International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-22
Available Online at http://acascipub.com/Journals.php

Research article

Robustification of Multivariate Non-zero correlated Gamma Mixture Distributed Multicollinear, Discrete Finite Count-related Heteroskedastic, Integer Values for Regressively Delineating Morbidity Statistics in Guatemala

Miriam F. Escobar[a], Toni Panaou[b], Samuel Alao[c], Benjamin G. Jacob[c*],

[a]Institute for the Study of Latin America and the Caribbean, School of Interdisciplinary Studies, College of Arts and Sciences, University of South Florida, Tampa, FL 33620

[b]Department of Civil and Environmental Engineering, College of Engineering, University of South Florida, Tampa, FL 33620, USA

[c]Department of Global Health, College of Public Health, University of South Florida, Tampa, FL 33612, USA

*Corresponding Author Email: bjacob1@health.usf.edu

Abstract

Historically, statistical, forecast, vulnerability, epidemiological morbidity, mapping has involved the analysis of disease incidence employing prevalence as a response variable. Often these variables occur as discrete, finite, aggregate counts over a georeferenceable, provincial geographical region subdivided by administrative departmental boundaries. Unfortunately, regression models have not been generated from empirical, department-level, georeferenecable, morbidity datasets to determine optimally parameterizable covariates related to prevalence rates. In this research, demographic explanators (e.g. age, number of department hospitals etc.,) and landscape goemorphologically stratified, empirical measures were forecastes employing regression specifications and temporal data collections for optimally targeting departments that had higher prevalence rates in Guatemala. The model followed the standard $y = mx + b$. The significance level for the study was chosen before data collection, and set to 5%.The significance level defined for the epidemiological study, a, was the probability of the study rejecting the null hypothesis [i.e., georeferenecable, department-level, geo-morphologically stratified explanator, associated to morbidity can be optimally regressed with normal error distributions (i.e., non-heteroskedastic, non-multicollinear, residual effects] given that it were true; and the p-value of a result, p, was the probability of obtaining a result at least as extreme, given that the null hypothesis were true. The distribution was Gaussian. The model variance

implied a substantial variability in the regressed covariates associated with the morbidity across the georeferenced departments. Since the morbidity model revealed a normal distribution of the, diagnostic, epidemiological, morbidity forecasts, and the likelihood that a relationship between two or more variables in the department-level, empirical, estimator datasets was caused by something other than random chance. Compilation of additional and accurate geo-spatiotemporal, department-level, morbidity data in Guatemala may allow continual updating of frequentist and non-frequentist (e.g., iterative Bayesian) random effects term estimates. In so doing research intervention teams may be able to bolster the quality of regressors in vulnerability, geo-morphologically stratified, epidemiological datasets for future modeling efforts.

Keywords: Morbidity, Guatemala, GIS, Regression Modeling, Gaussian.

1. Introduction

What are the influencing factors to morbidity in developing countries? What is the causation and why does morbidity occur at a given geographical location (henceforth geolocation) over another area? Distinguishing predictive probability distribution in regression models may aid in localizing statistically significant optimal, parameterizable covariates associated with morbidity statistics. Morbidity statistics is a type of statistics employed largely in the public health epidemiology (www.cdc.gov). These data focuses on the disease progression of a given population or sub-population, and or geolocation whilst quantitating the significance, prevalence, or expanse of disease in a group.

Currently, exponential distributional algorithms are available in the literature for quantitatively regressing morbidity–related georeferenceable descriptors. For example, predictors of post-operative mortality, morbidity, and long-term survival in patients with stage IV colorectal cancer was examined by Stillwell and Boden (2011) [1]. This study aimed to identify independent explanatorial regressors of postoperative mortality and morbidity as well as independent inferential covariates of long-term survival. The study was planned as a retrospective single-institution review. This study took place at the Department of Surgery at the Royal Brisbane and Women's Hospital, in Australia, between 1984 and 2004.Prospectively collected data were extracted from the records of 1,867 patients undergoing treatment for colorectal cancer. The outcomes for 379 patients undergoing surgical resection of their primary colon or rectal tumor in the presence of unresectable synchronous metastases were analyzed.

Independent predictive factors for postoperative mortality and morbidity as well as long-term survival were assessed by use of logistic regression and Cox regression analysis. Thirty-five (9.2%) patients died in the postoperative period and morbidity was 48.3%. Median survival was 11 months. Thirty-day postoperative mortality was independently associated with medical complications ($P < .001$), emergency operations ($P = .001$), female sex ($P = .002$), and age (≥ 70; $P = .007$) on regression analysis. Elderly (≥ 70) patients with either advanced local disease or extrahepatic metastases were at a particularly high risk. Preoperative predictors of surgical morbidity included male sex ($P = .028$) and advanced local disease ($P = .036$). Preoperative explanatory predictors of medical complications included repeat operations ($P < .001$), elevated urea levels ($P = .017$), and emergency operations ($P = .003$). Independent factors associated with poor overall survival included medical complications ($P < .001$), nodal stage (N2) ($P = .004$), poor tumor differentiation ($P = .006$), and apical lymph node involvement ($P = .042$). A subgroup of patients with advanced nodal disease (N2) and a poor tumor differentiation had a significantly poorer prognosis. The vulnerability, forecast, epidemiological, regression model found that elderly patients with advanced local disease or extrahepatic metastases were at high risk of 30-day postoperative mortality.

In statistical, predictive, vulnerability, epidmeiological, modeling, regression analysis is a statistical process for estimating the relationships amongst various variables (e.g, sociodemiographic, time series etc.) [2]. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between an explanatory, dependent variable and one or more independent variables ('prognosticators'). More specifically, time series, regression analysis for a dataset of epidemiological, morbidity, geo-morphologically stratified, parameter estimators may aid in understanding understand how the typical value of the dependent variable (a 'criterion

department-level, morbidity prevalence statistic) changes when any one of the independent variables is varied, whilst the other independent covariates are held fixed.

Most commonly, the conditional mean y (the dependent variable) given the value of x (independent variable) in a geomorphological, stratifiable, forecast, vulnerability, morbidity, department-level, georeferenecable, epidemiological model may be assumed to be affine of x. Less commonly the median or some quantile of the conditional distribution of y given x is expressed in a regression analysis [Hosmer and Lemeshew 2002]. Based on probability theory, two time series dependent, morbidity events *R* and *B* may be conditionally independent given a third event *Y* precisely, if the occurrence of *R* and the occurrence of *B* are independent events in their conditional probability distribution given *Y* . If *R* and *B* are conditionally independent given *Y* in the model, knowledge of whether *R* occurs will provide no information on the likelihood of *B* occurring. Further, knowledge of whether *B* occurs in the model will provide no information on the likelihood of *R* occurring. In geometry, a transformation, between affine spaces preserves points, straight lines and planes [3].

An affine transformation, however, may not necessarily preserve angles between lines or distances between, explanatorial, morbidity-related, georeferenceable, capture points (geolocations of hyperendemic morbidity foci).The model may preserve ratios of distances between the capture points lying on a straight line in an empirical, department-level, epidemiological dataset of, geomorphological-stratified, geo-spatiotemporally geosampled, morbidity statistics, Linear regression focuses on line distribution which focuses on x rather than joint distribution of y and x, which is the domain of a multivariate analysis [4].

The Poisson distribution is popular for probability modelling the number of times an event occurs in an interval of time or space which may be robust when optimally regressively quantitating department-level, geo-morphologically stratified, geo-sampled, geo-spatiotemporal, morbidity-specified, vulnerability parameters. A Poissonian regression model is a generalized linear model used to y model count data and contingently [5]. In probability theory and statistics, the Poisson distribution , is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space, if these events occur with a known average rate and independently of the time since the last event [5]. The Poisson distribution can also be employed for determining the number of events based on other specified intervals such as distance, area or volume measurements [6]. Poissonian distribution may be a distribution sampling based on the number of frequency, time, that depends on only one parameter, the mean number of morbidity events given set times of the same span

Poisson models constructed from a generalizable dataset of hierarchical, morbidity, statistical model, geo-referenceable, parameterizable unbiased, frequency regressors will assume the response variable y has a Poissonian distribution. The logarithm of a probability, Poissonian, morbidity-related, department-level, geomorphological stratified, explanative or expected regressable, attribute value may be modeled by a linear combination of unknown parameters. Hence, a discrete probability distribution may be robustly, parsimoniously rendered from a Poissionian, morbidity-specified, forecast, vulnerability, algorithmic model that expresses the probability of a given number of potential, geo-referenceable, events or covariates occurring in a fixed interval of time and/or space, if these events occur with a known average rate and independently of the time since the last sample event or sampled, unbiased estimator. For example given a Poisson process, the probability of obtaining exactly n successes in N trials may be given by the limit of a binomial distribution $$P_p(n \mid N) = \frac{N!}{n!\,(N-n)!}\, p^n (1-p)^{N-n}.$$ [1.1]henst constructing a morbidity, department-level, forecast, vulnerability model. In probability theory and statistics, the binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes–no question, and each with its own boolean-valued outcome: a random variable containing single bit of information: success/yes/true/one (with probability p) or failure/no/false/zero (with probability q = 1 − p) [7].

Jacob et al. (2014) [8] viewed equation [1.1] as the distribution of a function of the expected number of successes $v = Np$instead of the sample size N which had a fixed p, which then optimally quantitated an empirical dataset of geo-morphologically stratfieid, multidrug resistant tuberculosis, (MDR-TB) estimators in a Poissonian

model employing $P_{\nu|N}(n \mid N) = \frac{N!}{n!\,(N-n)!}\left(\frac{\nu}{N}\right)^n\left(1-\frac{\nu}{N}\right)^{N-n}$, for deducing forecastable,time series, vulnerability covariates for San Juan de Lurigancho Lima, Peru. Letting the sample size N become artificially inflated, in the morbidity model the distribution approached $P_\nu(n) = \lim_{N\to\infty} P_p(n \mid N) \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!}\frac{\nu^n}{N^n}\left(1-\frac{\nu}{N}\right)^N\left(1-\frac{\nu}{N}\right)^{-n} =$ $\lim_{N\to\infty}\frac{N(N-1)\cdots(N-n+1)}{N^n}\frac{\nu^n}{n!}\left(1-\frac{\nu}{N}\right)^N\left(1-\frac{\nu}{N}\right)^{-n} = 1\cdot\frac{\nu^n}{n!}\cdot e^{-\nu}\cdot 1 = \frac{\nu^n e^{-\nu}}{n!}$, (i.e., Poisson distribution) (see Papoulis 1984, pp. 101 and 554; Papoulis, A. "Poisson Process and Shot Noise." Ch. 16 in *Probability, Random Variables, and Stochastic Processes, 2nd ed.* New York: McGraw-Hill, pp. 554-576, 1984.). Note that the sample size N completely dropped out of the probability function in the MDR-TB model, which had the same functional form for all values of ν.

Jacob et al. (2014) [8] then constructed multiple, georeferenced, hierarchical models accompanied by non-generalized predictive residual, uncertainty, non-normal, diagnostic tests employing multiple, covariate, coefficient estimates clinically-sampled at the epidemiological study site. A SAS-based, agglomerative, polythetic clustering algorithm was employed to cartographically delineate, high and low, MDR-TB clusters geo-morphologically 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. An iterative, non-frequentist, probabilistic, Bayesian, estimation matrix was then constructed employing normal priors for each of the error coefficient estimates which revealed both spatially structured and spatially unstructured effects in the MDR-TB geo-spatiotemporal, geo-sampled, geo-referenceable data. 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, time series, dependent cluster, single marital status and building material used for house construction were important predictors. The authors conclude that latent non-normal, (e.g., lepokurotic distributions, unquantitated heteroskedastic parameters), erroneous, propogagtional probabilities in empirically regressed MDR-TB, clinical-geosampled, parameterizable, Poissionized estimates can aid in distinguishing unbiased inferencial covariates and non-zero first-order lag autocorvariate error in frequentistic and non-frequentistic, optimally regressable Multi-Drug Resistant Tuberculosis time series estimators.

An explicative, Poissonian, morbidity, forecast, vulnerability, epidemiological, morbidity model geosampled process may occur at a constant rate λ per unit time. Suppose that an epidemiologist or a public health officer interprets the changes in a Poissoninan process from a parameterizable morbidity-specfied, estimator point of view, (i.e., a change in the Poisson process for defining a termination of a system,). This process would count the number of terminations as they occur in an empirical, departmental-level, geo-referenceable, geomorphological, stratified, epidemiological dataset in regression space. In so doing, the rate of change λ may be interpreted as a hazard rate (or failure rate or force of morbidity) in the vulnerability forecasts. With a constant force of morbidity, the time until the next change may be exponentially distributed in a department–level model.

The epidemiologist or public health officer may be able to optimally quantitate the hazard rate function in a more general setting in the probability paradigm. The hazard function (also known as the failure rate, hazard rate, or force of mortality) $h(x)$ is the ratio of the probability density function $P(x)$ to the survival function $S(x)$, given by $h(x) = \frac{P(x)}{S(x)} = \frac{P(x)}{1-D(x)}$, where $D(x)$ is the distribution function (Evans, M.; Hastings, N.; and Peacock, B. *Statistical Distributions, 3rd ed.* New York: Wiley, 2000. p. 13) In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random

variable to take on a given value [9]. However, the department-level, geo-morphological, stratified, explanatory process that counts the number of terminations may not have a constant hazard rate in a morbidity model specification, but instead may have a hazard rate function $\lambda(t)$, based on a function of time t.Such a counting process may be quantifiable at the department-level employing a non-homogeneous Poisson process, morbidity model.

A non-homogeneous Poisson process is similar to an ordinary Poisson process, except that the average rate of arrivals is allowed to vary with time. Many applications that generate random points in time are modeled more faithfully with such non-homogeneous processes. The mathematical cost of this generalization, however, is that a forecast, vulnerability, morbidity –specified, optimizable model may lose the property of stationary increments. Non-homogeneous Poisson processes are best described in measure-theoretic terms (see Appendix 1). A stochastic model constructed based on a nonhomogeneous Poisson process may reveal iteratively interpolative, covariates associated with a hyperendemic, department-level, and capture point. Further, the formulation may allow for test coverage and detection coverage for empirical geo-referenecable, geo-sampled, department-level, geo-spatiotemporal, epidemiological, geo-morphologically stratified, morbidity statistics whilst providing a new decomposition of the mean value function.

Consider a nonhomogeneous Poisson process on [0,T][0,T] with mean value function m(t)m(t) for t∈[0,T]t∈[0,T] in a forecast, vulnerability, morbidity model. If an epidemiologist or public health officer lets X1X1 denote the time of the first termination event It may be shown that (X1|N(T)=1)(X1|N(T)=1) has the following cdf: F(x)=m(x)m(T),x∈[0,T]F(x)=m(x)m(T),x∈[0,T]. In probability theory and statistics, given two jointly distributed random variables X and Y, the conditional probability distribution of Y given X is the probability distribution of Y when X is known to be a particular value {Freedman 2005] In this case x in the model may would refer to the morbidity-related event (X1|N(T)=1)(X1|N(T)=1) not just F(x)=m(T)F(x)=m(T). In so doing, robust vulnerability georeferenceable forecastable morbidity statistics may be procured from an epidemiological, empirical dataset of vulnerability parameterizable, unbiased estimators.

Jacob et al. [8] employed a Monte Carlo simulation to assess the statistical properties of some Bayes where only a few time series, vulnerability data parameter estimators on a system was governed by a nonhomogeneous Poisson process where there was only imprecise prior information available. In particular, two Bayes procedures were analyzed employing truncated data. The first model employed a uniform prior probability distribution function (PDF) for the power law and a noninformative prior PDF for alpha, while the other employed a uniform PDF for the power law while assuming an informative PDF for the scale parameter obtained by employing a gamma distribution for the prior knowledge of the mean number of failures in a given time interval. For both cases, point and interval estimation of the power law and point estimation of the scale parameter were exploited. Comparisons were given with the corresponding capture point and interval maximum-likelihood estimates for sample sizes of 25 and 100. The Bayes procedures were computationally much more onerous than the corresponding maximum-likelihood ones, since they in general required a numerical integration. In the case of small sample size, however, their use may be justifiable by the exceptionally favorable statistical properties shown when compared with the classical ones. In particular, their robustness with respect to a wrong assumption on the prior beta mean was interesting based on the Poissonian non-homogeneous distributed model Hence various characterizations of the ordinary Poisson process, in terms of the inter-arrival times, the arrival times, and the counting process, and their characterizations may involve the counting process leading to the most natural generalization of a non-homogeneous processes for a departmental –level, geosampled, geomorphological stratified, geo-classified, landscape date feature attribute.

Consider a morbidity process that generates random points in time in a Possionian vulnerability, morbidity model. Then let N denote the number of explanatorial random points in the interval (0,t] for t≥0, so that N={Nt:t≥0} is the counting process. More generally, N(A)would denote the number of random morbidity, explanatorial, points in a measurable A⊆[0,∞) in the model so N would be random counting measure. As before, t↦Nt would be a random distribution function and A↦N(A)would be the random measure associated with the morbidity probability

distribution function. Suppose now that r:[0,∞)→[0,∞) is measurable in the model and dm:[0,∞)→[0,∞) by m(t)=∫(0,t]r(s)dλ(s). From properties of the integral in the forecast, vulnerability, morbidity, Poissonian model increasing right-continuously on [0,∞) a robust distribution function would be rendered. The positive measure on [0,∞) associated with a geosampled, department-level morbidity statistic may be optimally identifiable on measurable A⊆[0,∞) by m(A)=∫Ar(s)dλ(s)m. Thus, m(t)=m(0,t] and for s,t∈[0,∞) with s<ts<t, m(s,t]=m(t)−m(s)m(s,t]=m(t)−m(s) should render statistically significant, geomorphologically stratifiable, geo-spatiotemporally dependent, georeferenceable, demographic and landscape, paremeterizable covariates.

Recall that mean and variance of Poisson distribution are the same; e.g., E(X) = Var(X) = λ. However in practice, the observed variance is usually larger than the theoretical variance and in the case of Poisson, larger than would be the mean. This is known as overdispersion, an important concept that occurs with discrete data. Here we assumed that each term in an forecast, vulnerability , morbidity, epidemiological Poisson, model has the same probability Analyses assuming binomial, Poisson or multinomial distributions are sometimes invalid because of overdispersion [5].

We constructed a regression models (Poissonian) employing geo-morphological stratified, explanative geo-referenceable datasets of morbidity covariates ge-osampled in Guatemala. Regression is based on the quality of the predictor variable as an outcome variable and which selected variables have the heaviest weight on the dependent variable [10]. Our covariates were synthesized from existing morbidity time series, geo-sample datasets at the departmental level in Guatemala. Our assumption was that a frequentist, covariate at the departmental-level would aid in determining causation of morbidity in Guatemala. Our objectives were; 1) to construct a Poisson model with a 95% confidence interval employing multiple demographic and landscape explanatory, independent variables 2) to tease out noisy (e.g., variables that do not have a normal error distribution) feature data attributes in the vulnerability forecasts; and, 3) to check any violations of assumptions (non-homoskedasctic residuals) in a department–level, geomorphological stratified, morbidity, regression analyses for Guatemala.

2. Materials and Method

2.1 Study site

Guatemala is located in Central America. The country is bordered by Mexico on the west and north, Belize to the north east, El Salvador and Honduras on the south east. To the southern coast is the Pacific Ocean with the Caribbean Sea on the mid-east gap between Belize and Honduras. Guatemala is the most populated county in Central America, with an estimated 15.8 million. Guatemala is broken down by departments, totaling twenty-two. Within those departments consists of municipalities and varies by department [11].

Guatemala takes up a total of 108,889 km^2. It has three main regions: the highlands, the Pacific coast, and Peten region. The northern portion of Guatemala, which is the department of Peten, is imperceptibly populated. Most major cities are along the southern coastal regions of the country (Figure 1).

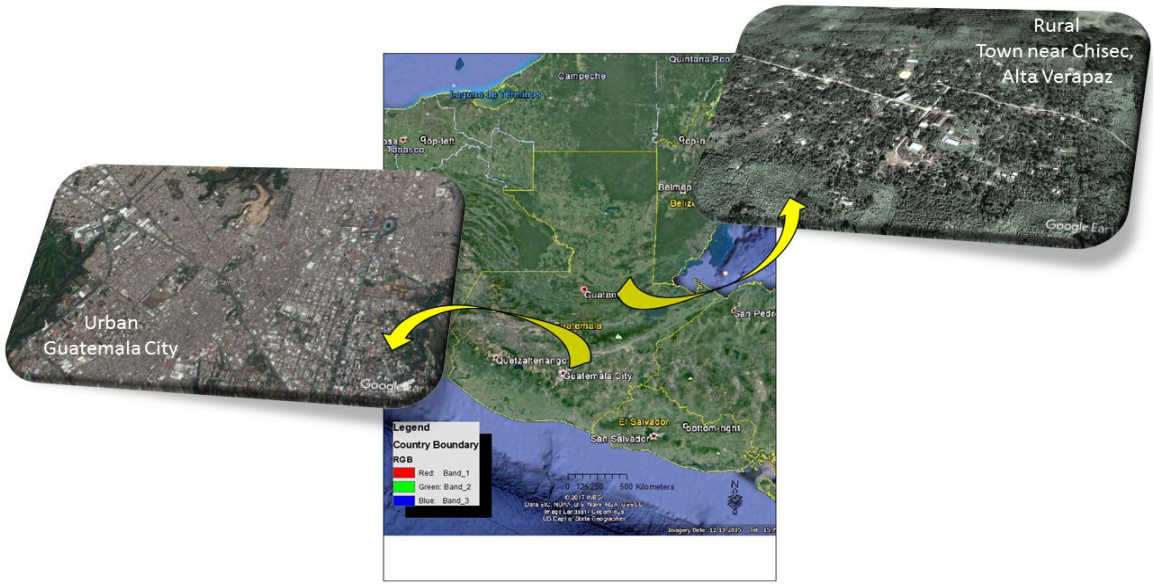

Figure 1 – The Guatemala study site

2.2. Data Gathering

The source for Guatemalan mortality data was obtained from the National Institute of Statistics Guatemala (INE Guatemala) with Microsoft Excel spreadsheets [12]. We worked with data collected for 2014 since year, there was the most data completed. The excel spreadsheet was prepared for Statistical Analysis System (SAS). Using only data from 2014 at the department level, death count, gross mortality rate, and mortality rate by malnutrition (E40-E46) for every 100,000 inhabitants were lined up as the dependent variables (observations). Rates for all three were lined up by all twenty-two departments (variables). Mortality rates cause externally for 100,000 habitants, rate of mortality from malnutrition for every 100,000 inhabitants for 2014, and gross mortality rate of mortality from malnutrition for every 100,000 inhabitants for 2014 were analyzed for Guatemala's twenty-two departments. The gross mortality rate for every 100,000 inhabitants for 2014 was independent variable was applied to the model.

Specific demographic data were selected as primary targets for this study. These included ethnicity of age groups of dead by department, sex of dead by residential department, and geographic area type of dead and ethnicity type of dead by department. Although there were more data provided from the database, some were omitted because they were not attached to a geographical location and several pieces of data were missing. This was the case for data labeled "ignorado" (ignored), "extranjero" (foreigner). There are many recognized ethnicities in Guatemala, the main groups which were collected were Maya, Garifuna, Xinka, and Mestizo/Ladino. For this analysis, other, ignored, Garifuna, and Xinka were excluded.

The figures below are of Guatemala in ArcMap. Figure 2 shows Guatemala with its departmental boundaries. Figure 3 shows Guatemala City with several data layers. Point of interest and places are historical and tourist sites. Artificial land use is land used for commercial and residential purposes. Natural land use is land use as parks. Figure 4 is a map of the northern region of Guatemala, the entire department Peten. We noticed that as we have the same layers applied as in Figure 3, Peten is noticeably less populated than the southern region of Guatemala.

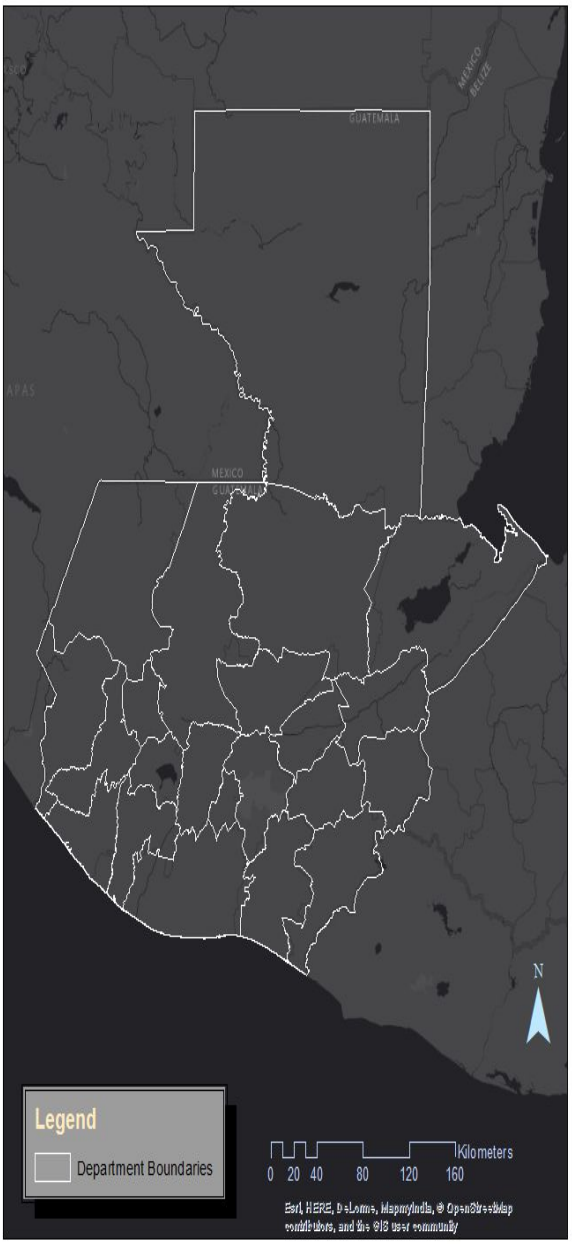

Figure 2 - Departments in Guatemala [23]

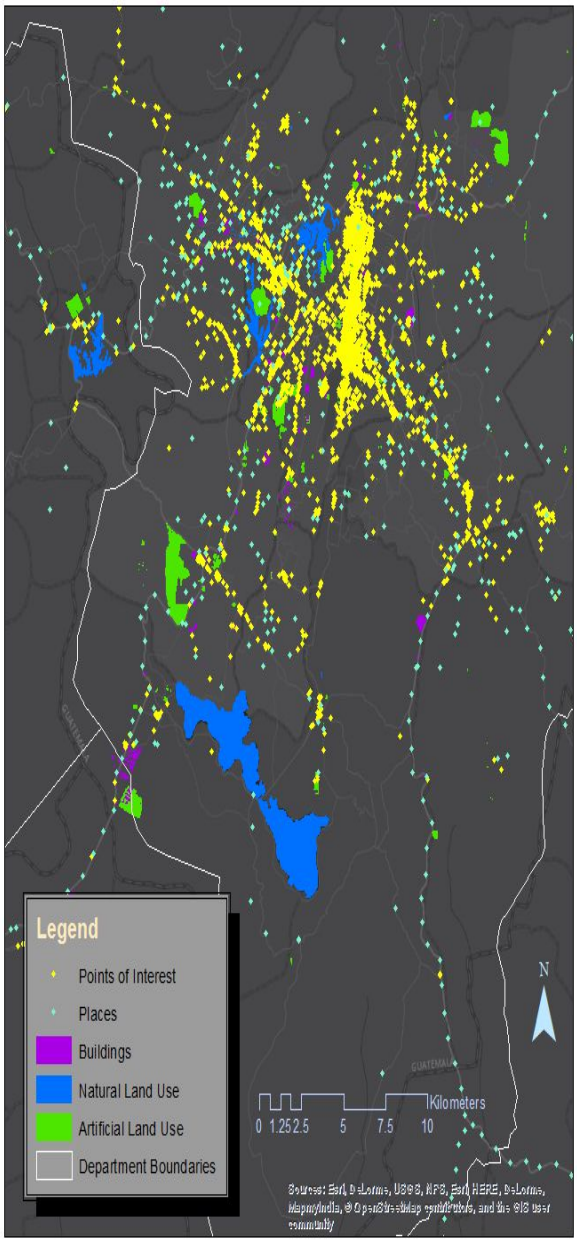

Figure 3 - Guatemala City [23][24]

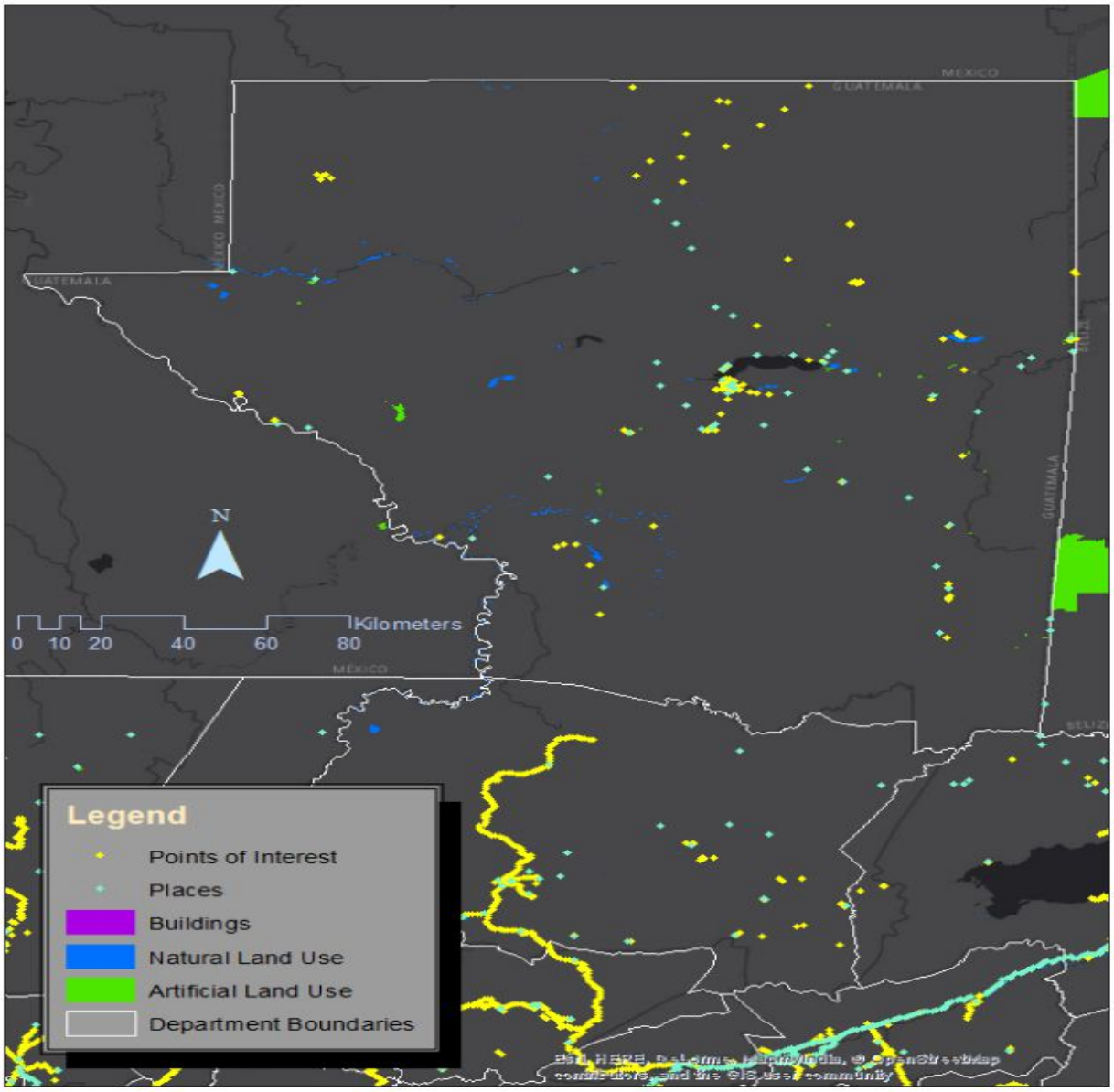

Figure 4 - Region/Department of Peten [23][24]

2.3 Regression analyses

Although logistical regression modeling with co-binary outputs are used for reviewing morbidity statistics, we employed a Poisson probability regression as the Guatemalan departmental datasets had count data. Unlike a binary logistic regression model which employs independent variables where each covariate is log-transformed so as to render a dichotomous outcome (0= absence , 1=presence), a Poissonian probability paradigm uses actual count data; hence a more robust non-deflated pseudo R^2 may be rendered [5]. Poisson probability, regression analyses were employed to infer the relationship between the geo-sampled, time-series, explanatorial, endemic, count, data variables and the archived department-level anthropogenic characteristics (i.e., independent variables). Further, since we had continuous variables that represented the explanatory regressors, the Poisson model was the more logistical application.

The relationship between the geo-sampled, department -level, endemic, socio-demographic, and landscape covariates was investigated by single variable regression analysis in PROC REG. Since prevalence data are binomial fractions, a regression model was employed, as is standard practice for the analysis of the department-level data.

The regression analyses assumed independent counts (i.e., *Ni*), taken at multiple, geo-sampled, geo-referenced, department-level $i = 1, 2, \square, n.$ The geo-spatiotemporal-related, department-level counts were then described by a set of variables denoted by matrix $\mathbf{X}_i$, where a $1 \times p$ was a vector of covariate coefficient indicator values for a interpretively geo-sampled, endemic transmission-oriented, explanative foci *i*. The expected value of these data was given by $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i)\exp(\mathbf{X}_i\beta),$ where β was the vector of parameterizable non-redundant covariates in the endemic, transmission-oriented, time-series, operationalizable, epidemiological, prognosticative, department-level, geomorphological-stratified, risk model where the Poisson rates were given by $\lambda_i(\mathbf{X}_i) = \mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i).$ The rates parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for each geo-sampled, endemic, morbidity, georeferenceable, departmental geolocation *i*. The dependent variable was department-level mortality. The Poisson regression model assumed that the geosampled, operationalizable, expository predictors was equally dispersed, that is, that the conditional variance equaled the condition mean. Partial correlations were then defined after introducing the concept of conditional distributions.

We initially restricted ourselves to just the explanatorial, conditional distributions obtained from the multivariate normal distributions. We noted an $n \times 1$ random vector **Z** which we partitioned into two random vectors **X** and **Y** where **X** was an $n_1 \times 1$ vector and **Y** was an $n_2 \times 1$ vector in the equation $Z = (XY).$ The conditional distribution properties of the regressed endemic morbidity operationalizable, asymptotical, endemic, transmission-oriented, parameterizable, department-level, socidemiographic covariate coefficients were then defined. Thereafter, we partitioned the mean vector and covariance matrix in a corresponding manner. That is, $\mu = (\mu 1 \mu 2)$ and $\sum = (\sum_{11}\sum_{21}\sum_{12}\sum_{22})$. In so doing, μ_1 optimally rendered the means for the regressed, time-series, explanatorily geo-spatiotemporal, geo-sampled, interpretively interpolative, asymptotically normalized, endemic, georeferenced, socio-demiographic prognosticative variables in the set $x_1,$ and $\sum_{11}$ along with the variances and covariances for set $x_1.$ The matrix $\sum_{12}$ provided the covariances between the observations [e.g., x_1 and set x_2] as did matrix $\sum_{21}.$ Any distribution for a subset of variables from multivariate normal, conditional on known infectious disease, geo-classified, eco-epidemiological values for another subset of variables has a multivariate normal distribution [3,8]. We noted that the conditional distribution of x_1 given the known values for x_2 was multivariate normal with a explanatorial, time series mean vector covariance $\text{matirx} = \mu 1 + \sum_{12}\sum -122(x_2 - \mu 2)\sum_{11} - \sum_{12}\sum -122\sum_{21}.$ The procedure employed a maximum likelihood estimation to find the operationalized, time-series, dependent, regression coefficients. The data was then log-transformed before analyses to normalize the distribution and minimize standard error.

The Poisson regression model can be generalized by introducing an unobserved heterogeneity term for observation *i*. Thus, the individuals are assumed to differ randomly in a manner that is not fully accounted for by the observed covariates. This is formulated as $E(y_i|\mathbf{x}_i, \tau_i) = \mu_i\tau_i = e^{\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ is independent of the vector of regressors $\mathbf{x}_i$. Then the distribution of y_i conditional on $\mathbf{x}_i$ and τ_i is Poisson with conditional mean and conditional variance $\mu_i\tau_i$: $f(y_i|\mathbf{x}_i, \tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$. We let $g(\tau_i)$ be

the probability density function (PDF) of τ_i. Then, the distribution $f(y_i|\mathbf{x}_i)$ (no longer conditional on τ_i) is obtained by integrating $f(y_i|\mathbf{x}_i, \tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i$

The Poisson regression morbidity model was generalized by introducing an unobserved heterogeneity term for observation *i*. Thus, the individuals were assumed to differ randomly in a manner that is not fully accounted for by the observed covariates. This was formulated as $E(y_i|\mathbf{x}_i, \tau_i) = \mu_i \tau_i = e^{\mathbf{x}_i'\boldsymbol{\beta}+\varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ is independent of the vector of regressors $\mathbf{x}_i$. Then the distribution of y_i conditional on $\mathbf{x}_i$ and τ_i is Poisson with conditional mean and conditional variance $\mu_i \tau_i$: $f(y_i|\mathbf{x}_i, \tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$.

We let $g(\tau_i)$ be the PDF of τ_i in the, department-level, stratified, morbidity model. Then, the distribution $f(y_i|\mathbf{x}_i)$ was no longer conditional on τ_i hence the distribution was obtained by integrating $f(y_i|\mathbf{x}_i, \tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i$

An analytical solution to this integral exists when τ_i was assumed to follow a gamma distribution. This solution was a binomial distribution. When the model contains a constant term, it is necessary to assume that $E(e^{\varepsilon_i}) = E(\tau_i) = 1$ in order to identify the mean of the distribution [3]. Thus, we assumed that τ_i follows a gamma (θ, θ) distribution with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$, $g(\tau_i) = \dfrac{\theta^\theta}{\Gamma(\theta)} \tau_i^{\theta-1} \exp(-\theta\tau_i)$ where $\Gamma(x) = \int_0^\infty z^{x-1} \exp(-z) dz$ is the gamma function and θ is a positive parameter. Then, the density of y_i given $\mathbf{x}_i$ is derived as

$$\begin{aligned} f(y_i|\mathbf{x}_i) &= \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i \\ &= \frac{\theta^\theta \mu_i^{y_i}}{y_i! \Gamma(\theta)} \int_0^\infty e^{-(\mu_i+\theta)\tau_i} \tau_i^{\theta+y_i-1} d\tau_i \\ &= \frac{\theta^\theta \mu_i^{y_i} \Gamma(y_i+\theta)}{y_i! \Gamma(\theta)(\theta+\mu_i)^{\theta+y_i}} \\ &= \frac{\Gamma(y_i+\theta)}{y_i! \Gamma(\theta)} \left(\frac{\theta}{\theta+\mu_i}\right)^\theta \left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i} \end{aligned}$$

Making the substitution $\alpha = \frac{1}{\theta}$ $(\alpha > 0)$, in the morbidity, department-level,epidemiological model the distribution was then rewritten as $f(y_i|\mathbf{x}_i) = \dfrac{\Gamma(y_i+\alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\dfrac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}} \left(\dfrac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}, \quad y_i = 0, 1, 2, \ldots$ Thus, the department-level, stratified, morbidity distribution was derived as a gamma mixture of Poisson random variables. It had a conditional mean $E(y_i|\mathbf{x}_i) = \mu_i = e^{\mathbf{x}_i'\boldsymbol{\beta}} V(y_i|\mathbf{x}_i) = \mu_i[1 + \frac{1}{\theta}\mu_i] = \mu_i[1 + \alpha\mu_i] > E(y_i|\mathbf{x}_i)$ and conditional variance.

3. Results

A Poisson regression analyses was constructed in PROC REG to determine the relationship between the morbidity, count data and the departments. The Poisson models were built by employing the, time-series, explanatory, endemic, department level, stratified, morbidity, parameterizable, demographic and landscape covariate

coefficients. We assumed that the log of the mean μ was a linear function of independent variables, [i.e. $\log(\mu) = \text{intercept} + b1^*X1 + b2^*X2 + \square + b3^*Xm$] in the department-level morbidity model which implied that μ was the exponential function of the independent variables when $\mu = \exp(\text{intercept} + b1^*X1 + b2^*X2 + \square + b3^*Xm)$. This, resulted in the distribution of the geo-sampled, parameter estimates (i.e., *Y*).

The Poisson regression morbidity probability model was generalized by introducing an unobserved heterogeneity term for observation *i*. Hence, the individuals were assumed to differ randomly in a manner that is not fully accounted for by the specified morbidity covariates. This was formulated as $E(y_i|\mathbf{x}_i, \tau_i) = \mu_i\tau_i = e^{\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was the independent of the vector of regressors $\mathbf{x}_i$. Then the distribution of y_i conditional on $\mathbf{x}_i$ and τ_i was Poisson with conditional mean and conditional variance $\mu_i\tau_i$: $f(y_i|\mathbf{x}_i, \tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$ We let $g(\tau_i)$ be the PDF of τ_i. Then, the distribution $f(y_i|\mathbf{x}_i)$ (no longer conditional on τ_i) was obtainable by integrating $f(y_i|\mathbf{x}_i, \tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i)g(\tau_i)d\tau_i$.

In the regression analyses, of the department-level, geo-referenceable, explanatory, endemic, morbidity, demographic interpolative, asymptotically normalized data the null hypothesis was: $H_0 : k = 0$ and the alternative hypothesis was: $H_a : k > 0$. We recorded the log-likelihood (i.e., LL) and the likelihood ratio (LR) test to compute the LR statistic using –2(LL) (Poisson) and the LL (i.e., negative binomial). The asymptotic distribution of the LR statistic had a probability mass of one half at zero and one half - Chi-square distribution with 1 df. To test the null hypothesis at the significance level α, the critical value of Chi-square distribution corresponding to significance level 2α, whereby there was a rejection of H_0 if LR statistic $> \chi^2(1-2\alpha, 1\text{df})$. The log of the mean, μ was generated using a linear function of the independent variables whereby, $\log(\mu) = \text{intercept} + b1^*X1 + b2^*X2 + \square + b3^*Xm$, in the regressed, time-series, endemic, transmission-oriented, predictive, model which implied that μ was the exponential function of the explanatory, diagnostic, department-level, independent variables when $\mu = \exp(\text{intercept} + b1^*X1 + b2^*X2 + \square + b3^*Xm)$. The SAS model data was then log-transformed and run.

As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1, since $\sum_{n=0}^{\infty} P_\nu(n) = e^{-\nu}\sum_{n=0}^{\infty}\frac{\nu^n}{n!} = e^{-\nu}e^{\nu} = 1$. The ratio of probabilities in the explicative, morbidity, forecast, vulnerability, department-level model was given by $\dfrac{P_\nu(n=i+1)}{P(n=i)} = \dfrac{\frac{\nu^{i+1}e^{-\nu}}{(i+1)!}}{\frac{e^{-\nu}\nu^i}{i!}} = \dfrac{\nu}{i+1}$.

The characteristic function for the Poisson distribution was $\phi(t) = e^{\nu(e^{it}-1)}$ (Papoulis 1984, pp. 154 and 554), and the cumulant-generating function was $K(h) = \nu(e^h - 1) = \nu\left(h + \frac{1}{2!}h^2 + \frac{1}{3!}h^3 + \ldots\right)$, so $\kappa_i = \nu$, in the morbidity

model The mean deviation of the Poisson distribution was given by $\mathrm{MD} = \frac{2\, e^{-\nu}\, \nu^{\lfloor \nu \rfloor + 1}}{\lfloor \nu \rfloor !}$. The Poisson distribution was also expressable in terms of $\lambda \equiv \frac{\nu}{x}$, the rate of changes, so that $P_\nu(n) = \frac{(\lambda x)^n e^{-\lambda x}}{n!}$. The moment-generating function of a Poisson distribution in two variables is given by $M(t) = e^{(\nu_1 + \nu_2)(e^t - 1)}$ [Haight 1967].If the independent variables x_1, x_2, ..., x_N have Poisson distributions with parameters μ_1, μ_2, ..., μ_N, then $X = \sum_{j=1}^{N} x_j$ has a Poisson distribution with parameter $\mu = \sum_{j=1}^{N} \mu_j$. [Freedman 2005].In the morbidity, forecast, vulnerability, epidemiological, department-level model, this parameter was the cumulant-generating function which was quantitated as $K_j(h) = \mu_j \left(e^h - 1\right)$. In probability theory and statistics, the cumulants κ_n of a probability distribution are a set of quantities that provide an alternative to the moments of the distribution. [Freedman 2005]

We let $M(h)$ be the moment-generating function, then the cumulant generating function was given by $K(h) = \ln M(h) = \kappa_1 h + \frac{1}{2!} h^2 \kappa_2 + \frac{1}{3!} h^3 \kappa_3 + \ldots$, $K \equiv \sum_j K_j(h) = \left(e^h - 1\right) \sum_j \mu_j = \mu \left(e^h - 1\right)$. Where κ_1, κ_2, ..., were the cumulants. If $L = \sum_{j=1}^{N} c_j x_j$ is a function of N independent variables, then the cumulant-generating function for L is given by $K(h) = \sum_{j=1}^{N} K_j(c_j h)$. [13]. In probability theory and statistics, the moment-generating function of a real-valued random variable is an alternative specification of its probability distribution. Thus, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions [14].

The tables below are the SAS outputs running PROC REG. The results fit the Poissonian model. See Tables 1 and 2 for output.

Table 1 - Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	18	1.3116	0.0729
Scaled Deviance	18	1.3116	0.0729
Pearson Chi-Square	18	1.2298	0.0683
Scaled Pearson X2	18	1.2298	0.0683
Log Likelihood		69.3634	
Full Log Likelihood		-38.7283	
AIC		85.4565	
AICC		87.8094	
BIC		89.8207	

Table 2 - Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.1146	0.4345	0.2360	1.9662	6.58	0.0103
County	1	-0.0076	0.0153	0.0377	0.0224	0.25	0.6177
Malnutrition	1	0.0047	0.0248	-0.0438	0.0533	0.04	0.8483
External	1	0.0072	0.0039	-0.0003	0.0148	3.53	0.0604
Scale	0	1.0000	0.0000	1.0000	1.0000		

4. Discussion

We constructed a Poissonian model to determine which explanatory covariate at a 95 percentile confidence interval was associated with morbidity at the department-level in Guatemala. The probability model revealed robust covariates "malnutrition" "external causes" gross mortality rates" were significant at a 95% confidence interval.

We checked all Poissonian assumptions in our model. The Poisson distribution is an appropriate model if the following assumptions are true 1) k is the number of times an event occurs in an interval and k can take values 0, 1, 2, ….; 2) the occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently; 3) the rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals; 4) two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur; (5) The probability of an event in a small sub-interval is proportional to the length of the sub-interval; or 6) the actual probability distribution is given by a binomial distribution and the number of trials is sufficiently bigger than the number of successes one is asking about [5].

The morbidity-related, Poissonian distribution reached a maximum when $\frac{d\,P_{\nu}(n)}{d\,n} = \frac{e^{-\nu}\,n\,(\gamma - H_n + \ln \nu)}{n!} = 0,$ where γ was the Euler-Mascheroni constant and H_n is a harmonic number, leading to the transcendental equation [] $\gamma - H_n + \ln \nu = 0,$ which could not be solved exactly for n. Euler-Mascheroni constant γ, sometimes also called 'Euler's constant' or 'the Euler constant' (but not to be confused with the constant $e = 2.718281\ldots$) is defined as the limit of the sequence $\lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right)$ y= $\lim_{n=n\to\infty}(H_n - \ln n),$ where H_n is a harmonic number [15]. A harmonic number is a number of the form $H_n = \sum_{k=1}^{n}\frac{1}{k}$ arising from truncation of the harmonic series. A harmonic number can be expressed analytically as $H_n = \gamma + \psi_0\,(n+1),$ where γ is the Euler-Mascheroni constant and $\Psi\,(x) = \psi_0\,(x)$ is the digamma function. [i.e., special function which is given by the logarithmic derivative of the gamma function (or, depending on the definition, the logarithmic derivative of the factorial] [5].

A generalization of the Poisson distribution may be employed to model the observed clustering of geo-referenceable, geo-morphologically stratified, morbidity statistics. The form of this distribution may be given by $f_b\,(N) = \frac{\overline{N}\,(1-b)}{N!}\left[\overline{N}\,(1-b) + N\,b\right]^{N-1}\,e^{\overline{N}\,(1-b)-N\,b},$ where N is the number of geo-spatiotemporally dependent, explanative, morbidity-related paramterizable covariates geo-sampled at an epidemiological intervention study site Letting $b = 0$ in the model may render $f_0\,(N) = \frac{e^{-N}\,\overline{N}^{N}}{N!},$ which indeed would be a Poisson distribution with $\nu = \overline{N}$. Similarly, letting $b = 1$ would render $f_1\,(N) = 0$

Fortunately, the Poissonian morbidity model was not over dispersed (whenst variance was equal to the mean). Hence, there was no requirement to use a negative binomial regression model with a non-homogenous gamma distributed mean to compensate for any outliers or other Poissonian noise. Commonly, in Poissonian infectious diseases models, there are propagational uncertainties due to assumptions of regression modeling [2] [16].

Extra-Poison variation in a department level, epidemiological, time series dependent, morbidity model may rest in its ability to utilize the specialized generalized linear model (GLM) it and residual statistics that come with the majority of GLM software. This may allow an epidemiologist or a public health officer to deduce the means for quantitatively testing different, geo-sampled, district-level morbidity, forecastable, vulnerability model estimators with tools built into the GLM algorithm. For example, further extensions to the respective Poissonian probability model constructed from the empirical geo-sampled, geo-spatiotemporal, department –level, morbidity data at the Guatemala study site may be customized depending on the type of underlying problem that is being addressed. These extended models can include, handling excessive response zeros (e.g., zero-inflated Poisson, zero-inflated negative binomial); hurdle models for handling responses having no possibility of zero counts (e.g., zero-truncated Poisson and zero-truncated negative binomial); morbidity models having responses with structurally absent values (e.g., truncated and censored Poisson) and models having longitudinal or clustered department-level morbidity data (e.g., fixed, random, and mixed effects negative binomial).

Further, negative binomial generalized estimating departmental-level morbidity linear-based equations may also be devised for situations when the sampled data can be split into two or more distributional subsets if the models violate the assumption that the mean is equal to the variance. This capability is rarely available with morbidity models estimated using full maximum likelihood or full quasi-likelihood methods.

The ratio in our epidemiological, geo-spatiotemporal, morbidity model indicated that the probability distribution can determine the hazard rate function. In fact, the tabulated ratio was the usual definition of the hazard rate function in our model. That is, the hazard rate function was optimally definable as the ratio of the density and the survival function (one minus the conditional density function).

Given two jointly distributed, geo-morphological, optimally stratifiable department-level, time series, morbidity-related, exploratory, random variables X and Y at the Guatemala epidemiological, study site the conditional probability distribution of Y given X was the probability distribution of Y when X was known to be a particular value in our model. In some cases the conditional probabilities may be expressable as functions containing the unspecified value x of X as a parameter. When both "X" and "Y" are categorical department-level morbidity-related, geomorphological, stratified explanators, a conditional probability table may be typically employed to represent the conditional probability. The conditional distribution may contrast with the marginal distribution of a random, specified, morbidity-related, predictive variable, which may reveal a distribution without reference to the value of the other geo-sampled department-level variables.

If the conditional distribution of Y given X is a continuous probability distribution, in a forecast, vulnerability, department-level, geomorphological, stratified, morbidity model then its' PDF may be known the conditional density function. When both "X" and "Y" are categorical variables, a conditional probability table is typically used to represent the conditional probability. The conditional distribution contrasts with the marginal distribution of a random variable, which is its distribution without reference to the value of the other variable. In probability theory and statistics, the marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. In a forecast, vulnerability, epidemiological morbidity-related, model may render the probabilities of various geo-spatiotemporal, morbidity frequency values of the variables in the subset without reference to the values of the other geosampled variables.

Suppose that two random geo-referenceable, morbidity-related, department-level, geo-morphological, stratified, geo-sampled, forecast, vulnerability, geo-spatiotemporal variables X and Y has a joint density function $f(x, y)$. The joint probability distribution can be expressed either in terms of a joint cumulative distribution

function or in terms of a joint probability density function (in the case of continuous morbidity geo-sampled geo-spatiotemporal variables) or joint probability mass function (in the case of discrete variables) [17]. If $f_X(x) > 0$, in the model, then an epidemiologist or public health officer may define the conditional density function $f_{Y|X}(y|x)$ given $X = x$ by $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$. Similarly an epidemiologist or public health manager can define the conditional density function $f_{X|Y}(x|y)$ given $Y = y$ by $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ if $f_Y(y) > 0$. Then, clearly we have the following relation $f(x,y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y)$. The properties of a conditional distribution, such as the moments, are often referred to by corresponding names such as the conditional mean and conditional variance [6].

More generally, an epidemiologist or other public health officer in Guatemala may refer to the conditional distribution of an empirically regressed, epidemiological, morbidity-related, department-level, geomorphological, stratified, time series subset of a dataset of more than two variables; this conditional distribution may be contingent on the values of all the remaining variables, and if more than one variable is included in the subset then this conditional distribution may be the conditional joint distribution of the included variables.

An epidemiologist or public health officer in Guatemala may also recover the survival function in an departmental-level, geomorphological stratified, epidemiological, morbidity, forecaast, vulnerability model employing multiple, temporally dependent, landscape and socio-demiographic covariates. For example, whenever $\lambda(x) = \frac{f_X(x)}{1 - F_X(x)}$, is optimally derivable in a vulnerability, morbidity regression forecast, model, an epidemiologist or public health officer should be able to solve $S_X(x) = e^{-\int_0^t \lambda(y)dy}$ The function $\Lambda(t) = \int_0^t \lambda(y)dy$ may be definable by a the cumulative hazard rate function. In so doing, the cumulative hazard rate function would be an alternative way of representing the hazard rate function in a department-level, geomorphologically stratifiable, morbidity, forecast, and, vulnerability model.

The probability $P[N_y = 0]$ for a robust, discrete, department-level, geomorphological, stratified, epidemiological, morbidity, explanatorial geo-referenecable variable N_y may be also optimally derivable from the non-homogeneous Poisson process. The continuous random variable T, may be the time until the first change in the morbidity model may be related to N_t in the model renderings. Hence the equation $S_T(t) = P[T > t] = P[N_t = 0] = e^{-\int_0^t \lambda(y)dy}$ may be able to quantitate the predicted probability in a morbidity model employing landscape and demographic, geo-spatiotemporal regressors. The distribution function and density function may also be derivable accordingly. In so doing, the hazard rate function $\lambda(t)$ may be equivalent to each of the following: $\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}$ and $\lambda(t) = \frac{-S'_T(t)}{S_T(t)}$.

A non-homogeneous Poisson process as described in this research, may determine, the hazard rate function $\lambda(t)$ whilst specifying the probability distribution of a department-level, geo-spatiotemporal, geo-morphological stratifiable, explanatorial, forecast, vulnerability morbidity, model T (i.e., the time until the first change). Once the rate of change function $\lambda(t)$ is known in the non-homogeneous Poisson process, an epidemiologist or public health officer can use it to generate the survival function $S_T(t)$. Examples of department-level stratifiable, exploratory,

morbidity models may be constructed by assuming the functional form of the hazard rate function. The result may hold even outside the context of a non-homogeneous Poisson process, that is, given the hazard rate function $\lambda(t)$, an epidemiologist or public health officer may be able to derive three distributional items $S_T(t)$, $F_T(t)$, $f_T(t)$.

Note that a measured covariate in a forecast, vulnerability, morbidity epidemiological geomorphological, stratifiable, model with respect to λ, and r may be absolutely continuous. In such circumstances a department-level epidemiological, forecast, vulnerability model may determine the random distribution function and the deterministic distribution function in the model. A process that produces random points in time (i.e., a non-homogeneous Poisson process) with rate function r may also define the counting process N which may satisfy the following properties in a forecast vulnerability, morbidity model:

a. If {Ai:i∈I}is a countable, disjoint collection of measurable related subsets of [0,∞) then {N(Ai):i∈I}is a collection of independent random variables.
b. If A⊆[0,∞)A⊆[0,∞) is measurable geo-sampled, geo-referenceable, capture point (department-level, hyperendemic geolocation) [e.g., then N(A) has the Poisson distribution with parameter m(A)m(A)].

5. Limitations

Historically, the indigenous people of Guatemala have been the backbone of its very economy. Today, they are employed in the agricultural and textile sector given hunger wages [11]. With the historical stigma of being of indigenous ancestry, the population today considers itself with a now increased 60% Mayan. Those that mostly identify with Mayan ancestry as well as other indigenous community live in rural and in the highlands of Guatemala. With Spanish being the official language of Guatemala, many indigenous people do not speak it. This creates barriers to education, training, healthcare, and other public services. Distinctively, there is a small community of people without an association to the Maya, the Garifuna and Xinka [11].

There have been many readings as Guatemala as the case study. Ramírez et al found connections made with violent deaths in men associated when pay days coincide with holidays [18]. Guberek and Hedstrom found that there have been changes in how deaths have been classified [19]. Cerón et al found that the indigenous suffer from abuse and discrimination in public health care facilities [20]. Poder and He associated social inequality affecting child health, income, and maternal education [21]. Hernandez et al found out that in health care, support is needed in mid-level workers [22]. Although there is much qualitative value from these studies, quantifiable data is needed to know the gravity of morbidity in Guatemala.

Although morbidity model revealed a covariate of statistical significance output at a 95% confidence interval (malnutrition), our data set is limited temporally. In future, forecast, vulnerability, morbidity regression, endemic model construction, larger geo-spatiotemporal, geomorphological-stratified, geo-classified, empirical, parameterizable, estimator dataset should be utilized as the independent variable. Further geospatially weighted frequency autocorrelation model should be constructed using geo-classifiable, departmental level, geo-referenceable, data feature attributes in AUTOREG. In so doing, clustering tendencies in empirical time series morbidity-related, departmental level, and morbidity statistics can be quantitatively assessed. Additionally, an autocorrelation model may tease out pseudo-replicated morbidity prognosticators in geographic space. In so doing, departmental-level, covariates may target specific provincial regions in Guatemala and prioritize for resource allocation. Figures 5 and 6 are pictures of one of Guatemala City's major hospital, Roosevelt Hospital.

Figure 5 [25] - Roosevelt Hospital in Guatemala City, Guatemala

Figure 6 [27] - Emergency Room Entranc

References

[1] J. Stillwell and P. Boden. (2011) UK internal and international migration in the twenty-first century. Chapter 6 in Stillwell J and Clarke M (eds.) Population Dynamics and Projection Methods: Essays in Honour of Philip Rees. Series: Understanding Population Trends and Processes - Volume 4. Berlin: Springer.

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

[3] R. Schultz. 2017. Affine transformations and convexity. [online] Available: http://math.ucr.edu/~res/math145A-2014/affine+convex.pdf

[4] M.H. Kutner, C.J. Nachtsheim, J. Netter (2004) Applied Linear Regression Models, 4th edition. McGraw-Hill/Irwin, Boston.

[5] F. A. Haight, Handbook of The Poisson Distribution. John Wiley and Sons, Inc., 1967.

[6] N. L. Johnson, A. W. Kemp, and S. Kotz. (2005) Univariate Discrete Distributions, 3rd edition. Wiley-Interscience, Hoboken.

[7] G. P. Wadsworth (1960). Introduction to Probability and Random Variables. New York: McGraw-Hill.

[8] B. G.Jacob, W. Gu, E. X. Caamano. (2014) Developing operational algorithms using linear and non-linear square estimation in Python for the identification of Culex pipiens and Culex restuans in a mosquito abatement district (Cook County, Illinois, USA) Geospatial Health 2009:3(2):157-176

[9] D. Stirzaker. (2007). Elementary Probability. Cambridge University Press.

[10] (2013) Statistical Solutions. What is Linear Regression? [online] Available: http://www.statisticssolutions.com/what-is-linear-regression/.

[11] H. E.Vanden and G. Prevost, Politics of Latin America: The Power Game. 5th Edition. New York: Oxford University Press, 2015.

[12] Instituto Nacional de Estadistica: Guatemala. (2017) [online] Available: https://www.ine.gob.gt/index.php/estadisticas/.

[13] J. Borwein and D. Bailey. (2003) Mathematics by Experiment: Plausible Reasoning in the 21st Century. A. K. Peters, Wellesly.

[14] E. Lukacs. (1970) Characteristic Functions, 2nd Edition. Griffin, London.

[15] Graham et al (1994)

[16] B. G.Jacob, W. Gu, E. X. Caamano, Developing operational algorithms using linear and non-linear square estimation in Python for the identification of Culex pipiens and Culex restuans in a mosquito abatement district (Cook County, Illinois, USA) Geospatial Health 2009:3(2):157-176

[17] M. Hazewinkle. (2001) The algebra of quasi-symmetric function is free over integers. Advances in Mathematics. [online] Available: http://www.sciencedirect.com/science/article/pii/S0001870801920171.

[18] D. E., Ramírez, C.C Branas, T. S. Richmond. (2017) The relationship between pay day and violent death in Guatemala: a time series analysis, Injury Prevention. [online] Available: http://injuryprevention.bmj.com/content/23/2/102.

[19] T. Guberek and M. Hedstrom. (2017) On or off the record? Detecting patterns of silence about death in Guatemala's National Police Archive. Springer Science Business Media Dordrecht. [online] Available: https://link.springer.com/article/10.1007/s10502-017-9274-3.

[20] A. Cerón, A. L. Ruano, S. Sanchez. (2016) Abuse and discrimination towards indigenous people in public health care facilities: experiences from rural Guatemala. International Journal for Equity in Health.

[online] Available: https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-016-0367-z.

[21] T. G. Poder, J. He. (2015). The Role of Ethnic and Rural Discrimination in the Relationship Between Income Inequality and Health in Guatemala. International Journal of Health Services, 2015. [online] Available: http://journals.sagepub.com/doi/pdf/10.1177/0020731414568509.

[22] A. R. Hernandez, Hurtig, A-K., Dahlblom, K. (2015) Integrating views on support for mid-level health worker performance: a concept mapping study with regional health system actors in rural Guatemala. International Journal for Equity in Health. [online] Available: http://umu.diva-portal.org/smash/get/diva2:862115/FULLTEXT01.pdf.

[23] (2017) Departments of Guatemala. ESRI shapefile.

[24] (2017) Point of Interests, Places, Buildings, Natural Land Use, and Artificial Land Use. OpenStreetMaps. [online] Available: www.mapcruzin.com.

[25] (2017) [online] Available: http://www.prensalibre.com/hemeroteca/historia-de-los-hospitales-de-guatemala.

[26] (2017) [online] Available: https://directorio.guatemala.com/listado/hospital-roosevelt.html.

Appendix 1

Non-homogeneous Poisson processes are best described in measure-theoretic terms. For exampled a basic measure space may be illustratable as $[0, \infty)$ with the σ-algebra of Borel measurable subsets (named for Émile Borel). As usual, λ denotes Lebesgue measure on this space, named for Henri Lebesgue. Recall that the Borel σ-algebra is the one generated by the intervals, and λ is the generalization of length on intervals.

Of all of our various characterizations of the ordinary Poisson process, in terms of the inter-arrival times, the arrival times, and the counting process, the characterizations involving the counting process leads to the most natural generalization to non-homogeneous processes. Thus, consider a process that generates random points in time, and as usual, let N_t denote the number of random points in the interval $(0, t]$ for $t \ge 0$, so that $\bs{N} = \{N_t: t \ge 0\}$ is the counting process. More generally, $N(A)$ denotes the number of random points in a measurable $A \subseteq [0, \infty)$, so N is our random counting measure. As before, $t \mapsto N_t$ is a (random) distribution function and $A \mapsto N(A)$ is the (random) measure associated with this distribution function.

Suppose now that $r: [0, \infty) \to [0, \infty)$ is measurable, and define $m: [0, \infty) \to [0, \infty)$ by $$m(t) = \int_{(0, t]} r(s) \, d\lambda(s)$$ From properties of the integral, m is increasing and right-continuous on $[0, \infty)$ and hence is distribution function. The positive measure on $[0, \infty)$ associated with m (which we will also denote by m) is defined on a measurable $A \subseteq [0, \infty)$ by $$m(A) = \int_A r(s) \, d\lambda(s)$$ Thus, $m(t) = m(0, t]$, and for $s, t \in [0, \infty)$ with $s < t$, $m(s, t] = m(t) - m(s)$. Finally, note that the measure m is absolutely continuous with respect to λ, and r is the density function. Note the parallels between the *random* distribution function and measure N and the *deterministic* distribution function and measure m. With the setup involving r and m complete, we are ready for our first definition.

A process that produces random points in time is a *non-homogeneous Poisson process* with rate function \(r \) if the counting process \(N \) satisfies the following properties:

1. If \(\{A_i: i \in I\} \) is a countable, disjoint collection of measurable subsets of \([0, \infty) \) then \(\{N(A_i): i \in I\} \) is a collection of independent random variables.
2. If \(A \subseteq [0, \infty) \) is measurable then \(N(A) \) has the Poisson distribution with parameter \(m(A) \).

Property (a) is our usual property of independent increments, while property (b) is a natural generalization of the property of Poisson distributed increments. Clearly, if \(r \) is a positive constant, then \(m(t) = r t \) for \(t \in [0, \infty) \) and as a measure, \(m \) is proportional to Lebesgue measure \(\lambda \). In this case, the non-homogeneous process reduces to an ordinary, homogeneous Poisson process with rate \(r \). However, if \(r \) is not constant, then \(m \) is not linear, and as a measure, is not proportional to Lebesgue measure. In this case, the process does not have stationary increments with respect to \(\lambda \), but does of course, have stationary increments with respect to \(m \). That is, if \(A, \, B \) are measurable subsets of \([0, \infty) \) and \(\lambda(A) = \lambda(B) \) then \(N(A) \) and \(N(B) \) will not in general have the same distribution, but of course they will have the same distribution if \(m(A) = m(B) \).

In particular, recall that the parameter of the Poisson distribution is both the mean and the variance, so \(\E\left[N(A)\right] = \var\left[N(A)\right] = m(A) \) for measurable \(A \subseteq [0, \infty) \), and in particular, \(\E(N_t) = \var(N_t) = m(t) \) for \(t \in [0, \infty) \). The function \(m \) is usually called the *mean function*. Since \(m^\prime(t) = r(t) \) (if \(r \) is continuous at \(t \)), it makes sense to refer to \(r \) as the rate function. Locally, at \(t \), the arrivals are occurring at an average rate of \(r(t) \) per unit time.

As before, from a modeling point of view, the property of independent increments can reasonably be evaluated. But we need something more primitive to replace the property of Poisson increments. Here is the main theorem.

A process that produces random points in time is a non-homogeneous Poisson process with rate function \(r \) if and only if the counting process \(\bs{N} \) satisfies the following properties:

1. If \(\{A_i: i \in I\} \) is a countable, disjoint collection of measurable subsets of \([0, \infty) \) then \(\{N(A_i): i \in I\} \) is a set of independent variables.
2. For \(t \in [0, \infty) \), \begin{align} &\frac{\P\left[N(t, t + h] = 1\right]}{h} \to r(t) \text{ as } h \downarrow 0 \\ &\frac{\P\left[N(t, t + h] > 1\right]}{h} \to 0 \text{ as } h \downarrow 0 \end{align}

So if \(h \) is small the probability of a single arrival in \([t, t + h) \) is approximately \(r(t) h \), while the probability of more than 1 arrival in this interval is negligible.

Suppose that we have a non-homogeneous Poisson process with rate function \(r \), as defined above. As usual, let \(T_n \) denote the time of the \(n \)th arrival for \(n \in \N \). As with the ordinary Poisson process, we have an inverse relation between the counting process \(\bs{N} = \{N_t: t \in [0, \infty)\} \) and the arrival time sequence \(\bs{T} = \{T_n: n \in \N\} \), namely \(T_n = \min\{t \in [0, \infty): N_t = n\}\), \(N_t = \#\{n \in \N: T_n \le t\} \), and \(\{T_n \le t\} = \{N_t \ge n\} \), since both events mean at least \(n \) random points in \((0, t] \). The last relationship allows us to get the distribution of \(T_n \).

For \(n \in \N_+ \), \(T_n \) has probability density function \(f_n \) given by \[f_n(t) = \frac{m^{n-1}(t)}{(n - 1)!} r(t) e^{-m(t)}, \quad t \in [0, \infty) \]

Proof:

Using the inverse relationship above and the Poisson distribution of N_t, the distribution function of T_n is $P(T_n \le t) = P(N_t \ge n) = \sum_{k=n}^{\infty} e^{-m(t)} \frac{m^k(t)}{k!}, \quad t \in [0, \infty)$ Differentiating with respect to t gives $f_n(t) = \sum_{k=n}^{\infty} \left[-m^\prime(t) e^{-m(t)} \frac{m^k(t)}{k!} + e^{-m(t)} \frac{k m^{k-1}(t) m^\prime(t)}{k!}\right] = r(t) e^{-m(t)} \sum_{k=n}^{\infty} \left[\frac{m^{k-1}(t)}{(k - 1)!} - \frac{m^k(t)}{k!}\right]$ The last sum collapses to $m^{n-1}(t) \big/ (n - 1)!$.

The functional form of this probability density function is clearly similar to the gamma distribution, and indeed, T_n can be transformed into a random variable with a gamma distribution. This amounts to a *time change* which will give us additional insight into the non-homogeneous Poisson process.

Let $U_n = m(T_n)$ for $n \in \N_+$. Then U_n has the gamma distribution with shape parameter n and rate parameter 1

Proof:

Let g_n denote the PDF of U_n. Since m is strictly increasing and differentiable, we can use the standard change of variables formula. So letting $u = m(t)$, the relationship is $g_n(u) = f_n(t) \frac{dt}{du}$ Simplifying gives $g_n(u) = u^{n-1} e^{-u} \big/ (n - 1)!$ for $u \in [0, \infty)$.

Thus, the time change $u = m(t)$ transforms the non-homogeneous Poisson process into a standard (rate 1) Poisson process. Here is an equivalent way to look at the time change result.

For $u \in [0, \infty)$, let $M_u = N_t$ where $t = m^{-1}(u)$. Then $\{M_u: u \in [0, \infty)\}$ is the counting process for a standard, rate 1 Poisson process.

Proof:

1. Suppose that $(u_1, u_2, \ldots)$ os a sequence of points in $[0, \infty)$ with $0 \le u_1 \lt u_2 \lt \cdots$. Since m^{-1} is strictly increasing, we have $0 \le t_1 \lt t_2 \lt \cdots$, where of course $t_i = m^{-1}(u_i)$. By assumption, the sequence of random variables $\left(N_{t_1}, N_{t_2} - N_{t_1}, \ldots\right)$ is independent, but this is also the sequence $\left(M_{u_1}, M_{u_2} - M_{u_1}, \ldots\right)$.
2. Suppose that $u, \, v \in [0, \infty)$ with $u \lt v$, and let $s = m^{-1}(u)$ and $t = m^{-1}(v)$. Then $s \lt t$ and so $M_v - M_u = N_t - N_s$ has the Poisson distribution with parameter $m(t) - m(s) = v - u$.

Equivalently, we can transform a standard (rate 1) Poisson process into a a non-homogeneous Poisson process with a time change.

Suppose that $\bs{M} = \{M_u: u \in [0, \infty)\}$ is the counting process for a standard Poisson process, and let $N_t = M_{m(t)}$ for $t \in [0, \infty)$. Then $\{N_t: t \in [0, \infty)\}$ is the counting process for a non-homogeneous Poisson process with mean function m (and rate function r)).

Proof:

1. Let $(t_1, t_2, \ldots)$ be a sequence of points in $[0, \infty)$ with $0 \le t_1 \lt t_2 \lt \cdots$. Since m is strictly increasing, we have $0 \le m(t_1) \lt m(t_2) \lt \cdots$. Hence $\left(M_{m(t_1)}, M_{m(t_2)} - M_{m(t_1)}, \ldots\right)$ is a sequence of independent variables. But this sequence is simply $\left(N_{t_1}, N_{t_2} - N_{t_1}, \ldots\right)$.
2. If $s, \, t \in [0, \infty)$ with $s \lt t$. Then $N_t - N_s = M_{m(t)} - M_{m(s)}$ has the Poisson distribution with parameter $m(t) - m(s)$.