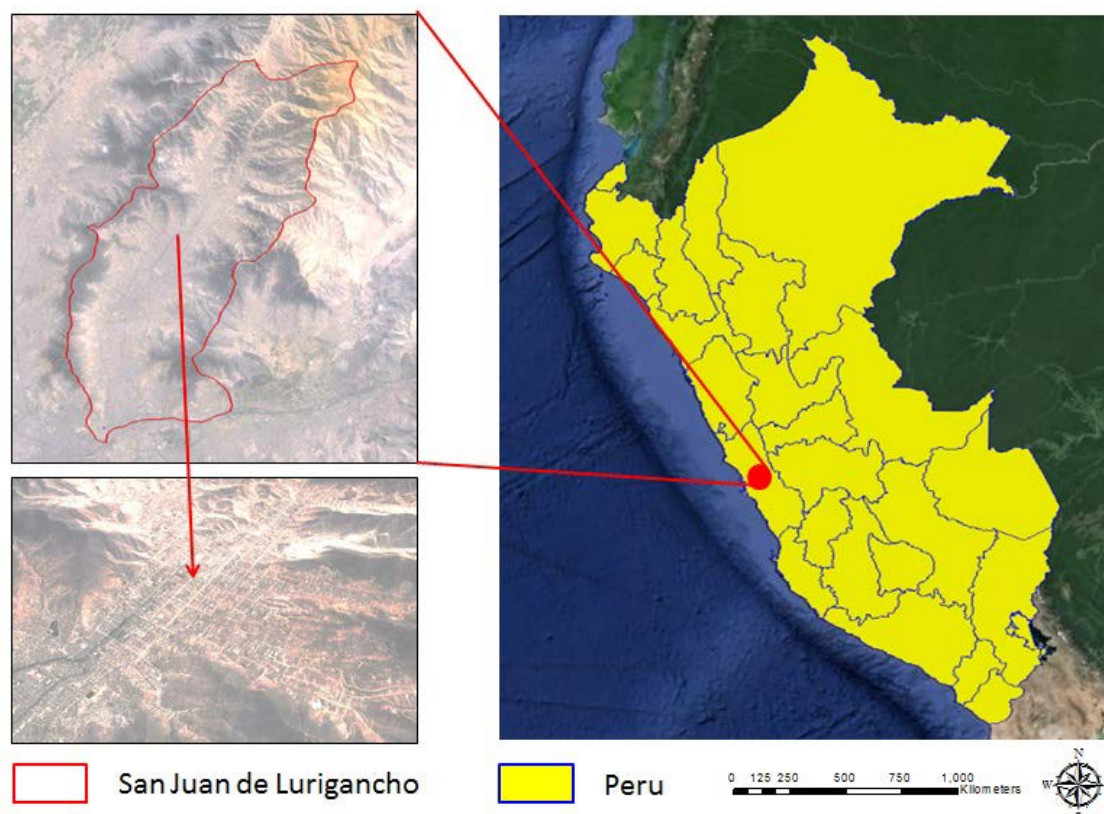
Figure 1. San Juan de Lurancho study site

2. Material and Methodology

Study Site: San Juan de Lurigancho (SJL) is the largest district in Lima, located in the Northeast of the province of Lima. (Figure 1). With a current population exceeding one million people and a total surface area of 131.3 km^2, constituting 4.91% of the total area of the province of Lima, it is the country's most populous district. SJL is bordered by the districts of Carabayllo and San Antonio in the Huarochirí Province to the north, by the Comas, Independencia and Rímac districts to the west and by Lurigancho to the east. The Rímac River marks the district's border with downtown Lima and El Augustino to the south. The most important urban areas in the district are Mangomarca, Zárate, Las Flores, Canto Grande and Bayovar. One of the first urban areas in SJL is Caja de Agua, which is located at the entrance of the district. Caja de Agua is surrounded by the San Cristobal and Santa Rosa hills from south to west. The altitude of SJL ranges from 2,240 meters (m) above mean sea level (AMSL.) at the peaks of Cerro Colorado Norte to 200 m AMSL at the level of the Rimac River. Urban areas have been developed in a longitudinal direction from the river border up to 350 m AMSL. Lima has a mild climate, although it is situated in the tropics. The weather in Lima is influenced by the cold offshore Humboldt Current, which ensures that summer temperatures hover in between 16-18°C and only a few degrees lower in June and July.

Subjects and setting: This research used data acquired from a retrospective study of a cohort of patients diagnosed with pulmonary TB and MDR-TB enrolled over an 18 month period. All patients underwent a complete evaluation, including drug susceptibility for first line drugs. This was a prospective multi-center observational study comparing the use of several investigational techniques with standard methods to assess the *in vitro* antimicrobial susceptibility of *M. tuberculosis,* either directly from patient specimens or from culture isolates. One thousand two hundred and fifty adults with pulmonary tuberculosis cultures were confirmed. After collection of baseline samples and completion of initial measurements, including susceptibility testing by conventional and research methods, all subjects started anti-TB chemotherapy as dictated by the standard of care at the site of enrollment. Subjects were recruited, among patients presenting smear positive pulmonary tuberculosis, to diagnostic and treatment sites in the following Health Centers: San Fernando, La Huayrona, Canto Grande, Jose Carlos Mariátegui, Huáscar XV, Huáscar II, Ganímedes, Cruz de Motupe, Piedra Liza, Bayóvar, Jaime Zubieta, San Juan, San Benito, Mangomarca, San Hilarion, Campoy, 15 de Enero, La Libertad, Juan Pablo II, Ascarruz Alto, 10 de Octubre, Sta. Fe de Totoritas, Proyectos Especiales, Santa Rosa, Ayacucho, Zarate, Medalla Milagrosa, Campoy Alto, Montenegro, Santa Maria, Tupac Amaru II and Caja de Agua.

Geographic mapping: Field sampling was conducted from July 2005 to July 2007. Thirty-one health centers in the SJL study site were mapped and classified using a CSI-Wireless differentially corrected global positioning systems (DGPS) Max receiver. This remote technology relies on the OmniStar L-Band satellite signal yielding a positional error of. 179 m (+/-. 392 m) (Jacob et al., 2007). Individual health centers and their associated land cover attributes were identified from the satellite imagery and entered into a VCMS relational database software product.

Data from the characterization of each health center was recorded on a Mobile Vector Control Management System (VCMS™) electronic data recording device. The field sampling was extended to a 1 km distance from the external boundary of a health center study site.

Remote sensing data: QuickBird (www.digitalglobe.com) images were acquired in March 11th 2008, for the SJL study site. QuickBird multispectral products provided four discrete non-overlapping spectral bands covering a range from 0.45 micrometer (µm) to 0.72 µm, with an 11-bit collected information depth with a spatial resolution of 0.61 m. QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* V.8.7™. The images were co-registered manually, using ground control points and georectified images from the QuickBird data. The satellite images were co-registered by applying a first order polynomial algorithm with a nearest neighbor resampling method. The Universal Transverse Mercator (UTM) Zone 37S datum WGS-84 projection was used for all of the spatial datasets.

Environmental parameters: Variables recorded included, MDR-TB prevalence rates, distance between individual Health Centers, population data, and aspects of land-surface in the SJL study site such as elevation and slope per sampled site. Distance measures were recorded in ArcGIS 9.2$^{®}$ with QuickBird data and by field sampling. The distance between health centers was categorized into numerous classes (e.g., 1: 0–5 km, 2: 5–10 km, and so on). The number of individuals cases of MDR-TB at each georeferenced individual health center was calculated and recorded. All variables used in this research are listed in Table 1.

Table 1. Clinical and environmental MDR-TB data sampled in the SJL study site

Variable in Database	Description of variable
ESTAB	Health care center
FENAC	Birth date
EdadA	Age
SEXO	Sex
TIPOVIV	Home
NUMHAB	Number of bedrooms
MATVIV	Building material
NUMPER	Number of persons living in the house
ELECTRIC	Electricity supply at home
AGUAPOT	Home access to potable water
DESAGUE	Wastepipe connected to the public network
ECIVIL	Marital status
OCUPA	Ocupation
TRAESTS	Do you work in any health care center?
TIEMTRA	Time of employment
INGMEN	Salary/Income per month
LJINH	Sensitivity test to isoniazid in LJ medium
LJIRIF	Sensitivity test to rifampin in LJ medium
LJIETB	Sensitivity test to ethambutol in LJ medium
LJISTM	Sensitivity test to streptomycin in LJ medium
MDR	Multidrug resistant

Description of Study Area: Quality of living conditions, employment status, dwelling characteristics, overcrowding status, and sensitivity to different TB drugs were analyzed and correlated. For the correlation analysis, 21 records were removed from the data set due to missing field information, leaving 764 records available. Some records had multiple attributes missing making the total number of records removed less than the number of affected records. All statistical results were calculated using R statistical software version 2.15.2. Table 2 summarizes the number of records with missing data from particular fields:

Table 2. Reason for record removal from study set

Field with missing data	Number of records affected
Address Type	6
Number of people in home	6
Profession	3
Waste connected to public network	3
Number of bedrooms in house	2
Access to electricity	1
Home type	1
Building material	1
Multi-drug resistance and sensitivity to rifampin/ethambutol/streptomycin	1

Grid-based algorithm: A 10 m x 10 m grid-based algorithm was overlaid on the base maps of the study site, in ArcGIS 9.2$^{®}$, to generate spatial sampling units. A 1 km buffer was placed around each health center. A unique identifier was placed in each gridded buffer. Each spatial cluster was then stratified by MDR-TB prevalence rates (Figure 1).

Hierarchical agglomerative polythetic cluster model: Initially, a 1 km buffer was placed around each health center using the QuickBird data in ArcGIS 9.2$^{®}$. FLEXIBLE|FLE in SAS 9.2$^{®}$ (Carey, North Carolina) was then used to request the flexible-beta method. The PROC CLUSTER statement then started the procedure, which specified a clustering method and optionally specified details for data processing. This technique resulted in the narrowest distance range for the clinical and environmental- sampled MDR-TB predictor covariate coefficient estimates. The flexible-beta method began by specifying METHOD=FLEXIBLE in SAS. By default, *b* was the value of the Beta. In this research Beta was set at −100. PROC CLUSTER then displayed the clustering process, showing statistics useful for estimating the number of clusters in the sampled datasets. PROC CLUSTER created an output dataset that revealed a cluster hierarchy of the health centers in the SJL study site based on the covariate coefficient indicator measurement values. Since in this research the georeferenced parameters were deemed to be equally important, we used the STD option in PROC CLUSTER to standardize the cluster-based covariate coefficient estimates to mean 0 and standard deviation 1. Explanatory predictor covariate coefficient estimates with large variances tend to have more effect on the resulting geographic clusters than variables with small variances but if all georeferenced observations are considered equally important, the STD option in PROC CLUSTER can standardize the sampled data (www.sas.com). In this research the STDIZE procedure standardized the spatiotemporal-sampled predictor covariate coefficient estimates in the SAS

dataset by subtracting the georeferenced location measures and dividing them by a scale measure. Finally, a unique identifier was incorporated for each cluster representing the data. In order to compute meaningful standardized rates, the individual sampled data was then aggregated geographically into high-low stratified clusters using the QuickBird data based on sampled MDR-TB prevalence rates (Figure 1).

The SJL study site was then examined extensively using longitude, latitude and altitude data. These criteria involved the centrographic measures of spatial mean, distance between the sampled georeferenced MDR-TB data and the distance from sample site to the nearest human habitation (Figure 2). Cumulative overcrowding distribution by people in household and bedrooms in house was then determined (Figure 3). The data was also comprised of individual observations of the sampled data together with a battery of categorical predictor covariate attribute measures which were expanded into multiple indicator variables. Histograms for these groups were generated using the sampled data (Figure 4).
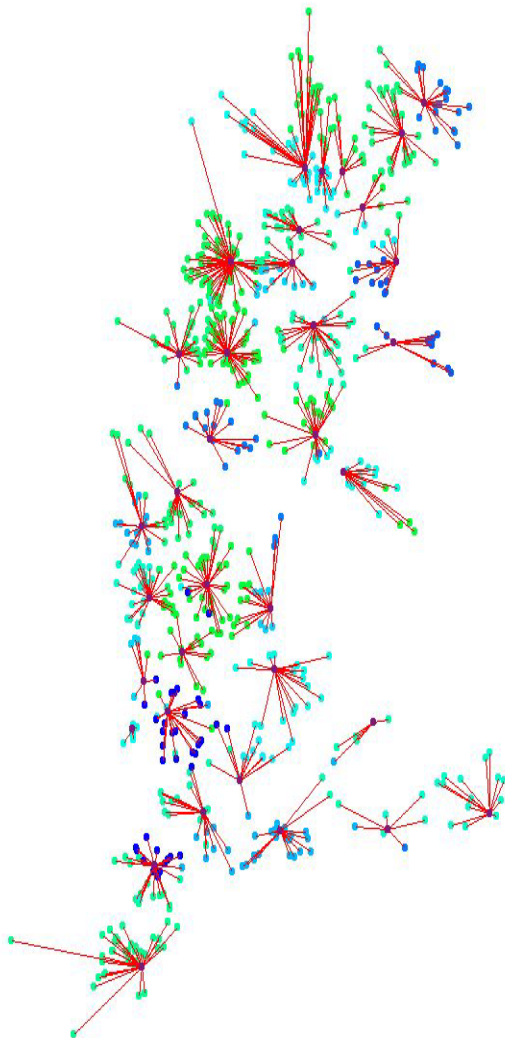

Figure 2. The distribution of the health centers, and the allocation of infected individuals to centers

*Red lines denote allocation to the nearest center.
*Graduated green-blue-purple denote the actual allocation of individuals to centers.

Regression analyses: Initially, a Poisson regression with statistical significance was calculated by a 95% confidence in SAS GEN MOD. The Poisson process in

our analyses was provided by the limit of a binomial distribution of the sampled district-level explanatory MDR-TB covariate coefficient estimates using

$$P_p\left(n|N\right) = \frac{N!}{n!(N-n)!} p^n \left(1-p\right)^{N-n} \qquad (2.1)$$

We viewed the distribution as a function of the expected number of count variables using the sample size N for quantifying the fixed p in equation (2.1), which was then transformed into the linear equation:

$$P_{\frac{v}{N}}\left(n|N\right) = \frac{N!}{n!(N-n)!}\left(\frac{v}{N}\right)^n \left(1-p\right)^{N-n} .$$ Based on the

sample size N, the distribution approached $\lim_{n\to\infty} P_p\left(n|N\right)$

$$= \lim_{N\to\infty} \frac{N(N-1)...(N-n+1)}{n!} \frac{v^n}{N^n}\left(1-\frac{v}{N}\right)^N \left(1-\frac{v}{N}\right)^{-n}$$

$$= \lim_{N\to\infty} \frac{N(N-1)...(N-n+1)}{n!} \frac{v^n}{n!}\left(1-\frac{v}{N}\right)^N \left(1-\frac{v}{N}\right)^{-n}$$

$$= 1 \cdot \frac{v^n}{n!}$$

$P_v\left(n\right)$ was $e^{-v} \cdot 1 = \frac{v^n e^{-v}}{n!}$.

The GENMOD procedure then fit a generalized linear model (GLM) to the sampled MDR-TB data by maximum likelihood estimation of the parameter vector β. In this research the GENMOD procedure estimated the seasonal-sampled parameters of each model numerically through an iterative fitting process. The dispersion parameter was then estimated by the residual deviance and by Pearson's chi-square divided by the degrees of freedom (d.f.). Covariances, standard errors, and *p*-values were then computed for the sampled covariate coefficients based on the asymptotic normality derived from the maximum likelihood estimation.

Note, that the sample size N completely dropped out of the probability function, which in this research had the same functional form for all the sampled district-level MDR-TB parameter estimator indicator values (i.e., ν). As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1. The ratio of probabilities was then determined by

$$\sum_{n=0}^{\infty} P_v\left(n\right) = e^{-v}\sum_{n=0}^{\infty} \frac{v^n}{n!} = e^{-v}e^v = $$ which was then

subsequently expressed as $\dfrac{P_v\left(n=i+1\right)}{P\left(n=i\right)} = \dfrac{\dfrac{v^{i+1}e^{-v}}{(i+1)!}}{\dfrac{e^{-v}V!}{i!}} = \dfrac{v}{i+1}$.

The Poisson distribution revealed that the explanatory covariate coefficients reached a maximum when

$$\frac{dP_v\left(n\right)}{dn} = \frac{e^{-v}n\left(\gamma - H_n + \ln v\right)}{n!} = 0,$$ where γ was the

Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation $\gamma - H_n + \ln v = 0$. The regression model also revealed that the Euler-Mascheroni constant arose in the integrals as

$$\gamma = -\int_0^\infty e^{-x} \ln x \, dx = -\int_0^1 \ln \ln \left(\frac{1}{x}\right) dx$$

$$= \int_0^\infty \left(\frac{1}{1-e^x} - \frac{1}{x}\right) e^{-x} dx = \int_0^\infty \frac{1}{x}\left(\frac{1}{1+x} - e^{-x}\right) dx \qquad (2.2)$$

Commonly, integrals that render γ in combination with temporal sampled constants include $\int_0^\infty e^{-x^2} \ln x \, dx = -\frac{1}{4}\sqrt{\pi}\left(\gamma - 2\ln 2\right)$ which is equal to $\int_0^\infty e^{-x}\left(\ln x\right)^2 dx = \gamma^2 + \frac{1}{6}\pi^2$ Thereafter, the double integrals in the MDR-TB regression model included $\gamma = \int_0^1 \int_0^1 \frac{x-1}{(1-xy)\ln(xy)} dx\, dy$.

In this research an interesting analog of equation (2.2) in the regression-based model was then calculated as $\ln\left(\frac{4}{\pi}\right) = \sum_{n=1}^\infty (-1)^{n-1}\left[\frac{1}{n} - \ln\frac{n+1}{n}\right]$. This solution $= \int_0^1 \int_0^1 \frac{x-1}{(1+xy)\ln(xy)} dx\, dy = 0.241564\ldots\gamma$ was also provided by incorporating Mertens theorem [i.e., $e^\gamma = \lim_{n\to\infty} \frac{1}{\ln p_n}\prod_{i=1}^n \frac{1}{1-\frac{1}{p_i}}$ where the product was aggregated over the district-level sampled values found in the empirical ecological datasets. IMertens' 3rd theorem: $\lim_{n\to\infty} \ln n \prod_{p\le n}\left(1-\frac{1}{p}\right) = e^{-\gamma}$ is related to the density of prime numbers where γ is the Euler–Mascheroni constant[(Hosmer and Lemeshew 2000].By taking the logarithm of both sides in the MDR-TB model, an explicit formula for γ was then derived employing

$$\gamma = \lim_{x\to\infty}\left[\sum_{p\le x} \ln\left(\frac{1}{1-\frac{1}{p}}\right) - \ln\ln x\right].$$ This expression was also rendered coincidently by quantifying the data series employing Euler, and equation (2.2) by first replacing $\ln n_b \ln(n+1)$, in the equation $\gamma = \sum_{k=1}^\infty \left[\frac{1}{k} - \ln\left(1+\frac{1}{k}\right)\right]$ and then generating $\lim_{n\to\infty}\left[\ln(n+1) - \ln n\right] = \lim_{n\to\infty}\left(1+\frac{1}{n}\right) = 0$. We then substituted the telescoping sum $\sum_{k=1}^\infty \ln\left(1+\frac{1}{k}\right) for \ln(n+1)$ which then generated $\ln\left(1+\frac{1}{k}\right) = \ln(k+1) - \ln k$. Thereafter, our product was

$$\lim_{n\to\infty}\left[\sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \ln\left(1+\frac{1}{k}\right)\right]_\gamma = \lim_{n\to\infty}\sum_{k=1}^n\left[\frac{1}{k} - \ln\left(1+\frac{1}{k}\right)\right].$$

Additionally, other series in our spatiotemporal MDR-TB regression model included the equation ($\diamondsuit$) where

$$\gamma = \sum_{n=2}^\infty (-1)^n \frac{\zeta(n)}{n} = \ln\left(\frac{4}{\pi}\right) + \sum_{n=1}^\infty \frac{(-1)^{n-1}\zeta(n+1)}{2^n(n+1)} \quad \text{and}$$

$\zeta(z)$ was $\gamma = \sum_{n=1}^\infty (-1)^n \frac{[\lg n]}{n}$ plus the Riemann zeta function. The Riemann zeta function $\zeta(s)$ is a function of a complex variable$s$ that analytically continues the sum of the infinite series $\sum_{n=1}^\infty \frac{1}{n^s}$ which converges when the real part of s is greater than 1 where lg is the logarithm to base 2 and the [x] is the floor function (Hastie, and Tibshirani 1990). Jacob et al. (2012) earlier provided a series equivalent to $\gamma = 1 - \sum_{n=1}^\infty \sum_{k=2^{n-1}}^{2^n-1} \frac{n}{(2k+1)(2k+2)}$ and, thereafter $\frac{1}{(2k+1)(2k+2)} = \frac{1}{2k+1} - \frac{1}{2k+2}$ which was then added to $0 = -\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$ to render Vacca's formula in a district-level malaria-related model. The authors then used used the sums

$$\gamma = \sum_{n=1}^\infty \sum_{k=2^n}^\infty \frac{(-1)^k}{k} = \sum_{k=1}^\infty \frac{1}{2^{k+1}}\sum_{j=0}^{k-1}\binom{2^{k-j}+j}{j}^{-1}$$

with k-j by replacing the undefined I and then rewrote the equation as a double series for applying the Euler's series transformation to each of the sampled time-series dependent explanatory covariate coefficient estimates.

In this research $\frac{n}{k}$ was used as a binomial coefficient, rearranged to achieve conditionally convergent series in our spatiotemporal MDR-TB linear model. The plus and minus terms were first grouped in pairs of the sampled covariate coefficient estimates employing the resulting series based on the actual observational covariate coefficient 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 general formulas due to $J_0\left(\sqrt{z^2-y^2}\right) = \frac{1}{\pi}\int_0^\pi e^{v\cos\theta}\cos(z\sin\theta)d\theta$ where $J_0(z)$ is a Bessel function of the first kind [3]. The Bessel function is a function $Z_n(x)$ defined in a robust regression model by using the recurrence relations $Z_{n+1} + Z_{n-1} = \frac{2n}{x}Z_n$ and $Z_{n+1} + Z_{n-1} = -2\frac{dZ_n}{dx}$ [2] which more recently has been defined as solutions in linear models using the differential equation $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(x^2 - n^2\right)y = 0$ [6].

In this research 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* $y=1+z'$. In mathematics, Bessel functions are canonical solutions y(x) of Bessel's differential equation: $x^2\dfrac{d^2y}{dx^2}+x\dfrac{dy}{dx}+\left(x^2-\alpha^2\right)y=0$ for an arbitrary real or complex number α (i.e., the order of the Bessel function); the most common and important cases are for α an integer or half-integer (Hosmer and Lemeshew 2000). Thereafter, to quantify the equivalence in the spatiotemporal malarial regression-based parameter estimators, we expanded $1/(1+x)$ in a geometric series and multiplied the district-level sampled MDR-TB data feature attributes by x^{2^n-1}, and integrated the term wise as in Sondow and Zudilin [6]. Other series for γ then included $\gamma=\dfrac{3}{2}-\ln 2-\sum_{m=2}^{\infty}(-1)^m\dfrac{m-1}{m}\left[\zeta(m)-1\right]$ and

$$\gamma=\frac{2^n}{e^{2n}}\sum_{m=0}^{\infty}\frac{2^{mn}}{(m+1)!}\sum_{t=0}^{m}\frac{1}{t+1}-n\ln 2+o\left(\frac{1}{2^n e^{2n}}\right) \quad . \quad \text{A}$$

rapidly converging limit for γ was then provided by

$$\gamma=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\left(\frac{1}{k}-\frac{\zeta(1-k)}{n^k}\right)\right]$$

$$=\lim_{n\to\infty}\left[\frac{2n-1}{2n}-\ln n+\sum_{k=2}^{n}\frac{1}{k}\left(1+\frac{B_k}{n^k}\right)\right]$$

where B_k was a Bernoulli number. Another limit formula was then provided

by the equation $\gamma=-\lim_{n\to\infty}\left[\dfrac{\Gamma\left(\dfrac{1}{n}\right)\Gamma(n+1)n^{1+1/n}}{\Gamma\left(2+n+\dfrac{1}{n}\right)}-\dfrac{n^2}{n+1}\right]$.

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 [(Hastie, and Tibshirani 1990).]. Jacob et al. [1] found if m and n are sampled values and f(x) is a smooth sufficiently differentiable function in a seasonal malarial-related 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)

$$S=\frac{1}{2}f(m)+f(m+1)+\ldots+f(n-1)+\frac{1}{2}f(n) \quad . \quad \text{The}$$

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 [Hosmer and Lemeshew 2000]. Thereafter, for the district-level MDR_TB -sampled values, p, we had

$$S-I=\sum_{k=2}^{p}\frac{B_k}{k!}\left(f^{(k-1)'}(n)-f^{(k-1)'}(m)\right)+R \text{ where B1}=$$

−1/2, B2 = 1/6, B3 = 0, B4 = −1/30, B5 = 0, B6 = 1/42, B7 = 0, B8 = −1/30, and R which was an error term. Note in this research $-B_1\left(f(n)+f(m)\right)=\dfrac{1}{2}\left(f(n)+f(m)\right)$. Hence, we re-wrote the regression-based MDR-TB formula as follows:

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

then rewrote the 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 district-level MDR-TB regression model was then rendered by $\gamma=\lim_{x\to\infty}\zeta\left(\zeta(z)\right)-2^x+\left(\dfrac{4}{3}\right)+1$ where $\zeta(z)$ was the Riemann zeta function. The Bernoulli numbers appear in the Taylor series expansions of the tangent and hyperbolic tangent functions, in formulas for the sum of powers of the first positive integers, in the Euler–Maclaurin formula and in expressions for certain values of the Riemann zeta function [(Hastie, and Tibshirani 1990).

Another connection with the primes was provided by $d(n)=\sigma_0(n)$ for the sampled time series MDR-TB numerical values from 1 to n in the spatiotemporal sampled dataset which in this research was found to be asymptotic to $\dfrac{\sum_{k=1}^{n}d(k)}{n}\sim\ln n+2\gamma-1$. De laValléePoussin proved that if a large number n is divided by all primes $\leq n$, then the average amount by which the quotient is less than the next whole number is g [Hosmer and Lemeshew 2000]. An identity for g in our malaria regression-based model was then provided by $\gamma=\dfrac{S_0(z)-K_0(z)}{I_0(z)}-\ln\left(\dfrac{1}{2}z\right)$ where $I_0(z)$ was a modified Bessel function of the first kind, $K_0(z)$ was a modified Bessel function of the second kind, and

$$S_0(z)\equiv\sum_{K=0}^{\infty}\frac{\left(\dfrac{1}{2}z\right)^{2k}H_k}{(k!)^2}$$ where H_n was a harmonic

number. For non-integer α, $Y\alpha(x)$ is related to $J\alpha(x)$ by: $Y_\alpha(x)=\dfrac{J_\alpha(x)\cos(\alpha\pi)-J_{-\alpha}(x)}{\sin(\alpha\pi)}$. In the case of integer order n, the function is defined by taking the limit as a non-integer α tends to n: $Y_n(x)=\lim_{\alpha\to n}Y_\alpha(x)$. (Hastie, and Tibshirani1990). In this research, the Bessel functions of the second kind, were denoted by $Y\alpha(x)$, and by $N\alpha(x)$, 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

$$B_k = \frac{B_{k-1}n^2}{k^2} = A_k = \frac{1}{k}\left(\frac{A_{k-1}n^2}{k} + B_k\right)$$ and

$$= U_k K_k = U_{k-1} + A_k$$

$$V_k = V_{k-1} + B_k \quad A_0 = -\ln n B_0 = 1U_0 = A_0 \quad \text{and} \quad V_0 = 1.$$

Reformulating this identity rendered the limit

$$\lim_{n\to\infty}\left|\sum_{k=0}^{\infty}\frac{\left(\frac{n^k}{k!}\right)^2 H_k}{\sum_{k=0}^{\infty}\left(\frac{n^k}{k!}\right)^2} - \ln n\right| = \gamma \quad .$$ 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 superfactorials to the complex numbers which is related to the Gamma function[(Hastie, and Tibshirani 1990).]. 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 our time-series dependent MDR-TB regression-based 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 $\prod$ 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

$$\gamma = \lim_{s\to 1^+}\sum_{n=1}^{\infty}\left(\frac{1}{n^s} - \frac{1}{s^n}\right) \text{ and } \gamma = \lim_{x\to\infty}\left[x - \Gamma\left(\frac{1}{x}\right)\right].$$

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, $\psi_0(x)$ 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-Mascheroni constant ((Hastie, and Tibshirani 1990). In mathematics, the n-th harmonic number is the sum of the reciprocals of the first n natural number s [Hosmer and Lemeshew 2000]. The difference between the nth convergent in equation ($\diamondsuit$) and γ in our district-level MDR-TB regression-based model was then calculated by $\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x-|x|}{x^2}dx$ where [x] was the floor function which satisfied the inequality

$\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x-|x|}{x^2}dx$. The symbol g was then $\gamma' \equiv e^{\gamma} \approx 1.781072$. This led to the radical representation of the sampled explanatory MDR-TB 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 $\gamma = \sum_{n=1}^{\infty}\frac{1}{n}\sum_{k=0}^{n-1}(-1)^{k+1}\binom{n-1}{k}\ln(k+1) \text{ and } \binom{n}{k}$ a binomial coefficient. Thereafter, another proof of product in the our spatiotemporal district-level MDR-TB regression model was provided by the equation $\frac{\pi}{2} = \left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/3}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/8}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{16}$. 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

$$e = \left(\frac{2}{1}\right)^{1/1}\left(\frac{2^2}{1\cdot 3}\right)^{1/2}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/3}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/4} .$$

The outputs revealed that the linear MDR-TB models contained a constant term. As such, it was necessary to assume that $E\left(e^{\tau i}\right) = E(\tau_i) = 1$ in order to identify the mean of the MDR-TB linearized distributions. We assumed that τ_i followed a gamma (θ,θ) distribution in the models with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$:

$$g(\tau_i) = \frac{\theta^{\theta}}{\Gamma(\theta)}\tau_i^{\theta-1}\exp(-\theta\tau_i)$$ where

$\Gamma(x) = \int_0^{\infty}z^{x-1}\exp(-z)dz$ was the gamma function and θ was a positive parameter. Thus, the density of y_i in the models was X_i which in this research was

$$f(y_i|X_i) = \int_0^{\infty} f(y_i|X_i,\tau_i)g(\tau_i)d\tau_i$$

$$= \frac{\theta^{\theta}\mu_i^{y_i}}{y_i!\Gamma(\theta)}\int_0^{\infty} e^{-(\mu_i+\theta)\tau_i}\tau_i^{\theta+y_i-1}d\tau_i$$

$$= \frac{\Gamma(y_i+\theta)}{y_i!\Gamma(\theta)}\left(\frac{\theta}{\theta+\mu_i}\right)^{\theta}\left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i} \quad (2.4)$$

$$= \frac{\theta^{\theta}\mu_i^{y_i}\Gamma(y_i+\theta)}{y_i!\Gamma(\theta)(\theta+\mu_i)^{\theta+y_i}}$$

Unfortunately, extra-Poisson variation was detected in the residual variance estimates in our MDR-TB models. As such, we constructed negative binomial regression models in PROC REG with non-homogenous means by incorporating $\alpha = \frac{1}{\theta}(\alpha > 0)$ in equation 2.1. The distribution was then rewritten as

$$f\left(y_i\middle|X_i\right)=\frac{\Gamma\left(y_i+\alpha^{-1}\right)}{y_i!\,\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i},$$

$y_i = 0,1,2,\ldots$. Thus, the negative binomial distribution was derived as a gamma mixture of Poisson random variables in the MDR-TB models. The conditional mean was then $E\left(y_i\middle|X_i\right)=\mu_i=e^{X_i'\beta}$ and conditional variance was $V\left(y_i\middle|X_i\right)=\mu_i\left[1+\frac{1}{\theta}\mu_i\right]=\mu_i\left[1+\alpha\mu_i\right]>E\left(y_i\middle|X_i\right).$

To further estimate the MDR-TB models, we specified DIST=NEGBIN(p=1) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1, set p=1, had the variance function $V\left(y_i\middle|X_i\right)=\mu_i+\alpha\mu_i$, which was linear in the mean in the model. The log-likelihood function of the NEGBIN1 regression model was then given by $\mathscr{L}=\sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j+\alpha^{-1}\right)-\ln\left(y_i!\right)\right.$

$\left.=-\left(y_i+\alpha^{-1}\exp\left(X_i'\beta\right)\right)\ln\left(1+\alpha\right)+y_i\ln\left(\alpha\right)\right\}$. The gradient for each MDR-TB model was then

$$\frac{\partial\mathscr{L}}{\partial\beta}=\sum_{i=1}^{N}\left\{\left(\sum_{j=0}^{y_i-1}\frac{\mu_i}{\left(j\alpha+\mu_i\right)}\right)X_i-\alpha^{-1}\ln\left(1+\alpha\right)\mu_iX_i\right\} \text{ and}$$

$$\frac{\partial\mathscr{L}}{\partial\alpha}=\sum_{i=1}^{N}\left\{\begin{array}{l}-\left(\sum_{j=0}^{y_i-1}\frac{\alpha^{-1}\mu_i}{\left(j\alpha+\mu_i\right)}\right)-\alpha^{-2}\ln\left(1+\alpha\right)\\-\frac{\left(y_i+\alpha^{-1}\mu_i\right)}{1+\alpha}+\frac{y_i}{\alpha}\end{array}\right\}.$$

In this research, the negative binomial regression MDR-TB models with variance function $V\left(y_i\middle|X_i\right)=\mu_i+\alpha\mu_i^2$, were referred to as the NEGBIN2 model. To estimate this model, we specified DIST=NEGBIN $\left(p=2\right)$ in the MODEL statements. A test of the Poisson 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 negative binomial regression models. The log-likelihood function of the models (NEGBIN2) was then generated by the equation: $\mathscr{L}=\sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j+\alpha^{-1}\right)-\ln\left(y_i!\right)\right.$

$\left.=-\left(y_i+\alpha^{-1}\right)\ln\left(1+\exp\left(X_i'\beta\right)\right)+y_i\ln\left(\alpha\right)+y_iX_i'\beta\right\}$ where y was an integer when the gradient was $\frac{\partial\mathscr{L}}{\partial\beta}=\sum_{i=1}^{N}\frac{y_i-\mu_i}{1+\alpha\mu_i}X_i$ and the residual variance estimates were

$$\frac{\partial\mathscr{L}}{\partial\alpha}=\sum_{i=1}^{N}\left\{\begin{array}{l}-\alpha^{-2}\sum_{j=0}^{y_i-1}\frac{1}{\left(j+\alpha^{-1}\right)}\\-\alpha^{-2}\ln\left(1+\alpha\mu_i\right)-\frac{y_i-\mu_i}{\alpha\left(1+\alpha\mu_i\right)}\end{array}\right\}.$$

First-order serial autocorrelation analyses: We then constructed multiple MDR-TB risk-based spatially dependent models using an AR(1) framework. In this research the AR(1) MDR-TB models were constructed using the clinical and environmental explanatory predictor covariate coefficients estimates sampled at the SJL study site. The models were defined as $X_t=c+\sum_{i=1}^{p}\varphi_1X_{t-1}+\varepsilon_t$ where $\varphi_1,\ldots,\varphi_p$ were the sampled predictor variables, c was a constant and ε_t was white noise. The MDR-TB models revealed that continuous time random process $w(t)$ where $t\in\mathbb{R}$ was a white noise process if, and only if, its mean function and autocorrelation function satisfied the following: $\mu_w(t)=E\{w(t)\}=0$ and $R_{ww}\left(t_1,t_2\right)=E\{w\left(t_1\right)w\left(t_2\right)\}=\left(N_0/2\right)\delta\left(t_1-t_2\right).$ (i.e., it is a zero mean process for all time and has infinite power at zero time shift) since its autocorrelation function was the Dirac delta function. The Dirac delta function is a generalized function depending on a real parameter such that it is zero for all values of the parameter except when the parameter is zero, and its integral over the parameter from $-\infty$ to ∞ is equal to one (Cressie 1993). In this research the AR(1)-process was given by: $X_t=c+\varphi X_{t-1}+\varepsilon_t$ where ε_t was a white noise process with zero mean and variance σ_ε^2.

Thereafter, we used different classifications, whereby, the autoregressive parameters in the MDR-TB models processes were defined as wide-sense stationary, if $|\varphi|<1$ which was obtained as the output of stable filters whose input was white noise. On the other hand, in the models, if $\varphi=1$, then X_t had infinite variance and, therefore, was not wide-sense stationary. In mathematical sciences, a stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space (Cressie 1993), thus, statistics such as the mean and variance in predictive autoregressive distribution models do not change over time or position. Consequently, in this research we assumed $|\varphi|<1$, where the mean $E\left(X_t\right)$ was identical for all the sampled cluster based MDR-TB covariate coefficient values of t.

In this research the mean of the sampled clinical and environmental MDR-TB data were denoted by μ, thus it followed $E\left(X_t\right)=E\left(c\right)+\varphi E\left(X_{t-1}\right)+E\left(\varepsilon_t\right)$, such that $\mu=c+\varphi\mu+0$, and hence $\mu=\frac{c}{1-\varphi}$. The variance in the residual forecasts were then delineated by $\mathrm{var}\left(X_t\right)=E\left(X_t^2\right)-\mu^2=\frac{\sigma_\varepsilon^2}{1-\varphi^2}$, where σ_ε was the standard deviation of ε_t. This was revealed by noting that $\mathrm{var}\left(X_t\right)=\varphi^2\,\mathrm{var}\left(X_{t-1}\right)+\sigma^2$, and that the quantity rendered from the equation was a stable fixed point of this relation in each model. Additionally, for $B_n=E\left(X_{t+n}X_t\right)-\mu^2=\frac{\sigma_\varepsilon^2}{1-\varphi^2}\varphi^{|n|}$, the autocovariance

function decayed with time (i.e., time constant) of $\tau = -1/\ln(\varphi)$ in both models. In order to further define the autocovariate function we wrote $B_n = K\varphi^{|n|}$ where K was independent of n. Then we noted that $\varphi^{|n|} = e^{|n|\ln\varphi}$ and matched this to the exponential decay law $e^{-n/\tau}$. We also noticed that the spectral density function was the Fourier transform of the autocovariance function in the MDR-TB models. The Fourier transform is a mathematical operation that decomposes a signal into its constituent frequencies (Cressei 1993). In this research, the discrete-time Fourier transform in both models was expressed as

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} B_n e^{-i\omega n} = \frac{1}{\sqrt{2\pi}} \left(\frac{\sigma_\varepsilon^2}{1+\varphi^2 - 2\varphi\cos(\omega)} \right).$$

This expression was periodic due to the discrete nature of the X_j, which was manifested as the cosine term in the denominator. We assumed that the sampling time $(\Delta t = 1)$ was much smaller than the decay time (τ). By doing so we were able to successfully apply a continuum approximation to $B_n : B(t) \approx \frac{\sigma_\varepsilon^2}{1-\varphi^2}\varphi^{|t|}$ which yielded a Lorentzian profile for the spectral density

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}} \frac{\sigma_\varepsilon^2}{1-\varphi^2} \frac{\gamma}{\pi(\gamma^2 + \omega^2)} \quad \text{where} \quad \gamma = 1/\tau \text{ was}$$

the angular frequency associated with τ in the MDR-TB models. In mathematics, Lorentz distribution is closely related to the Poisson kernel, which is the fundamental solution for the Laplace equation in the upper half-plane (see Haight 1967). In this research the Lorentz distribution had the probability density function

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)\right]} = \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2 + \gamma^2}\right]$$

where x_0 was the sampled MDR-TB parameter, specifying the location of the peak of the distribution of the georeferenced data sampled and γ was the scale parameter estimate which specified the half-width at half-maximum. In our Lorentz distribution γ was also equal to half the interquartile range (i.e., the probable error in the models).

Thereafter, an alternative expression for X_t was derived by first substituting $c + \varphi X_{t-2} + \varepsilon_{t-1}$ for X_{t-1} in the defining equations. Continuing this process n times yielded $X_t = c\sum_{k=0}^{N-1}\varphi^k + \varphi^N X_{t-N} + \sum_{k=0}^{N-1}\varphi^k\varepsilon_{t-k}$. We noticed that for n approaching infinity, φ^N our models' residual estimates approached zero and $X_t = \frac{c}{1-\varphi} + \sum_{k=0}^{\infty}\varphi^k\varepsilon_{t-k}$. The residual estimates also revealed that X_t was white noise convolved with the φ^k kernel plus the constant mean. If the white noise ε_t is a

Gaussian process then X_t is also a Gaussian process (Cressie 1993). The regression residuals revealed that X_t was normally distributed when φ was close to one in both models.

The DW statistic was then generated to detect the presence of first-order autocorrelation error coefficients in the regression residuals. We generated the DW statistic to test for first-order serial correlation in the MDR-TB models. We used the DWPROB option in SAS to print the significance level (i.e. *p*-values) for the Durbin-Watson tests. The DW statistic was used to test the null hypothesis $H_o : \varphi_1 = 0$ against $H_1 : -\varphi_1 > 0$. The following procedure statements performed the Durbin-Watson test for autocorrelation in the ordinary least squares (OLS) residuals for orders 1 through 4 in both models.

In this research, the generalized DW statistic was written as: $DW_j = \frac{\hat{u}^1 A_j' A_j \hat{u}}{\hat{u}^1 \hat{u}}$ where $\hat{u}$ was a vector of OLS residuals and A_j was a $(T-j) \times T$ matrix. The generalized DW statistic DW_j was then rewritten as:

$$DW_j = \frac{Y'MA_j'A_jMY}{Y'MY} = \frac{\eta'(Q_1'A_j'A_jQ_1)\eta}{\eta'\eta} \quad \text{where}$$

$Q_1'Q_1 = I_{T-k}$, $Q_1'X = 0$, and $\eta = Q_1'u$. The marginal probability for the DW statistic was: $\Pr(DW_j < c) = \Pr(h < 0)$ where $h = \eta'(Q_1'A_j'A_jQ_1)\eta$. The *p*-value, or the marginal probability for the generalized DW statistic was then computed by numerical inversion of the characteristic function $\phi(u)$ of the quadratic form $h = \eta'(Q_1'A_j'A_jQ_1 - cI)\eta$. In the models the trapezoidal rule approximation to the marginal probability Pr(h<0) was

$$\Pr(h < 0) = \frac{1}{2} - \sum_{k=0}^{K} \frac{lm\left[\phi\left(\left(k+\frac{1}{2}\right)\Delta\right)\right]}{\pi\left(k+\frac{1}{2}\right)} + E_I(\Delta) + E_T(K)$$

where $lm[\phi(\cdot)]$ was part of the characteristic function and $E_I(\Delta)$ and $E_T(K)$ were integration and truncation errors, respectively. The trapezoidal rule is a way to calculate the definite integral (Cressie 1993).

A numerically efficient algorithm was then used to quantify the dependent error components in our first-order autocorrelation MDR-TB models, which required O(*N*) operations for evaluation of the characteristic function $\phi(u)$. In this research the characteristic function in the sampled predictors covariate coefficients estimates were denoted as: $\phi(u)\left|I - 2iu\left(Q_1'A_j'A_jQ_1 - cI_{N-k}\right)\right|^{-\frac{1}{2}}$ where $|v|^{-\frac{1}{2}}|x'v^{-1}x|^{-\frac{1}{2}}|x'x|^{-\frac{1}{2}}$

$v = (1 + 2iuc)I - 2iucA_j'A_j$ and $i = \sqrt{-1}$. By applying the Cholesky decomposition to the complex matrix V, we obtained the lower triangular matrix G which satisfied V=

GG'. Cholesky decomposition is a decomposition of a symmetric, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose (Cressie 1993). This generated $n \times n$ complex matrices (i.e., A) which were positive definite if $R\left[x^4 Ax\right] > 0$ for all nonzero complex vectors $x \in C^n$, where X^4 denoted the conjugate transpose of the vector X. The conjugate transpose of an *m x n* matrix A in this research was the *n x m* matrix defined by $A^H \equiv \overline{A}^T$ where A^T denoted the transpose of the matrix A and $\overline{A}$ denoted the conjugate matrix in both MDR-TB models. If *A* is a complex matrix, then the conjugate transpose $A*$ is the matrix $A^* = \overline{A}^T$, where $A\_\overline{A}$ is the complex conjugate of *A*, and A^T is the transpose of *A* (Cressie 1993). The conjugate transpose of a matrix A in this research was implemented in *Mathematica* as ConjugateTranspose [A]. The conjugate transpose is also known as the adjoint matrix, adjugate matrix, Hermitian adjoint, or Hermitian transpose (Strang 1988). If a matrix is equal to its own conjugate transpose, it is said to be self-adjoint and is called a Hermitian (Cressie 1993). In this research the conjugate transpose of the matrix product in both MDR-TB models was given by $(ab)_{ij}^H = \left[(ab)^T\right]_{ij}$. Using the identity for the product of transpose rendered $\left(\overline{b^T a^T}\right)$, $\overline{b_{ik}^T a_{kj}^T}$, $\left(\overline{b^T}\right)_{ik}\left(\overline{a^T}\right)_{kj}$, $b_{ik}^H a_{kj}^H$, $\left(b^H a^H\right)_{ij}$ where Einstein summations were then used here to sum over repeated indices.

The Einstein summation convention implies that when an index occurs more than once in the same expression, the expression is implicitly summed over all possible values for that index (Cressie 1993). Therefore, in order to use the summation in this research it had to be clear from the context over what sampled MDR-TB range indices had to be summed. The Einstein summation convention was performed by letting $\{ei\}ni=1$ be an orthogonal basis in $\mathbb{R}n$ in both models. In mathematics, particularly linear algebra, an orthonormal basis for inner product space *V* with finite dimension is a basis for *V* whose vectors are orthonormal (Strang 2006). For example, the standard basis for a Euclidean space $\mathbb{R}^n$ is an orthonormal basis, where the relevant inner product is the dot product of vectors (Cressie 1993). The image of the standard basis under a rotation or reflection (or any orthogonal transformation) is also orthonormal, and every orthonormal basis for $\mathbb{R}^n$ arises in this fashion. In functional analysis, the concept of an orthonormal basis can also be generalized to arbitrary (infinite-dimensional) inner product spaces (or pre-Hilbert spaces) (Lay 2006). In this research the inner product of the vectors generated from the sampled MDR-TB data was $u = uiei = (\sum_{ni=1})uiei$ and $v = viei = (\sum_{ni=1})viei$, where $u.v = = uivjei.ej\ \delta ijuivj$. For a general inner product space *V*, an orthonormal basis can be used to define normalized orthogonal coordinates on *V*. Under these coordinates, the inner product becomes dot product of vectors. Thus the presence of an orthonormality in our

models reduces the study of a finite-dimensional inner product space to the study of $\mathbb{R}^n$ under dot product. We let *V* be a vector space with basis $\{ei\}ni = 1$ and a dual basis $\{ei\}ni = 1$. Then, for a vector $v = viei$ and dual vectors $\alpha = \alpha iei$ and $\beta = \beta iei$, we had $(\alpha + \beta)(v) = \alpha ivi + \beta jvj = (\alpha i + \beta i)vi$. A vector space is a mathematical structure formed by a collection of vectors: objects that may be added together and multiplied ("scaled") by numbers, called scalars in this context (Cressie 1993). This revealed that the summation convention in both models "distributive" occurred in a natural way. We then let $F : \mathbb{R}m \to \mathbb{R}n, x \to (F1, \ldots, Fn)$, and $G : \mathbb{R}n \to \mathbb{R}p, y \to (G1(y), \ldots, Gp(y))$ be smooth functions. Then $\partial_{xj}\partial(GoF)i(x) = \partial_{Gi}\partial_{yk}(F(x))\partial_{xj}\partial Fk(x)$, where the right hand side was summed over $k = 1, \ldots, n$. An index which is summed is called a dummy index or dummy variable (Cressie 1993). In this research, *i* was a dummy index in *viei*. In our models the expression did not depend on a dummy index, (i.e., $viei = vjej$). This greatly simplified and shortened the equations For example, in this research using Einstein summation the MDR-TB models rendered, $a_i a_i \equiv \sum_i a_i a_i$ and $a_{ik} a_{ij} \equiv \sum_i a_{ik} a_{ij}$. Thereafter it followed that $(AB)^H = B^H A^H$.

In this research the linear system of equations generated from the MDR-TB parameters with a positive definite matrix was efficiently solved using the Cholesky decomposition. The positive definite matrices had at least one matrix square root in each model. Furthermore, exactly one of its matrix square roots was itself positive definite in each model residual estimate. In this research, the MDR-TB matrices were said to be positive definite when A^H of the Hermitian part $A_H \equiv \frac{1}{2}\left(A + A^H\right)$ denoted the conjugate transpose. The complex matrices were then was broken into a Hermitian part $A_H \equiv \frac{1}{2}\left(A + A^H\right)$ (i.e., A_H was a Hermitian matrix) and an antihermitian part $A_{AH} \equiv \frac{1}{2}\left(A - A^H\right)$ (i.e., A_{AH} is an antihermitian matrix). In our research, A^H denoted the adjoint. This meant that a real matrix A was positive definite in the models only if the symmetric part $A_S \equiv \frac{1}{2}\left(A + A^T\right)$ where A^T was the transpose, was positive definite. Furthermore, our square MDR-TB matrices A were antihermitian if they satisfied $A^H = -A$, where A^H was the adjoint. Our models revealed matrices where $\begin{bmatrix} i & 1+i & 2i \\ -1+i & 5i & 3 \\ 2i & -3 & 0 \end{bmatrix}$ which were antihermitian matrices.

In this research, the marginal probability for dj given c_0 in the MDR-TB models was

$$\Pr{ob}\left(\frac{\Sigma_{I=1}^{n}\lambda_{jl}\xi_l^2}{\Sigma_{l=1}^{n}\xi_l^2}<c_0\right)=\Pr{ob}\left(q_j<0\right) \qquad \text{where}$$

$q_j=\sum_{l=1}^{n}\left(\lambda_{jl}-c_0\right)\xi_l^2$. Additionally, when the null hypothesis $H_o:\varphi_1=0$ held, the quadratic form qj had the characteristic function

$$\phi_j(t)=\prod_{l=1}^{n}\left(1-2\left(\lambda_{jl}-c_0\right)it\right)^{-\frac{1}{2}}$$ in the residual

predictor covariate coefficients estimates. The distribution function was then uniquely determined by this characteristic function:

$$F(x)=\frac{1}{2}+\frac{1}{2\pi}\int_0^{\infty}\frac{e^{itx}\phi_j(-t)-e^{-itx}\phi_j(t)}{it}dt$$ in both

models. We then tested $H_o:\varphi_4=0$ given $\varphi_1=\varphi_2=\varphi_3=0$ against $H_o:-\varphi_4>0$, in each modelusing the marginal probability (p-value) and:

$$F(0)=\frac{1}{2}+\frac{1}{2\pi}\int_0^{\infty}\frac{\left(\varphi_4(-t)-\varphi_4(t)\right)}{it}dt \qquad \text{where}$$

$$\varphi_4(t)=\prod_{l=1}^{n}\left(1-2\left(\lambda_{4l}-d_4\right)it\right)^{-\frac{1}{2}}$$ and $\hat{d}_4$ was the

calculated value of the fourth-order DW statistic.

In AUTOREG, two alternative statistics (i.e., Durbin h and t) were also used to test for time varying residuals that were asymptotically equivalent). In this research, we used

the h statistic, which was written as: $h=\hat{p}\sqrt{\dfrac{N}{1-NV}}$

where $\hat{p}=\sum_{l=2}^{N}\dfrac{\widehat{V}_l\widehat{V}_{l-1}}{\Sigma_{l=1}^{N}\widehat{V}_l^2}$, and $\hat{V}$ was the least squares

variance estimatefor the coefficient of the lagged dependent variables in the MDR-TB models. Durbin's t test consists of regressing the OLS residuals $\widehat{V}_l$ on predictor variables and $\widehat{V}_{l-1}$ for testing the significance of the estimate for coefficient of $\widehat{V}_{l-1}$ (Cressie 1993).

In PROC AUTOREG, an estimation method was used to generate autoregressive error models using the Yule-Walker (YW) method. The YW method can be considered as generalizedleast squares using the OLS residuals to estimate the covariances (Cressie 1993).

The YW equations we used included

$$\gamma_m=\sum_{k=1}^{p}\varphi_k\gamma_{m-k}+\sigma_\varepsilon^2\delta_{m,0}, \qquad \text{where} \qquad m=0,...,p,$$

yielded p + 1 equations; where γ m was the autocorrelation function of X in the MDR-TB models, σ_ε was the standard deviation of the input noise process; and, $\delta_{m,0}$ was the Kronecker delta function. The Kronecker's delta, is a function of two variables, usually integers, which is 1 if, they are equal and 0 otherwise (Hosmer and Leneshew 2000). The equations were solved by representing the sampled clinical and environmental MDR-TB predictors as a matrix form > 0 , thus,

rendering equation

$$\begin{bmatrix}\gamma_1\\\gamma_2\\\gamma_3\\\vdots\end{bmatrix}=\begin{bmatrix}\gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots\\\gamma_1 & \gamma_0 & \gamma_{-1} & \cdots\\\gamma_2 & \gamma_1 & \gamma_0 & \cdots\\\vdots & \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}\varphi_1\\\varphi_2\\\varphi_3\\\vdots\end{bmatrix}$$

solving all φ . For m = 0 we had $\gamma_0=\sum_{k=1}^{p}\varphi_k\gamma_{-k}+\sigma_\varepsilon^2$ in

both model's residual estimates which allowed us to solve σ_ε^2. The full auto-correlation function was then derived

by recursively calculating $\rho(\tau)=\sum_{k=1}^{p}\alpha_k\rho(k-\tau)$ in the

estimates. In this research, the YW equations were $\gamma_1=\varphi_1\gamma_0+\varphi_2\gamma_{-1}$ and $\gamma_2=\varphi_1\gamma_1+\varphi_2\gamma_0$ where $\gamma-\mathrm{k}=\gamma\mathrm{k}$. The equations then yielded

$\rho_1=\gamma_1/\gamma_0=\dfrac{\varphi_1}{1-\varphi_2}$ and the recursion formula which

then yielded $\rho_1=\gamma_1/\gamma_0=\dfrac{\varphi_1^2-\varphi_2^2+\varphi_1}{1-\varphi_2}$ in both model

estimates.

The equation defining the AR processes in the MDR-TB

models was then $X_t=\sum_{i=1}^{p}\varphi_i X_{t-i}+\varepsilon_t$. Thereafter, we

multiplied both sides by Xt − m and imputed the expected values which yielded

$$\mathrm{E}[X_t X_{t-m}]=\mathrm{E}\left[\sum_{i=1}^{p}\varphi_i X_{t-i}X_{t-m}\right]+\mathrm{E}[\varepsilon_t X_{t-m}].$$ In this

research $\mathrm{E}[X_t X_{t-m}]=\gamma_m$ was the autocorrelation function in the models. The values of the noise function were independent on each other, and Xt − m was independent of εt where m was greater than zero. The residual estimates revealed form > 0 , $\mathrm{E}[\varepsilon\mathrm{t}X\mathrm{t}-\mathrm{m}]=0$.

$$\mathrm{E}[\varepsilon_t X_t]=\mathrm{E}\left[\varepsilon_t\left(\sum_{i=1}^{p}\varphi_i X_{t-i}+\varepsilon_t\right)\right]$$

For m = 0 , . This

$$=\sum_{i=1}^{p}\varphi_i\,\mathrm{E}[\varepsilon_t\,X_{t-i}]+\mathrm{E}[\varepsilon_t^2]=0+\sigma_\varepsilon^2,$$

rendered $\gamma_m=\mathrm{E}\left[\sum_{i=1}^{p}\varphi_i X_{t-i}X_{t-m}\right]+\sigma_\varepsilon^2\delta_m$. when m $\geq$ 0

in the MDR-TB models. Furthermore, we used

$$\mathrm{E}\left[\sum_{i=1}^{p}\varphi_i X_{t-i}X_{t-m}\right]=\sum_{i=1}^{p}\varphi_i\,\mathrm{E}[X_t X_{t-m+i}]=\sum_{i=1}^{p}\varphi_i\,\gamma_{m-i},$$

which yielded th equation $\gamma_m=\sum_{i=1}^{p}\varphi_i\gamma_{m-i}+\sigma_\varepsilon^2\delta_m$. and

$$\gamma_m=\gamma_{-m}=\sum_{i=1}^{p}\varphi_i\gamma_{|m|-i}+\sigma_\varepsilon^2\delta_m. \text{ for m < 0.}$$

In this research, we let φ represent the vector of the autoregressive parameters, $\varphi=(\varphi_1,\varphi_2,...,\varphi_m)'$, and we let the variance matrix of the error vector be

$v = (v_1, ..., v_N)'$ Σ, $E(vv\Sigma = \sigma 2v)$. If the vector of autoregressive parameters φ is known, the matrix v can be computed from the autoregressive parameters; Σ is then $\sigma^2 v$ (Cressie 1993). Given Σ the efficient estimates of regression parameters β were computed using generalized least squares (GLS) for both models. The GLS estimates then yielded the unbiased estimate of the variance σ^2 in the models.

Shapiro–Wilk diagnostic test: The Shapiro–Wilk test was then used to test the null hypothesis that the spatiotemporal-sampled clinical and environmental MDR-TB predictor covariate coefficient estimates $x_1, ..., x_n$ came from a normally distributed population. In SAS, the primary test statistics for detecting the presence of non-normality is the Shapiro-Wilk (www.sas.com). Shapiro-Wilk test checks the normal assumption by constructing W statistic, which is the ratio of the best estimator of the variance based on the square of a linear combination of the order statistics to the usual corrected sum of squares estimator of the variance (Cressie 1993). In this research, to perform the test, the W statistic was constructed by considering the regression of the ordered sampled parameter estimates on corresponding expected normal order statistics, which was linear in both MDR-TB models. W isa measure of the straightness of the normal probability plot, and small values indicate departures from normality (Cressie 1993). We used the test statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$ where $x_{(i)}$ (with parentheses enclosing

the subscript index *i*) was the *i*th order statistic, (i.e., the *i*th-smallest numbering the sampled clinical and environmental MDR-TB sampled datasets), where $x = (x_1 + ... + x_n)/n$ was the sample mean; the constants a_i were given by

$$(a_1, ..., a_n) = \frac{m^\top V^{-1}}{(m^\top V^{-1} V^{-1} m)^{1/2}}$$ where $m = (m_1, ..., m_n)^\top$

and $m_1, ..., m_n$ were the expected values of the order statistics of independently and identically distributed(i.d.d). random variables sampled from the standard normal distribution; and, *V* was the covariance matrix of those order statistics. In probability theory and statistics, a sequence or other collection of random variables is independent and identically distributed (i.i.d.) if each random variable has the same probability distribution as the others and all are mutually independent (Cressie 1993).We also computed $S^2 = (n-1)s^2$ where s^2 was the sample variance using the parameters sampled at the SJL study site.

In this research, the marginal probability for dj given c_0 in the MDR-TB models was

$$Prob\left(\frac{\sum_{I=1}^{n} \lambda_{jl} \xi_l^2}{\sum_{l=1}^{n} \xi_l^2} < c_0\right) = Prob(q_j < 0)$$ where

$q_j = \sum_{l=1}^{n} (\lambda_{jl} - c_0)\xi_l^2$. Additionally, when the null hypothesis $H_o : \varphi_1 = 0$ held, the quadratic form qj had the characteristic function $\phi_j(t) = \prod_{l=1}^{n} \left(1 - 2(\lambda_{jl} - c_0)it\right)^{-\frac{1}{2}}$ in the residual predictor covariate coefficients estimates. The distribution function was then uniquely determined by this characteristic function: $F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{itx}\phi_j(-t) - e^{-itx}\phi_j(t)}{it} dt$ in both models. We then tested $H_o : \varphi_4 = 0$ given $\varphi_1 = \varphi_2 = \varphi_3 = 0$ against $H_o : -\varphi_4 > 0$, in each modelusing the marginal probability (p-value) and:

$$F(0) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{(\varphi_4(-t) - \varphi_4(t))}{it} dt$$ where

$\varphi_4(t) = \prod_{l=1}^{n} \left(1 - 2(\lambda_{4l} - d_4)it\right)^{-\frac{1}{2}}$ and $\hat{d}_4$ was the calculated value of the fourth-order DW statistic.

Bayesian analyses: In the Bayes formulation, the specification of the MDR-TB models was performed by assigning priors to all unknown parameters. We used our dataset of spatiotemporal-sampled clinical and environmental observations $X = [x_{1...}, x_n]$; whereby, each x_i for $i = 1...n$ was assumed to be distributed according to some distribution $p(x_i | \theta)$. The posterior probability $Pr(M | D)$ of the MDR-TB models (i.e., *M*) was given by the sampled clinical and environmental data (i.e., *D*) which was given by Bayes' theorem: $Pr(M | D) = \frac{Pr(D | M)Pr(M)}{Pr(D)}$. Given a model selection problem in which we have to choose between two models, on the basis of observed data *D*, the plausibility of the two different models M_1 and M_2, parameterized by model parameter vectors θ_1 and θ_2 which was assessed by the Bayes factor *K* given by

$$K = \frac{Pr(D | M_1)}{Pr(D | M_2)} = \frac{\int Pr(\theta_1 | M_1)Pr(D | \theta_1, M_1)d\theta_1}{\int Pr(\theta_2 | M_2)Pr(D | \theta_2, M_2)d\theta_2}$$

where $Pr(D|M_i)$ was called the marginal likelihood for model *i*.

In this research θ was a parameter that was unknown and thus had to be inferred from the sampled georeferenced data. Our Bayesian procedure began by assuming that θ was distributed according to some prior distribution $p(\theta | \alpha)$, where the parameter α was a hyperparameter. The joint probability was then generated using: $p(X | \theta) = p(x_1, ..., x_n | \theta) = \prod_{i=1}^{n} p(x_i | \theta)$, whereby, the equations $p(X | \theta, \alpha) = p(X | \theta)$ and $p(x_i | \theta, \alpha) = p(x_i | \theta)$ were conditionally independent of the hyperparameter. We assumed that the two quantities are related by their conditional probability.

This conditional probability (i.e., likelihood function) was dependent on the modality in both models. The conditional probability of an event A assuming that B has occurred, denoted p(A|B), equals $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, which can be proven directly using $P(A \mid B)P(B) = P(A \cap B)$ which can be generalized to $P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B)$. The estimate was computed as a function of the posterior density.

In Bayesian statistics, the posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence is taken into account. Similarly, the posterior probability distribution is the distribution of an unknown quantity, treated as a random variable, conditional on the evidence obtained from an experiment or survey (Cressie 1993). In this reseach, the posterior probability is the probability of the sampled MDR-TB time series dependent parameters θ provided the evidence $C : p(\theta \mid X)$.

This value contrastses with the likelihood function, which in our model was provided by the probability of the evidence given the MDR-TB parameters: $p(X \mid \theta)$. In our model the probability distribution function was represented by $p(\theta)$ and the sampled observations X with the likelihood $p(X \mid \theta)$. Next, then the posterior probability was defined as $\Pr(\theta \mid X) = \dfrac{\Pr(\theta)\Pr(X \mid \theta)}{\Pr(X)}$.

The posterior probability of the MDR-TB model was then written in the memorable form as Posterior probability $\propto$ Prior probability $\times$ Likelihood .
The posterior probability distribution of one sampled MDR-TB random variable gave the value of another can be calculated with Bayes' theorem by multiplying the prior probability distribution by the likelihood function, and then dividing by the normalizing constant, as follows:

$$f_{X \| Y=y}(x)$$
$$= \frac{f_X(x) L_{X \| Y=y}(x)}{\int_{-\infty}^{\infty} f_X(x) L_{X \| Y=y}(x)\, dx}$$

which provided the posterior probability density function for a random variable X given the data $Y = y$, where: $f_X(x)$ was the prior density of X, $L_{X \mid Y=y}(x) = f_{Y \mid X=x}(y)$ was the likelihood function as a function of x, $\int_{-\infty}^{\infty} f_X(x) L_{X \mid Y=y}(x)\, dx$ was the normalizing constant, and $f_{X \mid Y=y}(x)$ was the posterior density of X as provided by the sampled MDR-TB data $Y = y$.

The specification of a prior density in addition to the likelihood function Bayesian inference then determined the posterior distribution of the sampled MDR-TB parameters $p(\theta \mid \mathbf{X}, \alpha)$ using

$$
\begin{aligned}
p(\theta \mid \mathbf{X}, \alpha) &= \frac{p(\theta, \mathbf{X}, \alpha)}{p(\mathbf{X}, \alpha)} \\
&= \frac{p(\theta, \mathbf{X}, \alpha)}{\int_{\theta} p(\theta, \mathbf{X}, \alpha) d\theta} \\
&= \frac{p(\mathbf{X} \mid \theta, \alpha)p(\theta \mid \alpha)}{\int_{\theta} p(\mathbf{X} \mid \theta, \alpha)p(\theta \mid \alpha)\,\mathrm{d}\theta} \\
&= \frac{p(\mathbf{X} \mid \theta)p(\theta \mid \alpha)}{\int_{\theta} p(\mathbf{X} \mid \theta)p(\theta \mid \alpha)\,\mathrm{d}\theta} \\
&= \frac{[\prod_{i=1}^{n} p(x_i \mid \theta)]p(\theta \mid \alpha)}{\int_{\theta}[\prod_{i=1}^{n} p(x_i \mid \theta)]p(\theta \mid \alpha)\,\mathrm{d}\theta}
\end{aligned}
$$

In this research we defined the deviance as $D(\theta) = -2\log(p(y \mid \theta)) + C$, where y was the sampled MDR-TB data, $p(y \mid \theta)$ was the likelihood function and C was a constant. The expectation $\bar{D} = \mathbf{E}^{\theta}[D(\theta)]$ was used a measure of how well the MDR-TB models fit the sampled data.

The residual revealed that the larger the expectation value, the worse the fit. The effective number of parameters for both models was then computed as $p_D = \bar{D} - D(\bar{\theta})$, where θ was the expectation of θ. The DIC was calculated in both models as $DIC = p_D + \bar{D}$.

We then used PROC MCMC for generating the multivariate density functions in the Bayesian estimation analysis. In PROC MCMC we used the logarithm of LOGMPDFWISHART for determining the Wishart distribution and the logarithm LOGMPDFIWISHART for the inverted-Wishart distribution. We let x be an *n*-dimensional random vector with mean vector μ and covariance matrix Σ. The density was

$$pdf(x, \mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n |\Sigma|}}$$ where

$|\Sigma|$ was the determinant of the covariance matrix Σ. The density function from the Wishart distribution in the MDR-TB models was:

$$pdf(x, \mu, \Sigma) = \frac{1}{C_n(\mu)}|\Sigma|^{-\frac{1}{2}}|x|^{\frac{\mu-n-1}{2}}\exp\left(-\frac{1}{2}tr\left(\Sigma^{-1}x\right)\right)$$

with $\mu > n$, and the trace of the square matrices A was:

$$tr(A) = \sum_i a_{ii}, \quad C_n(\mu) = 2^{\frac{\mu n}{2}}\Gamma_n\left(\frac{\mu}{2}\right),$$

Additionally, the

$$\Gamma_n(z) = \pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma\left(z - \frac{i-1}{2}\right)$$

density function from the inverse-Wishart distribution was

$$pdf(x, \mu, \Sigma) = \frac{1}{D_n(\mu)}|\Sigma|^{\frac{\mu-n-1}{2}}|x|^{-\frac{\mu}{2}}\exp\left(-\frac{1}{2}tr\left(\Sigma x^{-1}\right)\right)$$

for $\mu > 2n$, and $D_n(\mu) = 2^{\frac{(\mu-n-1)}{2}} \Gamma_n\left(\frac{\mu-n-1}{2}\right)$ for both models.

The marginal and conditional distributions from the inverse Wishart-distributed matrix were then further quantified using $A \sim W^{-1}(\psi, m)$. We partitioned the matrices for determining if Ψ was conformable with each other using: $A = \begin{bmatrix} A_{11} & A_{22} \\ A_{21} & A_{22} \end{bmatrix}, \Psi = \begin{bmatrix} \Psi_{11} & \Psi_{22} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$ where A_{ij} and Ψ_{ij} were $p_i \times p_j$ matrices. We then determined if A_{11} was independent of $A_{11}^{-1}A_{12}$ and $A_{22 \cdot 1}$, when $A_{22 \cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ was the Schur complement of A_{11} in $A_{11} \sim W^{-1}(\Psi_{11}, m - p_2)$ employing $A_{11}^{-1}A_{12} | A_{(22 \cdot 1)} \sim MN\left(p_1 x p_2\right)$ when of $\left(\psi_{11}^{-1}\psi_{12}, A_{(22 \cdot 1)} \otimes \psi_{11}^{-1}\right)$ $MN_{pxq}(\cdot, \cdot)$ was a matrix normal distribution generated from the spatiotemporal-sampled clinical and environmental parameters; and, $A_{22 \cdot 1} \sim W^{-1}(\Psi_{11}, m)$. In linear algebra and the theory of matrices, the Schur complement of a matrix block (i.e., a submatrix within a larger matrix) is commonly defined using $p \times p$, $p \times q$, $q \times p$ and $q \times q$ matrices, where D is invertible and $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ so that M is a $(p+q) \times (p+q)$ matrix(Cressie 1993).If the sampled clinical and environmental MDR-TB observations $X = x_1, \ldots, x_n$ were independent p-variety Gaussian variables drawn from a distribution, then the conditional distribution had a $W^{-1}(A + \Psi, n + m)$ distribution, where $A = XX^T$ is n times the sample covariance matrix. Because the prior and posterior distributions are the same family, the inverse Wishart distribution was the conjugate to the multivariate Gaussian generated from the sampled georeferenced MDR-TB explanatory predictor covariate coefficient estimates.

Model data input was also conducted in WinBUGS® but the number of chains had to bespecified before compilation. WinBUGS® is statistical software used for Bayesian modeling which incorporates an iterative estimation algorithm that starts from arbitrary initial values that can be generated from priors based on frequentist estimates (Gilks 1996). A Markov chain was generated using a sequence of s $X_1, X_2, X_3, \ldots$ with the Markov property, namely that, given the present state, the future and past states were independent. The following definition appliesn $\mathbb{R}^n$ -valued stochastic process $X = (X_t : t \in I)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to possess the Markov property if, for each $A \in \mathcal{B}(\mathbb{R}^n)$ and s,t$\in$I,s<t $\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | \sigma(X_s))$, where $\{\mathcal{F}\}_{t \in I}$ is the natural filtration and $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel sigma-algebra on $\mathbb{R}^n$.

In the case that the process takes discrete values and is indexed by a discrete time, this was reformulated as follows; $\begin{aligned} &\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1} \ldots X_0 = x_0) \\ &= \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \end{aligned}$ such that that $X = (X_t : t \geq 0)$ was a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with natural filtration $\{\mathcal{F}_t\}_{t \geq 0}$. Then X is said to have the strong Markov property if, for each stopping time τ, conditioned on the event $\{\tau < \infty\}$, the process $X_{\tau+\cdot}$ (which maybe needs to be defined) is independent from $\mathcal{F}_\tau := \{A \in \mathcal{F} : \tau \cap A \in \mathcal{F}_t, t \geq 0\}$ and $X_{\tau + t} - X_\tau$ has the same distribution as X_t for each t ≥ 0 . The strong Markov property is a stronger property than the ordinary Markov property, since by taking the stopping time $\tau = t$, the ordinary Markov property can be deduced.

Alternatively, also the Markov property was formulated as follows; $E[f(X_t) | F_s] = E[f(X_t) | \sigma X_s]$ for all $t \geq s \geq 0$ and $f : \mathbb{R}^n \to \mathbb{R}$ bounded and measurable. Both the spatiotemporal MDR-TB model residual estimates revealed $\begin{aligned} &\Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\ &= \Pr(X_{n+1} = x | X_n = x_n). \end{aligned}$. In this research the probability of going from state i to state j in n time steps was $p_{ij}^{(n)} = \Pr(X_n = j | X_0 = i)$ and the single-step transition was $p_{ij} = \Pr(X_1 = j | X_0 = i)$ For quantifying the external values in the Markov chains we used $p_{ij}^{(n)} = \Pr(X_{k+n} = j | X_k = i)$ and $p_{ij} = \Pr(X_{k+1} = j | X_k = i)$. The n-step transition probabilities satisfied the Chapman–Kolmogorov equation, that for any k such that $0 < k < n$, $p_{ij}^{(n)} = \sum_{r \in S} p_{rj}^{(k)} p_{rj}^{(n-k)}$.

When the stochastic process under consideration is Markovian, the Chapman–Kolmogorov equation is equivalent to an identity on transition densities (Spiegelhalter 2002). In our Markov models we assumed that $i_1 < \ldots < i_n$. Then, because of the Markov property our MDR-TB models rendered, $\begin{aligned} &p_{i_1, \ldots, i_n}(f_1, \ldots, f_n) \\ &= p_{i_1}(f_1) p_{i_2; i_1}(f_2 | f_1) \ldots p_{i_n; i_{n-1}}(f_n | f_{n-1}) \end{aligned}$ where the conditional probability $p_{i;j}(f_i | f_j)$ was the transition probability between the times $i > j$

In this research the Chapman–Kolmogorov equation took the form $p_{i_3; i_1}(f_3 | f_1) = \int_{-\infty}^{\infty} p_{i_3; i_2}(f_3 | f_2) p_{i_2; i_1}(f_2 | f_1) df_2$ where S was the state space of the Markov chain in both modelse let $p_{i_1, \ldots, i_n}(f_1, \ldots, f_n)$ be the joint probability density function f the values of the random variables f_1 to f_n. Then, the Chapman–Kolmogorov equation generated by the sampled random variables was $p_{i_1, \ldots, i_{n-1}}(f_1, \ldots, f_{n-1}) = \int_{-\infty}^{\infty} p_{i_1, \ldots, i_n}(f_1, \ldots, f_{n-1}) df_n$ using a straightforward marginalization over the nuisance variables. Note, that we did not assume anything about the

temporal or any other ordering of the sampled clinical and environmental MDR-TB random variables in the equation, thus, the estimates were applied equally to the marginalization of any parameter. When the stochastic process under consideration was Markovian, the Chapman–Kolmogorov equation was equivalent to an identity on transition densities in the MDR-TB models. In the Markov chain setting, one assumed that $i_1 < ... < i_n$. Then, because of the Markov property,

$$p_{i_1,...,i_n}(f_1,...,f_n) = p_{i_1}(f_1) p_{i_2;i_1}(f_2 \mid f_1)...p_{i_n;i_{n-1}}(f_n \mid f_{n-1})$$

the conditional probability $p_{i;j}(f_i \mid f_j)$ was the transition probability between the times $i > j$ in the MDR-TB models The Chapman–Kolmogorov equation generated using the sampled MDR-TB data then took the form

$$p_{i_3;i_1}(f_3 \mid f_1) = \int_{-\infty}^{\infty} p_{i_3;i_2}(f_3 \mid f_2) p_{i_2;i_1}(f_2 \mid f_1) df_2 .$$ Our models revealed that when the probability distribution on the state space of a Markov chain was discrete and the Markov chain was homogeneous, the Chapman–Kolmogorov equations can be expressed in terms of (possibly infinite-dimensional) matrix multiplication, thus: $P(t+s) = P(t)P(s)$ where $P(t)$ was the transition matrix of jump t, (i.e., $P(t)$ is the matrix such that entry (i,j) contains the probability of the chain moving from state i to state j in t steps). Additionally it followed that to calculate the transition matrices of jump t, it was sufficient to raise the transition matrix of jump one to the power of t, that is $P(t) = P^t$ in both MDR-TB models. The marginal distribution $\Pr(X_n = x)$ was the distribution over states at time n in the residuals. The initial distribution was $\Pr(X_0 = x)$ in both models. The evolution of the process through each step was then described by

$$\Pr(X_n = j)$$
$$= \sum_{r \in S} p_{rj} \Pr(X_{n-1} = r)$$
$$= \sum_{r \in S} p_{rj}^{(n)} \Pr(X_0 = r)$$

Additionally, in this research we extended this analyses to show that that distance between the nth step transition probability and the invariant probability measure in the MDR-TB models was bounded by $p * [a + bg(x)]$ for the constant $a > b > 0$ and $p < 1$. The invariant was then used to obtain convergence rates to quantify the transition probabilities for autoregressive processes in the models using a random walk on a half line. In this research a random walk with reflecting zone on the nonnegative integers generated from the sampled predictor covariate coefficients was a Markov chain whose transition probabilities $q(x,y)$ were those of a random walk [i.e., $q(x,y) = p(y-x)$] which was outside a finite set $\{0,1,2,...,K\}$. As such, the distribution $q(x,\cdot)$ stochastically dominated $p(\cdot - x)$ for every $x \in \{0,1,2,...,K\}$. Under mild hypothesis, it is proved that

when $\sum xp_x \rangle 0$, the transition probabilities satisfy $q^n(x,y) \sim C_{xy} R^{-n} n^{-\frac{3}{2}}$ as $n \to \infty$, and when $\sum xp_x \rangle = 0$, $q^n(x,y) \sim C_{xy} n^{-\frac{1}{2}}$. We did so to extend and strengthen the MDR-TB models' residuals in countable state space. Inference for MCMC simulation techniques was then based on weighted OLS proposals and on latent utility representations of multi-categorical MDR-TB regression-based models.

We then let δ_{ij} denote the drug resistant indicator of drug j for subject i, $i = 1...,n$, and $j = 1,...4$.

We assumed that δ_{ij} is determined by a latent variable z_{ij}. Specifically, $\delta_{ij} = I(z_{ij} > 0), z_i = (z_{i1},...,z_{i4})^T \sim N(\mu_i, I_4)$ where $I(A)$ is equal to 1 if A is true. The parameter $\mu_i = (\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4})^T$ and I_4 is a 4 by 4 identity matrix. We decompose μ_{ij} as spatial random effects $f(x_i, y_i)$ and subject effect η_i, i.e., $\mu_{ij} = f_i(x_i, y_i) + \eta_i$, $f_i \sim GP(f_0, \sigma_j^2 \kappa)$, $\eta_i \sim N(0, \sigma^2)$ where x_i and y_i represent the latitude and the longitude. The spatial random effects f_i's are modeled using Gaussian Process. The prior distributions for the parameters are given below: $\sigma^2 \sim Inv - Gamma(a_0, b_0)$, $\sigma_j^2 \sim Inv - Gamma(a_0, b_0)$, $f_0 \sim GP(0, \sigma_0^2 \kappa)$ where $\kappa(a,b) = \exp(-\|a-b\|/\rho)$ and $\rho \sim Uniform(c,d)$. We decompose $\kappa(a,b) = \sum_{l=1}^{\infty} \lambda_l \phi_l(a) \phi_l(b)$ as $\sum_{l=1}^{L} \lambda_l \phi_l(a) \phi_l(b)$, then we have a representation of f_i for $j = 1,..., 4$. We used Gibbs sampling to make inferences about the posterior distribution of key parameters and the irspatial effects.

The estimations are carried out in the statistical software package R (Version 2.15.0).

3. Results

The data comprised of 785 individual observations of patients with tuberculosis, together with a battery of 25 attribute measures some of which were categorical and expanded into multiple indicator variables. One of the explanatory variables was number of people in an individual's home (NOP). Because the only population at risk count available was the NOP (i.e., no data were available for households with no members infected with tuberculosis), these figures were used to construct a standardized rate of infected individuals per 100. Some variable fell out of the model. (see Table 2) The correlation between quality of living indicators and employment status showed generally weak $(r < 0.09)$ correlations between employment variables and access to utilities as can be seen in Table 3. The only strong correlation was found to be between the availability of
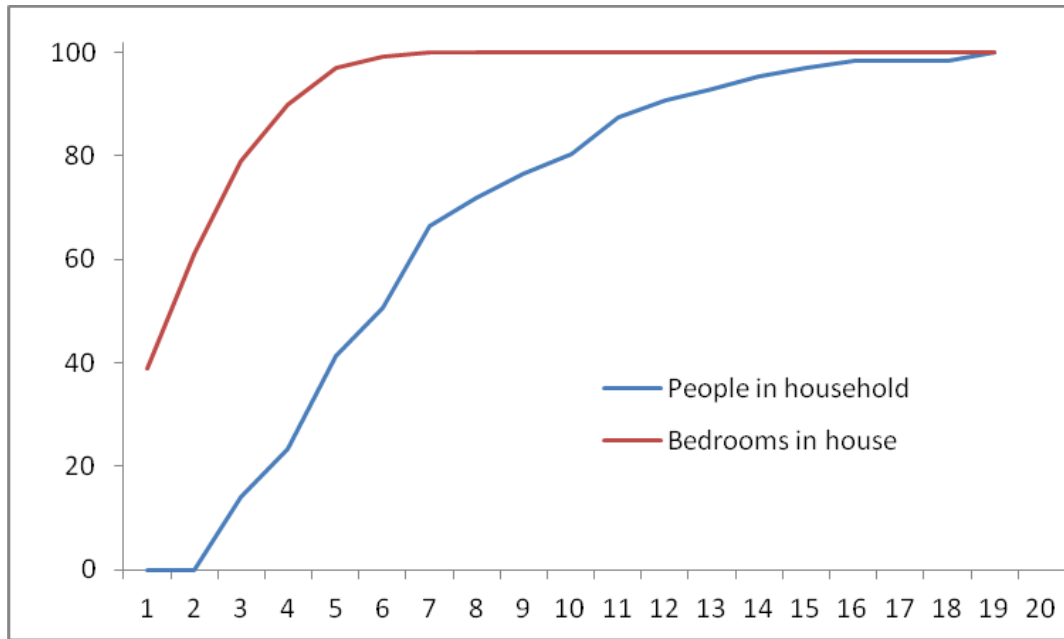
potable water and home waste connected to the public network $(r = 0.77)$. The correlation between these two variables and home access to electricity was weak $(r < 0.37)$.

Table 3. Correlation between various quality of living and employment status characteristics (p-values)

	Potable Water	Waste	Working Status[1]	Health Center Worker
Electricity	0.36 (2E-16)	0.29 (1E-14)	0.02 (8E-1)	-0.02 (2E-2)
Potable Water		0.77 (2E-16)	0.01 (8E-1)	0.04 (5E-5)
Waste			-0.01 (8E-1)	0.05 (2E-7)
Working Status				-0.08 (2E-2)

[1] Students, unemployed, and retired individuals were grouped together into the unemployed category.

The correlation between the number of people living in a home and the number of bedrooms was found to be 0.51. It is generally accepted that the overcrowding standard is more than 2 people per bedroom (Blake, Kellerson, & Simic, 2007). The same report also states that although neither a house with 4 people and 2 bedrooms nor one with 8 people and 4 bedrooms would be considered crowded by definition, the latter house is considered more crowded. Table 4 shows the frequency of overcrowding by number of bedrooms per house. Of the total study population, 16.8% lives in overcrowded conditions compared to 2.7% in the U.S. in 2005(Blake et al., 2007). However, it should be noted that the most frequent form of overcrowding occurs when the number of people living in a house is one or two more than the threshold necessary to qualify living conditions as overcrowded. While an extra person living in a 1- or 2-bedroom house increases overcrowding significantly, the effect of an additional resident is minimized as the number of bedrooms in a house increases. Furthermore, as Table 5 shows, the frequency of overcrowded households decreases quickly as the number of rooms in a house increases due to the fact that there are very few households containing more than 9 individuals. Figure 3 shows the cumulative distribution of overcrowding by people and bedrooms per house. It can be seen that more than 50% of overcrowded conditions occur in houses that have 1 or 2 bedrooms while the same number of overcrowded conditions occur in houses that have 7 or less people. Therefore, overcrowding is largely a problem of smaller dwellings with slightly above overcrowding conditions rather than grossly overcrowded small dwellings or slightly overcrowded large houses.

Table 4. Frequency of households with a number of people (across) living in a given number of bedrooms (down)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20
1	21	8	18	12	9	4	6	0	0	0	0	1	0	0	0	0	0	0	0
2	3	7	20	28	14	8	4	1	0	0	0	0	0	0	0	0	0	0	1
3	0	9	15	19	25	11	10	6	1	2	1	1	0	1	1	0	0	0	0
4	2	5	9	18	19	18	17	13	5	3	3	1	1	1	0	0	0	0	0
5	2	3	9	18	17	17	12	9	5	2	5	1	2	0	0	0	0	0	1
6	0	3	5	11	18	21	19	8	8	2	1	1	0	1	1	1	0	0	0
7	0	1	2	8	10	8	12	8	5	1	3	3	0	1	0	1	0	0	0
8	0	0	2	4	8	3	7	5	7	3	4	1	0	0	3	0	0	0	0
9	0	0	1	2	0	2	2	1	4	3	2	0	1	1	0	0	1	0	0
10	0	0	0	0	2	4	4	0	2	0	1	2	0	0	1	0	0	0	0
11	0	0	1	1	1	1	3	2	1	1	0	0	1	0	3	0	0	0	1
12	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
20	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1

Figure 3. Cumulative overcrowding distribution by people in household (blue) and bedrooms in house (red)
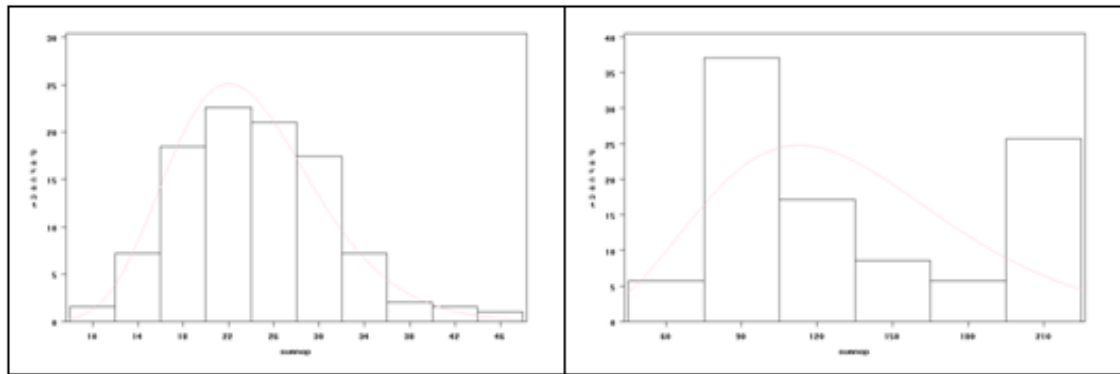

Figure 4. Histograms of geographic group sizes with corresponding superimposed gamma distributions for (a): the high MDR-TB prevalence cluster; and,(b): the high MDR-TB prevalence cluster

Table 5. Overcrowding (more than 2 people per bedroom) by number of bedrooms

Number of Bedrooms	Overcrowded, n (%)
1	50 (63.3%)
2	28 (32.6%)
3	23 (22.6%)
4	14 (12.2%)
5	9 (8.7%)
6	3 (3%)
7	1 (1.6%)
8+	0 (0.0%)

Cement and brick materials are the most commonly used building materials (85.0%) followed by other (7.7%) for houses in this study (see Table 6). Not surprisingly, cement and brick houses are the most commonly used building materials for permanent houses (65.8%),proper (36.5%) and family (33.9%) homes. The most commonly used material for squatter and temporary dwellings is rush mat. Only 9 houses are built from adobe while other building materials were used to build a large amount of houses (58 total). The majority of houses are permanent structures (77.4%) and only 9 houses are defined as squatter home types.

Table 6. Frequency of home type, building material, and address type

	Building material				Address type	
	Cement and bricks, n(%)	Adobe, n (%)	Rush mat, n (%)	Other, n (%)	Permanent, n (%)	Temporary, n (%)
Home type						
Proper	279 (36.5%)	5 (0.6%)	25 (3.2%)	36 (4.7%)	341 (44.6%)	4 (0.5%)
Rented	110 (14.4%)	2 (0.3%)	2 (0.3%)	2 (0.3%)	21 (2.8%)	95 (12.4%)
Family Home	259 (33.9%)	2 (0.3%)	13 (1.7%)	18 (2.3%)	227 (29.7%)	65 (8.5%)
Squatter	2 (0.3%)	0 (0.0%)	7 (0.9%)	2 (0.3%)	2 (0.3%)	9 (1.2%)
Address type						
Permanent	503 (65.8%)	5 (0.7%)	33 (4.3%)	50 (6.6 %)		
Temporary	147 (19.2%)	4 (0.5%)	14 (1.8%)	8 (1.1%)		

In order to compute meaningful standardized rates, the individual data were aggregated geographically into geographic groups. These groups were constructed with a hierarchical cluster analysis using the Lance-Williams flexible-beta method (SAS PROC CLUSTER; beta = -100). The model revealed that X and Y were the highest and lowest MDR-TB cluster stratified by prevalence rate in the SJL study site. The predictor variables used for clustering were: longitude, latitude, and altitude; the criteria involved the centrographic measures of spatial mean and standard distance. (Table 7) Histograms for these groups constructed in SAS were roughly characterized by a gamma distribution with respective K-S goodness-of-fit probabilities of 0.141 and 0.006. A goodness-of-fit criterion can be a measure of within-cluster homogeneity versus among-cluster heterogeneity, often measured by the distance of each plot to the center of the cluster to which it belongs, compared to the average distance to other clusters (Kulldorff et al., 2005).

Table 7. MDR-TB parameter estimates with clusterings used to specify repeated covariate measures

covariate	With no random effects		With random effects	
	Parameter estimate	Standard error	Parameter estimate	Standard error
High MDR-TB prevalence cluster				
intercept	-1.5555	0.1417	-1.5487	0.1228
age	0.0126	0.0032	0.0124	0.0028
# bedrooms	-0.1459	0.0127	-0.1509	0.0108
time on the job	0.2222	0.0347	0.2449	0.0303
rental home	0.6090	0.0774	0.5230	0.0675
single	0.3334	0.0693	0.2743	0.0600
random effects	0		1	
scale	3.4338		2.9112	
Low MDR-TB prevalence cluster				
intercept	-1.4404	0.1507	-1.4119	0.1463
age	0.0133	0.0032	0.0121	0.0031
# bedrooms	-0.1420	0.0128	-0.1415	0.0123

All aggregated attribute variables were converted to percentages, and all aggregated interval/ratio variables were converted to means.

We then derived MDR-TB-related OLS Estimators using $y_i = \beta_0 + \beta_1 x_i + u_i$ where y_i was a dependent variable, x_i was an independent right-hand side (RHS) variable, u_i was the error term (unobservable), β_0 *and* β_1 were coefficients. The ordinary least squares procedure minimizes the sum of squares error (SSE) (Hosmer and Lemeshew 2000). The minimization problem was given as follows:

$$Minimize\ SSE\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$$

$$= \sum_{i=1}^{n} \widehat{u}_i^2 = \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2$$

$$F.O.C.\quad \frac{\partial SSE}{\partial \widehat{\beta}_0} \qquad (1\text{-}a)$$

$$= 2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)(-1)$$

$$\frac{\partial SSE}{\partial \widehat{\beta}_1} = 2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)(-x_i) = 0 \qquad (1\text{-}b)$$

Thereafter, we derived the MDR-TB related OLS estimate of $\beta_0\left(\widehat{\beta}_0\right)$. We then divided equation (1-a) by -2 which rendered $\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\widehat{\beta}_0 - \sum_{i=1}^{n}\widehat{\beta}_1 x_i = 0$. This equation was quantitated as $\sum_{i=1}^{n} y_i - n\widehat{\beta}_0 - \sum_{i=1}^{n}\widehat{\beta}_1 x_i = 0$ by dividing both sides by *n*. We then arranged $\overline{y} - \widehat{\beta}_0 - \widehat{\beta}_1 \overline{x} = 0$ and $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$ to solve the OLS estimate $\widehat{\beta}_0$. We then attained the OLS estimate of $\beta_1, \left(\widehat{\beta}_1\right)$ which we then rearranged as

$$\sum_{i=1}^{n} y_i = n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_{i=1}^{n} x_i \text{ for (1-a) (3.1) see Table 1.}$$

We then rearranged equation (1-b) as

$$\sum_{i=1}^{n} x_i y_i = \widehat{\beta}_0 \sum_{i=1}^{n} x_i + \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 \qquad (3.2)$$

We multiplied equation (3.1) by the sum of x_i and equation (3.2) by *n*. Subsequently we attained

$$\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = n\widehat{\beta}_0 \sum_{i=1}^{n} x_i + \widehat{\beta}_1\left[\sum_{i=1}^{n} x_i\right]^2$$

$$n\sum_{i=1}^{n} x_i y_i = n\widehat{\beta}_0 \sum_{i=1}^{n} x_i + n\widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 .$$ We then employed solved these equations which then rendered:

$$n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = n\widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \widehat{\beta}_1\left[\sum_{i=1}^{n} x_i\right]^2$$ Solving for this MDR-TB oriented predictive equation also yielded the OLS estimate $\beta_1\left(\widehat{\beta}_1\right): \widehat{\beta}_1 = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left[\sum_{i=1}^{n} x_i\right]^2}$.

Then:

$$\hat{\beta}_1 = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left[\sum\limits_{i=1}^{n} x_i\right]^2} = \frac{n\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{n\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \quad (3.3)$$

$$= \frac{Cov(x,y)}{Var(x)}$$

The numerator of equation (3.3) was then re-written as follows:

$$n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)$$

$$= n\left(\sum_{i=1}^{n} x_i y_i\right) - n\overline{x}\left(\sum_{i=1}^{n} y_i\right) - n\overline{y}\left(\sum_{i=1}^{n} x_i\right) - n^2 \overline{x}\,\overline{y}$$

$$= n\left(\sum_{i=1}^{n} x_i y_i\right) - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i \quad (3.4)$$

$$= n\left(\sum_{i=1}^{n} x_i y_i\right) - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i$$

The denominator of equation (3.4) was then be re-written as follows:

$$n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = n\sum_{i=1}^{n} x_i^2 - 2n\overline{x}\left(\sum_{i=1}^{n} x_i\right) + n^2 \overline{x}^2$$

$$= n\sum_{i=1}^{n} x_i^2 - 2\left(\sum_{i=1}^{n} x_i\right)^2 + \left(\sum_{i=1}^{n} x_i\right)^2$$

$$= n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$$

The MDR-TB-reated regression model with autocorrelated disturbances was as follows: $y_t = x_t' \beta + v_t$, $v_t = \varepsilon_t - \varphi_1 v_{t-1} - \dots - \varphi_m v_{t-m}$ and $\varepsilon_t N\left(0, \sigma^2\right)$. In these equations, y_t are the dependent values, x_t is a column vector of regressor variables, β is a column vector of structural parameters, and ε_t is normally and independently distributed with a mean of 0 and a variance of σ^2. Note that in this parameterization, the signs of the autoregressive parameters are reversed from the parameterization documented in most of the literature.

PROC AUTOREG offers four estimation methods for the autoregressive error model. The default method, Yule-Walker (YW) estimation, is the fastest computationally. The Yule-Walker method used by PROC AUTOREG is described in Gallant and Goebel (1976). Harvey (1981) calls this method the two-step full transform method. The other methods are iterated YW, unconditional least squares (ULS), and maximum likelihood (ML). The ULS method is also referred to as nonlinear least squares (NLS) or exact least squares (ELS).

We let φ represent the vector of the MDR-TB RELATED autoregressive parameters, $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_m)'$ and let the variance matrix of the error vector $v = (v_1, \dots, v_N)'$ be Σ, $E(vv') = \Sigma = \sigma^2 V$. If the vector of autoregressive parameters φ is known, the matrix V can be computed from the autoregressive parameters. Σ is then $\sigma^2 V$. Given Σ, the efficient estimates of regression parameters β can be computed using generalized least squares (GLS). The GLS estimates then yieldED the unbiased estimate of the variance σ^2.

The Yule-Walker method alternates estimation of β using generalized least squares with estimation of φ using the Yule-Walker equations applied to the sample autocorrelation function. The YW method starts by forming the OLS estimate of β. Next, φ is estimated from the sample autocorrelation function of the OLS residuals by using the Yule-Walker equations. Then V is estimated from the estimate of φ, and Σ is estimated from V and the OLS estimate of φ^2. The autocorrelation corrected estimates of the regression parameters β are then computed by GLS, using the estimated Σ matrix. These are the Yule-Walker estimates.

If the ITER option is specified, the Yule-Walker residuals are used to form a new sample autocorrelation function, the new autocorrelation function is used to form a new estimate of φ and V, and the GLS estimates are recomputed using the new variance matrix. This alternation of estimates continues until either the maximum change in the $\hat{\varphi}$ estimate between iterations is less than the value specified by the CONVERGE= option or the maximum number of allowed iterations is reached. This produces the iterated Yule-Walker estimates. Iteration of the estimates may not yield much improvement.

The Yule-Walker equations, solved to obtain $\hat{\varphi}$ and a preliminary estimate of σ^2, are $R\varnothing = -r$ Here $r = (r_1, \dots, r_m)'$, where ri is the lag i sample autocorrelation. The matrix R is the Toeplitz matrix whose i,j th element is $^r|i-j|$. If you specify a subset model, then only the rows and columns of R and r corresponding to the subset of lags specified are used. If the BACKSTEP option is specified, for purposes of significance testing, the matrix [Rr] is treated as a sum-of-squares-and-crossproducts matrix arising from a simple regression with N-k observations, where k is the number of estimated parameters.

The Unconditional Least Squares and Maximum Likelihood Methods then defined the transformed error, e as e=L⁻¹en where in the MDR-TB model.. The unconditional sum of squares for the model, S, is $S = n'V^{-1}n = e'e$. The ULS estimates are computed by minimizing S with respect to the sampled parameters β and φ_i. The full log likelihood function for the autoregressive error model was

$$l = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln\left(|V|\right) - \frac{S}{2\sigma^2}$$

where |V| denotes determinant of V. For the ML method, the likelihood function is maximized by minimizing an equivalent sum-of-squares function. Maximizing l with

respect to σ^2 (and concentrating σ^2 out of the likelihood) and dropping the constant term $-\frac{N}{2}\ln(2\pi)+1-\ln(N)$ produces the concentrated log likelihood function $l_c = -\frac{N}{2}\ln\left(S\,|\,V\,|^{1/N}\right)$. Rewriting the variable term within the logarithm gives $S_{ml} = |L|^{1/N}\,e'e\,|L|^{1/N}$.

PROC AUTOREG then computed the ML estimates by minimizing the objective function $S_{ml} = |L|^{1/N}\,e'e\,|L|^{1/N}$. The maximum likelihood estimates may not exist for some data sets (Anderson and Mentz; 1980). The sample autocorrelation function is computed from the structural residuals or noise $n_t = y_t - x_t'b$, where b is the current estimate of β. The sample autocorrelation function of the MDR-TB risk model was the sum of all available lagged products of n_t of order j divided by $\ell + j$, where ℓ was the number of such products.. In this research the Toeplitz matrix of autocorrelations, R, was at least positive semidefinite. If there are missing values, these autocorrelation estimates of r can yield an R matrix that is not positive semidefinite (Griffith 2003). If such estimates occur, a warning message is printed, and the estimates are tapered by exponentially declining weights until R is positive definite.

The calculation of V from φ for the general AR(M) model was complicated, and the size of V was dependent on the number of clinical observations. Instead of actually calculating **V** and performing GLS in the usual way, in practice a Kalman filter algorithm was used to transform the data and compute the GLS results through a recursive process. In all of the estimation methods, the original data was transformed by the inverse of the Cholesky root of V. We let L denote the Cholesky root of V— that is, V=LL' with L lower triangular. For an AR(m) model, L⁻¹ is a band diagonal matrix with *m* anomalous rows at the beginning and the autoregressive parameters along the remaining rows (Griffith 2003). Since there was no missing values, after the first m-1 observations the MDR-TB data was transformed as $z_t = x_t + \varphi_1 x_{t-1} + ... + \varphi_m x_{t-m}$.

The transformation was carried out using a Kalman filter, and the lower triangular matrix L ws never directly computed. The Kalman filter algorithm, as it applies here, is described in Harvey and Phillips (1979) and Jones (1980). Although L was not computed explicitly, for ease of presentation the model terms were based on L. If there are missing values, then the submatrix of L consisting of the rows and columns with nonmissing values is used to generate the transformations (Griffith 2003). The ULS and ML estimates were then qunatiated employing a Gauss-Newton algorithm to minimize the sum of squares and maximize the log likelihood, respectively in the first-order MDR-TB model. The relevant optimization was performed simultaneously for both the regression and AR parameters. The OLS estimates of β and the Yule-Walker estimates of φ were used as starting values for these methods. The Gauss-Newton algorithm required the derivatives of e or $|L|^{1/N}$ e with respect to the sampled time series MDR-TB CLINICAL parameters. The

derivatives with respect to the parameter vector β were $\frac{\partial e}{\partial \beta'} = -L^{-1}X\frac{\partial |L|^{1/N}\,e}{\partial \beta'} = -|L|^{1/N}L^{-1}X$. These derivatives were computed by the transformation described previously. The derivatives with respect to φ were computed by differentiating the Kalman filter recurrences and the equations for the initial conditions.

For the Yule-Walker method, the estimate of the error variance, s^2, was the error sum of squares from the last application of GLS, divided by the error degrees of freedom (e.g., number of clincal MDR-TB observations *N* minus the number of free parameters). The variance-covariance matrix for the components of B was taken as $s^2\left(X'V^{-1}X\right)^{-1}$ for the Yule-Walker method. For the ULS and ML methods, the variance-covariance matrix of the parameter estimates was computed as $s^2\left(J'J\right)^{-1}$. For the ULS method, J was the matrix of derivatives of e with respect to the sampled parameters. For the ML method, J was the matrix of derivatives of $|L|^{1/N}$ e divided by $|L|^{1/N}$. Since φ was known, the estimate of the variance-covariance matrix of B was $s^2\left(X'V^{-1}X\right)^{-1}$. Park and Mitchell (1980) investigated the small sample performance of the standard error estimates obtained from some of these methods. In particular, simulating an AR (1) model for the noise term, they found that the standard errors calculated using GLS with an estimated autoregressive parameter underestimated the true standard errors. These estimates of standard errors in the MDR-TB risk model are the ones calculated by PROC AUTOREG with the Yule-Walker method.

For ULS or ML estimation, the joint variance-covariance matrix of all the regression and auto regression clinical MDR-TB time series explanatory parameters was computed The estimates of the standard errors calculated with the ULS or ML method took into account the joint estimation of the AR and the regression parameters and gave more accurate standard-error values than the YW method.. At the same values of the autoregressive parameters, the ULS and ML standard errors will always be larger than those computed from Yule-Walker (Griffith 2003). However, simulations of the models used by Park and Mitchell (1980) suggest that the ULS and ML standard error estimates can also be underestimates. Caution is advised, especially when the estimated autocorrelation is high and the sample size is small(Griffith 2003). For the Yule-Walker method, the variance-covariance matrix wascomputed only for the regression parameters.

The Yule-Walker estimation method is not directly appropriate for estimating models that include lagged dependent variables among the regressors. Therefore, the maximum likelihood method is the default when the LAGDEP or LAGDEP= option is specified in the MODEL statement. However, when lagged dependent variables are used, the maximum likelihood estimator is not exact maximum likelihood but is conditional on the first few values of the dependent variable.

In this research, ie_t was the <u>residual</u> associated with the observation at time t, then the Durbin Watson test statistic was $d = \dfrac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$, where T was the number of georeferenced clinical MDR-TB observations. Since in the autoregressive MDR-TB model d was approximately equal to $2(1 - r)$, when r was the sample autocorrelation of the residuals, $d = 2$ indicated no autocorrelation. The value of d always lies between 0 and 4 (Cressie 1993). In the model residual forecasts the Durbin–Watson statistic was substantially less than 2, indicating positive serial correlation. As a rough rule of thumb, if Durbin–Watson is less than 1.0, there may be cause for alarm. Small values of d indicate successive error terms are, on average, close in value to one another, or positively correlated (Griffith 2003). If $d > 2$, successive error terms are, on average, much different in value from one another, i.e., negatively correlated (Cressie 1993). In regressions, this can imply an underestimation of the level of statistical significance.

To validate for positive autocorrelation at significance α, the test statistic d in the MDR-TB model was compared to lower and upper critical values ($d_{L,\alpha}$ and $d_{U,\alpha}$): If $d < d_{L,\alpha}$, there was statistical evidence that the error terms are delineating positive spatial autocorrelation (PSA) (aagragtion of similar values in geospace) (Griffith 2003). If $d > d_{U,\alpha}$, there is **no** statistical evidence that the error terms are positively autocorrelated while applying $d_{L,\alpha} < d < d_{U,\alpha}$, the test would be inconclusive (Cressie 1993). Positive serial correlation is serial correlation in which a positive error for one observation increases the chances of a positive error for another observation. To validate our resultswe tested for negative autocorrelation at significance α. Thereafter the test statistic $(4 - d)$ was compared to lower and upper critical values ($d_{L,\alpha}$ and $d_{U,\alpha}$). If $(4 - d) < d_{L,\alpha}$, in the MDR-TB model then there was statistical evidence that the error terms are negatively autocorrelated. If $(4 - d) > d_{U,\alpha}$, in the model thenwe assumed was no statistical evidence and that the error terms were representing negative spatial autocorrelation (NSA) (i.e., aggregation of dissimilar values in geospace)while $d_{L,\alpha} < (4 - d) < d_{U,\alpha}$, the test is inconclusive (Cressie 1993). Negative serial correlation in an MDR-TB model implies that a positive error for one observation increases the chance of a negative error for another observation and a negative error for one observation increases the chances of a positive error for another (Jacob et al. 2013). In our model, the critical values, $d_{L,\alpha}$ and $d_{U,\alpha}$, varied by level of significance (α), the number of observations, and the number of predictors in the regression equation. Further if the design matrix X of the regression is known, exact critical values for the distribution of d under the null hypothesis of no serial correlation can be calculated. Under the null hypothesis d is distributed as

$\dfrac{\sum_{i=1}^{n-k} V_i \xi_i^2}{\sum_{i=1}^{n-k} \xi_i^2}$ where n are the number of sampled MDR-TB observations and k the number of regression variables; the ξ_i are independent standard normal random variables; and the V_i are the nonzero eigenvalues of $\left(I - X(X^T X)^{-1} X^T \right) A$ where A is the matrix that transforms the residuals into the d statistic, i.e. $d = e^T A e$. We then computed the Durbin h static and the MDR-TB parmater estimates.

The calculated Durbin h statsic for the autoregressively predicted MDR-TB parameter estimatess:

Statistic	Value	Prob	Label
Durbin h	2.7814	0.0027	Pr > h

The first-order MDR-TB parameter estimation then was qunatitated as:

Parameter Estimates				
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	1.5742	0.9300	1.69	0.0999
ylag	1	0.9376	0.0510	18.37	<.0001

In this research the pdf was

$$p(B \mid \Psi, m) = \frac{|\Psi|^{m/2} |B|^{-(m+p+1)/2} \exp\left(-tr\left(\Psi B^{-1}\right)/2\right)}{2^{mp/2} \Gamma_p(m/2)}$$

where $\Psi = \Sigma^{-1}$ and $\Gamma_p(\cdot)$ was the multivariate gamma function. The multivariate Gamma function, [i.e., $\Gamma_p(\cdot)$], is a generalization of the Gamma function which is useful in multivariate statistics, appearing in the pdf of the inverse Wishart distributions (Spiegelhater et al. 2002). We also noticed that the gamma function had two equivalent expressions. One was $\Gamma_p(a) = \int_{s>0} \exp(-trace(S))|S|^{a-(p+1)/2} dS$ where $S > 0$ meaning S was positive-definite. The other one, was $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left[a + (1-j)/2\right]$ from which we determined the recursive relationships in the sampled MDR-TB predictors using $\Gamma_p(a) = \pi^{(p-1)/2} \Gamma(a) \Gamma_{p-1}\left(a - \frac{1}{2}\right)$. Thus, $= \pi^{(p-1)/2} \Gamma_{p-1}(a) \Gamma\left[a + (1-p)/2\right]$ $\Gamma_1(a) = \Gamma(a)$, $\Gamma_2(a) = \pi^{1/2} \Gamma(a) \Gamma(a - 1/2)$ and $\Gamma_3(a) = \pi^{3/2} \Gamma(a) \Gamma(a - 1/2) \Gamma(a - 1)$.

Marginal and conditional distributions from an inverse Wishart-distributed matrix were then generated. In this research $A \sim W^{-1}(\Psi, m)$ had an inverse Wishart distribution. We partitioned the matrices A and Ψ conformably with each other using $A = \begin{bmatrix} A_{11} & A_{22} \\ A_{21} & A_{22} \end{bmatrix}$, $\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{22} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$ where A_{ij} and Ψ_{ij} were $p_i \times p_j$ matrices. We obtained A_{11} which was

independent of $A_{11}^{-1}A_{12}$ and $A_{22\cdot1}$, where $A_{22\cdot1} = A_{22} - A_{21}A_{11}^{-1}A_{12}A_{12}$ was the Schur complement of A_{11} in A. Commonly, a finite element problem is split into non-overlapping sub-domains in predictive distribution models and the unknowns in the interiors of the sub-domains are eliminated. In this research, the remaining Schur complement system on the unknowns associated with sub-domain interfaces was solved by the conjugate gradient method. In mathematics, the conjugate gradient method is an algorithm for the numerical solution of particular systems of linear equations, namely those whose matrix is symmetric and positive-definite (Hosmer and lenaeshew 2000).

In this research, the conjugate gradient method was unstable with respect to the perturbations in both the MDR-TB models (e.g., most directions were not in practice conjugate, and the exact solution was never obtained). Fortunately, the conjugate gradient method can be used as an iterative method as it provides monotonically improving approximations X_κ to the exact solution which can reach the required tolerance after a relatively small number of iterations (Cressie 1993). Our improvement was linear and its speed was determined by the condition number $\kappa(A)$ of the system matrix A; where, the larger the $\kappa(A)$ the slower the improvement in both MDR-TB models. Since some of our $\kappa(A)$ were large, preconditioning was used to replace the original system $Ax - b = 0$ with $M^{-1}(Ax - b) = 0$ so that $\kappa(M^{-1}A)$ got smaller than $\kappa(A)$. In most cases, preconditioning is necessary to ensure fast convergence of the conjugate gradient method (Cressie 1993). In this research, the preconditioned conjugate gradient method for the MDR-TB models took the following form:

$r_0 = b - Ax_0 : z_0 : M^{-1}r_0 : p_0 := z_0 : k := 0$ repeat:
$X_{k+1} := X_k + a_k p_k : r_{k+1} := r_k - \alpha_k Ap_k$: if r_{k+1} was sufficiently smaller than exit loop end if $z_{k+1} := M^{-1}r_{k+1}$,

$\beta_k := \dfrac{z_{k+1}^T r_{k+1}}{z_k^T r_k}$, $P_{k+1} : z_{k+1} + \beta_k P_k$ and $k := k+1$. In our models the above formulation was equivalent for applying the conjugate gradient method without preconditioning to the system where $E^{-1}A(E^{-1})^T \hat{x} = E^{-1}b$ and where $EE^T = M$ and $\hat{x} = E^T x$. The preconditioner matrix M, was symmetric positive-definite and fixed, (i.e., stationary from iteration to iteration) in both models. We then compared the derived estimates of the number of iterations with empirical distribution functions from the larval habitat model residual estimates. The empirical distribution function, is the cumulative distribution function associated with the empirical measure of the sample (Cressie 1993).

We also obtained $A_{11} \sim W^{-1}(\Psi_{11}, m - p_2)$, $A_{11}^{-1}A_{12} \mid A_{22\cdot1} \sim MN_{p_1 \times p_2}\left(\Psi_{11}^{-1}\Psi_{12}, A_{22\cdot1} \otimes \Psi_{11}^{-1}\right)$, where $MN_{p \times x}(\cdot, \cdot)$ was a matrix normal distribution and

$A_{22\cdot1} \sim W^{-1}(\Psi_{22\cdot1}, m)$. A conjugate distribution was then determined to make inference about a covariance matrix Σ whose prior $p(\Sigma)$ had a $W^{-1}(\Psi, m)$ distribution. The models revealed that the sampled clinical and environmental MDR-TB observations were independent p-variate Gaussian variables drawn from a $N(0, \Sigma)$ distribution in both models. The conditional distribution $p(\Sigma \mid X)$ of the sampled clinical and environmental data had a $W^{-1}(A + \Psi, N + M)$ distribution, where $A == XX^T$ was the number of sampled predictors times the sample covariance matrix. Due to its conjugacy to the multivariate Gaussian, it was possible to integrate out the Gaussian-based parameters using:

$$P(X \mid \Psi, m) = \int P(X \mid \Sigma) P(\Sigma \mid \Psi, m) d\Sigma$$

$$= \frac{|\Psi|^{\frac{m}{2}} \Gamma_p\left(\frac{m+n}{2}\right)}{\pi^{\frac{np}{2}} |\Psi + A|^{\frac{m+n}{2}} \Gamma_p\left(\frac{m}{2}\right)}.$$ The variance

of the diagonal used the same formula in the MDR-TB models with $i = j$, which was then simplified to:

$$\mathrm{var}(b_{ii}) = \frac{2\psi_{ii}^2}{(m-p-1)^2(mp-3)}.$$ The mean was then

$$E(B) = \frac{\Psi}{m-p-1}$$ in both models. Thereafter, we calculated the variance of each element of B in models which there after rendered

$$\mathrm{var}(b_{ij}) = \frac{(m-p+1)\psi_{ij}^2 + (m-p-1)\psi_{ii}\psi_{jj}}{(m-p)(m-p-1)^2(m-p-3)}.$$

In this research, we used the Conjugate distribution to make inference about a covariance matrix Σ whose prior had a $W^{-1}(\psi_{11}, m)$ distribution. A conjugate prior has the same functional form in q as the likelihood function which leads to a posterior distribution belonging to the same distribution family as the prior (Cressie 1993). In the models, the Beta(a_1, a_2) distribution had probability mass function $f(q)$ given by: $f(\theta) = \dfrac{\theta^{\alpha_1 - 1} \cdot (1-\theta)^{\alpha_2 - 1}}{\int_0^1 t^{\alpha_1 - 1} \cdot (1-t)^{\alpha_2 - 1} \cdot dt}$

(3.4). The denominator was a constant in both models so we rewrote the equations as: $f(\theta) \propto \theta^{\alpha_1 - 1} \cdot (1-\theta)^{\alpha_2 - 1}$. We estimated the true probability of p, the likelihood function $l(s, n; q)$ in both models. This was given by the binomial distribution probability mass function in the models which was written using q to represent the unknown parameter p (i.e., $l(s, n; \theta) = \binom{n}{s} \cdot \theta^s \cdot (-\theta)^{n-s}$

(3.5). Since the binomial coefficient $\binom{n}{s}$ was constant for the sampled MDR-TB parameters datasets (i.e. known n, s), we rewrote the equation as: $l(s, n; \theta) \propto \theta^s \cdot (1-\theta)^{n-s}$

By doing so the models revealed that the Beta distribution and the binomial likelihood function had the same functional form in q, (i.e. $q^a \cdot (1-q)^b$, where a and b were constants). Since the posterior distribution is a product of the prior and likelihood function, it too had the same functional form. By combining Equations 3.4 and 3.5 we then had: $f\theta|s,n) \propto \theta^{\alpha_1-1+s} \cdot (1-\theta)^{\alpha_2-1+n-s}$. We noticed that we had a Beta(a_1+s, a_2+n-s) distribution, so the posterior density was actually

$$f\theta|s,n) = \frac{\theta^{\alpha_1-1+s} \cdot (1-\theta)^{\alpha_2-1+n-s}}{\int_0^1 t^{\alpha_1-1+s} \cdot (1-t)^{\alpha_2-1+n-s} \cdot dt} \text{ in the models.}$$

The Beta(1, 1) distribution in the MDR-TB models was the same as a Uniform(0, 1) distribution, so we started with a Uniform(0, 1) prior for p, which then revealed that the posterior distribution could be rendered by Beta($s+1$, $n-s+1$). The Jeffrey's prior for a binomial probability then was calculated as a Beta(½, ½). In Bayesian probability, the Jeffrey's prior, is a non-informative (objective) prior distribution on parameter space that is proportional to the square root of the determinant of the Fisher information: $p(\vec{\theta}) \propto \sqrt{\det I(\vec{\theta})}$. It has the key feature that it is invariant under reparameterization of the parameter vector $\vec{\theta}$.

Note, although some infectious disease cluster modelers have used a Beta(0, 0) prior for quantifying within residual cluster based predictor covariate coefficient estimates it is mathematically undefined and therefore meaningless by itself, giving a posterior distribution of Beta(s, $n-s$) which has a mean of s/n: in other words it provides an unbiased estimate for the binomial probability but has a mode of $(s-1)/(n-2)$ which is not intuitive, and doesn't work if $s=0$ or n.

For an alternate parameterization $\vec{\varphi}$ in the MDR-TB models we derived $p(\vec{\varphi}) \propto \sqrt{\det I(\vec{\varphi})}$ from $p(\vec{\theta}) \propto \sqrt{\det I(\vec{\theta})}$ using the change of variables theorem, the definition of Fisher information, and that the product of determinants which in this research was calculated a the determinant of the matrix product using:

$$p(\vec{\varphi}) = p(\vec{\theta}) \left| \det \frac{\partial \theta_i}{\partial \varphi_j} \right|$$

$$\propto \sqrt{\det I(\vec{\theta}) \det^2 \frac{\partial \theta_i}{\partial \varphi_j}}$$

$$= \sqrt{\det \frac{\partial \theta_k}{\partial \varphi_k} \det E\left[\frac{\partial \ln L}{\partial \theta_k} \frac{\partial \ln L}{\partial \theta_l} \right] \det \frac{\partial \theta_l}{\partial \varphi_j}}$$

$$= \sqrt{\det E\left[\sum_{k,l} \frac{\partial \theta_k}{\partial \varphi_k} \frac{\partial \ln L}{\partial \theta_k} \frac{\partial \ln L}{\partial \theta_l} \frac{\partial \theta_l}{\partial \varphi_j} \right]}$$

$$= \sqrt{\det E\left[\sum_{k,l} \frac{\partial \ln L}{\partial \varphi_i} \frac{\partial \ln L}{\partial \varphi_j} \right]} = \sqrt{\det I(\vec{\varphi})}.$$

Thereafter, for a case of a single MDR-TB parameter space sampled variable we derived

$$p(\varphi) = p(\theta) \left| \frac{d\theta}{d\varphi} \right| \propto \sqrt{E\left[\left(\frac{d \ln L}{d\theta} \right) \right]^2 \left(\frac{d\theta}{d\varphi} \right)^2}$$

$$= \sqrt{E\left[\left(\frac{d \ln L}{d\theta} \frac{d\theta}{d\varphi} \right) \right]^2} = \sqrt{E\left[\left(\frac{d \ln L}{d\varphi} \right) \right]^2} = \sqrt{I(\varphi)}$$

Because the spread of tuberculosis may be governed by similarities as well as dissimilarities in geographic locations, eigenvectors portraying both PSA and NSA were included in the analysis. The clusterings used to identify geographic groups in the stratified clusters were treated as repeated measures as to exploit spatial autocorrelation effects. Model parameters were generated employing variance decomposition estimates. Stepwise selected predictor covariate coefficient estimates accounted for the most variance, with the common set of selected coefficient estimates being age, number of bedrooms, time on the job, and rental home account. We found that several coefficient estimates covariates were specific to the clusterings used for aggregation purposes. All random effects had both a spatially structured SSRE and a SURE component. This made some difference in the Poisson regression model parameter estimation, but failed to account for much of the detected over dispersion in the MDR-TB models. The SURE's tended to account for a decreasing amount of variance with increasing geographic aggregation. Meanwhile, spatial structuring tended to increase in importance with increasing geographic aggregation. In the models accounting for spatial structuring yielded a SURE that better conformed to a bell-shaped curve, while the SSREs included a mixture of PSA and NSA components, with weak NSA being dominant. The two finer geographic aggregations were characterized by a mixture whose net spatial autocorrelation was close to 0.

A stepwise logistic regression analysis of the sampled MDR-TB data identified the following six covariates, in the spatial analysis of the geographically aggregated parameters in using a 10% selection criterion: number of bedrooms, time on the job (in months), latitude, rented home type, testing sensitive to isoniazid in an LJ medium, and testing sensitive to streptomycin in an LJ medium. These six covariates rendered a pseudo-R^2 value of 0.3249, suggesting that the estimates accounted for about 32% of the geographic variation across the 195 groups.

A Bayesian specification of the problem was also solved using WinBUGS®, employing normal priors for each of the six logistic regression coefficients. This solution had posterior mean regression coefficients and standard errors that were almost identical to those for a frequentist solution. 160,000 MCMC replications were executed. The first 10,000 were discarded as a burn-in set, and the resulting 150,000 were weeded such that only every third replication result was retained. The final MCMC dataset contained 50,000 replications. These replications conformed closely to a normal distribution, had no trend in the time series plot, and displayed no serial correlation. The estimated coefficients for testing sensitive to isoniazid and testing sensitive to streptomycin in an LJ medium had MCMC chains containing marked serial

correlation. Weeding each of these chains by 200 adjusted for this correlation, but reduced the sample sizes to 750 replications. The resulting estimates and standard errors were very similar to those for the 50,000 replications This Bayesian problem was then respecified to include a random effects term, which contained both SSRE and SURE components; the latter not conforming very closely to a normal distribution (the diagnostic Shapiro-Wilk statistic had a null hypothesis probability of P(S-W) = 0.0025). The random effects increased the pseudo-R^2 value to 0.9877. The SSRE accounted for about 56% of the random effects, contained eigenvectors representing both PSA and NSA, and represented an overall map pattern characterized by negligible autocorrelation (i.e., MC = -0.0544). The final MDR-TB model output detailing the sequential decomposition of the variance. The logistic regression model mean response, which was estimated with quasi-likelihood techniques because of the presence of severe under dispersion (i.e., deviance = 0.0247), comprised the estimates.

The linear analysis of geographically aggregated MDR-TB data in indicated that: (1) as the number of bedrooms in a house in which infected persons reside increases, and as the percentage of isoniazid-sensitive infected persons increased, the standardized rate of MDR-TB tended to decrease; and, (2) as the average working time, as the percentage of rental house types, and as the percentage of streptomycin-sensitive persons increased, the standardized rate of MDR-TB tended to increase. These results also indicated that spatial autocorrelation plays a key role in the aggregated pattern of the standardized rates of MDR-TB, with a very weak tendency toward geographic clustering in terms of a gradient as a function of latitude, combined with a compensating mixture of PSA and NSA. Meanwhile, the SURE indicated that a sizeable amount of variability was unaccounted for by variables other than those contained in the clinical and environmental-sampled dataset which most likely played an important role in the geographic distribution of individuals infected with MDR-TB.

A stepwise logistic regression analysis of these data identified the following two covariates, using a 10% selection criterion marital status and adobe building material. These covariate coefficient estimates rendered a pseudo-R^2 value of 0.2966, suggesting that they accounted for about 30% of the geographic variation across the sampled MDR-TB data. The Bayesian estimation matrix then was respecified to include a random effects term, which contained both SSRE and SURE components, the latter conforming to a normal distribution (the diagnostic Shapiro-Wilk statistic had a null hypothesis probability of P(S-W) = 0.1572). The random effects increased the pseudo-R^2 value to 0.9973. The SSRE accounted for about one third of the random effects, contained eigenvectors representing both PSA and NSA, and represented an overall MDR-TB map pattern characterized by weak NSA (i.e., MC = -0.2240). The final model output detailing sequential decomposition of variance was generated. The logistic regression model mean response, which was estimated with quasi-likelihood techniques because of the presence of severe underdispersion (i.e., deviance = 0.0017), contained the estimates.

These results indicated that as the percentage of houses of adobe construction in which infected persons resided increased, the standardized rate MDR-TB tended to increase; whereas, as the percentage of infected persons whose marital status was single increased, the standardized rate of MDR-TB tended to decrease. These results also indicated that spatial autocorrelation plays a key role in the aggregated pattern of the standardized rates of MDR-TB, with a weak tendency toward geographic clustering, but a concomitant stronger tendency toward geographic dispersion (a mixture of positive and dominant NSA was present). Meanwhile, the SURE indicated that a sizeable amount of variability was unaccounted for by the sampled variables other than those contained in the dataset, and these unknown variables most likely played an important role in the geographic distribution of individuals infected with MDR-TB. Patterns in these maps reflected the groupings. The SURE maps contained a more geographically mixed set of values; whereas, similar conspicuous patterns appear in the SSRE maps.

We then qunatiated presents the pairwise spatial correlations for the four drug resistance outcomes. We see a strong spatial correlation among INH, RIF, and EMB, implying that the drug resistance to these three drugs are similar over the region. Meanwhile, we see a small spatial correlation of resistance to SM with the other three drugs. Figure 5 and Figure 6 shows the spatial effects of four drug resistance outcomes. Since INH, RIF, and EMB have strong spatial correlations, the resistance pattern are similar for these three drugs. Large spatial effects were found in the West side boundary and southeast corner, implying increased probability of drug resistance to these three drugs in these regions. For the drug SM, large spatial effects were spotted in several regions and lowest in the north.

4. Discussion

In the high prevalence hierarchical intra-cluster-based MDR-TB regression model number of bedrooms, time on the job (in months), rented home type, testing sensitive to isoniazid in an LJ medium, and testing sensitive to streptomycin in an LJ medium. Overall resistance to one or more drugs Praharaj et al. (2004) used 1378 isolates from HIV negative cases and 68 isolates from AIDS cases for determining drug susceptibility to first line antitubercular drugs which revealed that 13.78% with streptomycin resistance was the highest. Similar resistance to one or more drugs has been reported by Jena et al. (1995) and Sonnenberg et al. (2000) who had found approximately 12.7% and 11% respectively. Drug susceptibility testing for *Mycobacterium tuberculosis* is especially required in difficult cases of tuberculosis (TB) chemotherapy and in cases of MDR-TB; combined resistance to isonizid and rifampicin with or without resistance to any other drug. For example, Agatha et al. (2003), used 70 pulmonary isolates of *M tuberculosis* [Indirect drug susceptibility testing (DST)] and 20 sputum (10 acid fast bacilli [AFB] positive and 10 AFB negative) specimens (direct DST) using 0.2μg isoniazid (INH), 2μg ethambutol (EMB), 40 μg rifampicin (RIF) and 4 μg streptomycin STR) which revealed that the sensitivity and specificity for isoniazid was 100% and for streptomycin was 91.8%. The rapid and accurate susceptibility testing of *M. tuberculosis* is essential for effective patient

treatment and to prevent transmission of the disease. In the low MDR-TB stratified clusters being single was an important covariate. Urban crowding and crowded group living situations in poorly ventilated spaces has led to increased disease transmisssion (www.dhpe.org/infect/tb.html). Incidence rates of MDR-TB in prisons, juvenile detention centers, and homeless shelters are higher than that in the general population. TB bacteria also can flourish in crowded nursing homes because older adults often have immune systems weakened by illness or aging (http://bhealthblog.com/tuberculosis).
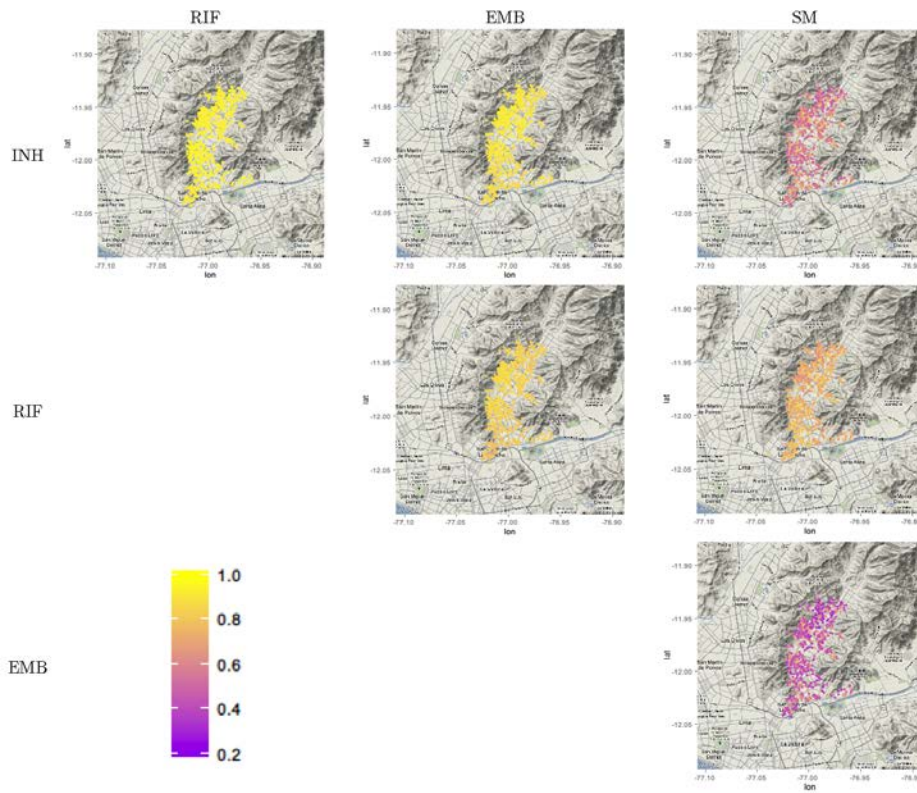
Figure 5. Bayesian random effect components for the clusters of aggregated individuals based INH, RIF, and EMB for a SSRE high MDR-TB prevalence clusters
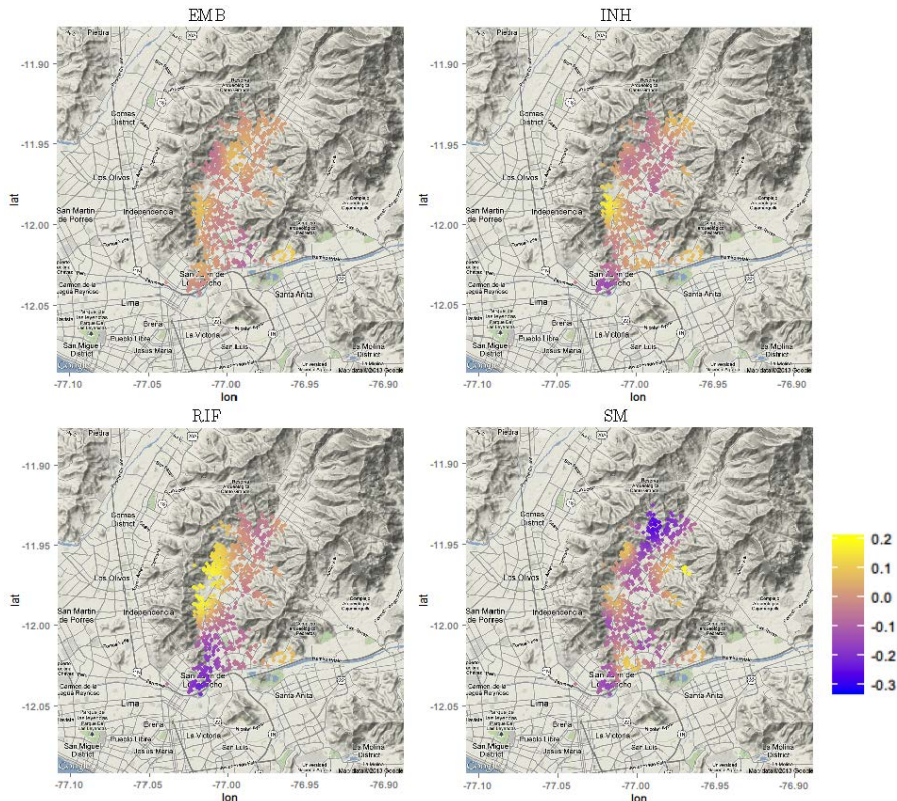
Figure 6. Bayesian random effect components for the clusters of aggregated individuals based INH, RIF, and EMB for a SSRE for the MDR-TB prevalence cluster

We then generated DW statistics to test for first-order autocorrelation error coefficients in the regression models. The tests were easy to compute under standard assumptions and possessed optimal power properties for identifying the serial dependence in the MDR-TB endemic transmission-oriented models. Although the power function of the DW tests were dependent upon the regressor vectors, useful upper and lower bounds for the power estimates were established. The bounds were obtained from irregularities on the roots of non-definitive symmetric matrices. The d statistics contained comparisons between the clinical and environmental predictor variables in each model separated by lags greater than one. Unfortunately, closing the gaps modified the structure of the regression matrices, consequently, we could not apply specialized tables for the regressors using seasonal dummy variables. We did however, quantify the serial dependence scheme against the DW gauged to obtain other tests (e.g., Yulk-Walker). These tests revealed that there was marginal levels of first-order positive autocorrelation in the models. Finite sample power of the DW test can quantitate fractionally integrated first-order serial disturbances in hierarchical MDR-Bluster-based regression models.

A Bayesian specification was then solved employing respecified priors for each of the regression-based coefficients which allowed flexible model fitting and estimation and mapping of all "high risk" MDR-TB geolocations generated from the linear and non-linear estimation models. Our Bayesian matrices offered extensive statistical treatment of the sampled clinical and environmental explanatory predictor covariate coefficient estimates including inference about each sampled MDR-TB parameter value in the SJL study site and calculation of confidence intervals for model predictions. In this research, an extremely wide range of fit statistics were defined based on the distribution of predictions used in constructing the models including the calculation of p values, describing the probability that the sampled parameter arose by chance given the model assumptions. Bayesian and non-linear optimization MDR-TB models can provide efficient asymptotical estimators that are more efficient than simple linear-based regression covariate coefficient estimates For example, Bayesian inference can express the actual credibility related in the analyses of the sampled MDR-TB parameters whereas, the frequentist confidence interval would include only the estimator in a given percentage of the data sampled.

In this research, we found that WinBUGS® was able to recognize conjugate specifications which was updated via direct sampling using standardized algorithms. In the estimation matrix, multiple MCMC replications were executed. Our replications conformed closely to a normal distribution and had no trend in the time series plot. The estimated coefficients for testing the sampled georeferenced cluster-based explanatory predictor covariate coefficient estimates had MCMC chains containing marked serial correlation. Weeding each of the chains generated by the iterations adjusted for this correlation, but reduced the sample sizes in each MDR-TB model. The resulting variance estimates in both the models were very similar to those for the replications. Posterior predictive simulations were then explicitly accounted for in the parametric uncertainty estimates.

Equilibrium probabilities of the Markov processes in the models were also computed by multiplying the probabilities of the embedded chain by the mean times spent in the various states. The geometrically ergodic Markov chains were shown to have a positive extremal index as soon as the drift function in the MDR-TB models were satisfied. Our Markov chains indicated that geometric ergodicity was a key requirement to consistent variance estimation of the asymptotic normal distribution in the models. In this research, the Markov chains also guaranteed the consistency of a batch means estimate of the asymptotic variance in the sampled clinical and environmental parameters, which in turn allowed for the construction of asymptotically valid standard errors in the models. The Bayesian matrices were then respecified to include a random effects term which contained both SSRE and SURE components; the latter not conforming very closely to a normal distribution in either model. The random effects were modeled via Bayesian specifications for quantifying spatial heterogeneity globally in the within-cluster based explanatory predictor covariate coefficient estimates. The random effects increased pseudo-R^2 values. The SSRE accounted for the random effects in the high and low MDR-TB stratified clusters. The clusters contained eigenvectors representing both PSA and NSA but, the overall MDR-TB map pattern in the models residuals was characterized by weak NSA.

The models generated in this research captured the NSA in the residual intra-cluster correlation analyses which may have been attributable to the competitive locational processes, negative spatial externalities, the construction of spatial correlograms, the spectrum (i.e., eigenvalues) of a geographic weights matrix, the calculation of linear regression residuals and the computation of local indicators of spatial autocorrelation (LISA) statistics. Fortunately, NSA can be detected in empirical analyses of spatiotemporal-sampled MDR-TB predictor variables. For example, a Moran's scatterplot for NSA detection can be easily constructed by graphing Cartesian pairs of sampled clinical and environmental cluster based explanatory predictor covariate coefficient attribute value z-scores (i.e., summation of Z_l scores) in SAS®. Focusing on the mean response specification in a spatial filter logistic model using a geographic weighted matrix can then capture NSA in predictive autoregressive MDR-TB distribution model cluster based residual estimates.

Several worthwhile implications can be drawn from this research. For example, a general problem in MDR-TB modeling concerns aggregation of sampled predictor variables in order to calculate rates while avoiding small valued covariate coefficient measurement estimates (see Gandhi et al. 2006). In this research, for example, clinical and environmental sampled individual patient data were treated as repeated measures, as well as aggregated into clusters. Tendencies were detectable between these two treatments of individuals, and across the clusterings. Those variables that furnished statistical explanation in all but the coarsest geographic aggregation (low MDR-TB prevalence cluster) were: number of bedrooms, time on job, and rental home. The stepwise selected covariates tended to account for about 30% of the variance in all cases. All random effects contained a spatially structured component. Those spatially unstructured components were

estimated by employing individual data and then the clusters were treated as repeated measures which accounted for a decreasing amount of variance with increasing geographic aggregation coarseness (i.e., 20%, 11%, and 3%). In contrast, having no repeated measurements but employing a Bayesian analysis in which a prior was attached to the random effects yielded what appeared to be overestimated components (i.e., the residual and the spatially unstructured random effects could not be differentiated): all SURE estimates accounted for about 50% of the variance; whereas, the residual estimates were less than 1.5%. In other words, attaching a prior distribution in a Bayesian analysis failed to furnish sufficient ancillary information for differentiating between these two components. Nevertheless, eigenvector spatial filtering methodology allowed estimation of the SSRE. The MDR-TB model estimates indicated a concomitant strong tendency toward geographic dispersion. A mixture of positive and dominant negative spatial autocorrelation was present in the SSRE using the clinical and environment-sampled MDR-TB parameters.

In this paper a Bayesian model selection procedure was investigated using a MCMC approach. Bayesian model selection, particularly the MCMC method considered in this paper has many advantages over traditional methods for time series explanatory endemic transmission-orieentd MDR-TB predictive risk modeling analyses using closed form approximations. Following the tuning of the prior distributions, the next step in fitting the multinomial model was to program the likelihood function. When fitting this sample model with PROC NLMIXED, Chen and Kuo (2001) offer an alternative specification of the model, a Poisson non-linear time series explanatory epidemiological risk model. To fit the Poisson non-linear MDR-TB epidemiologicalendemic transmission-oriented risk model, texperimenters may transpose the data set so that, within every clinical sampled parameter estimator . As of SAS 9.3 (Stokes, 2011), PROC MCMC supports multivariate distributions such as the multinomial distribution. Therefore the model can be programmed directly as it appears above. The following code illustrates the PROC MCMC statement within a macro, which will allow running multiple chains more easily. Multiple chains are necessary for Gelman-Rubin convergence diagnostics.

Gelman and Rubin (1992) propose a general approach to monitoring convergence of MCMC output in which $m > 1$ parallel chains are run with starting values(high MDR-TB prevalence cluster-relatd explanatory covariates) that are over dispersed relative to the posterior distribution. Convergence is diagnosed when the chains have `forgotten' their initial values, and the output from all chains is indistinguishable. The gelman.diag diagnostic is applied to a single MDR-tb-related predictor variable from the chain. It is based a comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance. There are two ways to estimate the variance of the stationary distribution ina a time series MDR-TB epidemiological endemic transmission-oriented risk model: the mean of the empirical variance within each chain, *W*, and the empirical variance from all chains combined, which can be expressed as *sigma.hat^2 = (n-1)W/n + B/n* where *n* is the number of iterations and *B/n* is the empirical between-chain variance. If the chains have

converged, in the risk model then both estimates are unbiased. Otherwise the first method will underestimate the variance, since the individual chains in the MDR-TB risk model would not have had time to range all over the stationary distribution, and the second method will overestimat*e* the variance, since the starting sampled clinical points were chosen to be overdispersed. The convergence diagnostic is based on the assumption that the target distribution is normal. A Bayesian credible interval can be constructed theerfater using a t-distribution with mean *mu.hat = Sample mean of all chains combined* and variance *V.hat=sigma.hat2 + B/(mn)*and df estimated by the method of moments *d = 2*V.hat^2/Var(V.hat)*.Use of the *t*-distribution accounts for the fact that the mean and variance of the posterior distribution are estimated. The convergence diagnostic for the model would be *R=sqrt((d+3) V.hat /((d+1)W)*. Values substantially above 1 indicate lack of convergence (Gelman, and Rubin, 1992). If the chains have not converged, Bayesian credible intervals based on the t-distribution may be too wide.

Though a stochastic search of the model space, modern computational techniques, such as Reverse Jump MCMC (RJMCMC) (Green, 1995),can allow model selection in cases where there is a large number of sampled clinical MDR-TB explanatory covariates under consideration. This is typically difficult in the frequentist MDR-TB-related frameworks. In addition, inference for time series, explanatory, MDR-TB–related, endemic, transmission-oriented coefficients, accounting for model uncertainty, may be a byproduct of the RJMCMC approach. In a robust Bayesian MDR-TB endemic transmission-oriented explanatory endemic trasnmission-oriented paradigm, uncertainty may have a straightforward probabilistic interpretation. Model uncertainty is accounted for in the Bayesian paradigm by allowing the model to vary as a random quantity (Clyde and George, 2004). Traditionally, predictive time series explanatory MDR-TB related epidemiological risk models, or regression coefficients, are given a certain amount of *a priori* weight.

For example, suppose an experimeneter is given a sequence $(x_1, \ldots, x_n)$ of IID $N\left(\mu, \delta_v^2\right)$ random variables and an a priori distribution of μ given by $N\left(\mu_0, \delta_m^2\right)$. We wish to find the maximum a posteriori probability (MAP estimate of μ .) The MAP can be used to obtain a sampled clinical MDR-TB sample point estimate of an unobserved quantity on the basis of empirical dataset. The MAP may be closely related to Fisher's method of maximum likelihood (ML), but employs an augmented optimization objective which incorporates a prior distribution over the quantity to estimate. The function to be maximized is then given by

$$f\left(\mu\right)f\left(x|\mu\right) = \pi\left(\mu\right)L\left(\mu\right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_m}\exp\left(-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_m}\right)^2\right)\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{1}{2}\left(\frac{x_j-\mu}{\sigma_v}\right)^2\right)$$

which is equivalent to minimizing the following function of μ : $\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma_v}\right)^2 + \left(\frac{\mu-\mu_0}{\sigma_m}\right)^2$. Thus, we see that the

MAP estimator for μ is given by

$$\hat{\mu}_{MAP} = \frac{n\sigma_m^2}{n\sigma_m^2 + \sigma_v^2}\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) + \frac{\sigma_v^2}{n\sigma_m^2 + \sigma_v^2}\mu_0$$ which turns

out to be a linear interpolation between the prior mean and the sample mean weighted by their respective covariances. The case of $\sigma_m \to \infty$ is called a non-informative prior and leads to an ill-defined a priori probability distribution; in this case $\hat{\mu}_{MAP} \to \hat{\mu}_{ML}$.

The MDR-TB RISK model may then be updated via Bayesian learning just as the parameters are in the classic Bayesian parameter estimation framework to obtain the posterior distribution of the model. The prior weighting of the sampled explanatory MDR-TB time series coefficients is another benefit over frequentist methods such as Akaike information criterion (AIC) which a measure of the relative quality of a statistical model for a given set of data. AIC deals with the trade-off between the goodness of fit of the model and the complexity of the model quantity (Cressie 1993). Bayesism is founded on information theory as it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data (Griffith 2003).

Information theory is a branch of applied mathematics, electrical engineering, and computer science involving the quantification of information. Information theory was developed by Claude E. Shannon to find fundamental limits on signal processing operations such as compressing data and on reliably storing and communicating data ny other areas, including statistical inference, natural language processing, cryptography, neurobiologythe evolutionand functionof molecular codes, model selection in ecology, thermal physicsquantum computing, plagiarism detectionand other forms of data analysis Suppose that a sampled empirical dataset of MDR-TB related clinical paramter estimators is generated by some unknown process *f*. An experimenter may then consider two candidate models to represent *f*: g_1 and g_2. If the experimenter knew *f*, then he or she could find the sampled MDR-TB information lost from using g_1 to represent *f* by calculating the Kullback–Leibler divergence, $D_{KL}(f \| g_1)$; similarly, the information lost from using g_2 to represent *f* could be found by calculating $D_{KL}(f \| g_2)$.

In probability theory and information theory, the Kullback–Leibler divergence (also information divergence, information gain, relative entropy, or KLIC; here abbreviated as KL divergence) is a non-symmetric measure of the difference between two probability distributions P and Q. Specifically, the Kullback–Leibler divergence of Q from P, denoted $D_{KL}(P \| Q)$, is a measure of the information lost when Q is used to approximate P The KL divergence measures the expected number of extra bits required to code samples from P when using a code based on Q, rather than using a code based on P. Typically P represents the "true" distribution of data, observations, or a precisely calculated theoretical distribution. The measure Q typically represents a theory, model, description, or approximation of P. Although it is often intuited as a metric or distance, the KL divergence is not a true metric — for example, it is not symmetric: the KL divergence from P to Q is generally not the same as

that from Q to P. However, its infinitesimal form, specifically its Hessian, is a metric tensor: it is the Fisher information metric.KL divergence is a special case of a broader class of divergences called *f*-divergences

The Kullback–Leibler divergence is always non-negative, $D_{KL}(P \| Q) \geq 0$ a result known as Gibbs' inequality, with $D_{KL}(P \| Q)$ zero if and only if $P = Q$ almost everywhere. The entropy $H(P)$ thus sets a minimum value for the cross-entropy $H(P,Q)$, the expected number of bits required when using a code based on Q rather than P; and the KL divergence therefore represents the expected number of extra bits that must be transmitted to identify a value x drawn from X, if a code is used corresponding to the probability distribution Q, rather than the "true" distribution P. The Kullback–Leibler divergence remains well-defined for continuous distributions, and furthermore is invariant under parameter transformations. For example, if a transformation is made from variable x to variable $y(x)$, then, since $P(x)dx = P(y)dy$ and $Q(x)dx = Q(y)\,dy$ the Kullback–Leibler divergence may be rewritten:

$$D_{KL}(P \| Q) = \int_{x_a}^{x_b} P(x)\log\left(\frac{P(x)}{Q(x)}\right)dx$$

$$= \int_{y_a}^{y_b} P(y)\log\left(\frac{P(y)dy/dx}{Q(y)dy/dx}\right)dy = \int_{y_a}^{y_b} P(y)\log\left(\frac{P(y)}{Q(y)}\right)dy$$

where $y_a = y(x_a)$ and $y_b = y(x_b)$. Although it was assumed that the transformation was continuous, this need not be the case. This also shows that the Kullback–Leibler divergence produces a dimensionally consistent quantity, since if x is a dimensioned variable, $P(x)$ and $Q(x)$ are also dimensioned, since e.g. $P(x)\,dx$ is dimensionless. The argument of the logarithmic term is and remains dimensionless, as it must. It can therefore be seen as in some ways a more fundamental quantity than some other properties in information theory[9] (such as self-information or Shannon entropy), which can become undefined or negative for non-discrete probabilities. The Kullback–Leibler divergence is additive for independent distributions in much the same way as Shannon entropy. If P_1, P_2 are independent distributions, with the joint distribution $P(x,y) = P_1(x)P_2(y)$, and Q, Q_1, Q_2 likewise, then $D_{KL}(P \| Q) = D_{KL}(P_1 \| Q_2) + D_{KL}(P_2 \| Q_2)$. The Kullback–Leibler divergence between two multivariate normal distributions of the dimension k with the means μ_0, μ_1 and their corresponding nonsingular covariance matrices Σ_0, Σ_1 is:

$$D_{KL}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\left(\begin{array}{c} tr\left(\Sigma_1^{-1}\Sigma_0\right) + (\mu_1 - \mu_0) \\ -k - \ln\left(\frac{\det\Sigma_0}{\det\Sigma_1}\right) \end{array}\right)$$. The logarithm

in the last term must be taken to base *e* since all terms apart from the last are base-*e* logarithms of expressions that are either factors of the density function or otherwise arise naturally. The equation therefore gives a result

measured in nats. Dividing the entire expression above by $\log_e 2$ yields the divergence in bits.

By so doing An experimenter choose then cloose the candidate model that minimized the information loss. We cannot choose with certainty, because we do not know *f*. Akaike (1974) showed, however, that we can estimate, via AIC, how much more (or less) information is lost by g_1 than by g_2. It is remarkable that such a simple formula for AIC results. The estimate, though, is only valid asymptotically; if the number of data points is small, then some correction is often necessary. AIC does not provide a test of a model in the sense of testing a null hypothesis; i.e Certain covariates can be given more or less weight in determining the most appropriate model. Methods such as AIC selection weight all covariates equally. The posterior distribution of interest in Bayesian model inference is the joint distribution of the model and the parameters for each model. A sample from this distribution is obtained

from the RJMCMC sampler and inference concerning regression parameters and the model itself can be extracted from this sample. The RJMCMC approach also has one other major advantage over AIC and Bayesian closed form approximations, it is directly extendable to

spatial generalized linear mixed models (GLMM). This implies the RJMCMC approach can be an all purpose tool for geostatistical regression inference for Gaussian and non-Gaussian.

Another candidate approach for linearly quantifying clinical and environmental MDR covariate coefficient estimates in the future is to make use of locally polynomial regression modeling. Locally linear approaches to modeling large and complex data sets have been successfully demonstrated in a variety of applications, most of which include a separation of the data into blocks based on prior knowledge about the structure of the data being analyzed. This suggests that nonlinear and non-monotone response surfaces may be handled by local high-order polynomial models constructed using spatiotemporal MDR-TB parameters. In order to also account for linearly dependent regressors in spatiotemporal-sampled MDR-TB data and inter-correlations between the responses, more alternatives to OLS, can be generated. Bi-linear methods based on estimated latent variables, [e.g. Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR)] may also reveal considerable success in spatiotemporal cluster-based MDR-TB parameter analysis. Martens and Martens (1986) demonstrated the use of PLSR as an alternative to ANOVA to facilitate the interpretation of multi-response residual within cluster-based data from designed experiments. PLSR maximizes the explained covariance between the regressors and the responses (Cressie 1993). In contrast to most other linear regression methods, PLSR also utilizes inter-correlations between the response variables for model stabilization, and does not require the regressor variables to be linearly independent. PLSR is efficient for compressing inputs, intermediate states and output variables into their most relevant subspaces [i.e., spanned by the estimated latent variables, also called PLS components (PCs)], and, hence may provide a versatile means for MDR-TB cluster-based data compression by reducing the rank of both regressors (i.e., X) and responses (i.e.,Y). This, in turn, may provide an effective approach to identification of important features in complex MDR-TB sampled clinical and environmental predictors. PLSR is equivalent to OLS when the regressor rank is not reduced, that is, when all PLS components are included (Hosmer and Leneshew 2000). Modeling MDR-TB based on such estimated latent variables represented by so-called X- and Y- score vectors also may have the advantage of being suited for graphical visualization, inspection and interpretation via their associated sets of coefficients describing the relationship between the score vectors and the original sampled MDR-TB parameters. Campbell et al. (2006) has shown that metamodels based on subspaces found by PLSR, when compared to Legendre polynomials and PCA, gave the simplest and most predictive basis for sensitivity analysis for a set of computational models. In Martens et al. (2009) the suitability of PLSR for interpretation of complex biological systems and use of PLSR in sensitivity analysis was demonstrated. This motivates probing the versatility of a new variant of local MDR-TB modeling, [i.e., Hierarchical Cluster-based PLS regression (HC-PLSR)], which will assume no prior knowledge about the sampled clinical and environmental data structure. Therefore besides quantitating HC-PLSR MDR-predictor covariate coefficient estimates, the metamodelling performance of HC-PLSR, global PLSR and global ordinary least squares regression for a dynamic spatiotemporal hierarchical cluster-based MDR-TB regression framework can be constructed. These test beds can also encompass large classes of dynamic models. We can also compare the HC-PLSR MDR-TB model approach in terms of explained variance and prediction accuracy increases with the degree of nonlinearity and the presence of positive feedback loops with other hierarchical cluster-based regression model outputs.

In conclusion, a regression analysis of the clinical and environmental MDR-TB predictor variables identified multiple covariate coefficient estimates associated with the sampled data. Thereafter, we developed a practical approach to diagnosing the existence of a latent stochastic process in the mean of the regression model. The asymptotic distribution of standard generalized linear model estimators were derived for helping to quantiate where an autocorrelated latent process was present. Simple formulae for the effect of autocovariance on standard errors of the regression coefficients were also provided. Methods for adjusting for the severe bias in the proposed estimators of autocovariance were derived and their behaviour was investigated. The model filtered out the latent autocorrelation patterns from the error-covariance matrix. Incorporation of all relevant eigenvectors in the MDR-TB model left the remaining residual component spatially uncorrelated. A Bayesian matrix was then generated using an MCMC algorithm. The model was respecified to include a random effects error term, which contained both SSRE and SURE components. The random error effects in the Bayesian coefficients increased the pseudo-R^2 values in the models. The SSRE accounted for about one third of the random effects error in the sampled MDR-TB data, which contained eigenvectors representing both PSA and NSA. The final model revealed an overall map pattern characterized by weak NSA in the models. Quasi-likelihood techniques in a regression equation and Bayesian prior distributions can quantify intra-cluster

correlations using sequential decomposition of variance estimates from clinical and environmental-sampled MDR-TB explanatory variables for identifying at-risk populations.

References

[1] Anselin, L. 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In M. Fischer, H. Scholten, and D. Unwin (eds.), *Spatial Analytical Perspectives on GIS*. London: Taylor and Francis.

[2] Becerra, M.C., Bayona, J., Freeman, J., Farmer, P.E. and Kim, J.Y., 2000. Redefining MDR-TB transmission 'hot spots'. Int J Tuberc Lung Dis, 4(5): 387-94.

[3] Besag, J., Green, P., Higdon, D. and Kengersen, K., 1995. Bayesian computation and stochastic system, Stat. Sci., pp. 3-66.

[4] Besag, J., York, J. and Mollie, A., 1991. Bayesian image restorarion, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43(1).

[5] Besag, J., Newell J. The detection of clusters in rare diseases. Journal of the Royal Statistics Society A. 1991; 154: 143-155.

[6] Blake, K. S., Kellerson, R. L., & Simic, A. (2007). Measuring Overcrowding in Housing: U.S. Department of Housing and Human Development.

[7] CDC, 1999. Reported TB in the United States, 1998: TB Surveillance report, Atlanta.

[8] Cegielski, C.G., Hall, D.J. and Rebman, C., 2006. Enterprise resource planning systems implementation success. International Journal of Information Systems and Change Management, 1(3): 301-317.

[9] Clayton, D.G. and Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics, 43(3): 671-681.

[10] Fahrmeir, L. and Lang, S., 2001. Bayesian semiparametric regression analysis of multicategorical time-space data. Annals of the Institute of Statistical Mathematics, 53(1): 11-30.

[11] Ferreira, J.T.A.S., Denison, D.G.T. and Holmes, C.C., 2002. Partition Modelling.

[12] Frieden, T.R. et al., 1993. Emergence of vancomycin-resistant enterococci in New York City. Lancet, 342(8863): 76-9.

[13] Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing, 7(1): 57-68.

[14] Gandhi, N. et al., 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. The Lancet, 368(9547): 1575-1580.

[15] Gelman, A., Chew, G.L. and Shnaidman, M., 2004. Bayesian Analysis of Serial Dilution Assays. Biometrics, 60(2): 407-417.

[16] Getis, A. and Griffith, D.A., 2002. Comparative spatial filtering in regression analysis. Geographical Analysis, 34: 130-140.

[17] Ghosh, S. et al., 1999. Type 2 diabetes: evidence for linkage on chromosome 20 in 716 Finnish affected sib pairs. Proc Natl Acad Sci U S A, 96(5): 2198-203.

[18] Godoy, P. et al., 2004. Characteristics of tuberculosis patients with positive sputum smear in Catalonia, Spain. Eur J Public Health, 14(1): 71-5.

[19] Goodchild M (1986) Spatial autocorrelation (CATMOG 47). GeoBooks, Norwich.

[20] Griffith, D., 2004. A spatial filtering specification for the autologistic model. Environment and Planning A

[21] Griffith, D. and Peres-Neto, P.R., 2006. Spatial modeling in ecology: The flexibility of eigenfunction spatial analysis. Ecology, 87(10): 2603-2613.

[22] Griffith, D.A., 2000. A linear regression solution to the spatial autocorrelation problem. J of Geogr Syst, 2(2): 141-156.

[23] Griffith, D.A., 2003. Spatial autocorrelation on spatial filtering. Springer.

[24] 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 Geographics, 4: 18.

[25] Hastie, T.J. and Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall.

[26] Hosmer, D.W. and Lemeshow, S., 2000. Applied logistic regression. Wiley.

[27] Jacob, B.G. et al., 2007. Environmental abundance of Anopheles (Diptera: Culicidae) larval habitats on land cover change sites in Karima Village, Mwea Rice Scheme, Kenya. Am J Trop Med Hyg, 76(1): 73-80.

[28] Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R. and Mostashari, F., 2005. A space-time permutation scan statistic for disease outbreak detection. PLoS Med, 2(3): e59.

[29] Kulldorff, M. and Nagarwalla, N., 1995. Spatial disease clusters: Detection and inference. Statistics in Medicine, 14: 799-810.

[30] Le, N.D. Petkau, A.J., Rosychuk, R.J. Surveillance of clustering near point sources. Statistics in Medicine 1996;15**:**727-740.

[31] Le Gallo J, Ertur C (2003) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. Papers in Regional Science 82:175-201

[32] Oeltmann, J.E. et al., 2008. Multidrug-resistant tuberculosis outbreak among US-bound Hmong refugees, Thailand, 2005. Emerg Infect Dis, 14(11): 1715-21.

[33] Pearson, M.L. et al., 1992. Nosocomial transmission of multidrug-resistant *Mycobacterium tuberculosis*. Annals of Internal Medicine, 117(3): 191-196.

[34] Rosychuk, R.J., Huston, C, Prasad, NGN. Spatial event cluster detection using a compound Poisson distribution. Biometrics. 2006; 62: 465–470

[35] Rushton, G. and Lolonis, P., 1996. Exploratory spatial analysis of birth defects rates in an urban population. Statistics in Medicine, 15(7-9): 717-726.

[36] Shah, H.N., Jain, P. and Chibber, P.J., 2006. Renal tuberculosis simulating xanthogranulomatous pyelonephritis with contagious hepatic involvement. Int J Urol, 13(1): 67-8.

[37] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(4): 583-639.

[38] Smith T (2001) Aggregation bias in maximum likelihood estimation of spatial autoregressive processes. Paper presented to the North American Regional Science Association, Charleston, 15-17, 2001 November.

[39] Waller, L.A. and Zelterman, D., 1997. Log-linear modeling with the negative multinomial distribution. Biometrics, 53(3): 971-82.

[40] Zhang, T. and Lin, G., 2009. Spatial scan statistics in loglinear models. Computational Statistics & Data Analysis, 53(8): 2851-2858.

[41] Zignol, M. et al., 2006. Global incidence of multidrug-resistant tuberculosis. J Infect Dis, 194(4): 479-85.