School of Computing & Communications | Lancaster University

MSc Data Science

Country-wide Rapid Epidemiological Mapping of Onchocerciasis (REMO) in Gabon

CHIC563 — Geostatistical Models

Lancaster University

15th December 2023

Abstract

River blindness (onchocerciasis) causes severe itching, skin lesions and vision impairment including blindness. Most cases (>99%) are located in sub-Saharan Africa. The prevalence of infection in local communities is used as the basis for country-wide repartition, as per REMO principles developed by the World Health Organisation (WHO). The levels of endemicity were evaluated in 59 villages, from this an onchocerciasis risk of infection map was created, in order to focus ivermectin distribution programmes.

1 Introduction

River Blindness, or Onchocerciasis, is spread via the filarial worm *Onchocerca Volvulus* which is transmitted by female black flies of the genus *Simulium*. Onchocerciasis was previously endemic in sub-Saharan Africa, accounting for over 99% of global cases [1]. In these countries, Onchocerciasis was a severe public health problem resposible for visual impairment and debilitating skin diseases, this had serious socio-economic consequences, including depopulation of fertile river valleys [1]–[3]. The Onchocerciasis Control Programme in West Africa (OCP) has successfully controlled onchocerciasis by vector control in the savanna belt of nine West African countries [4]. However, in the remaining endemic areas in Africa, vector control was not feasible or cost-effective. This led to the registration and introduction of ivermectin for the treatment of human onchocerciasis in 1987.

Figure 1: Participating countries in the African Programme for Onchocerciasis Control (APOC) [src:[5]]

An international coalition of Non-Governmental Development Organisations(NGDOs) spearheaded ivermectin distribution efforts [6] and in 1995 the African Programme for Onchocerciasis Control (APOC) was created with the mandate to support the establishment of community directed treatment with ivermectin in all remaining areas where onchocerciasis was an ongoing public health problem [7].

The first challenge for APOC was to determine the areas to prioritise when distributing ivermectin. Historical information on the spread of onchocercasis in the 20 APOC countries as well as surveys in villages were interpreted by WHO and maps were drawn of the approximate distribution of onchocerciasis in Africa - however, this information was either incomplete or reliable to give a confident model of distribution for targeting ivermectin treatment. The distribution of different vector species and the location of breedinbg sites were not known, whilst the main diagnostic method was a parasitological method based upon the microscopic examination of skin biopsies; an invasive and time-consuming method. In response to these problems, a rapid assessment method for the Rapid Epidemiological Mapping of Onchocerciasis (REMO) was developed in 1993 and successfully tested at scale in Cameroon and Nigeria [8]. The spatial epidemiology of onchocerciasis is closely related to the distribution of local river systems, their suitability for simulium breeding and the flight range of the vector when seeking a blood meal. REMO is defined in the WHO Manual for Rapid Epidemiological Mapping of Onchocerciasis [9].

This paper will create infection risk maps with relation to elevation and proximity of water.

2 Methods

The data collected via REMO was the responsibility of the Ministry of Health of each APOC country in collaboration with its partners in onchocerciasis control and with technical and financial support from APOC. Gabon has a total of 60 survey villages (shown in Figure 2), reporting their respective coordinates, tests performed and positive results. The analysis of this data is performed using Geographic Information System (GIS) concepts and geostatical analysis. The geographic information used for the analysis includes

- Admininstrative Boundaries, Rivers and Lakes
- Topography
- Geographic Coordinates and Prevalence of Surveyed Villages

The prevalence of each village survey represents the size of the node in Figure 2, where prevalence is the ratio of positive tests compared to the overall tests performed. The average prevalence of all tested sites is 1.84% with a standard deviation of 3.18, a histogram of this data is shown in Figure 3. Previous surveys have indicated areas of high risk to be where prevalence is above 20% and observing prevalence values above 60%, however as ivermectin treatment has been effective the maximal observed prevalence is only 11.8% hence the thresholds have been lowered to above 5%.

Figure 2: Test Sites

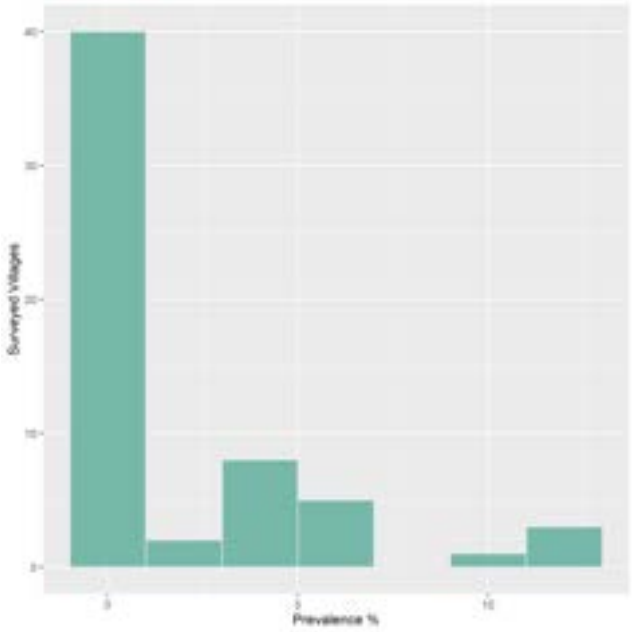


Figure 3: Histogram of Prevalence in Surveyed Villages

As previously mentioned, focusing on the impact that elevation and water proximity have on the risk of infection, Figure 4 demonstrates the raster covariate layers to be explored.

2.1 Kriging

Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial plane. It differs from similar methods such as Linear Regression, Gaussian decays or Inverse

(a) Elevation (b) Water Proximity

Figure 4: Gabon Topological Data

Distance Weighted Interpolation in that kriging the interpolation is based on the spatial arrangement of the empirical observations rather than on a presumed model of spatial distribution, whilst also generating estimates of uncertainty surrounding each interpolated value.

This is a two step process

- Spatial Covariance Structure of sampled points via variogram
- Interpolate points via Weights Derived from Covariance Structure

This process relies on the assumption of spatial autocorrelation among the sampled data points - the values must co-vary in space. Kriging can be used to maintain spatial variability where it would be lost when using a simpler method [10].

This report will utilise variograms as a visual depiction of the covariance exhibited between a pair of points in the sampled data. For each pair of points the γ or "semi-variance" is plotted against the distance or "lag" between them.

Kriging Weights

Kriging weights are calculated such that points nearby to the location are given more weight than those far away. Neighbourhood clustering is also taken into account so that clusters of points do not adversely influence each other to help reduce bias in the predictions. The kriging process also generates an expected value for any actually sampled location where it must equal the observed value, it is proven that all interpolated values will be the Best Linear Unbiased Predictor (BLUP).

Weights for each interpolated point are calculated according to the interpolated location in reference to all sampled points. The weights are determined from the variogram based upon the spatial structure of the data, derived by the following formula;

$$z(\hat{x}_0) = \sum_{N}^{i=1} \lambda_i z(x_i) \tag{1}$$

where $\hat{z}$ is the value of the predicted point at location x_0, this is equal to the sum of the value of each sampled points x_i and weighted using λ_i.

Assumptions

The two key assumptions of this method to provide accurate and usable data are those of stationarity and isotropy;

- Stationarity - The joint probability distribution does not vary over the plane. Therefore, parameters such as mean, range and sill do not vary. That is, the same variogram model is valid across the plane.
- Isotropy - uniformity in all directions

Limitations

The kriging method is heavily dependant on the variogram, so the model is sensitive to misspecification of the variogram, this is exacerbated when second-order stationarity is not met (which is difficult in real-world applications). This can be addressed with some newer adaptations such as Bayesian approaches. Additionally the accuracy of interpolation is limited if the number of sampled observations is small, the data is limited in spatial scope, or the data are not sufficiently correlated spatially.

Empirical Bayesian Kriging

Empirical Bayesian Kriging (EBK) extends upon Kriging by accounting for second-order staionarity problems, additionally EBK automatically tunes hyperparameters through a process of subsetting and simulations. As previously discussed, standard kriging calculates the variogram from known data locations and use this single variogram to make predictions at unknown locations; this process implicitly assumes that the estimated variogram is the true variogram for the interpolation region. By not taking the uncertainty of variogram estimation into account, other kriging methods underestimate the standard errors of prediction.

Variogram Estimation

Unlike standard kriging (which use weighted least squares), the variogram parameters in EBK are estimated using restricted maximum likelihood (REML). Due to the computational limitations of REML for large datasets, the input data is first divided into overlapping subsets of a specified size (defaulted to 100 points per subset). In each subset, variograms are estimated in the following way:

1. Estimate variogram
2. Using this variogram, new data is simulated
3. Estimate new variogram
4. Repeat

This process creates a large number of variograms for each subset, and when they are plotted together, the result is an empirical distribution of variograms that are shaded by density (the darker the blue color, the more variograms pass through that region).

2.2 Cross Validation

The cross validation is based upon a "leave one out" method, which allows the determination of how well an interpolation model fits the data. By removing a single point from the dataset and using all remaining points to predict the location the point that was removed. The predicted value is then compared to the measures value.

3 Results

The output of each geostatistical model is probability of risk of infection from onchocerciasis, this is categorised into low to severe risk, represented in the colour gradients of the map, where blue is low risk, to red which is severe risk.

3.1 Kriging - Elevation Model

The standard kriging model is shown in Figure 5, using 12 lags, each with a size of 0.204. The variogram is shown in Figure 6a, with its map in Figure 6b.

The cross validation of this model gives a mean score of 0.002, showing the model is consistent with the observed samples. Additionally, the Root-Mean-Square-Standardised error is 1.03, showing this model is ok, but leaves some accuracy to be desired. This model result is undesireable due to the inability to deal with the variogram change that the ocean presents, it is noticeable that according to this basic model the sea is at a medium risk of oncocerciasis - which is obviously false.

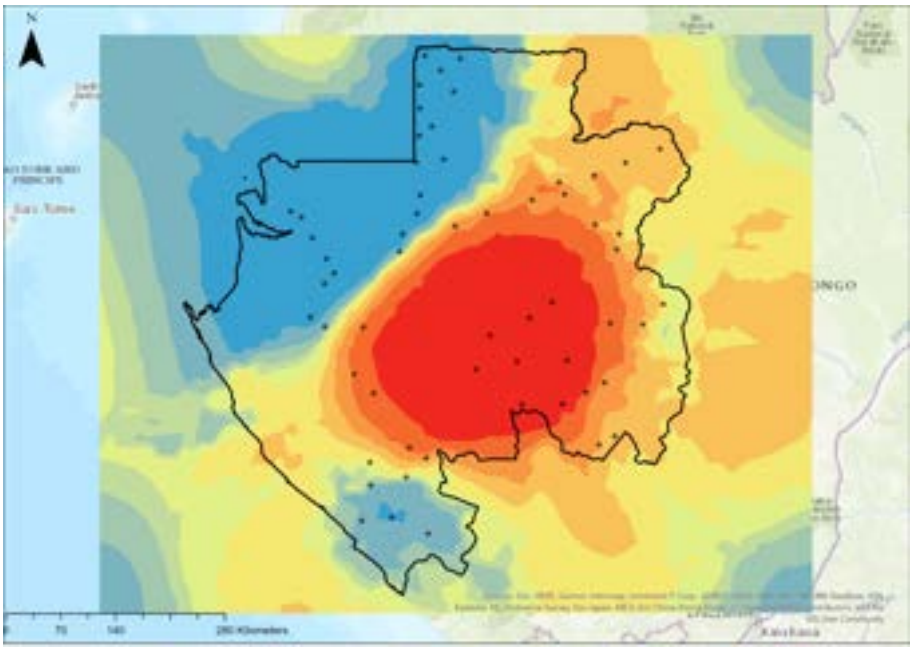
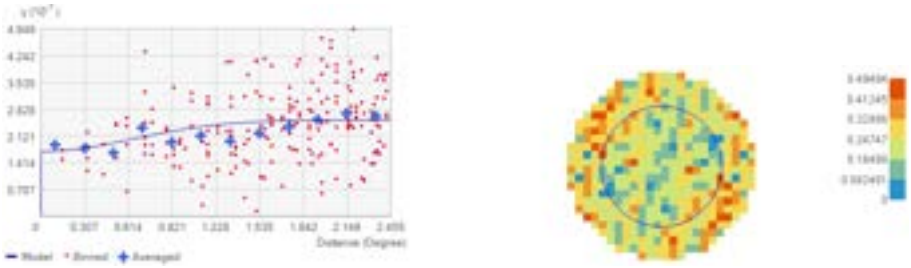


Figure 5: Bayesian Kriging Probability



(a) Model (b) Map

Figure 6: Kriging Semivariogram

3.2 Empirical Bayesian Kriging - Elevation Model

The EBK Elevation model is shown in Figure 7, it demonstrates far more complicated boundaries between risk-categories, using 100 simulations with a stopping condition of 95% minimum cumulative variance. Additionally using a subset size of 40, the surface type is the probability of exceeding a prevalence of 5.8% (the mid-range of observed samples, chosen manually for best output). Using an empirical transformation, where an non-parametric kernal mixture is applied to the dependent as it is not normally distributed. The variogram is shown in Figure 8, using an K-Bessel model, which allows the variogram to decrease the spatial autocorrelation error term at a higher accuracy than alternative models such as exponential. This variogram is based upon the nugget and sill, shown in Figure 9, wit the variance and transformation applied in Figure 10. The nature of the ability to use multiple variograms is demonstrated in the model recognising the ocean is of beyond low risk, the high risk area is clustered tightly, with the majority of the remaining area being low risk. The cross validation applied shows a very tight qq-plot, Figure 11a, representing this distribution is very close to the observed prevalence, however the predicted output, Figure 11b deviates negatively perhaps indicating this model under-reports risk.

3.3 Empirical Bayesian Kriging - Elevation and Water Distance Model

The final model experiments with integrating distance from water as a known covariate, as described earlier, the distance from water is calculated as euclidean distance from nearest water source - whereby elevation is also taken into account. The distance raster is shown in Figure 4b, where red is closest to water sources, and green is furthest away - they are represented numerically as a matrix similar to elevation.This model is setup in the same way as previous, with the exception of using a secondary covariate layer. The model output is shown in Figure 12, this looks very similar to the output of the previous EBK model with the clear distinction that the risk bands appear to follow river patterns, as well as a larger high risk area - perhaps due to a large cluster of slow running water at high elevations.

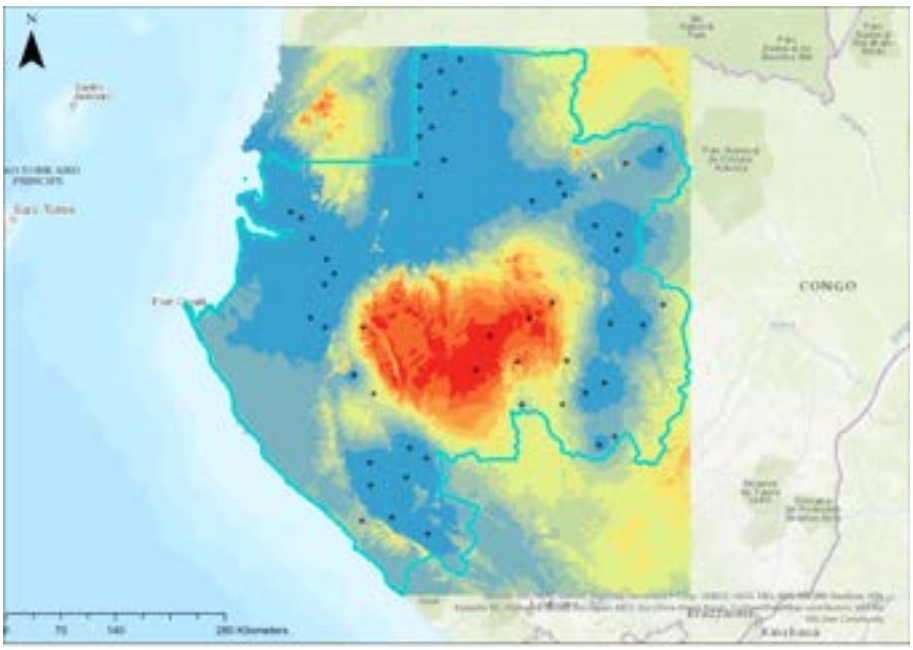


Figure 7: EBK Probability

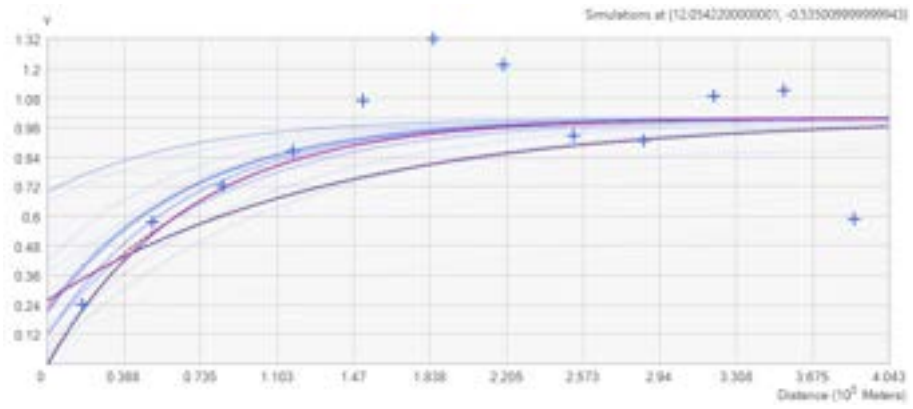
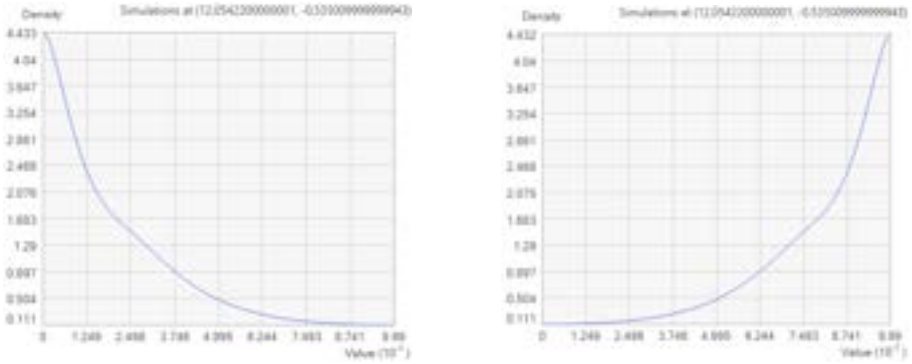


Figure 8: EBK Variogram

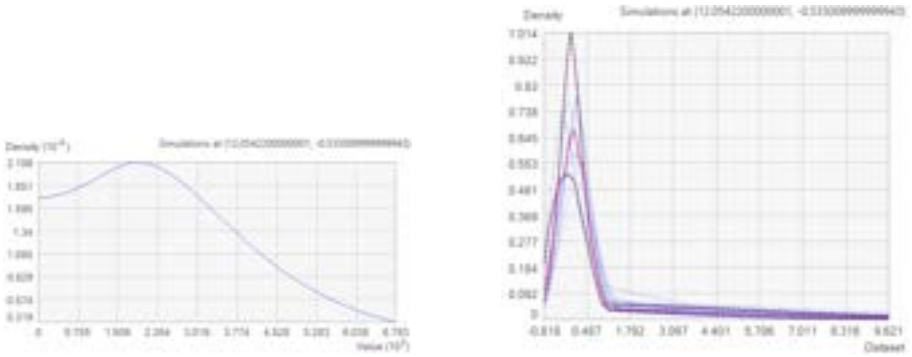
4 Discussion

The present results demonstrate the progress of country-wide epidemiological mapping of onchocerciasis using the rapid assessment method. That is the REMO maps have played a significant role in endemic onchocerciasis and the control of it in APOC countries. The delineation of high risk areas has led to the creation of more than 100 ivermectin programmes. This report has outlined the areas which require further support, especially in the central region where prevalence is distinctly higher than the remaining areas which see very few to no cases. Future work could incorporate population density as a further raster layer to either relate population density to prevalence or to use population



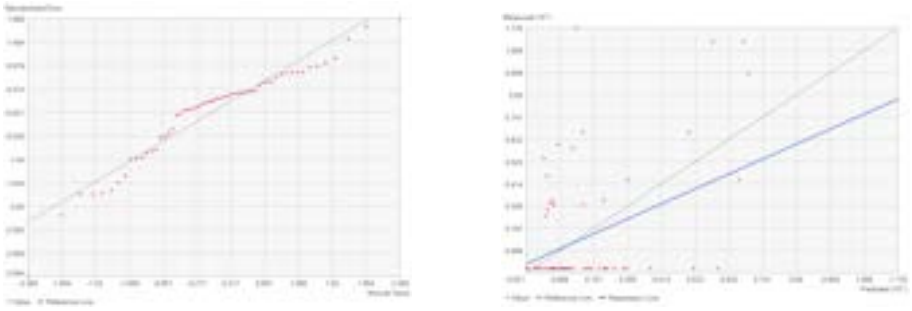
(a) Nugget (b) Sill

Figure 9: Empirical Bayesian Kriging Model Parameters



(a) Variance (b) Transformation

Figure 10: Empirical Bayesian Kriging Model Parameters



(a) QQ-Plot (b) Predicted Output

Figure 11: Empirical Bayesian Kriging Cross Validation via Leave One Out

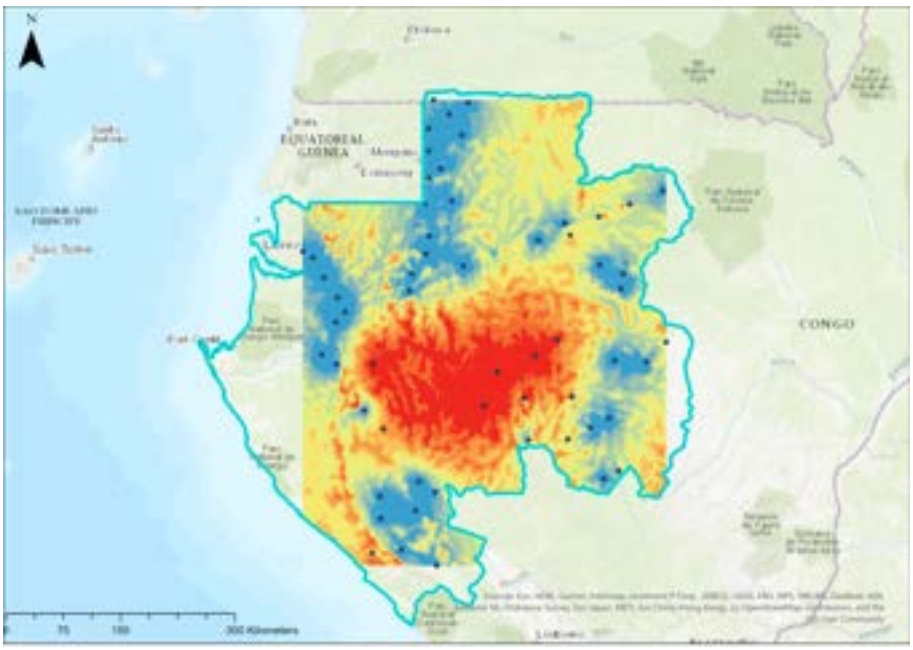


Figure 12: EBK Probability Mixed Model

density to order where new ivermectin programmes should be rolled out.

References

[1] WHO Expert Committee on Onchocerciasis Control (1993 : Geneva, S. and Organization, W. H., *Onchocerciasis and its control : Report of a who expert committee on onchocerciasis control*, 1995.

[2] Murdoch, M. E., Asuzu, M. C., Hagan, M., Makunde, W. H., Ngoumou, P., Ogbuagu, K. F., Okello, D., Ozoh, G. and Remme, J., 'Onchocerciasis: The clinical and epidemiological burden of skin disease in africa,' *Annals of Tropical Medicine & Parasitology*, vol. 96, no. 3, pp. 283–296, 2002.

[3] Benton, B., 'Economic impact of onchocerciasis control through the african programme for onchocerciasis control: An overview,' *Annals of tropical medicine and parasitology*, vol. 92 Suppl 1, S33–9, Apr. 1998.

[4] Boatin, B., 'Onchocerciasis control programme in west africa (ocp),' *Annals of tropical medicine and parasitology*, vol. 102 Suppl 1, pp. 13–7, Sep. 2008.

[5] Noma, M., Zouré, H. G., Tekle, A. H., Enyong, P. A., Nwoke, B. E. and Remme, J. H., 'The geographic distribution of onchocerciasis in the 20 participating countries of the african programme for onchocerciasis control: (1) priority areas for ivermectin treatment,' *Parasites & Vectors*, vol. 7, no. 1, p. 325, 2014.

[6] Bush, S. and Hopkins, A., 'Public–private partnerships in neglected tropical disease control: The role of nongovernmental organisations,' *Acta tropica*, vol. 120 Suppl 1, S169–72, Apr. 2011.

[7] Alley, W. S., Oortmarssen, G. J. van, Boatin, B. A., Nagelkerke, N. J., Plaisier, A. P., Remme, J. H., Lazdins, J., Borsboom, G. J. and Habbema, J. D. F., 'Macrofilaricides and onchocerciasis control, mathematical modelling of the prospects for elimination,' *BMC Public Health*, vol. 1, no. 1, Nov. 2001.

[8] Ngoumou, P., Walsh, J. F. and Mace, J. M., 'A rapid mapping technique for the prevalence and distribution of onchocerciasis: A cameroon case study,' *Annals of Tropical Medicine & Parasitology*, vol. 88, no. 5, pp. 463–474, 1994.

[9] Ngoumou, P., Walsh, J. F., Prevention of Blindness, W. P. for the, Research, U. B. S. P. for and Tropical Diseases, T. in, *A manual for rapid epidemiological mapping of onchocerciasis / p. ngoumou and j. f. walsh*, 1993.

[10] Auchincloss, A. H., Roux, A. V. D., Brown, D. G., Raghunathan, T. E. and Erdmann, C. A., 'Filling the gaps: Spatial interpolation of residential survey data in the estimation of neighborhood characteristics,' *Epidemiology*, vol. 18, no. 4, pp. 469–478, Jul. 2007.

Word Count : 1963