Application of Area to Point Kriging to Buruli Ulcer Incidence in Ashanti and Brong Ahafo Regions of Ghana

Bonyah Ebenezer*, Owusu-Sekyere Ebenezer and Ossei Linda

Abstract
Buruli ulcer (BU) is the third most common mycobacterium disease after tuberculosis and leprosy. The disease eats through the skin, muscle and bone, leaving victims with disfiguring and debilitating craters. Ghana is the second most endemic country globally, after Cote d’Ivoire with over 1,048 cases with the most endemic regions being the Ashanti, Greater Accra, Central and the Brong Ahafo. The goal of the paper is to use Area to Point Kriging (ATP) method to model the spatial distribution of Buruli ulcer incidence in the Ashanti and Brong Ahafo Regions of Ghana. The ATP method used consist of three steps; filtering of noise in the data based on Poisson kriging, the mapping of the corresponding risk at a fine scale and estimating geographical clustering of the disease at the administrative units. This paper focused on the spatial analysis of Buruli ulcer incidence in the Ashanti and Brong Ahafo region in terms of sex. The research revealed that there is a large range of spatial autocorrelation in males than in females in the various administrative units. The administrative units in Brong Ahafo close to Ashanti region have high BU incidence than the units far away from the Ashanti. The clustering analysis revealed that only Amanse West district is statistically significant for both sexes.

Keywords: Area to point kriging; Buruli ulcer; Area to area kriging; Mycobacterium ulcerans

Introduction
Buruli ulcer (BU) is a serious skin infection caused by Mycobacterium ulcerans [1]. BU has been described as a neglected emerging disease that affects neglected rural people in neglected areas of the world [2]. Buruli ulcer is rated as the third most common mycobacterium infection after Tuberculosis and Leprosy [3]. Although there is a lot of literature on the possible causes of Buruli ulcer (BU), no one is sure where the bacterium lives in the environment and therefore the arguments have been purely speculative [4]. It is also a mystery how the mycobacterium enters the human body, although it is clear the bacterium is unable to do so by itself [5]. However, current research has shown the disease is commonly associated with rapid environmental change to the landscape including deforestation, eutrophication, dam construction, irrigation, farming, mining, and habitat fragmentation [6]. The proximity to slow moving water such as swamps, man-made lakes, dams, creeks, and living in lower elevation areas appear to have a higher risk for Buruli ulcer [7]. The disease has been reported in many countries of West Africa, Indonesia, Malaysia, Mexico, France, Australia and Peru. In most of these countries, BU is known to afflict impoverished inhabitants living in remote areas where amenities of modern medical science are not available or expensive [8].

In Africa, all ages and sexes are affected, but most cases of the disease occur in children between the ages of 4–15 years [9].

Ghana is the second most endemic country in the world after Cote d’Ivoire with about 1,048 Buruli Ulcer case in 2011 [10]. Africa tops the list of most affected regions with Cote d’Ivoire leading the rate with 2,670 patients. According to a recent WHO statistics, the total population of Buruli Ulcer cases recorded globally including that of Ghana is 5,076 [10]. There have been a lot of studies on the disease in Ghana. However many of the studies have concentrated on small geographical areas of high prevalence. There have been several studies by Duker et al. in [7] and on the relationship between arsenic concentrations and the mean Buruli ulcer prevalence in settlements along arsenic-enriched drainage pathways and arsenic-enriched farmland. Again, Asiedu and Etauful [9] concluded that poverty was a major contributing factor to the high prevalence of the disease in the District. Owusu-Sekyere [4] conducted a research on the socio-economic effects of Buruli ulcer in the Amanse West District and the findings revealed that the disease could have a profound effect on education and human welfare because of the high incidence of stigmatization. Bonyah and Owusu-Sekyere [11] used Geospatial Modeling to examine Buruli Ulcer Prevalence in Amanse West District. Whereas all these studies are significant in expanding the knowledge towards the understanding of the causes and effects of the disease, there has not been any attempt to study the two endemic regions in the northern sector of Ghana, which are the Ashanti and Brong Ahafo Regions.

The aim of this paper is to examine the spatial distribution of Buruli ulcer by sex in the Ashanti and Brong Ahafo Regions using Area to Point Kriging method. The geostatistical analysis we applied consists of three steps: filtering of noise in the data based on Poisson kriging, mapping of the corresponding risk at a fine scale and estimating geographical clustering of the disease at the administrative units. This application was first employed by Christian Lajaunie [12] to account for spatial heterogeneity in the population of children to estimate the semivariogram of the “risk of developing cancer” from semivariogram of observe mortality rates. Oliver et al. [13] applied binomial cokriging to produce a map of the risk of childhood cancer in the west midlands of England. Following the successful application of this methodology, Goovaerts and Jacquez [14] employed the same methodology to map lung cancer mortality across the US. Cressie [15] examined the spatial distribution of the account of sudden–infant–death syndromes for 100 countries of North California. In his approach a two–step transformation of the data was taken into account by removing first, mean variance dependence of the data and next the heteroscedasticity. Not long ago geostatistics was applied to map the number of low birth weight (LBW) at the census tract level. This was...
to provide information for county-level data and covariates measured over different spatial supports [16]. In order to address a real data problem in mapping of disease, Monestiez et al. [17,18] proposed an approach called Poisson Kriging that filters the data before mapping. Poisson Kriging is capable to be combined with stochastic simulation to come out multiple realizations of the spatial distribution of disease risk. There has been remarkable geostatistical framework given to the analysis of areal data such as medical geography [19]. This has been implemented by several authors including Gotway and Young [16] and Kyriakidis [19] to predict area values. This approach is referred to as “area-to-point” (ATP) or “area to area” (ATA) kriging according to Kyriakidis [19]. The unique feature about ATP kriging is that it allows the mapping of variability within geographical unit (polygon) and at the same time ensuring the coherence of the prediction. For instance, disaggregated estimates of count data are non-negative and the sum is equal to the original aggregated count. Kerry et al. [20] applied ATP and ATA for analyzing the geography of offenses and for identifying significant clusters of crimes on car-related thefts in the Baltic States. Shao et al. [21] applied ATP to introduce sex for the cancer rates, and see the difference between age-adjusted rates and age-sex-adjusted rates. Montero et al. [22] also used kriging strategy (area-to-point kriging) to solve the potential mismatch between environmental information that is both objective and subjective.

Methodology

Study area

For the purpose of this study, the Ashanti and Brong Ahafo regions were compacted together as one geographical unit for easy analysis and interpretation. The study area is drained by rivers such as Tano and Oda. The area is home to the Kintampo waterfalls and the Bui Dam. Lake Volta flows along the eastern edge of Brong-Ahafo and the port of Yeji. The area has moist-semi deciduous forest, mostly in the southern and south eastern parts and a guinea savannah, which is pre-dominant in the North-eastern portion. The area has a tropical climate, with high temperatures averaging 23.9°C (750F) and a double maxima rainfall pattern. Rainfall ranges, from an average of 1000 millimeters in the northern parts to 1400 millimeters in the southern parts. Agriculture is the Dominant economic activity because of the rich agricultural lands. There is also small scale mining, timber and logging and petty trading. The study area has a population of about 7,007,174 million made up of 3,449,862 males and 3,557,312 females (GSS, 2010). There are two referral hospitals, the Komfo Anoye teaching hospital located in Kumasi in the Ashanti Region and the Sunyani government hospital located in Sunyani in the Brong Ahafo Region. With the exception of the two referral hospitals, the rest of the districts lack access to quality health care, thus aggravating the disease situation in the area (Figure 1).

Data sources

Data for the study was obtained from Disease Control Units (DCU) of the two regions. The confirmed cases of Buruli ulcer disease as of the time of the study were 1590 for males and 1356 for Females. The data classification on the basis of sex was to find out the incidence rates between the two sex groups at the various administrative units. Population data obtained from Ghana Statistical Service was used in computing the raw rates of Buruli ulcer disease. Raw rates were calculated as the number of BU cases in each district divided by the estimated Population in 2010. In order to better, appreciate the risk of the disease; the raw rates were rescaled by multiplying it by a factor of 100,000. This expresses the raw rates as per 100,000 people.

Geostatistical approach

Area to-area poisson kriging: For a given number N of geographical units \( v_i \) (e.g., counties), denote the observed mortality rates (areal data) as \( z(v_i) = d(v_i)/n(v_i) \), where \( d(v_i) \) is the number of recorded mortality Buruli ulcer cases and \( n(v_i) \) is the size of the population at risk. The disease count \( d(v_i) \) is construed as a realization of a random variable \( D(v_i) \) that adopts a Poisson distribution with one parameter (expected number of counts) that is the product of the population size \( n(v_i) \) by the local risk \( R(v_i) \). The noise-filtered mortality rate for a given area \( v_i \), called mortality risk, is forecasted as a linear combination of the kernel rate \( z(v_i) \) and the rates observed in \((K-1)\) neighboring entities

\[
\hat{r}(v_i) = \sum_{j=1}^{K} \lambda_j z(v_j)
\]

The weights \( \lambda_j \) given to the \( K \) rates are worked out by solving the following system of linear equations, known as the “Poisson Kriging” system

\[
\sum_{j=1}^{K} \lambda_j \left( C_R (v_j, v_j) + \delta_j \frac{m^*}{n(v_j)} \right) + \mu(v_i) = C_R (v_i, v_0) i=1,\ldots, K
\]

\[
\sum_{j=1}^{K} \lambda_j = 1
\]

where \( \delta_{ij} = 1 \) if \( i=j \) and 0 otherwise. \( m^* \) is the population-weighted mean of the N rates. The “error variance” term, \( m^*/n(v_i) \), leads to smaller weights for less reliable data (e.g., rates measured over smaller populations). In addition to the population size, the kriging system accounts for the spatial correlation among geographical units through the area-to-area covariance terms \( C_R (v_i, v_j) = \text{Cov}(R(v_i), R(v_j)) \) and \( C_R (v_i, v_0) \).

Those covariances are numerically approximated by averaging the point-support covariance \( CR(h) \), computed between any two locations discretizing the areas \( v_i \) and \( v_j \).
\[
\begin{align*}
\bar{C}_R(v_i, v_j) &= \frac{1}{P} \sum_{s=1}^{P} \sum_{i=1}^{P} \sum_{s=1}^{P} \text{wss}^t C_R(u_s - u_s) \\
(3)
\end{align*}
\]

where \( P \) and \( P' \) are the number of points used to discretize the two areas \( v_i \) and \( v_j \), respectively.

The weights \( \text{wss}^t \) are computed as the product of population sizes assigned to each discretizing point \( u_s \) and \( u'_s \) with \( n(u_s) = n(u'_s) \).

The uncertainty of the Buruli ulcer mortality risk persisting within the geographical unit \( v_i \) can be modeled using the conditional cumulative distribution function (ccdf) of the risk variable \( R(v_i) \). Under the assumption of normality of the prediction errors, that ccdf is defined as

\[
F(v_i, r|K) = \text{Pr} \{ R(v_i) \leq r | K \} = G \left( \frac{r - \overline{R}_i}{\hat{\sigma}(v_i)} \right) \\
(4)
\]

where \( G(.) \) is the cumulative distribution function of the standard normal random variable, and \( \hat{\sigma}(v_i) \) is the square root of the kriging variance estimated as

\[
\sigma^2(v_i) = \overline{C}_R(v_i, v_i) - \sum_{i=1}^{K} \lambda_i \overline{C}_R(v_i, v_i) - \mu(v_i) \\
(5)
\]

The weights \( \lambda_i \) are the within-area covariances that is computed according to (3) with \( v_i = v_j = v_i \). The notation \( "|K" \) expresses conditioning to the local information, say, \( K \) neighboring observed rates. The function (4) gives the probability that the unknown risk is no greater than any given threshold \( r \). It is modeled as a Gaussian distribution with the mean and the variance corresponding to the Poisson kriging estimate and variance.

**Area-to-point (atp) poisson kriging:** A particular case of ATA kriging is when the prediction support is so small that it can be assimilated to a point \( u_s \) leading to the following area-to-point Poisson kriging estimator and kriging variance

\[
\hat{r}_{PK}(u_s) = \sum_{i=1}^{K} \lambda_i \overline{C}_R(v_i, v_j) \\
(6)
\]

\[
\sigma^2_{PK}(u_s) = C_R(0) - \sum_{i=1}^{K} \lambda_i \overline{C}_R(v_i, v_i) - \mu(u_s) \\
(7)
\]

The kriging weights and the Lagrange parameter \( \mu(u_s) \) are computed by solving the following system of linear equations

\[
\begin{align*}
\sum_{j=1}^{K} \lambda_j (u_s) \left[ \overline{C}_R(v_j, v_j) + \delta \frac{m^*}{n(v_j)} \right] + \mu(u_s) &= \overline{C}_R(v_i, v_j), i = 1, ..., K \\
\sum_{j=1}^{K} \lambda_j (u_s) &= 1
\end{align*}
(8)
\]

The ATP kriging system is similar to the ATA kriging system (2), except for the right-hand-side term where the area-to-area covariances \( \overline{C}_R(v_i, v_j) \) are replaced by area-to-point covariances \( \overline{C}_R(v_i, u_s) \) that are approximated as

\[
\overline{C}_R(v_i, u_s) = \frac{1}{P} \sum_{s=1}^{P} \sum_{s=1}^{P} \text{wss}^t C_R(u_s - u_s) \\
(9)
\]

where \( P \) is the number of points used to discretize the area \( v_i \) and the weights \( \text{wss}^t \) are computed as for expression (3). ATP kriging can be conducted at each node of a grid covering the study area, resulting in a continuous (isopleth) map of mortality risk and reducing the visual bias that is typically associated with the interpretation of choropleth maps. Another interesting property of the ATP kriging estimator is its coherence. The population-weighted average of the risk values estimated at the \( P \) points \( u_s \) discretizing a given entity \( v_i \) yields the ATA risk estimate for this entity

\[
\hat{r}_{PK}(u_s) = \frac{1}{n(v_i)} \sum_{s=1}^{P} n(u_s) \overline{r}_{PK}(u_s) \\
(10)
\]

Constraint (10) is satisfied if the same \( K \) area data are used for the ATP kriging of the \( P \) risk values

**Deconvolution of the semivariogram of the risk:** Both ATA and ATP kriging require knowledge of the point support covariance of the risk \( CR(h) \), or equivalently the semivariogram \( \gamma R(h) \). This function cannot be estimated directly from the observed rates, since only areal data is available. Thus, only the regularized semivariogram of the risk can be estimated as:

\[
\hat{\gamma}_R(h) = 1 - \frac{1}{2} \sum_{a,b} \frac{N(\text{h})}{n(v_a) + n(v_b)} \left\{ \frac{n(v_a) + n(v_b)}{(n(v_a) + n(v_b))} \right\} \left\{ \left[ \hat{r}(v_a) - \hat{r}(v_b) \right]^2 + m^* \right\} \\
(11)
\]

where, \( N(\text{h}) \) is the number of pairs of areas \( (v_a, v_b) \) whose population-weighted centroids are separated by the vector \( h \). The different spatial increments \( \left[ \hat{r}(v_a) - \hat{r}(v_b) \right]^2 \) are weighted by a function of their respective population sizes, \( n(v_a) + n(v_b) / (n(v_a) + n(v_b)) \), which is inversely proportional to their standard deviations [18]. Derivation of a point-support variogram \( \gamma(h) \) from the variogram \( \gamma_R(h) \) fitted to areal data is called the deconvolution.

Derivation of a point-support variogram from the experimental semivariogram \( \gamma_R(h) \) computed from areal data is called “deconvolution” an operation that has been the topic of much research [16,19]. In this paper, we adopted the iterative procedure introduced for rate data measured over irregular geographical units [23], whereby one seeks the point-support model that, once regularized, is the closest to the model fitted to areal data. The experimental variogram was fitted using weighted least square in SpaceStats developed by Biomedware in USA. The procedure for which the theoretical variogram was fitted into experimental variogram was based on the deconvolution as explained earlier.

**Cluster analysis**

A common task in disease analysis is to examine administrative units in adjacent geographical locations that are significantly similar or different. Similarity between the Buruli ulcer incidence rate observed within area \( v_R \) and those recorded in the \( j(v_a) \) neighboring areas \( v_a \) can be computed by the local Moran statistic [24] as:

\[
\text{Citation: Ebenezer B, Ebenezer OS, Linda O (2013) Application of Area to Point Kriging to Buruli Ulcer Incidence in Ashanti and Brong Ahafo Regions of Ghana. Geoinfor Geostat: An Overview 1:1.}
\]

\[
doi: http://dx.doi.org/10.4172/2327-4581.1000102
\]
where, $m$ and $s$ are the mean and standard deviation of the set of N area incident rates respectively. This local indicator of spatial association (LISA) is simply the product of the kernel rate and the average of the neighboring rates.

The distribution of the local Moran statistic under the null hypothesis of complete spatial randomness is usually obtained through a random of shuffling all the count(s) except at $v_i$ each time calculating (12) to get the distribution of simulated LISA values.

The empirical values of (12) are compared with this distribution to compute the $P$ value for the rest. This randomization ignores the population size associated with each areal unit [14].

Results and Discussions

The Figures 2 and 3 indicate the omnidirectional variogram of Buruli ulcer for females and males using the risk computed from district-level rates, using estimator (11). The experimental variogram was fitted using a Cubic model with a range of 37.5 km for females and 41.5 km for males. However, BU incidence for males has better range of spatial autocorrelation than incidence for females in each administrative unit. Each model was deconvoluted using the iterative procedure.

The deconvoluted variogram model was then used to compute aggregated risk values at the district level in both regions using ATA and ATP kriging, see Figure 4. In all cases, the estimation was based on the $K=32$ closest observations, which were selected according to the population-weighted districts, for ATA kriging. All the kriging maps of Buruli ulcer incidence for females were smoother than the map for the raw rates because the noise, due to the population size associated with each areal unit [14].

The BU incidence rate for females at the various administrative units (Figure 4A) shows that the disease is more endemic in the southern districts of the study area than the other districts. Incidentally, the southern portions represent the Ashanti region. It can therefore be implied from the results that the Ashanti region has more endemic districts than the Brong Ahafo region as evidenced in Figure 4C which represents the ATP risk map of Buruli ulcer for females in the study area. Within the ATP risk map for females (Figure 4C) some isolated districts in Brong Ahafo region which shares border with the Ashanti region such as Tano North and South, Asunafo North and South also have high risk of BU. This revelation by this paper is not to say that the other districts are free from the disease incidence. It is not surprising that the southern districts of the ATP risk map shows high incidence of the disease because these areas are characterized by massive agricultural and artisanal mining activities that disturb the environment and contributes to the spread of the disease. These findings are consistent with a study conducted by the Ministry of Health in 2010 which concluded that for the past five years, the incidence of the disease in the two regions have been very high due to artisanal mining activities which have destroyed most of the water bodies in the area. This is the condition for which Mycobacterium ulcerans (MU) thrive well. Again, these communities lack access to proper health care and therefore, majority of the women with the disease do not report for medical attention. The unreported cases are also fueled by the high cost of treatment and the level of stigmatization, Sekyere, [4]. The northern part of Brong Ahafo such as Kintampo, Jaman North and South are less endemic, probably because these are dry areas with less stagnant waters, pounds, rivers and streams. The mycobacterium does not do well in such environment.

Contrary to the studies by Amofa et al. [25] in the Amanfie West that females were more infected than their male counterparts, this study found the opposite. This research revealed that more males suffered the disease in the study area than their female counterparts, Figure 5F. Again, Figure 5F further suggests that the incidence of Buruli ulcer for males is quite wide spread unlike their female counterparts that tended to concentrate only on the southern portion of the study area. There are majority of cases among males in the Northern districts of the study area which happens to be districts in the Brong Ahafo region. The high prevalence of the disease among
males (Figure 5D) could be attributed to the changing nature of livelihood activities which is more male oriented such as the mining activities.

Similarly, districts that did not report of female case now reports high males Buruli ulcer incidence, one of such districts is Kintampo North, see Figure 5F. This situation could be attributed to the mobility of men in search of greener pastures.

As was the case with the female sexes, there were more males with the Buruli ulcer disease in the southern districts of the study area. These areas represent the southern part of the Brong Ahafo region and the Ashanti region. The endemic districts are the Amansie Westmost endemic in Ghana, Tano North and South. These districts have had the worst environmental degradation for the past five years EPA [26].

Finally, the estimated geographical clustering of the disease at the administrative units showed that only two administrative units, the Amansie West District and the Bekwai Municipal were significant, Figure 7K.

The ATP kriged risk variance map for both males and females (Figure 6G and Figure 6H) is relatively different from ATP kriged risk map (Figure 4C and Figure 5F). However, it can be observed from both maps the spatial distribution of the BU disease. Again, the ATA kriged variance map for males and females (Figure 6I and Figure 6J) is very similar to that of (Figure 4B and Figure 5E) and therefore, yield better prediction. The kriged variance map which normally shows the extent of deviation comes about due to noise in data set and software problems thus affecting the exact interpolation.

The Local Moran statistic (Figure 7L) shows that only Amansie West District is significant. This administrative unit by implication has the highest Buruli ulcer incidence in the area under study. The disease is clustered around the major rivers which are Offin, Punpun, and Oda and the most affected communities are Tontokrom, Edubia and Kaniago which are all found around the major rivers Bonyah and Sekyere [11]. The fact that the other administrative units are not significant (p-value>0.05), does not imply they are free from the incidence of the Buruli ulcer disease. The high incidence of the disease for males in the District has been blamed on the activities of artisanal mining MOH [27]. The mining activities have led to the contamination all the water bodies with arsenic. The high levels of Arsenic (As) concentrations could cause BU occurrence to increase. This is supported by Duker et al. [7] that the influence of arsenic on gold activities enhances the bacteria called water bug which is believed to host Mycobacterium ulcerans and when bitten by these insects may contribute to the spread of BU. According to Duker et al. [7] arsenic may play a vital role in the spatial distribution of BU.

Conclusion

This study has demonstrated how geostatical method can be used to model Buruli ulcer incidence by sex distribution. The Area to Point Kriging (ATP) method used in this study has given an insight into more localized potential “hot spots” for the Buruli ulcer disease that may not be evident when non geostatical methods are employed. ATP kriging is used to create a continuous risk surface that reduces the visual bias associated with large administrative units. The research showed a large range of spatial autocorrelation in males than that of females in the distribution of BU disease. The research further showed that the risk associated with Buruli ulcer was centered in the southern districts of the study area which co-existingly, represents the Ashanti region and few districts in the Brong Ahafo region. The research revealed that the worst endemic areas had suffered massive environmental degradation through agricultural and artisanal mining activities. The artisanal mining for example had elevated arsenic content in soils and water bodies thus
exposing the population to the disease causing organisms. This study recommends that further modeling would benefit from additional inputs to identify other confounding factors that contribute to Buruli ulcer morbidity. Broader socio-economic data, age, sex and improved geospatial coverage would enable considerable refinements of the model used in this research.

References


