Loa Loa in Central African Countries: A Comparison of Spatial Models

Carson Keeter

December 2019

Introduction

Concerning Common Central African Diseases

There are many infectious diseases that plague third world countries, especially those in tropical regions. These infectious diseases are caused by pathogens such as viruses, bacteria, protozoa, and helminths, where the former will be the focus of this essay. The infectious diseases caused by these pathogens are divided into two groups; diseases that are HIV/AIDS, tuberculosis, & malaria and those that are not.

The diseases that are not the "Big Three" are commonly known as **Neglected Tropical Diseases** (or **NTDs**, for short). The concern surrounding these NTDs is manifold. These diseases affect no less than 1.4 billion people around the world. Of that 1.4 billion, 500 million are children[1]. Fortunately, the treatment for many of these diseases is relatively inexpensive. However, when cheap treatments are considered magnified to the billions, cost increases. It is estimated that control of NTDs would require funding of between \$2 billion and \$3 billion (USD) over a period of five to seven years[2]. NTD solution efforts are also frequently overshadowed by the efforts to reduce HIV/AIDS, tuberculosis, & malaria rates. Additionally, there is no celebrity "figure head" representing the NTD treatment efforts. There are no worldwide concerts or charity auctions championing the solution. NTDs have nothing. Lastly (and of most concern), many of the worldwide organizations that concern themselves with global health cannot agree upon a set list of NTDs[3,4,5]. This disagreement concerning what is or is not a NTD may seem minuscule, but this inconsistency is the platform on which declension stands.

The **World Health Organization** (WHO) and the **Center for Disease Control** (CDC) agree upon a set list of 20 NTDs. It so happens that these diseases get the most attention in terms of funding efforts. However, many scientific journals such as *PLoS* consider far more widespread NTDs. Yet, these journals only publish research and rarely have the funds to mitigate the prevalence of these diseases. Now, this is not to say that the efforts of WHO and the CDC are in vain but the dichotomy between organizations that should be concerned with this global health endemic is concerning.

This brings us to the topic of this essay.

Concerning *Loa Loa*

The concern of this essay is not the morphology of parasites and the symptoms associated therein. However, it is important to understand the significance of this disease and how it affects 12-13 million people worldwide[6].

Loiasis is a disease brought on by *Loa loa*, a species of roundworm, that causes localized swelling typically near the joints. The larvae of this species of roundworm develop in horseflies which infect humans by biting them. After bites from these infected flies, the roundworm larvae mature in 5-6 months then travel in the peripheral blood during the day and migrate into the lungs at night[7].

Although the disease is seemingly benign, *Loa loa* can migrate to the eye and cause extreme pain and often times blindness. These symptoms usually take anywhere from 4 months to years to appear, making efforts to eliminate infected horseflies nearly impossible. This is the reason why spatial analysis is so imperative for this topic (and most other NTDs).

Mentioned previously, it is estimated that there are 13-14 million people across 11 countries affected by this disease. The highest prevalence is seen in Cameroon, Republic of the Congo, Democratic Republic of Congo, Central African Republic, Nigeria, Gabon, and Equatorial Guinea. The rates of *Loiasis* are lower but still present in Angola, Benin, Chad and Uganda. Detection of this disease is simple yet unpractical in endemic regions and treatment is complicated, frequently contraindicated, and under-researched[8].

The sole purpose of this analysis is to determine if there is a spatial component concerning *Loa loa* prevalence along with vegetation and elevation. The results of this analysis may prove useful in the efforts to treating individuals with the disease *or* locating areas for increased pesticide use. Whatever the results yield, I will maintain the purpose stated above.

Exploratory Analysis

Concerning the Data

These data contain seven variables: *longitude (degrees), latitude (degrees), number of individuals tested, number of positive tests, elevation (ft), maximum normalized difference vegetation index (NDVI), and standard deviation of NDVI*. The data is mostly contained in **Cameroon** with a few data points in **Nigeria** and **Central African Republic** (denoted in the red rectangle in figures 1 & 2). To simplify the model making process, as well as interpretability, *prevalence* of L. loa was calculated ($\frac{Positive\ Tests}{Total\ Tests}$).

Lastly, there is no variable concerning population density of the area tested. However, the distance between the two highest populated cities could be a reasonable estimate (to be discussed further). Using the *geosphere* R package, the distance was calculated using the Haversine equation based on the coordinates of these two cities:

- Yaounde (capital): 1.9 million (2005)[9]
- Douala: 1.8 million (2005)[9]

It is important to note that these population estimates were also recorded around the same time as when the L. loa prevalence data were collected.

Considering all of these variables, summary data can be found in table 1 (with the addition of *model residuals*, *predicted outcome variable*, and other model specific measures that will be addressed shortly). The following short-hand corresponds to the names in the dataset:

- *lon*: Longitude (degrees)
- *lat*: Latitude (degrees)
- *number_tested*: Total number of individuals tests
- *number_positives*: Number of positive tests
- *elevation*: Elevation in meters (obviously)
- *max_NDVI*: Maximum value for Normalized Difference in Vegetation Index
- *sd_NDVI*: Standard deviation for Normalized Difference in Vegetation Index (seasonal variance)
- *perc_pos*: Disease prevalence (percent positive)
- *cap_dist*: Distance from Yaounde (kilometers)
- *cap2_dist*: Distance from Douala (kilometers)

Lastly, figure 3 shows the prevalence of L. loa in the countries of interest. One can notice clear spatial clustering in the western region as well as further north. From here, the following models attempt to define the relationship between these two variable and any latent spatial features.

Model Selection

There are many ways to make spatial models. However, this essay will be generally revolving around 4: *General linear model*, *Isotropic spatial model*, *Anisotropic spatial model*, and a *mixed-effects spatial model*. These models all answer the same question, yet approach it differently.

General Linear Model (OLS)

The general linear model (in this case, multiple regression) specifies the mean function. This is the basis of the spatial analysis. Whatever this model doesn't "pick up" will be picked up by the following spatial models.

After many models were constructed, the "leanest" mean function is the following:

$$Prevalence = -55.88 + 86.84(Maximum\ NDVI) - 0.045(Distance\ from\ Yaounde) + 0.063(Distance\ from\ Douala) - 0.005(Elevation) + \epsilon(s)$$

The comparison between the full model and the chosen mean function can be found in table 2. Based on the table, one can see all significant p-values below 0.05. Additionally, the R^2 value is 0.511.

Figure 4 shows the plotted residuals of this model in the area of interest. Similar to prevalence plot, one can notice quite a bit over under-estimation (high residual) in similar places where there is high prevalence of L. loa.

Lastly, figures 5 & 6 shows the distribution of the OLS residuals. Based on the apparent normal distribution, the following spatial models should be sufficient.

Isotropic Spatial Model

Isotropic models do *not* consider both distance and orientation/direction, only distance as a function of spatial correlation.

Using the *geostatsp* package, an isotropic model is constructed with the previously specified mean function (table 3). Note that the predictor variables are scaled (via z-score) in order to put the interpretation in terms of standard deviations.

The output is clunky but packed with information. First, it is important to note that the anisotropy measures are the opposite of anisotropic. The "shape" of the function is close to exponential with a $0°$ angle. Once can also see that the data become independent around 0.26$°$ (which is approximately 28.86km). Lastly, one sees the relevant estimates for the beta coefficients and the respective p-values. Unlike OLS regression, these beta coefficients are not as useful in this spatial context.

Again, this model assumes that there is no directionality in the correlation. The following model does.

Anisotropic Spatial Model

Unlike isotropic models, anisotropic models account for directionality and orientation when estimating spatial correlation.

Table 4 shows the anisotropic model summary. One can see similar values as the isotropic model. However, the shape and anisotropic parameters are considered and estimated. The "anisoAngleDegrees" estimates the minor axis of the spatial correlation in degrees. The major axis, intuitively, is calculated by adding or subtracting 90 to the the minor axis estimate. Comparisons can be seen below.

Mixed-effects Spatial Model (Non-normal)

The previous models all assume that the residuals of the data are normally distributed. However, this may not be the case. Using the *spaMM* package, a non-normal spatial effects model can be constructed.

Table 5 shows the output from the non-normal spatial model using the specified mean function. Notice the fairly large t-values (indicating the number of conditional standard errors away from 0, which can be translated into p-values if need be). The effective range is approximately 74.50 km $((3/rho) * 111km)$.

Additional Information for Model Parameter Estimates and Construction

A few preparatory steps were left out from the above analysis in order to maintain conciseness. The following list details what was omitted:

- Projection: *Universal Transverse Mercator (UTM) Zone 30*
- Coordinate Reference System: *EPSG 26391*
- Nugget: approximately 0.5 (Figure 7)
- Libraries used :
    - *geoR*: Geostatistical analysis
    - *tidyverse*: Data wranglin', cleanin', and figure makin'
    - *sf*: Support for simple features, a standardized way to encode spatial vector data
    - *sp*: Map plotting and retrieving coordinates
    - *gstat*: Spatial modelling and prediction
    - *ggfortify*: Concise residual plots for OLS
    - *maps*: Map creation
    - *maptools*: Map creation
    - *tmap*: Map creation
    - *geostatsp*: Geostatistical analysis
    - *RColorBrewer*: Contains color palettes for graphics
    - *ggthemes*: ggplot2 extension for map data
    - *geosphere*: Calculate Haversine distance
    - *spaMM*: Non-normal spatial modelling

Model Comparison

Each model reports similar information. However, between these four models, one is more "correct". Table 6 shows a comparison of estimated parameters and AIC between each model. Based on AIC alone, the non-normal model performs the best in both known (conditional) and unknown (marginal) estimations. In fact, the other models do not have marginal or conditional estimates.

In addition to AIC, the effective range (converted from degrees to kilometers) shows the estimated distance where the prevalence of L. loa becomes independent of the mean function and spatial effects (when assuming non-stationarity).

The predictive power of the spatial models vary greatly. Figures 8 and 9 show the predicted prevalence and predicted standard deviation between the isotropic and anisotropic models across the area of interest, respectively. One can notice that the anisotropic model fits ellipses rather than perfect circles, which is expected. However, one can also notice that the strength of these models greatly falters in areas where there is no recorded data. This is an issue that can be remedied by the non-normal spatial model.

Figure 10 shows the predicted prevalence of L. loa in the area of interest when it is modeled non-normally. The improvement from the normal isotropic and anisotropic is stark. One notices that the predicted prevalence is still the most consistent around areas where there is data. However, the marginal estimation (where there is no data) does a significantly better job at estimating prevalence. One thing to note: as the predicted prevalence goes farther away from areas where there is data, the prediction regresses back to the mean (which is approximately 20%). Since the non-normal model performs the best, further interpretation will be conducted.

Since the prevalence of L. loa is bionomially distributed, the interpretation is put on the logit scale. For example, based on the non-normal model, for every 1 standard deviation increase in NDVI (~ 0.057), the odds of the prevalence of L. loa increase by $e^{0.4776}$.

Figures 11 through 15 show the observed prevalence against the predicted prevalence subsetted into the different continuous variables. One can see that the predictive "power" of model is the best around small values of prevalence while being slight more variant around larger prevalences. It does not seem that there are any major differences in predictive power when consider the intensity of the predictor levels.

Lastly, table 6 again shows a table comparing all the models made in this analysis and their AIC & effective range.

Discussion

From this analysis, it can be seen that the non-normal spatial model does the "best" job at modelling and predicted prevalence of L. loa. This is for a few reasons (possibly). First, the prevalence of L. loa is not totally normally distributed. Therefore, the model that does not assume normality (in this case, a multivariate normal distribution), the model is not as "restricted" as the other models. However, it is not perfect.

Figure 16 shows the latent spatial effects in the region of interest based on the non-normal model. One can notice that, generally, the non-normal model does a fairly good job of "covering" the areas near cities with high population (which were specifically modeled in this analysis). However, there are a few areas with a large spatial latent effect that the model is not picking up (denoted in the darker shades). When looking at satellite images of the area, there is a fairly large lake (read: standing water) that may explain the latent spatial. In future analysis, it may be worth recorded these areas of standing water.

The sole purpose of this analysis was to determine if there is a spatial component concerning *Loa loa* prevalence along with vegetation and elevation. It can be determined, therefore, that this goal was accomplished. The prevalence of L. loa was modeled fairly well, especially with the consideration of new variables (when compared to other analysis[10]. As stated above, the results of this analysis may prove useful in the efforts to treating individuals with the disease *or* locating areas for increased pesticide use.

Figures

Figure 1: Map of Africa with Region of Interest Highlighted in Red

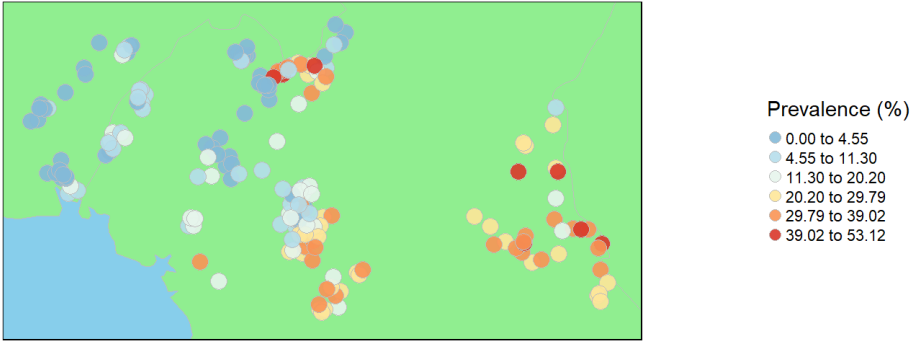Figure 2: Region of Interest (Cameroon, Nigeria, and Central African Republic)

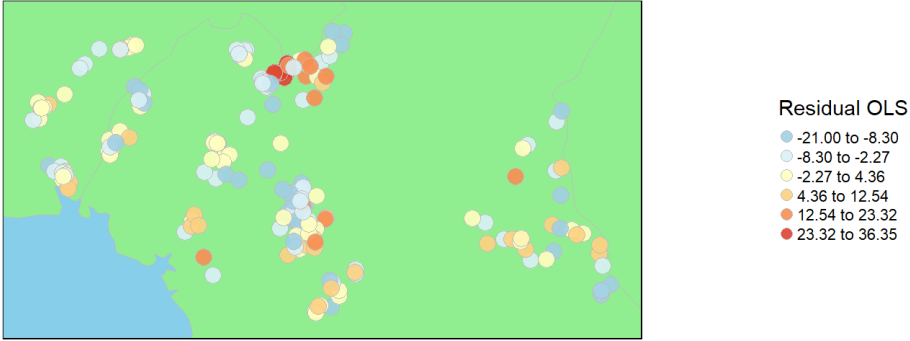Figure 3: Prevalence of Loa Loa in Cameroon, Nigeria, and C.A.R.

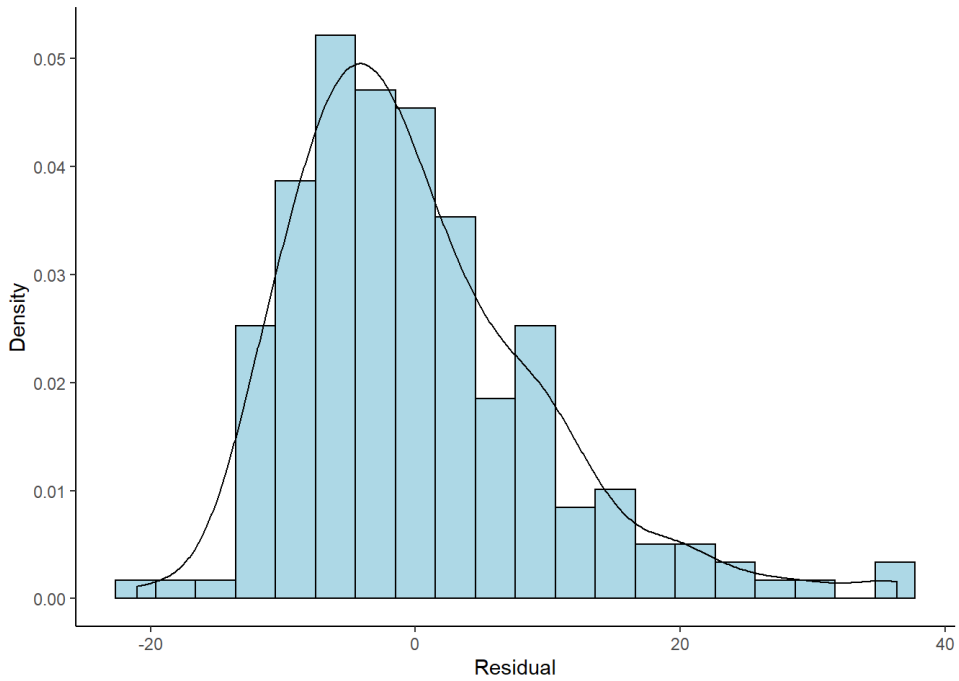Figure 4: Residuals from OLS Model

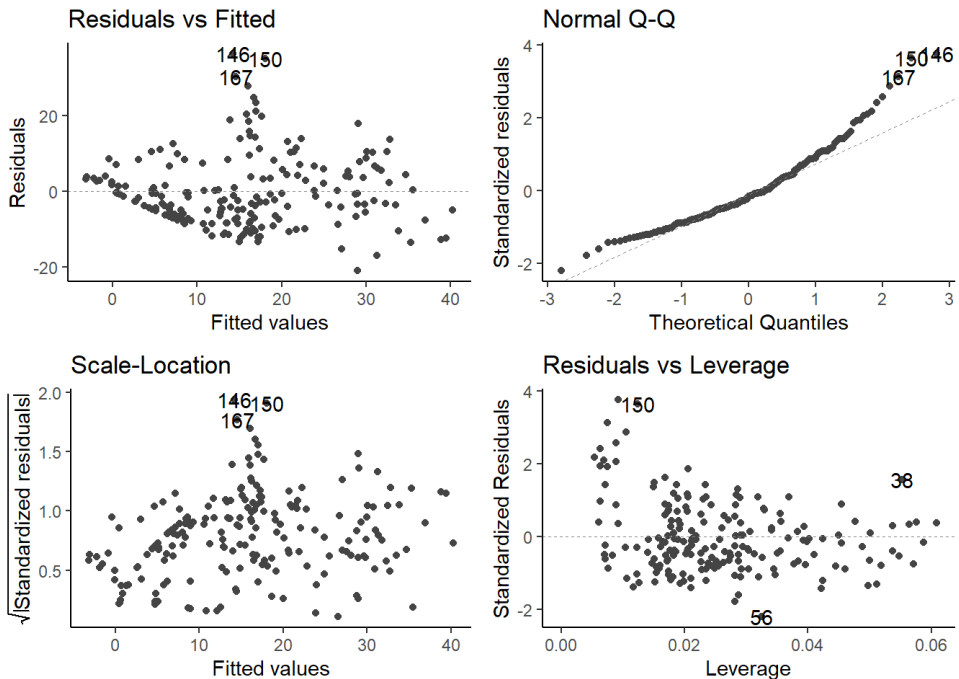Figure 5: Density Plot of Residuals from OLS Model

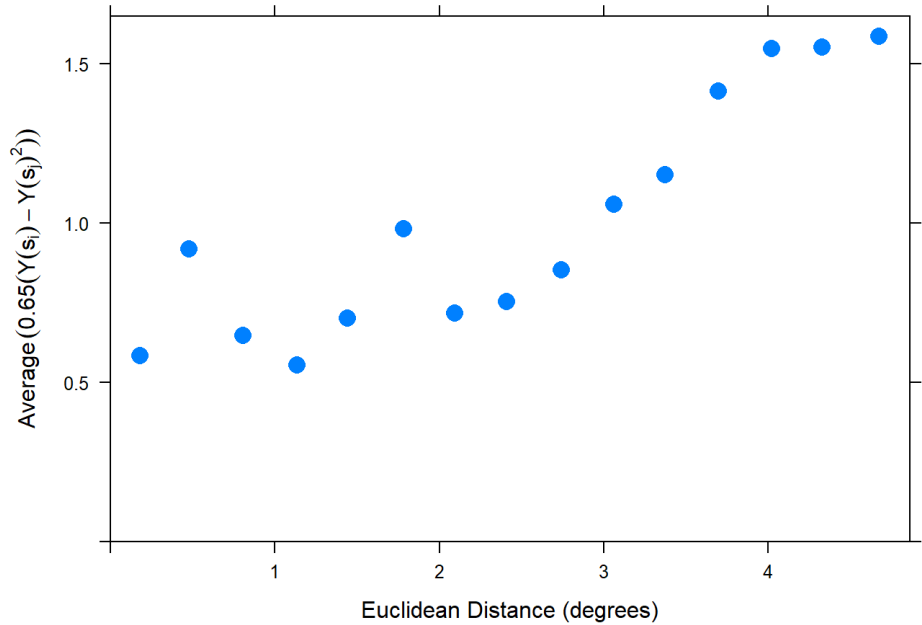Figure 6: Residuals Plots for Normality from OLS Model

Figure 7: Variogram for Nugget Estimation

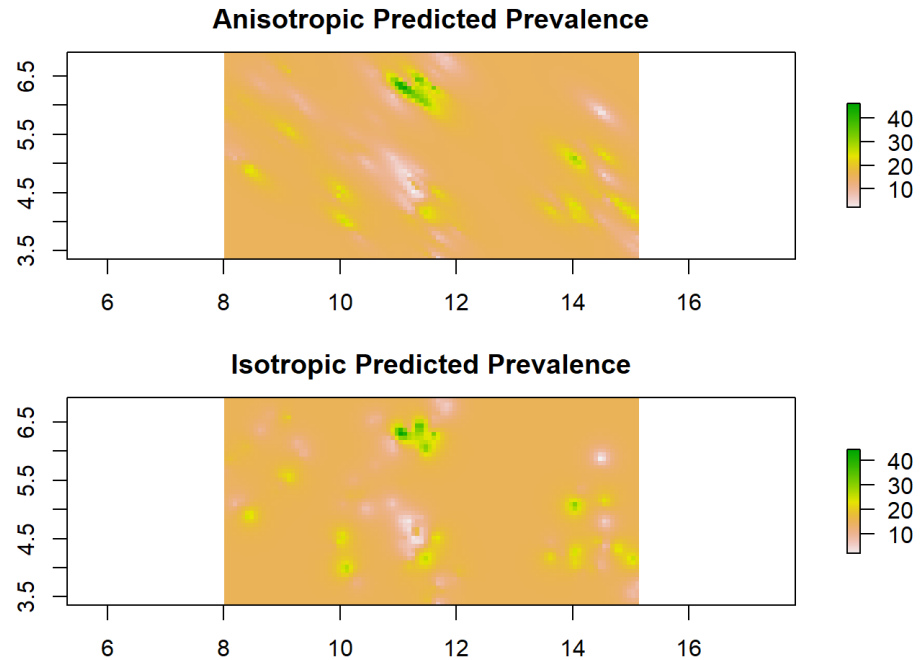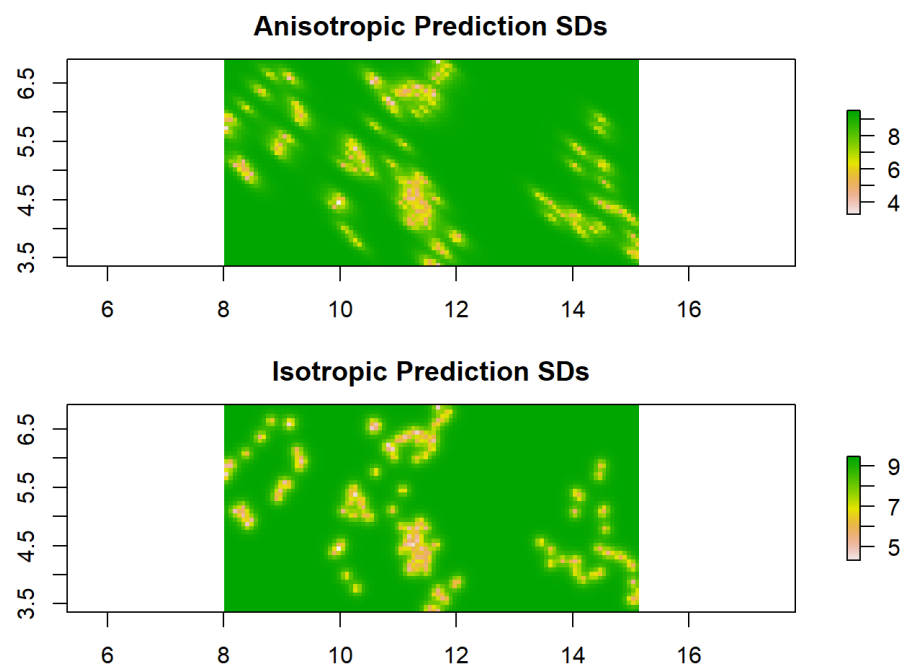Figure 8: Predicted Prevalence between Isotropic and Anisotropic Spatial Models

Figure 9: Predicted Standard Deviation of Prevalence between Isotropic and Anisotropic Spatial Models
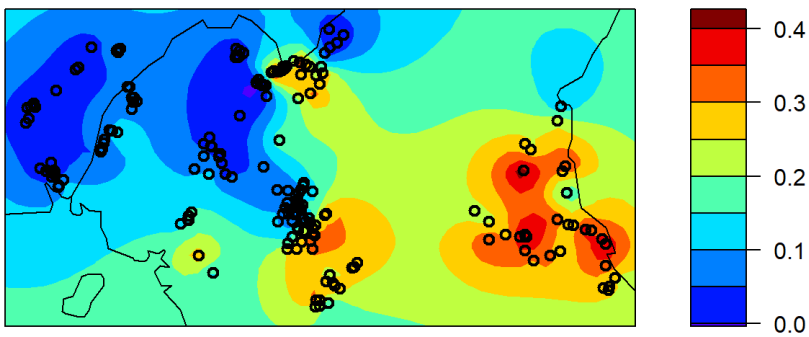
Figure 10: Predicted Prevalence of Non-normal Spatial Model
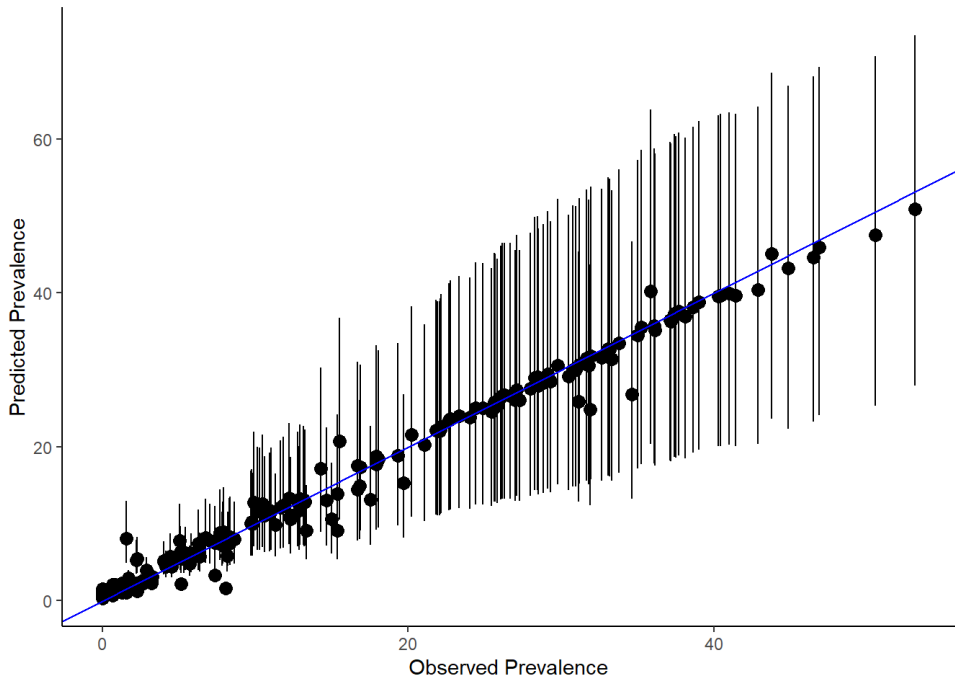
Figure 11: Observed Prevalence vs Predicted Prevalence (non-normal)

Figure 12: Observed Prevalence vs Predicted Prevalence with Elevation (m) (non-normal)

Figure 13: Observed Prevalence vs Predicted Prevalence with Distance to Yaounde (km) (non-normal)
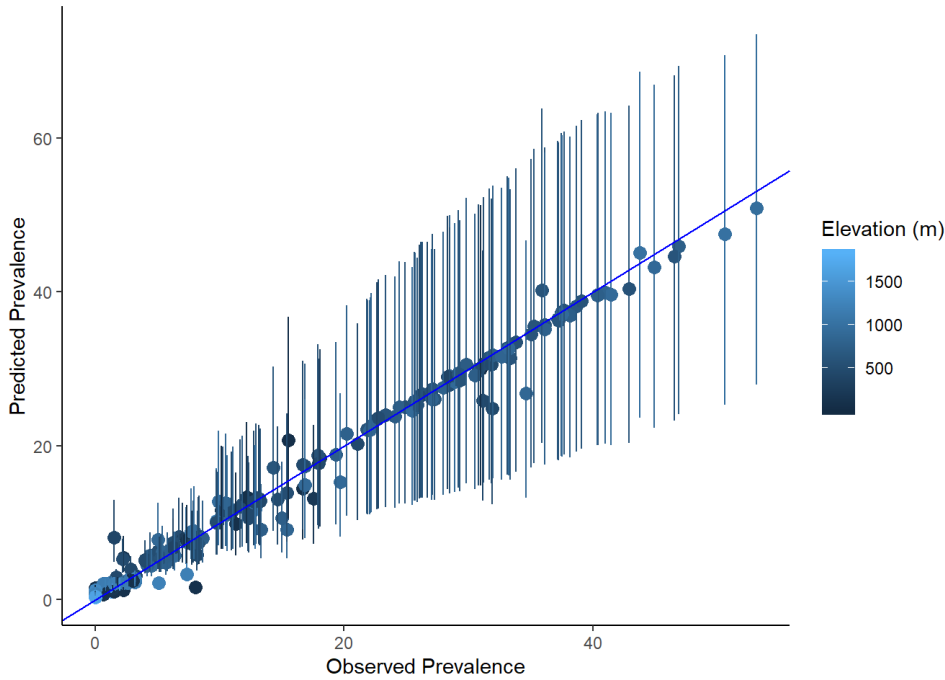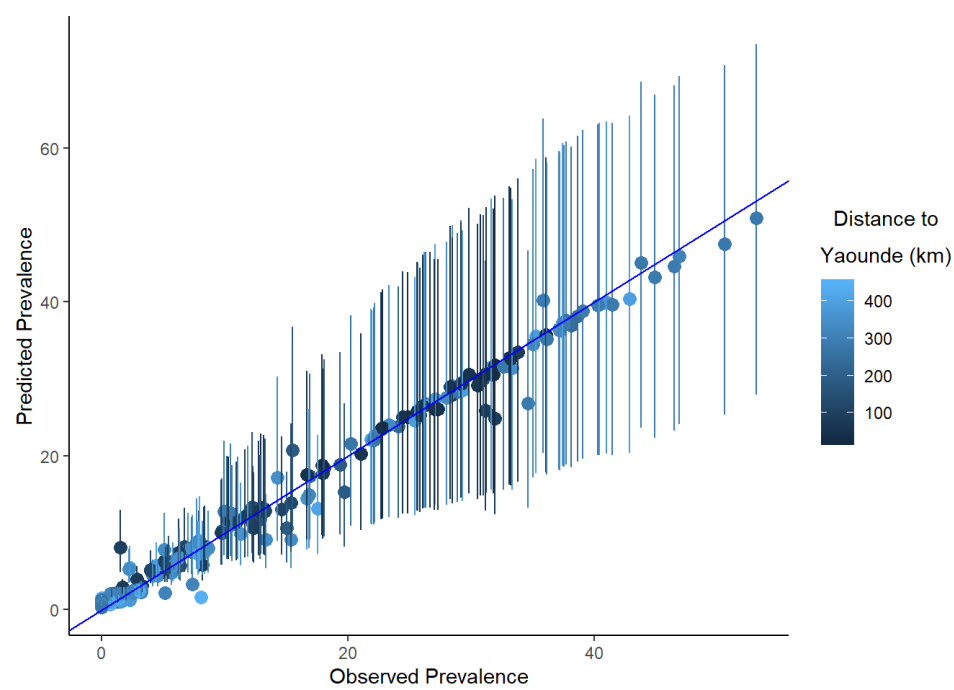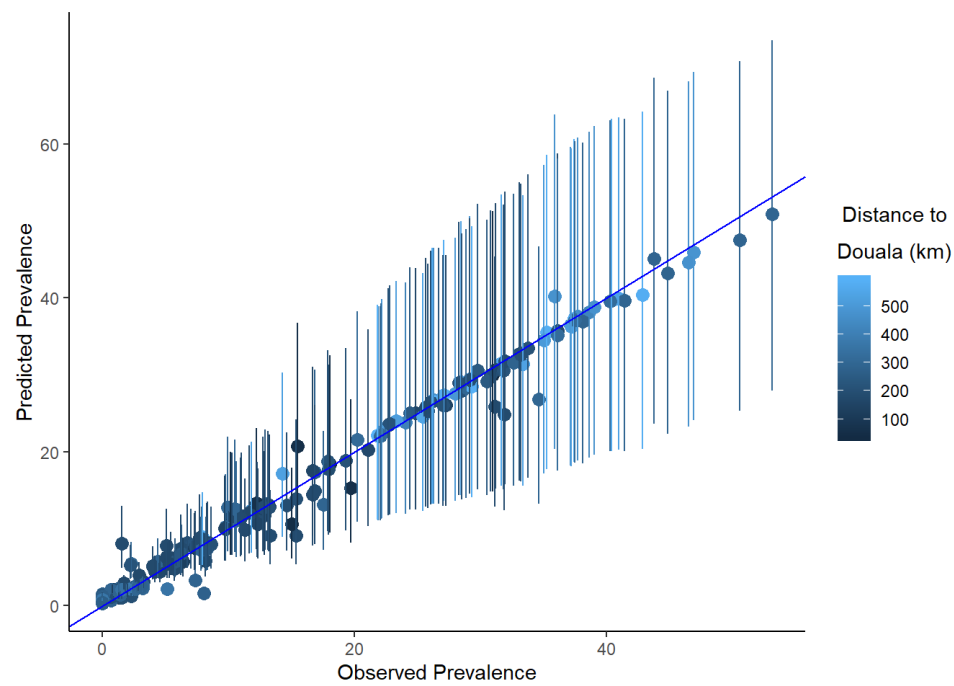

Figure 14: Observed Prevalence vs Predicted Prevalence with Distance to Daoula (km) (non-normal)
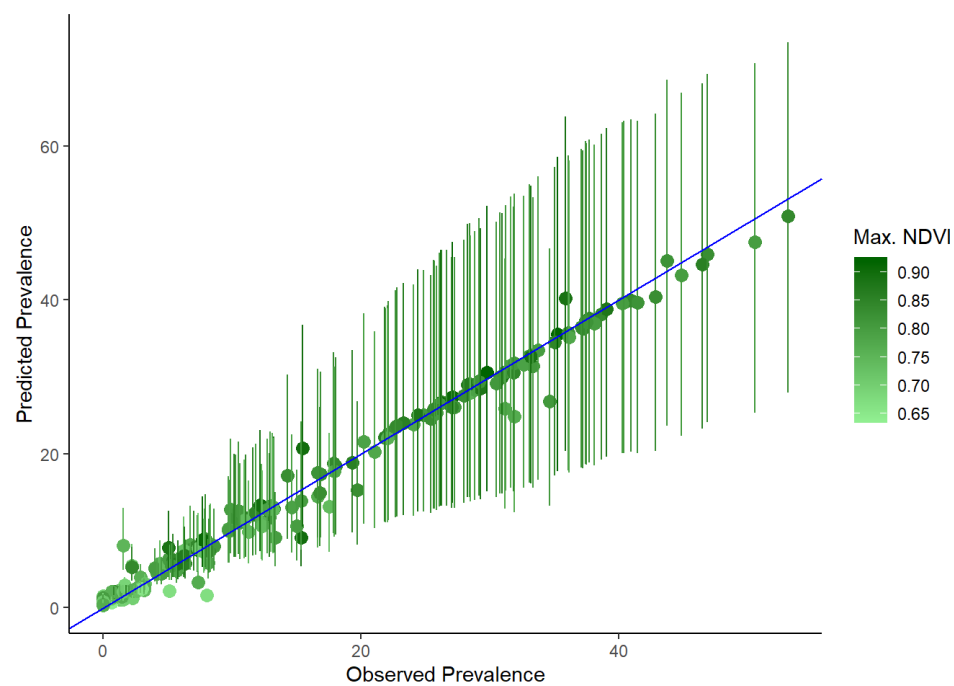

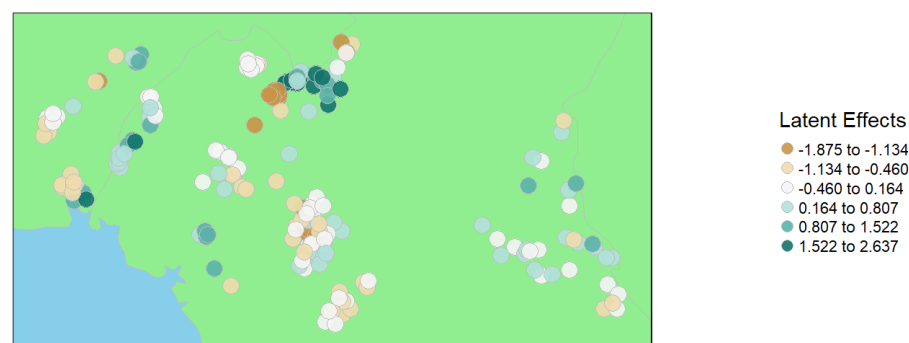Figure 15: Observed Prevalence vs Predicted Prevalence with Vegetation Index (non-normal)

Figure 16: Non-normal Spatial Latent Effects

Tables

Table 1: Summary Statistics

	lon	lat	number_tested	number_positives	elevation	max_NDVI	sd_NDVI	perc_pos	cap_dist	cap2_dist	resid	latent	predicted	re
	Min. : 8.004	Min. :3.350	Min. : 24.0	Min. : 0.00	Min. : 4.0	Min. :0.6400	Min. :0.1037	Min. : 0.000	Min. : 23.05	Min. : 35.17	Min. :-20.996	Min. :-1.87512	Min. : 0.2961	Mi
	1st Qu.:10.049	1st Qu.:4.389	1st Qu.: 94.0	1st Qu.: 4.00	1st Qu.: 456.0	1st Qu.:0.7600	1st Qu.:0.1443	1st Qu.: 3.012	1st Qu.:106.28	1st Qu.:181.08	1st Qu.: -6.652	1st Qu.:-0.58008	1st Qu.: 2.8933	1s 2.(
	Median :11.180	Median :5.058	Median :132.0	Median : 13.00	Median : 610.0	Median :0.8000	Median :0.1552	Median :11.842	Median :281.58	Median :222.54	Median : -1.590	Median :-0.03814	Median :11.6849	Me 6.5
	Mean :11.100	Mean :5.151	Mean :135.3	Mean : 21.83	Mean : 676.5	Mean :0.7947	Mean :0.1556	Mean :15.750	Mean :242.90	Mean :267.31	Mean : 0.000	Mean : 0.03977	Mean :15.6145	Me
	3rd Qu.:11.583	3rd Qu.:6.093	3rd Qu.:154.0	3rd Qu.: 35.00	3rd Qu.: 878.0	3rd Qu.:0.8300	3rd Qu.:0.1666	3rd Qu.:27.072	3rd Qu.:337.70	3rd Qu.:299.31	3rd Qu.: 4.361	3rd Qu.: 0.55563	3rd Qu.:26.6437	3r Qu
	Max. :15.136	Max. :6.878	Max. :432.0	Max. :162.00	Max. :1829.0	Max. :0.9200	Max. :0.2067	Max. :53.125	Max. :447.33	Max. :595.47	Max. : 36.352	Max. : 2.63650	Max. :50.9097	Ma

Table 2: OLS Model Comparison

	Full Model			Chosen Mean Function		
Coefficient	Estimates	Conf. Int (95%)	P-Value	Estimates	Conf. Int (95%)	P-Value
(Intercept)	-57.53	-79.17 – -35.89	**<0.001**	-55.88	-76.58 – -35.17	**<0.001**
elevation	-0.00	-0.01 – -0.00	**0.011**	-0.01	-0.01 – -0.00	**0.005**
max_NDVI	83.07	53.87 – 112.26	**<0.001**	86.84	61.31 – 112.37	**<0.001**
sd_NDVI	27.38	-74.69 – 129.46	0.597			
cap_dist	-0.04	-0.06 – -0.03	**<0.001**	-0.04	-0.06 – -0.03	**<0.001**
cap2_dist	0.06	0.05 – 0.08	**<0.001**	0.06	0.05 – 0.08	**<0.001**
Observations	197			197		
R^2 / R^2 adjusted	0.522 / 0.509			0.521 / 0.511		

Table 3: Isotropic Model Summary

	estimate	stdErr	ci0.005	ci0.995	ci0.025	ci0.975	ci0.05	ci0.95	ci0.1	ci0.9	pval	Estimated
(Intercept)	15.3940669	0.1369724	15.0412493	15.7468846	15.1256059	15.6625280	15.1687673	15.6193666	15.2185297	15.5696042	0	TRUE
scale(max_NDVI)	3.6584737	0.0965354	3.4098150	3.9071324	3.4692678	3.8476796	3.4996871	3.8172603	3.5347586	3.7821888	0	TRUE
scale(elevation)	-1.9437119	0.1164668	-2.2437105	-1.6437133	-2.1719826	-1.7154411	-2.1352827	-1.7521410	-2.0929701	-1.7944537	0	TRUE
scale(cap_dist)	-7.3147875	0.1759172	-7.7679199	-6.8616550	-7.6595787	-6.9699962	-7.6041454	-7.0254295	-7.5402343	-7.0893406	0	TRUE
scale(cap2_dist)	8.4180039	0.1414579	8.0536326	8.7823753	8.1407516	8.6952563	8.1853265	8.6506814	8.2367184	8.5992895	0	TRUE
sdNugget	3.3414605	NA	1.1674668	9.5637476	1.5011940	7.4376518	1.7072831	6.5398399	1.9802436	5.6383762	NA	TRUE
sdSpatial	8.8366646	NA	6.8146634	11.4586206	7.2514478	10.7684209	7.4856467	10.4315158	7.7650660	10.0561466	NA	TRUE
range	0.2610502	NA	0.1068101	0.6380219	0.1322536	0.5152766	0.1475321	0.4619143	0.1673505	0.4072125	NA	TRUE
shape	0.5760410	NA	-0.5693828	1.7214648	-0.2955188	1.4476008	-0.1553951	1.3074771	0.0061587	1.1459233	NA	TRUE
anisoRatio	1.0000000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	FALSE
anisoAngleRadians	0.0000000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	FALSE
anisoAngleDegrees	0.0000000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	FALSE
boxcox	1.0000000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	FALSE

Table 4: Anisotropic Model Summary

	estimate	stdErr	ci0.005	ci0.995	ci0.025	ci0.975	ci0.05	ci0.95	ci0.1	ci0.9	pval	Estimated
(Intercept)	15.5041178	0.1395259	15.1447228	15.8635127	15.2306520	15.7775836	15.2746181	15.7336175	15.3253081	15.6829274	0	TRUE
scale(max_NDVI)	3.6784497	0.0899185	3.4468349	3.9100645	3.5022126	3.8546868	3.5305468	3.8263525	3.5632144	3.7936849	0	TRUE
scale(elevation)	-1.9222312	0.1119576	-2.2106149	-1.6338475	-2.1416641	-1.7027983	-2.1063851	-1.7380773	-2.0657107	-1.7787517	0	TRUE
scale(cap_dist)	-7.6295550	0.1743880	-8.0787487	-7.1803613	-7.9713492	-7.2877608	-7.9163977	-7.3427123	-7.8530422	-7.4060678	0	TRUE
scale(cap2_dist)	8.2061745	0.1421592	7.8399968	8.5723523	7.9275477	8.4848014	7.9723435	8.4400055	8.0239903	8.3883588	0	TRUE
sdNugget	2.1578792	NA	0.0700586	66.4649794	0.1589895	29.2877322	0.2418077	19.2568038	0.3921222	11.8749774	NA	TRUE
sdSpatial	9.2754259	NA	7.0196715	12.2560615	7.5032816	11.4661198	7.7634667	11.0818439	8.0746606	10.6547545	NA	TRUE
range	0.1848685	NA	0.0718386	0.4757377	0.0900549	0.3795057	0.1010941	0.3380647	0.1155113	0.2958701	NA	TRUE
shape	0.4310706	NA	-0.3225340	1.1846753	-0.1423517	1.0044930	-0.0501605	0.9123018	0.0561300	0.8060113	NA	TRUE
anisoRatio	2.8983313	NA	1.0040302	8.3666049	1.2936773	6.4933691	1.4728159	5.7035804	1.7103490	4.9114679	NA	TRUE
anisoAngleRadians	-0.8986199	NA	-1.2363107	-0.5609290	-1.1555709	-0.6416689	-1.1142599	-0.6829798	-1.0666311	-0.7306086	NA	TRUE
anisoAngleDegrees	-51.4871254	NA	-70.8353860	-32.1388648	-66.2093331	-36.7649177	-63.8423916	-39.1318592	-61.1134596	-41.8607912	NA	TRUE
boxcox	1.0000000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	FALSE

Table 5: Non-normal Spatial Model Summary

```
formula: cbind(number_positives, number_tested - number_positives) ~ scale(max_NDVI) +
    scale(elevation) + scale(cap_dist) + scale(cap2_dist) + Matern(1 |
    lon + lat)
Estimation of corrPars and lambda by Laplace REML approximation (p_bv).
Estimation of fixed effects by Laplace ML approximation (p_v).
Estimation of lambda by 'outer' REML, maximizing p_bv.
Family: binomial ( link = logit )
 ------------ Fixed effects (beta) ------------
                Estimate Cond. SE t-value
(Intercept)      -2.3209  0.16497 -14.069
scale(max_NDVI)   0.4776  0.09388   5.087
scale(elevation) -0.3665  0.11794  -3.108
scale(cap_dist)  -0.9647  0.20678  -4.666
scale(cap2_dist)  0.8185  0.16143   5.070
 -------------- Random effects --------------
Family: gaussian ( link = identity )
               --- Correlation parameters:
   1.nu    1.rho
0.50000 4.46978
            --- Variance parameters ('lambda'):
lambda = var(u) for u ~ Gaussian;
   lon + lat  :   0.8873
# of obs: 197; # of groups: lon + lat, 197
 ------------ Likelihood values  ------------
                      logLik
p_v(h) (marginal L): -660.9773
  p_beta,v(h) (ReL): -666.3830
```

Table 6: Comparison of Models

Model	AIC	Effective Range (km)
OLS	1460.45	—
Isotropic	1366.46	28.9766
Anisotropic	1363.17	20.5204
Non-normal	1335.95 (marginal)	74.5003
Non-normal	1121.96 (conditional)	—

Citations and the like

1. "DNDi – Best Science for the Most Neglected". www.dndi.org. Archived from the original on 13 March 2018. Retrieved 5 May 2018.

2. Hotez PJ (January 2010). "How To Cure 1 Billion People? — Defeat Neglected Tropical diseases". Scientific American. 302 (1): 90–94, 96. doi:10.1038/scientificamerican0110-90 (doi:10.1038/scientificamerican0110-90). PMID 20063641. Archived from the original on 6 August 2014.

3. "Neglected tropical diseases". World Health Organization. Archived from the original on 22 February 2014. Retrieved 24 November 2015.

4. "CDC – Neglected Tropical Diseases – Diseases". www.cdc.gov. Archived from the original on 4 December 2014. Retrieved 30 October 2016.

5. "PLoS Neglected Tropical Diseases: A Peer-Reviewed Open-Access Journal". journals.plos.org. Archived from the original on 18 November 2016. Retrieved 30 October 2016.

6. Boussinesq, M., Gardon, J., Gardon-Wendel, N., and J. Chippaux. 2003. "Clinical picture, epidemiology and outcome of Loa-associated serious adverse events related to mass ivermectin treatment of onchocerciasis in Cameroon". Filaria Journal 2: 1–13.

7. Turkington, C., & Ashby, B. (2007). Encyclopedia of Infectious Diseases. New York: Facts on File.

8. Wanji, S., Tendongfor, N., Nji, T., Esum, M., Che, J. N., Nkwescheu, A., Alassa, F., Kamnang, G., Enyong, P. A., Taylor, M. J., Hoerauf, A., and D. W. Taylor. 2009. "Community-directed delivery of doxycycline for the treatment of onchocerciasis in areas of co-endemicity with loiasis in Cameroon". Parasites & Vectors. 2(39): 1–10.

9. "Overall total population" – World Population Prospects: The 2019 Revision". population.un.org. United Nations Department of Economic and Social Affairs, Population Division.

10. Diggle, et al. "Spatial modelling and the prediction of Loa loa risk: decision making under uncertainty." PLOS Neglected Tropical Diseases, vol. 10, no. 12, 2016, doi:10.1371/journal.pntd.0005157 (doi:10.1371/journal.pntd.0005157)

Full Code

```r
setwd("C:/Users/keete/Documents/Fall 2019/Spatial Stats/Final Project/Loa")

#### Data Clean & Variable Creation ####

## Libraries

library(geoR)
library(tidyverse)
library(readxl)
library(sf)
library(sp)
library(gstat)
library(ggfortify)
library(maps)
library(maptools)
library(sf)
library(tmap)
library(geostatsp)
library(RColorBrewer)
library(ggthemes)
library(geosphere)
library(spaMM)


#### Read in data
loa <- read.table("loaloa.txt", header = F)

#### Rename variables
names(loa) <- c(
  "lon",
  "lat",
  "number_tested",
  "number_positives",
  "elevation",
  "max_NDVI",
  "sd_NDVI"
)

summary(loa)

##### Create % positive variable
names(loa)

##### Create Distance from major cities variable

loa <- loa %>%
  mutate(perc_pos = (number_positives/number_tested)*100) %>%
  rowwise() %>%
  mutate(cap_dist = distHaversine(c(lon, lat), c(11.501346, 3.844119)),  # Yaounde, Cam.
         cap_dist = cap_dist/1000,
         cap2_dist = distHaversine(c(lon, lat), c(9.786072, 4.061536)),  # Douala, Cam.
         cap2_dist = cap2_dist/1000,
         cap_dist = as.numeric(cap_dist),
         cap2_dist = as.numeric(cap2_dist)
  )


summary(loa)
max_lon <- max(loa$lon)
min_lon <- min(loa$lon)

max_lat <- max(loa$lat)
min_lat <- min(loa$lat)

correction <- .25
### Note: Yaounde is capital and 1st largest at 1.9 mil.
###       Douala is 2nd largest city at 1.8 mil. (as of 2005 Census)

##### Spatial points dataframes

loa_sp <- SpatialPoints(cbind(loa$lon, loa$lat), proj4string = CRS("+proj=longlat"))

loa_sp <- spTransform(loa_sp, CRS("+init=epsg:26391"))     # Needed for maps

loa_sf <- st_as_sf(
  loa,
  coords = c("lon", "lat"),
  crs = CRS("+proj=utm +zone=30")
)

d <- data.frame(
  st_coordinates(
    loa_sf
  ),
  loa[, -1:-2]
)

loa_model_sp  <- SpatialPointsDataFrame(
  coords = d[, 1:2],
  data = d,
  proj4string = CRS(
    "+proj=utm +zone=30"
  )
)

### End of data clean ###


##### OLS model
model <- lm(
  perc_pos ~ max_NDVI + cap_dist + cap2_dist + elevation,
  data = loa
)

summary.lm(model)

##### Obtain residuals

loa$resid <- resid(model)

autoplot(
  model,
  smooth.colour = NA
) +
  theme_classic(

  )

full_model <- lm(
  perc_pos ~ . - resid - lat - lon  - number_positives - number_tested,
```

```r
  data = loa
)

summary.lm(full_model)

ggplot(
  data = loa,
  aes(
    x = resid
  )
) +
  geom_histogram(
    aes(
      y=..density..
      ),
    colour = "black",
    fill = "lightblue",
    bins = 20
    ) +
  geom_density(
    alpha = .2
    ) +
  theme_classic(

  ) +
  labs(
    y = "Density",
    x = "Residual"
  )


##### Cutoff and for variogram
cutoff <- .65*max(
  dist(
    cbind(
      loa$lon,
      loa$lat
    )
  )
)

bins <- 15

###### Variogram/Covariogram (with scale variables)
variogram <- variogram(
  scale(perc_pos) ~ 1,
  locations = ~lon + lat,
  data = loa,
  cutoff = cutoff,
  width = cutoff/bins
)

plot(
  variogram,
  pch = 16,
  cex = 1.5,
  ylab = expression(
    paste(
      "Average ", (0.65*(Y(s[i]) - Y(s[j])^2))
      )
    ),
  xlab = "Euclidean Distance (degrees)"
)


covariogram <- variogram(
  scale(perc_pos) ~ 1,
  locations = ~lon + lat,
  data = loa,
  cutoff = cutoff,
  width = cutoff/bins,
  covariogram = T,
  map = T
)

plot(
  covariogram,
  contour = T
)


##### Spatial Models #####


###### Anisotropic Model (with scaled variables)
fit_aniso <- lgm(
  data = loa_model_sp,
  perc_pos ~ scale(max_NDVI) + scale(elevation) + scale(cap_dist) + scale(cap2_dist),
  grid = 100,
  shape = .5,
  fixShape = FALSE,
  nugget = .5,
  fixNugget = FALSE,
  aniso = TRUE,
  reml = TRUE
)

summary(fit_aniso)

AIC(fit_aniso)

##### Isotropic model (with scaled variables)
fit_iso <- lgm(
  perc_pos ~ scale(max_NDVI) + scale(elevation)+ scale(cap_dist) + scale(cap2_dist),
  data = loa_model_sp,
  grid = 100,
  shape = .5,
  fixShape = FALSE,
  nugget = .5,
  fixNugget = FALSE,
  aniso = FALSE,
  reml = TRUE
)

summary(fit_iso)

AIC(fit_iso)

##### Non-normal Spatial Model
```

```r
spamm_fit <- fitme(
  cbind(number_positives, number_tested - number_positives) ~ scale(max_NDVI) + scale(elevation) + scale(cap_dist) + scale(c
ap2_dist) +
    Matern(1|lon + lat),
  fixed = list(
    nu = .5
  ),
  method = "REML",
  data = loa,
  family = "binomial"
)

spamm_fit

AIC(spamm_fit)

latent <- ranef(spamm_fit)[[1]]

loa$latent <- latent

##### Figures for An- and Isotropic Model (Predicted SD and Predicted Prev.)

par(mfcol=c(2,1), mai=c(0.5,0.5,0.5,0.5))

plot(fit_aniso$predict[["predict"]],main='Anisotropic Predicted Prevalence')

plot(fit_iso$predict[["predict"]],main='Isotropic Predicted Prevalence')


par(mfcol=c(2,1), mai=c(0.5,0.5,0.5,0.5))

plot(fit_aniso$predict[["krigeSd"]],main='Anisotropic Prediction SDs')

plot(fit_iso$predict[["krigeSd"]],main='Isotropic Prediction SDs')


##### Figure for non-normal model

filled.mapMM(
  spamm_fit,
  map.asp = 3,
  nlevels = 10,
  gridSteps = 50,
  add.map = TRUE,
  axes = F,
  yrange = c(
    min_lat - correction,
    max_lat + correction
  ),
  xrange = c(
    min_lon - correction,
    max_lon +  correction
    )
  )


#### Observered ~ fitted figure (observed vs predicted)

predicted <- as.numeric(predict.HLfit(spamm_fit)*100)

loa$predicted <- predicted

int <- get_intervals(spamm_fit)*100

loa <- data.frame(loa, int)

## Observed vs Fitted

ggplot(
  data = loa,
  aes(
    x = perc_pos,
    y = predicted
    )
  ) +
  geom_point(
    size = 3
  ) +
  geom_linerange(
    aes(
      ymin = respVar_0.025,
      ymax = respVar_0.975
    )
  ) +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "blue"
  ) +
  labs(
    x = "Observed Prevalence",
    y = "Predicted Prevalence"
  ) +
  theme_classic(

  )

## w/ Elevation

ggplot(
  data = loa,
  aes(
    x = perc_pos,
    y = predicted,
    col = elevation
  )
) +
  geom_point(
    size = 3
  ) +
  geom_linerange(
    aes(
      ymin = respVar_0.025,
      ymax = respVar_0.975
    )
  ) +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "blue"
```

```
  ) +
  labs(
    x = "Observed Prevalence",
    y = "Predicted Prevalence",
    col = "Elevation (m)"
  ) +
  theme_classic(

  )

# w/ Distance from Cap.

ggplot(
  data = loa,
  aes(
    x = perc_pos,
    y = predicted,
    col = cap_dist
  )
) +
  geom_point(
    size = 3
  )+
  geom_linerange(
    aes(
      ymin = respVar_0.025,
      ymax = respVar_0.975
    )
  ) +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "blue"
  ) +
  labs(
    x = "Observed Prevalence",
    y = "Predicted Prevalence",
    col = expression(atop("Distance to", paste("Yaounde (km)")))
  ) +
  theme_classic(

  )


# w/ Distance to Daoula

ggplot(
  data = loa,
  aes(
    x = perc_pos,
    y = predicted,
    col = cap2_dist
  )
) +
  geom_point(
    size = 3
  )+
  geom_linerange(
    aes(
      ymin = respVar_0.025,
      ymax = respVar_0.975
    )
  ) +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "blue"
  ) +
  labs(
    x = "Observed Prevalence",
    y = "Predicted Prevalence",
    col = expression(atop("Distance to", paste("Douala (km)")))
  ) +
  theme_classic(

  )

# w/ NDVI

ggplot(
  data = loa,
  aes(
    x = perc_pos,
    y = predicted,
    col = max_NDVI
  )
) +
  geom_point(
    size = 3
  ) +
  geom_linerange(
    aes(
      ymin = respVar_0.025,
      ymax = respVar_0.975
    )
  ) +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "blue"
  ) +
  labs(
    x = "Observed Prevalence",
    y = "Predicted Prevalence",
    col = "Max. NDVI"
  ) +
  scale_color_gradient(
    low="lightgreen",
    high="darkgreen"
    ) +
  theme_classic(

  )


#### Model comparison

spamm_aic_1 <- format(as.numeric(AIC(spamm_fit)[1])[1], digits = 6, format = "f")
spamm_aic_2 <- format(as.numeric(AIC(spamm_fit)[2])[1], digits = 6, format = "f")
spamm_ef <- format(3/(as.numeric(spamm_fit$corrPars$`1`[2]))*111, digits = 6, format = "f")
```

```
spamm_aic_1 <- paste(spamm_aic_1, "(marginal)")
spamm_aic_2 <- paste(spamm_aic_2, "(conditional)")


comparison <- data.frame(
  "Model" = c("OLS", "Isotropic", "Anisotropic", "Non-normal", "Non-normal"),
  "AIC" = c(format(AIC(model), digits = 6, format = "f"), format(AIC(fit_iso), digits = 6, format = "f"), format(AIC(fit_ani
so),digits = 6, format = "f"), spamm_aic_1, spamm_aic_2),
  "Effective Range km" = c("---", format((fit_iso$parameters[1])*111, digits = 6, format = "f"), format((fit_aniso$parameter
s[1])*111,digits = 6, format = "f"), spamm_ef, "---"))

comparison


### End of Models ###

#### Maps ####


##### Maps of Africa
africa <- map(
  database = "world",
  regions = c("Nigeria", "Cameroon", "Central African Republic", "Chad"),
  fill = T,
  plot = F
)

africa_fill <- map2SpatialPolygons(
  africa,
  IDs = africa$names,
  proj4string = CRS(
    "+proj=utm +zone=30"
  )
)

world_ref1 <- map(
  database = "world",
  fill = T,
  plot = F
)

world_ref1 <- map2SpatialPolygons(
  world_ref1,
  IDs = world_ref1$names,
  proj4string = CRS(
    "+proj=utm +zone=30"
  )
)


##### World Reference Map

ggplot(
  world_ref1,
  aes(
    x = long,
    y = lat,
    group = group
  )
) +
  geom_polygon(
    fill="lightgreen",
    colour = "grey"
  ) +
  coord_equal(
    xlim = c(-25, 60),
    ylim = c(-35, 45)
  ) +
  theme_map(

  ) +
  theme(
    plot.background = element_rect(
      fill = 'skyblue'
    )
  ) +
  annotate(
    "rect",
    xmin = min(loa$lon) - correction,
    xmax = max(loa$lon) + correction,
    ymin = min(loa$lat) - correction,
    ymax = max(loa$lat) + correction,
    alpha = .25,
    fill = "red"
  )


names <- data.frame(
  country = c("Nigeria", "Cameroon", "Central African Republic", "Chad"),
  lat = c(7.602539, 10.070801, 15.395507, 15.820313),
  long = c(5.207658, 4.477856, 5.047171, 8.566537)
)

##### Central Africa Reference map
ggplot(
  world_ref1,
  aes(
    x = long,
    y = lat,
    group = group
  )
) +
  geom_polygon(
    fill="lightgreen",
    colour = "grey"
  ) +
  coord_equal(
    xlim = c(min(loa$lon) - 4, max(loa$lon)) + 2,
    ylim = c(min(loa$lat) - 2, max(loa$lat) + 2)
  ) +
  theme_map(

  ) +
  theme(
    plot.background = element_rect(
      fill = 'skyblue'
    )
  ) +
  annotate(
```

```r
      "rect",
      xmin = min(loa$lon) - correction,
      xmax = max(loa$lon) + correction,
      ymin = min(loa$lat) - correction,
      ymax = max(loa$lat) + correction,
      alpha = .05,
      fill = "red"
    ) +
    annotate(
      "text",
      label = c("Nigeria", "Cameroon", (expression(atop("Central", paste("African Republic")))), "Chad"),
      x = c(7.602539, 11.370801, 16.2, 16.4),
      y = c(8, 5.2, 5.8, 8.566537),
      size = 5
    ) +
    annotate(
      "text",
      label = c("Yaounde", "Douala"),
      x = c(11.501346, 9.786072) + .5,
      y = c(3.844119, 4.061536) - .15
    ) +
    annotate(
      "point",
      color = "black",
      x = 9.786072,
      y = 4.061536
    ) +
    annotate(
      "text",
      label = "*",
      x = 11.501346,
      y = 3.844119,
      size = 8.5
    )


##### L. Loa Prevalence Map

tm_shape(
  africa_fill,
  ylim = c(
    min_lat - correction,
    max_lat + correction
  ),
  xlim = c(
    min_lon - correction,
    max_lon +  correction
  )
) +
  tm_fill(
    col = "lightgreen"
  ) +
  tm_borders(
    lwd = .5,
    col = "grey"
  ) +
  tm_shape(
    loa_sf
  ) +
  tm_symbols(
    col = "perc_pos",
    title.col = "Prevalence (%)",
    palette = "-RdYlBu",
    n = 6,
    style = "jenks",
    border.lwd = 0.1,
    border.col = 'gray',
    alpha = 0.9,
    jitter = 0.1,
    scale = 0.85
  )  +
  tm_layout(
    bg.color = "skyblue",
    saturation = 1,
    legend.position = c("right", "center")
  ) +
  tm_legend(
    legend.outside = TRUE,
    frame = F
  )


##### OLS Residual Map
tm_shape(
  africa_fill,
  ylim = c(
    min_lat - correction,
    max_lat + correction
  ),
  xlim = c(
    min_lon - correction,
    max_lon +  correction
  )
) +
  tm_fill(
    col = "lightgreen"
  ) +
  tm_borders(
    lwd = .5,
    col = "grey"
  ) +
  tm_shape(
    loa_sf
  ) +
  tm_symbols(
    col = "resid",
    title.col = "Residual OLS",
    palette = "-RdYlBu",
    n = 6,
    style = "jenks",
    border.lwd = 0.1,
    border.col = 'gray',
    alpha = 0.9,
    jitter = 0.1,
    scale = 0.85,
    midpoint = 0
  ) +
  tm_layout(
    bg.color = "skyblue",
    saturation = 1,
```

```r
      legend.position = c("right", "center")
    )  +
    tm_legend(
      legend.outside = TRUE,
      frame = F
    )

### Latent Spatial Effects

latent <- ranef(spamm_fit)[[1]]

loa$latent <- latent

loa_sp <- SpatialPoints(cbind(loa$lon, loa$lat), proj4string = CRS("+proj=longlat"))

loa_sp <- spTransform(loa_sp, CRS("+init=epsg:26391"))     # Needed for maps

loa_sf <- st_as_sf(
  loa,
  coords = c("lon", "lat"),
  crs = CRS("+proj=utm +zone=30")
)

tm_shape(
  africa_fill,
  ylim = c(
    min_lat - correction,
    max_lat + correction
  ),
  xlim = c(
    min_lon - correction,
    max_lon +  correction
  )
) +
  tm_fill(
    col = "lightgreen"
  ) +
  tm_borders(
    lwd = .5,
    col = "grey"
  ) +
  tm_shape(
    loa_sf
  ) +
  tm_symbols(
    col = "latent",
    title.col = "Latent Effects",
    palette = "BrBG",
    n = 6,
    style = "jenks",
    border.lwd = 0.1,
    border.col = 'gray',
    alpha = 0.9,
    jitter = 0.1,
    scale = 0.85,
    midpoint = 0
  ) +
  tm_layout(
    bg.color = "skyblue",
    saturation = 1,
    legend.position = c("right", "center")
  )  +
  tm_legend(
    legend.outside = TRUE,
    frame = F
  )

###### End of Maps ######

###### End of the whole damn thing ######
```

RColorBrewer Maroon Palette