Spatial Variation in Risk Factors of Lymphatic Filariasis in Hotspot Zones in Ghana

Efiba Vidda Kwarteng (lynnvid@yahoo.com)
Kwame Nkrumah University of Science and Technology College of Engineering https://orcid.org/0000-0001-7562-9063

Samuel Ato Andam-Akorful
Kwame Nkrumah University of Science and Technology College of Engineering

Alexander Kwarteng
Kwame Nkrumah University of Science and Technology College of Science

Da-Costa Boakye Mensah Asare
Kwame Nkrumah University of Science and Technology College of Engineering

Jonathan Arthur Quaye-Ballard
Kwame Nkrumah University of Science and Technology College of Engineering

Frank Babu Osei
University of Twente

Alfred Allan Duker
Kwame Nkrumah University of Science and Technology College of Engineering

Research article

Keywords: Lymphatic filariasis, machine learning, Ensemble modelling, Generalised boosted model (GBM), Random forest (RF), ecological niche modelling

Posted Date: December 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-116528/v1

Version of Record: A version of this preprint was published on January 28th, 2021. See the published version at https://doi.org/10.1186/s12889-021-10234-9.

Abstract

Lymphatic Filariasis (LF), a parasitic nematode infection, often resulting in disability poses huge economic burden to affected countries. To meet eradication deadlines in line with the global Neglected Tropical Diseases elimination and health system strengthening goals, novel strategies are needed to complement existing approaches. LF endemicity is localized and prevalence, spatially heterogeneous. Species distribution models (SDMs) can help identify subtle differences in risk factors that influences the transmission of LF in geographically distinct regions. Thus in this contribution, presence absence records of microfilaria (m)f in Ghana were stratified into Northern and Southern Zones and used to run SDMs, whilst climate, socioeconomic and land coverage variables provided explanation information. GLM (Generalized Linear model), GBM (Generalized Boosted Model), ANN (Artificial Neural Network), SRE (Surface Range Envelope), MARS (Multivariate Adaptive Regression Splines) and RF (Random Forests) algorithms were run for both study zones and also for the entire country for the purpose of comparison. Best model quality was obtained with RF and GBM algorithms with highest AUC between 0.98 and 0.95, respectively. The models predicted high suitable environments for LF transmission in the short grass savanna areas in the northern and along the coastal southern parts of Ghana. Mainly, land cover and socioeconomic variables such as, proximity to inland waterbodies and population density uniquely influenced LF transmission in the South while poor housing was a distinctive risk factors in the North. Precipitation, temperature, slope and poverty were common risk factors but with subtle variations in response values, which was confirmed by the countrywide model. This study has demonstrated that, an understanding of the geographic distinctness in risk factors is required to inform the development of area-specific transmission control systems towards LF elimination in Ghana and internationally.

Introduction

Lymphatic filariasis (LF) is one of the neglected tropical diseases (NTDs) which presents chronic disabling and disfiguring pathologies with occasional painful attacks on affected persons (Taylor, 2009). LF is a mosquito-borne infection caused by filarial nematodes which include: *Wuchereria bancrofti*, *Brugia timori* and *B. malayi* (Specht *et al*., 2019). These worms produce larvae i.e. microfilariae (*mf*) which are transmitted by mosquitoes in endemic areas, thus, reducing *mf* levels is significant towards LF eradicating (Famakinde, 2018). It is estimated that over 1.4 million individuals are at risk of the infection in 83 endemic countries (Cano *et al*., 2018). Currently, the mainstay eradication strategies which include Mass Drug Administration (MDA) and vector control have significantly interrupted LF transmission in many previously endemic settings (Dorkenoo *et al*., 2018). While these achievements are commendable, there is the need to adopt novel approaches especially in foci, where LF transmission is ongoing despite several years of implementing these control strategies.

In Ghana, studies on LF have shown differences in disease prevalence and multiplicity of symptoms in two geographically distinct regions, i.e., the northern and southern parts (Gyapong, Adjei and Sackey, 1996). The northern regions of the country exhibit higher prevalence compared to the southern regions, but the middle forest belt is relatively free from the infection (Gyapong, Adjei and Sackey, 1996: Dunyo *et al*., 1996). Elsewhere, a study has revealed some level of genetic variability in parasites strains in the two endemic areas (de Souza *et al*., 2014). In addition, Pi-Bansa *et al*., (2019), identified a vector of very high vectorial capacity specific to the coastal areas i.e. southern Ghana. These variations in the two regions could be due to the existence of different climatological, land cover and socioeconomic risk factors.

For instance, De Souza et al., (2010) reported that ecological and climatic variables such as elevations greater than 200 m, mean daily precipitation between 2.6 – 3.8 mm, and mean daily temperature range between 24.5–26.0^{o}C, influences the distribution of *Anopheles gambiae,* one of the vectors for LF transmission in Ghana. Eneanya *et al*., (2018)

compared the influence of similar factors on LF transmission in endemic and non-endemic areas in Nigeria. At the global scale, another study used climatic and environmental variables in a boosted regression tree (BRT) model to map the transmission limits of LF (Cano *et al.*, 2014), confirming the influence of geo-environmental risk factors on vector population and vectorial capacity (Eneanya *et al.*, 2018; De Souza *et al.*, 2012).

While these studies present very useful findings, their spatial scale of analysis obscures some micro-level risk factors (Manzoor, Griffiths and Lukac, 2018), which may be important for designing strategies for disease control especially in hotspots zones. According to Williams et al., (2012), the spatial scale for analysis should include the known environmental or geographic limits of the species under study for quality model predictions. In the West African sub-region, different geographical zones have been documented (de Souza et al., 2017). The south is characterized by wetlands, while the north is characterized by dry lands and sub-Sahelian climate (Pi-Bansa et al., 2019). In Ghana, the northern and the southern regions although both highly endemic for *W. bancroftian* infections, have distinct geographic characteristics in terms of land cover and climate. This distinction is likely to influence vector proliferation and transmission potential differently.

Therefore, to facilitate LF elimination in these two highly endemic areas, a local understanding of the environmental, climatic and socioeconomic factors that drives transmission is required to review existing control programmes. In line with this, the present study sought to map the environmental niches of LF and examine the behaviour of the diverse risk factors that drive transmission in the northern and southern zones of Ghana. Since the prevalence of LF is denoted by the presence or absence of microfilaria (*mf*) cases, data on *mf* survey from sentinel and spot-check sites across Ghana were stratified into Northern Zone (NZ) and Southern Zone (SZ) and used to run Species Distribution Models (SDMs), whilst climate, socioeconomic and land coverage variables were used as covariates. The analysis was then performed over the entire country for the purpose of comparison.

Materials And Methods

Study Area

The study was conducted in Ghana as shown in Figure 1. However, because LF appears to be localized in Northern and Southern Ghana, the study area was subdivided to only include endemic areas in these two regions. To investigate risk factors in the two highly endemic zones and how they compare with result from the entire country, three zonal analyses were performed: Countrywide (CW), Northern Zone (NZ) and Southern Zone (SZ). The area considered the SZ in this study included districts that lie along the coastal savannah, tropical rainforest and some portion of the moist semi-deciduous forest region of Ghana, while the NZ comprised of the Sudan savannah and some part of the Guinea savannah.

The SZ lies within the high rain forest ecological zone of the West African sub region, with strands of mangroves (de Souza et al., 2017) and lots of wetlands. The climate in this region is tropical, characterized by two distinctive seasonal rainfalls; major one between April and June and a minor one which occurs between September and October. Relative humidity is generally high, averaging between 75% to 85% in the rainy, and 70% to 80% in the dry seasons. The highest mean temperature is 34 ^{o}C whereas the lowest is 20 ^{o}C. In contrast, the NZ lies in the dry Guinea Savannah Ecological zone (Pi-Bansa et al., 2019) with a sub-Sahelian climate made up of a wet and a dry season. The wet season extends from April to October, with a mean annual rainfall of approximately 1365 mm. Similarly, the dry season is subdivided into the Harmattan occurring from November to mid-February and the dry hot season from mid-February to April. Monthly temperatures range from 20 °C to 40 °C.

LF prevalence data

Data on *mf* cases in Ghana was obtained from published article in peer-reviewed journals (Biritwum *et al.*, 2019: Gyapong *et al.*, 2002). The data spanning 2000 to 2014 contained information on the year samples were collected, number of years of MDA, number of people examined, and number of *mf* positive recorded for each study community. In all, 430 communities were surveyed for LF infections as part of a transmission assessment survey in Ghana. Details of this dataset is described by Biritwum *et al.* (2019). Spatial locations of these communities were extracted from multiple sources including Google Earth Pro, Open Street Map, directory of cities and towns (world database) and database of the Ghana National Identification Authority card registration projects. Figure 2 shows a map of the spatial distribution of *mf* cases in Ghana (Figure 2a) and the NZ (Figure 2b) and SZ (Figure 2c) zones.

Geo-environmental and climatological data source

To identify the combination of explanatory variables that create suitable environment for the transmission of lymphatic filariasis, land cover, socioeconomic and climatic predictors were obtained from various remotely-sensed datasets. Enhanced Vegetation Index (EVI) was generated from the Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite image, specifically MOD13Q1 v006 (NASA, 2020). This data is generated every 16 days at 250 meter (m) spatial resolution. Land surface temperature covariates were also computed from MOD11A2 data from this site.

From the United State Geological Surveys (USGS) earth explorer project (*US Geological Surveys*, 2020), a raster dataset of elevation produced by the Shuttle Radar Topography Mission (SRTM) and Slope covariate was derived. In addition, Landsat 7 ETM + 1 level 1 at 30 x 30m resolution of less than 1% cloud cover was downloaded from the same sight for Land Use/Land Cover (LULC) classification.

To determine rural and mostly poor areas in Ghana, Night-light emissivity captured by the Operational Linescan System instrument on board the Defence Meteorological Satellite Programme satellite, from 2000 to 2014 was used as a proxy (Elvidge *et al.*, 1999). This instrument measures visible and infrared radiation emitted at night time. The values range from 0 to 62, representing undetectable emissivity and maximum emissivity, respectively. Night-light emissivity has been shown to correlate with economic development in subnational regions of developing countries (Bruederle and Hodler, 2018). Another socioeconomic variable used was the prevalence of housing with improved drinking water and sanitation, sufficient living area, and durable construction across sub-Saharan Africa (Tusting *et al.*, 2019). The prevalence of houses built with finished materials are higher in urban areas than in rural areas showing 84% and 34% improvement respectively.

Precipitation and temperature variables were downloaded from the WorldClim database (WorldClim, 2020). This dataset provides a set of global climate layers obtained by interpolation of weather station datasets distributed across the world. Other covariates used in the SDMs with details on the sources are provided in Table 1. Input grids were resampled to a common spatial resolution of 1 km^2 using bilinear resampling for analysis performed with CW data, while a finer resolution of 250m was used for the NZ and SZ to capture detailed information (Williams *et al.*, 2012). Raster layers were coerced to the same boundary extent to enable stacking for analysis.

Table 1. Environmental variables used in the SDMs for *mf* occurrence and their sources

Variable	Variable Description	Source
Population	Population Density	WorldPop (Linard *et al.*, 2012)
Housing	Improved Housing	The malaria atlas project (Tusting *et al.*, 2019)
DEM	Digital elevation model	STRM (*US Geological Surveys*, 2020)
Waterbodies	Proximity to all water bodies and wetlands; swamps and marshes	
Slope	Derived from elevation	
LULC	Land use and land cover classes	Landsat 7 (*US Geological Surveys*, 2020)
Bio 1	Annual Mean Temperature	WorldClim (WorldClim, 2020)
Bio 12	Annual Precipitation	
Bio17	Precipitation of the direst quarter	
Bio 18	Precipitation of Warmest Quarter	
Bio 19	Precipitation of Coldest Quarter	
NTL	Distance to stable night light	
MeanDayLST	Mean Day Land Surface Temperature	MOD11A2 (NASA, 2020)
MaxDayLST	Maximum Day Land Surface Temperature	
MinDayLST	Minimum Day Land Surface Temperature	
MeanNightLST	Mean Night Land Surface Temperature	
MaxNightLST	Maximum Night Land Surface Temperature	
MinNightLST	Minimum Night Land Surface Temperature	
EVI	Enhanced Vegetation Index	MOD13Q1 (NASA, 2020)

Variable selection and model development

To identify optimal suite of covariates to include in the specie distribution models, the variables were grouped into three categories; land cover, socioeconomic and climatic variables (Moraga *et al.*, 2015). Within each group, a test for variable collinearity with the Variance Inflation Factor (VIF) diagnostic method was adopted. Since there are no formal criteria for deciding when a VIF is too large, a generic cutoff value of was used (Craney *et al.*, 2007). This approach reduces any potential collinearity and confounding effects such that for p - 1 independent variables,

$$VIF_i = \frac{1}{1-r_i^2}, \qquad i = 1, \dots\dots\dots, p-1, \qquad (1)$$

where r_i^2 is the coefficient of determination obtained by fitting a regression model for the ith independent variable on the other independent variables. After collinearity check, only Bio1 (Mean Annual Temperature) had collinearity problem.

Variable Relative Contribution

After strongly correlated variables were removed, the range of variables influencing the occurrence of *mf*, were identified using boosted regression trees (BRT). This method draws insights and techniques from both statistical and machine learning traditions. The advantage of this method over the others is its strong predictive performance and consistent identification of relevant variables and interactions. Here, the probability of *mf* occurrence, y = 1, in a sampled community with covariates X, is given as p(y = 1|X). This probability models via a logit function *f*(x) = p(y = 1|X).

Analytically, BRT regularization involves jointly optimizing the number of trees (*nt*), learning rate (*lr*), and tree complexity (*tc*). The optimal number of trees was estimated by the default 10-fold cross-validation (CV) method (Elith, Leathwick and Hastie, 2008). With a slow enough *lr* 0.01, the CV estimates of *nt* are reliable and close to those from independent data. To ensure modelling of possible interactions between predictors, a *tc* of 5 was selected. A *tc* of 1 fits an additive model while a *tc* of 2 fits a model with up to two-way interactions, and so on (Elith, Leathwick and Hastie, 2008). It has been proven that, stochasticity improves model performance, and fractions in the range of 0·5–0·75 have given best results for presence–absence responses (Elith, Leathwick and Hastie, 2008), therefore a bag fraction of 0·75 was used from here on and an error structure of Bernoulli.

Relative importance of variables were computed by measuring the number of times a predictor variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and an average over all the trees is then determined (Friedman, 2001). Expressing in mathematical terms, the relative influence, I_j of the input variables X_j for a collection of decision trees $\{T_m\}_1^M$, is given by

$$\hat{I}_j^2 = \frac{1}{M}\sum_{m=1}^{M} \hat{I}_j^2(T_m) \qquad (2)$$

where is the number of iteration. The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response. A threshold of 10% was set below which a variable is considered to have no substantial contribution to the model (Rogers, 2006). Variables dropped in both zones were EVI, DEM, maximum night land surface temperature, Bio19, Bio18, maximum day land surface temperature, LULC, mean and minimum night land surface temperature. In addition, Bio17, distance to inland water body, population density and mean day Land surface temperature had less than 10% contribution in the Northern Zone, whereas improved housing, Bio12, and minimum night land surface temperature had insignificant contribution to the model for the Southern Zone.

Model Selection

Six model classes i.e. generalized linear models (GLM) (Pearson *et al.*, 2006), multivariate adaptive regression splines (MARS) (Friedman, 1991), artificial neural networks (ANN) (Gant and Gant, 2001), generalized boosted models (GBM) (Elith, Leathwick and Hastie, 2008), Random Forests (RF) (Breiman, 2001), and one rectilinear envelope (SRE) (Booth *et al.*, 2014) were tested using Biomod2 package in R (Thuiller *et al.*, 2009). Out of these, the Random Forest and GBM were the best performing models for our data and were therefore used for modelling and prediction of LF. Hundred (100) model runs for each algorithm was performed iteratively, and the evaluation values of each run was stored and then averaged to make the final result more robust. Model evaluation was performed based on the area under the receiver operating characteristic (ROC) curve. This measure the ability of the final ensemble model to fit the presence-absence data and predict across unsampled locations.

Results

Distribution of *mf* in Ghana

The distribution of *mf* cases from 430 communities surveyed in Ghana shows that *mf* infection is mainly found in the Northern, the Southern and some parts of the middle belt of Ghana. The presence points indicated in red in Figure 2 showed *mf* occurred along coastal communities in Southern Ghana (i.e. Western and Central Regions). In the Northern sector, *mf* cases were widespread in most of the districts, which also had very high incidence as compared to Southern Ghana.

Model performance

Model performances of six model algorithms for the SZ and NZ are shown in Table 2. Judging by AUC (i.e. Area Under Receiver Operating Characteristic (ROC) curve), sensitivity (percentage of presences correctly predicted) and specificity (percentage of absences correctly predicted) values, RF and GBM models outperformed the ANN, SRE, MARS and GLM. AUC values between 0.5–0.6 indicate a failed model performance, whereas 0.6–0.7 represent poor model quality; 0.7–0.8 represent models with fair performance and 0.8–0.9; indicating a good model performance (Hanley and McNeil, 1982). Overall, the RF was of best quality for the country wide modeling as well as the northern and southern zones (Table 2). For further evaluation, results of model with AUC ≥ 0.8 only (i.e. RF and GBM) were considered for ensemble modelling.

Table 2. Calculated AUC, sensitivity and specificity values of different SDM algorithms for *mf* occurrence in CW, NZ and SZ.

Model	AUC CW	Sensitivity CW	Specificity CW	AUC NZ	Sensitivity NZ	Specificity NZ	AUC SZ	Sensitivity SZ	Specificity SZ
GBM	0.95	92.23	86.91	0.94	91.00	91.00	0.91	86.5	88.03
RF	0.97	94.99	93.64	0.98	95.51	95.55	0.95	92.5	97.04
GLM	0.83	88.55	71.55	0.84	79.63	82.81	0.82	72.4	87.35
ANN	0.82	84.61	73.27	0.71	86.38	52.16	0.79	64.95	89.36
SRE	0.67	76.57	58.12	0.64	7.74	51.2	0.44	97	3.08
MARS	0.87	84.73	80.82	0.82	80.22	77.72	0.77	61.8	91.73

Influence and Importance of Risk Variable in Northern and Southern Ghana

Variable importance was evaluated for environmental, socioeconomic and climatic variables as shown in Table 3. Here, it was observed that in all the two zones and also in comparison with the country analysis, distance to stable night light was an important variable although a weak one for both GBM and RF algorithms: CW (0.10 and 0.14, respectively), NZ (<0.01 and 0.17, respectively) and SZ (0.03 and 0.13, respectively). In addition, weak but important values were computed for variables such as, terrain slope for NZ (0.10 and 0.18) and SZ (0.04 and 0.10) and improved housing for CW (0.09 and 0.10) and NZ (0.13 and 0.16) for both algorithms, respectively. On the other hand, while Bio 12 was of highest importance for both algorithms in CW (0.43 and 0.31) and NZ (0.63 and 0.36, respectively), proximity to water bodies was given highest importance for the same algorithms (i.e. 0.59 and 0.41 respectively) in SZ. Finally, the following variables of varying importance i.e. fair to weak values were unique for the three study zones: CW (maximum

night land surface temperature and DEM), NZ (Minimum day land surface temperature), and SZ (Bio 17, Population density and Minimum day land surface temperature).

Table 3. Variables used in SDMs and difference in variable importance for the CW, NZ and SZ. The highest variable importance value for each model is highlighted with bold and underlined numbers.

Variables	GBM CW	RF CW	GBM NZ	RF NZ	GBM SZ	RF SZ
bio17	-	-	-	-	0.04	0.10
bio12	**0.43**	**0.31**	**0.63**	**0.36**	-	-
NTL	0.10	0.14	<0.01	0.17	0.03	0.13
Population	-	-	-	-	0.25	0.13
Slope	-	-	0.10	0.18	0.04	0.10
DEM	0.21	0.17	-	-	-	-
Proximity to Waterbodies	0.10	0.19	-	-	**0.59**	**0.41**
Improved Housing	0.09	0.10	0.13	0.16	-	-
Mean Day Land Surface Temperature	-	-	-	-	0.04	0.13
Minimum Day Land Surface Temperature	-	-	0.11	0.13	-	-
Maximum Night Land Surface Temperature	0.06	0.06	-	-	-	-

Partial Dependence Plots of Factors associated with mf transmission

Figures 3 and 4 shows the response plots of each covariate for RF and GBM models run with data from NZ and SZ. In Northern Ghana, high suitability for *mf* is negatively associated with annual precipitation (i.e. rainfall values greater than 1000 mm result in a decrease in *mf* occurrence), high terrain slope, increased distance to stable night light and minimum day land surface temperature (increased values above 23°C). There was a general increase of *mf* in areas with less improvement in housing in northern Ghana and appeared to decrease in areas with greater than 30% improved housing (Figure 3g).

In the south, high suitability values were associated with distance to water bodies and low values with terrain slope. It was observed that proximity to water bodies, population density, mean day land surface temperature showed a negative correlation with *mf* occurrence while increase in terrain slope and increased distance to stable night light showed a positive correlation with *mf* occurrence as shown in Figure 4. Response plot for CW model showed consistency in the occurrence of *mf* with covariates observed in the NZ and SZ such as precipitation, distance to stable nigh light (supplementary file). The behavior of the type of housing shows better in the CW model as mf occurrence decreases in areas with higher improvement in housing.

Probability Maps of LF Occurrence

The map with the CW dataset shown in figure 5a, presents a discrimination of endemic and non-endemic areas over Ghana. Probability maps of the two zones compared with the countrywide analysis highlighted areas with zero or little

occurrence probability (<0.5) in large forest regions of Ghana (Figure 5 and 6). The map shown in Figure 5a and 6a suggests that a larger portion of the north western Ghana was environmentally suitable and better able to drive *mf* transmission. Areas of endemicity rather shrinks sharply moving towards the north eastern part of Ghana showing high suitability in the extreme northern part of the country. Figure 5b is an exaggeration of the coastline for cartographic purposes and agreed with suitable areas demarcated in figure 6b. These correspond to mangrove ecosystems and freshwater swamps in the southern parts of the country.

Discussion

Despite several years of mass drug administration and vector control measures against human lymphatic filariasis in Ghana, some areas continue to serve as hotspot for its transmission. LF occurrence in Ghana appears to vary from one location to another. In Ghana, two major endemic zones are known for LF, i.e. northern and southern zones. The mid-section has only some few cases of *mf* believed to have been imported from the North (Gyapong *et al.*, 1996). To address these differences, and to provide appropriate elimination strategies, understanding the differential factors i.e., environmental, climatic and socioeconomic covariates that drive *mf* transmission in the two study zones is key.

Evaluation of model performances revealed that RF and GBM algorithms performed better for all three zonal *mf* datasets. Application of the two models in the current study showed that the AUC of success rate ranged from 0.95 to 0.98 and 0.91 to 0.95 for RF and GBM, respectively. This may be attributed to the fact that RF and GBM are better able to handle large covariates (Thuiller *et al.*, 2010) as provided in this study.

The probability of *mf* occurrence is influenced by different combination of variables in both northern and southern Ghana as observed in this study. In the North, the occurrence of *mf* was influenced by low values of annual precipitation, but decreased with high values above 1000 mm. The precipitation variable behaved differently in Southern Ghana with precipitation of the driest quarter sustaining LF transmission. In Ghana, heavy rainfall between April to June usually result in flooding in the Northern region (Lolig *et al.*, 2014). In the South, high rainfall pattern and low elevation particularly along the coast may result in surface water run-offs. These occurrences may sweep away breeding habitats reducing the survival of LF vector and subsequent transmission of *mf*. However, availability of rain especially in the coastal areas during the driest period of the year from late December to March can create pockets of stagnant water bodies to sustain mosquito breeding, therefore, increasing LF transmission. It implies that, whereas rainfall is needed for vector breeding, excessive rainfall could potentially result in flood and sweeping away of breeding sites in *mf* endemic areas (Dieng *et al.*, 2012). Findings from this study is consistent with previous study by Abiodun *et al.* (2016), Cano *et al.* (2014) and Eneanya *et al.* (2018).

Similarly, the occurrence of *mf* declined with high land surface temperature during the day. This is consistent with adult mosquito survival and larval development, which suggests that both adult and larvae are unable to survive at high temperatures (Lardeux and Cheffort, 2001). The response curve shows that, minimum day land surface temperature values between 23°C to 24.5^{o}C (GBM and RF respectively) may increase mortality rate of either larvae or adult mosquitoes in the North. Comparatively, in the South, vector survival is supported beyond these temperature ranges until temperatures between 27°C to 29°C. This may be as result of thick vegetation cover or tree canopy in southern Ghana which may create suitable conditions likely to sustain vector survival even at high temperatures. It was also observed that the probability of mf occurrence decreased with increasing terrain slope in both areas. This finding is consistent with studies by (Eneanya et al., 2018). What accounts for such observation is that steeper surfaces could lead to faster surface water runoff, thus decreasing collection of water in pockets and eventually reducing breeding sites for vectors associated with mf transmission. In addition to poverty, areas of poor housing support transmission in the North.

Distance to stable night light was an important covariate for both northern and southern Ghana as well as at the countrywide scale. In the North, the response curve shows that suitable areas for LF occurrence were generally rural and poor communities. In the South, some communities located in peri-urban to urban communities had high probability of LF occurrence. This is true because some coastal communities located in peri-urban areas in the Western region have high *mf* prevalence. Low population density which reflect communities around vegetated and underdeveloped areas in the South was also an important variable.

Despite suitability in the North and South, slight differences in suite of environmental variables at varying value ranges drive LF occurrence suggesting diverse risk factors. This implies that any efforts or strategy intended towards eliminating the disease should consider unique condition prevailing at relatively fine spatial scale. Limitation of the study was that, the model did not consider important risk factors at the community or individual level such as sanitation and other demographic factors that are likely to improve the predictions. Finally, larger sample size could lead to more precise predictions.

Conclusion

The probability of *lymphatic filiarisis* occurrence is influenced by different variable combinations in Northern and Southern Ghana. For both sectors, the transmission is prevalent in poor rural communities in low lying areas. For northern Ghana, it occurs in relatively warm, low lying rural communities with poor housing, especially those characterized by mud houses. In addition, a mean annual precipitation between 900 mm to 1000 mm provides a conducive environment for transmission. Similarly, rural, poor, low lying and mostly coastal communities in the South present the suitable environment for LF transmission. However, some peri-urban areas along the coast were observed to be suitable areas. Generally, the infection is efficiently transmitted in warm lowland communities within 2 km of inland water bodies such as mangroves, lagoons, and rivers in the South. Moreover, rainfall within the relatively warm part of the year was identified as an important risk factor as it likely contributes to formation of stagnant water bodies, suitable for mosquito breeding. Some intervention such as housing improvements may offer protection against transmission in the north as delineated suitable areas generally had poor housing, while improvement in sanitary conditions along coastal communities in the south, may be considered to support elimination strategies. The findings of the present study can be utilized by policy makers in advancing evidence-based strategies to eliminate LF in Ghana.

Data Sharing

The data sources are publicly available online and have been referenced in the manuscript. All maps presented in this study were generated by the authors using ArcGIS Desktop 10.6, and no permissions are required to publish them.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Funding

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

EVSK substantially contributed to the conception and design of the work, acquisition, analysis, interpretation of data and the original draft preparation of the manuscript.

SAA contributed by way of conception and design, interpretation of data and was a major contributor in writing the manuscript.

AK contributed by conception and design of the work, and was a major contributor in writing the manuscript.

JAQ substantively revised the manuscript.

DBA contributed to the conception and design of the work.

FBO substantively revised the manuscript.

AAD: contributed to the conception and design of the work, and substantively revised the manuscript.

All authors read and approved the final manuscript.

Acknowledgment

We acknowledge the individuals and groups who routinely collect and report this data to peer reviewed journals.

References

Abiodun, G. J. *et al.* (2016) 'Modelling the influence of temperature and rainfall on the population dynamics of Anopheles arabiensis', *Malaria Journal*. BioMed Central, 15(1), pp. 1–15. doi: 10.1186/s12936-016-1411-6.

Biritwum, N. K. *et al.* (2019) 'Progress towards lymphatic filariasis elimination in Ghana from 2000-2016: Analysis of microfilaria prevalence data from 430 communities', *PLoS Neglected Tropical Diseases*, 13(8), pp. 1–15. doi: 10.1371/journal.pntd.0007115.

Booth, T. H. *et al.* (2014) 'BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies', *Diversity and Distributionsversity*, pp. 1–9. doi: 10.1111/ddi.12144.

Breiman, L. (2001) 'ST4_Method_Random_Forest', *Machine Learning*, 45(1), pp. 5–32. doi: 10.1017/CBO9781107415324.004.

Bruederle, A. and Hodler, R. (2018) 'Nighttime lights as a proxy for human development at the local level', *PLoS ONE*, 13(9), pp. 1–22. doi: 10.1371/journal.pone.0202231.

Cano, J. *et al.* (2014) 'The global distribution and transmission limits of lymphatic filariasis: Past and present', *Parasites and Vectors*, 7(1), pp. 1–19. doi: 10.1186/s13071-014-0466-x.

Cano, J. *et al.* (2018) 'Identifying co-endemic areas for major filarial infections in sub-Saharan Africa: Seeking synergies and preventing severe adverse events during mass drug administration campaigns', *Parasites and Vectors*. Parasites &

Vectors, 11(1), pp. 1–11. doi: 10.1186/s13071-018-2655-5.

Craney, T. A. *et al.* (2007) 'Model-Dependent Variance Inflation Factor Cutoff Values Model-Dependent Variance Inflation Factor Cutoff Values', 2112. doi: 10.1081/QEN-120001878.

Dieng, H. *et al.* (2012) 'The effects of simulated rainfall on immature population dynamics of Aedes albopictus and female oviposition', *International Journal of Biometeorology*, 56(1), pp. 113–120. doi: 10.1007/s00484-011-0402-0.

Dorkenoo, M. A. *et al.* (2018) 'Molecular xenomonitoring for post- validation surveillance of lymphatic filariasis in Togo : no evidence for active transmission', *Parasites & Vectors*. Parasites & Vectors, pp. 1–9. doi: 10.1186/s13071-017-2611-9.

Dunyo, S. K. *et al.* (1996) 'Lymphatic filariasis on the coast of Ghana', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 90(6), pp. 634–638. doi: http://dx.doi.org/10.1016/S0035-9203%2896%2990414-9.

Elith, J., Leathwick, J. R. and Hastie, T. (2008) 'A working guide to boosted regression trees', *Journal of Animal Ecology*, 77(4), pp. 802–813. doi: 10.1111/j.1365-2656.2008.01390.x.

Elvidge, C. D. *et al.* (1999) 'Radiance calibration of DMSP-OLS low-light imaging data of human settlements', *Remote Sensing of Environment*, 68(1), pp. 77–88. doi: 10.1016/S0034-4257(98)00098-4.

Eneanya, O. A. *et al.* (2018) 'Environmental suitability for lymphatic filariasis in Nigeria', *Parasites and Vectors*. Parasites & Vectors, pp. 1–13.

Famakinde, D. (2018) 'Mosquitoes and the Lymphatic Filarial Parasites: Research Trends and Budding Roadmaps to Future Disease Eradication', *Tropical Medicine and Infectious Disease*, 3(1), p. 4. doi: 10.3390/tropicalmed3010004.

Friedman, J. H. (1991) 'Multivariate Adaptive Regression Splines', *The Annals of Statistics*, 19(1), pp. 1–67.

Friedman, J. H. (2001) 'Greedy Function Approximation: A Gradient Boosting Machine', *Annals of Statistic*.

Gant, V. and Gant, V. (2001) 'Clinical applications of artificial neural networks', (January). doi: 10.1017/CBO9780511543494.

Gyapong, J. O. *et al.* (2002) 'The use of spatial analysis in mapping the distribution of bancroftian filariasis in four West African countries', *Annals of Tropical Medicine and Parasitology*, 96(7), pp. 695–705. doi: 10.1179/000349802125001735.

Gyapong, J. O., Adjei, S. and Sackey, S. O. (1996) 'Descriptive epidemiology of lymphatic filariasis in Ghana', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 90(1), pp. 26–30. doi: 10.1016/S0035-9203(96)90466-6.

Hanley, J. A. and McNeil, B. J. (1982) 'The meaning and use of the area under a receiver operating characteristic (ROC) curve', *Radiology*, 143(1), pp. 29–36. doi: 10.1148/radiology.143.1.7063747.

Lardeux, F. and Cheffort, J. (2001) 'Ambient temperature effects on the extrinsic incubation period of Wuchereria bancrofti in Aedes polynesiensis: Implications for filariasis transmission dynamics and distribution in French Polynesia', *Medical and Veterinary Entomology*, 15(2), pp. 167–176. doi: 10.1046/j.0269-283X.2001.00305.x.

Linard, C. *et al.* (2012) 'Population Distribution , Settlement Patterns and Accessibility across Africa in 2010', 7(2). doi: 10.1371/journal.pone.0031743.

Lolig, V. *et al.* (2014) 'Households' coping strategies in drought- and flood-Prone communities in Northern Ghana', *Journal of Disaster Research*, 9(4), pp. 542–553. doi: 10.20965/jdr.2014.p0542.

Manzoor, S. A., Griffiths, G. and Lukac, M. (2018) 'Species distribution model transferability and model grain size-finer may not always be better', *Scientific Reports*. Springer US, 8(1), pp. 1–9. doi: 10.1038/s41598-018-25437-1.

Moraga, P. *et al.* (2015) 'Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: Integrated use of geostatistical and mathematical modelling', *Parasites and Vectors*. Parasites & Vectors, 8(1), pp. 1–16. doi: 10.1186/s13071-015-1166-x.

NASA (2020) *DAAC NL. NASA LP DAAC: MOD13Q1 Vegetation Indices 16-Day L3 Global 250m: NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota*.

Pearson, R. G. *et al.* (2006) 'Model-based uncertainty in species range prediction', *Journal of Biogeography*, 33(10), pp. 1704–1711. doi: 10.1111/j.1365-2699.2006.01460.x.

Pi-Bansa, S. *et al.* (2019) 'Potential factors influencing lymphatic filariasis transmission in "hotspot" and "control" areas in Ghana: the importance of vectors', *Infectious Diseases of Poverty*. Infectious Diseases of Poverty, 8(1), p. 9. doi: 10.1186/s40249-019-0520-1.

Rogers, D. J. (2006) 'Models for Vectors and Vector-Borne Diseases', *Advances in Parasitology*, 62(05), pp. 1–35. doi: 10.1016/S0065-308X(05)62001-5.

de Souza, D. K. *et al.* (2014) 'The epidemiology of lymphatic filariasis in Ghana, explained by the possible existence of two strains of Wuchereria bancrofti', *Pan African Medical Journal*, 17, pp. 1–7. doi: 10.11604/pamj.2014.17.133.3370.

De Souza, D. K. *et al.* (2012) 'Diversity and transmission competence in lymphatic filariasis vectors in West Africa, and the implications for accelerated elimination of Anopheles-transmitted filariasis', *Parasites and Vectors*, 5(1). doi: 10.1186/1756-3305-5-259.

Specht, S. *et al.* (2019) 'Elimination of lymphatic filariasis in South East Asia', *BMJ (Online)*, 364, pp. 1–4. doi: 10.1136/bmj.k5198.

Taylor, M. J. (2009) 'Current Practices in the Management of Lymphatic Filariasis', *Expert Rev Anti Infect Ther*, 7, pp. 595–605.

Thuiller, W. *et al.* (2009) 'BIOMOD - A platform for ensemble forecasting of species distributions', *Ecography*, 32(3), pp. 369–373. doi: 10.1111/j.1600-0587.2008.05742.x.

Thuiller, W., Lafourcade, B. and Miguel, A. (2010) 'Presentation for BIOMOD', *http://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/inst/doc/Biomod_Presentation_Manual.pdf?revision=218&root=biomod&pathrev=218*.

Tusting, L. S. *et al.* (2019) 'Mapping changes in housing in sub-Saharan Africa from 2000 to 2015', *Nature*. Springer US, 568(7752), pp. 391–394. doi: 10.1038/s41586-019-1050-5.

Tustingi, L. S. *et al.* (2020) 'Housing and child health in sub-Saharan Africa: A cross-sectional analysis', *PLoS Medicine*, 17(3), pp. 1–18. doi: 10.1371/JOURNAL.PMED.1003055.

US Geological Surveys (2020). Available at: https://earthexplorer.usgs.gov/.

Williams, K. J. *et al.* (2012) 'Which environmental variables should I use in my biodiversity model?', *International Journal of Geographical Information Science*, 26(11), pp. 2009–2047. doi: 10.1080/13658816.2012.698015.

WorldClim (2020) *Global Climate Data. https://www.worldclim.org/*.

Figures

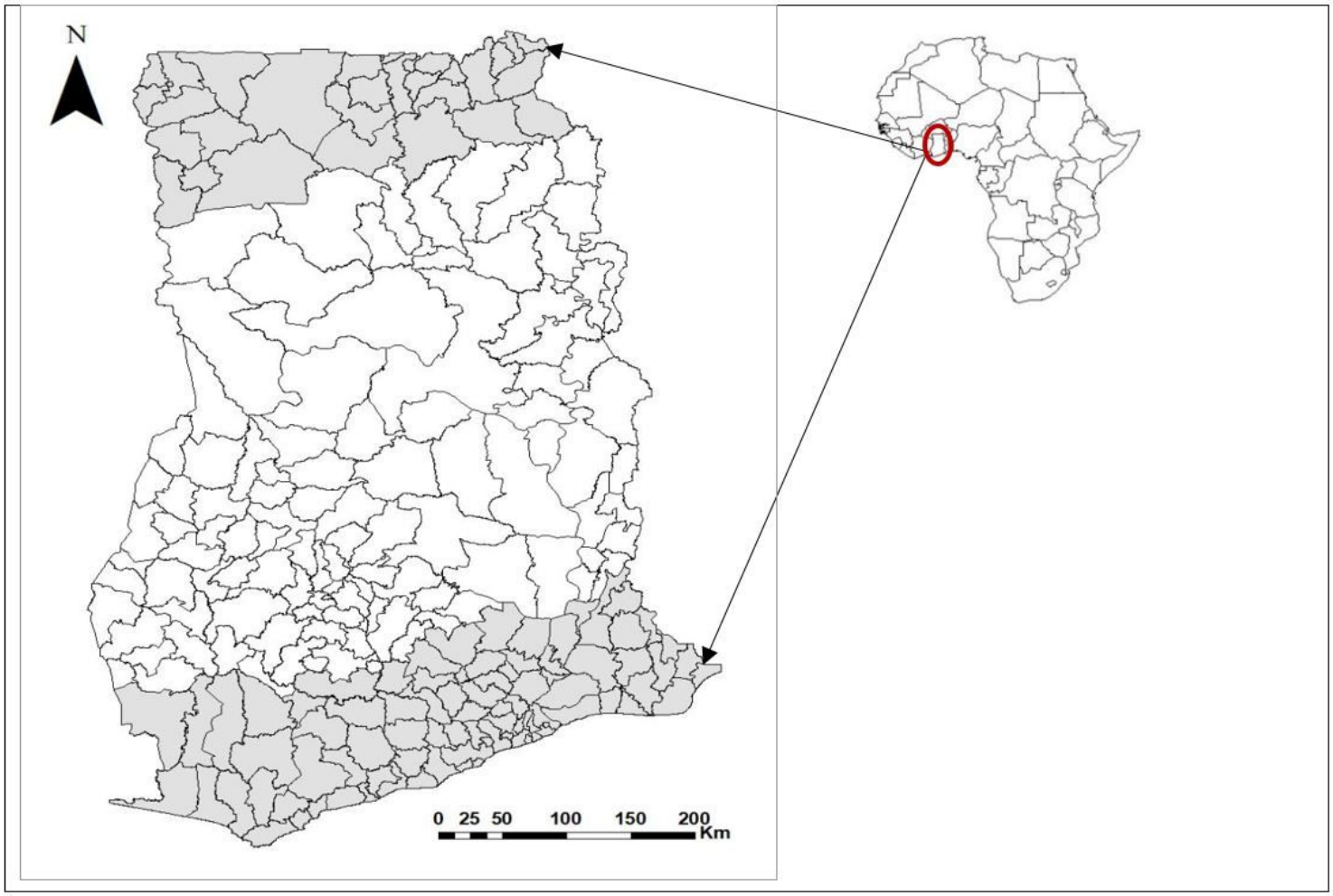

Figure 1

Map of Ghana showing the districts included in the two study zones, NZ and SZ shaded in grey

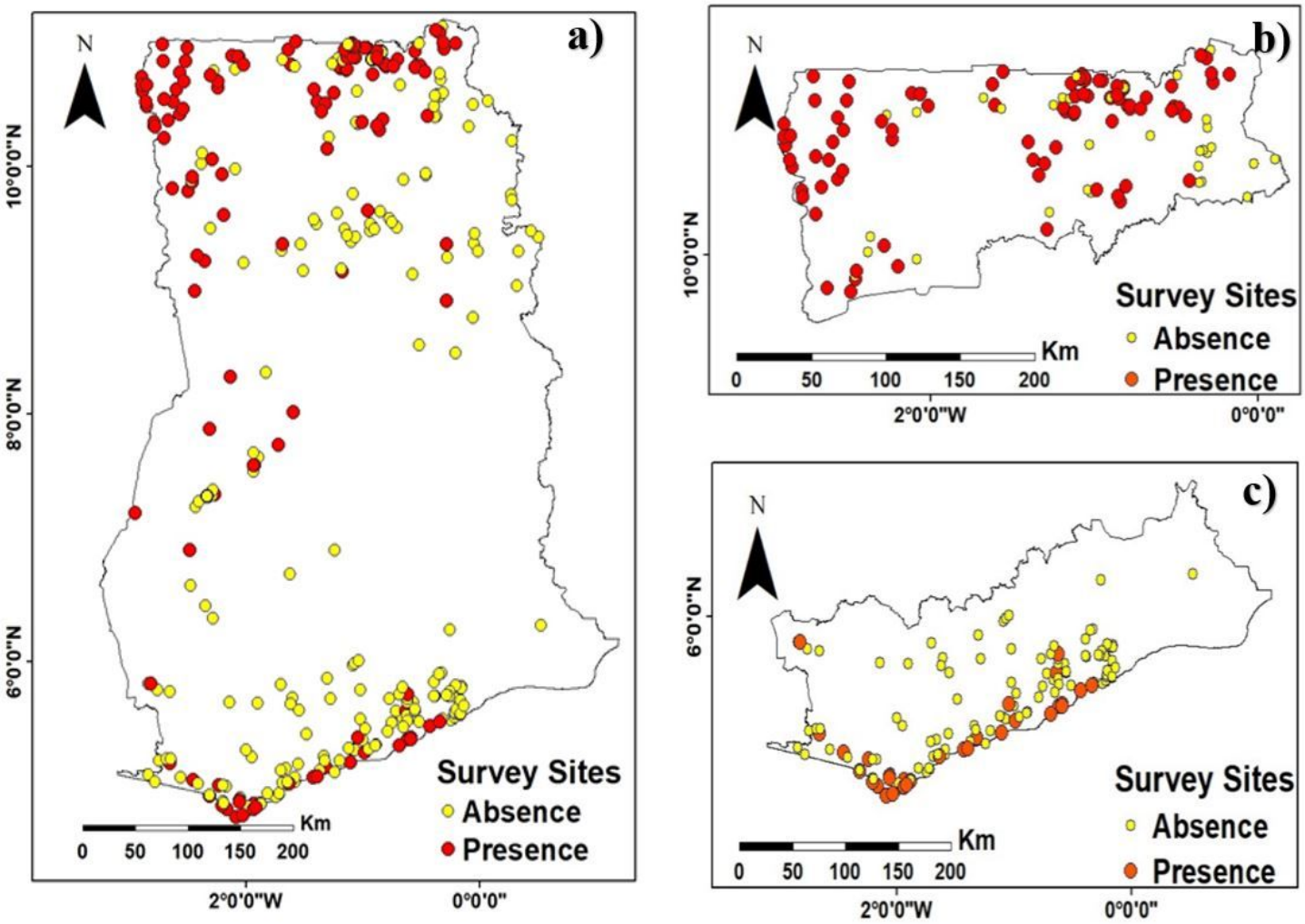

Figure 2

mf cases for surveyed communities from 2000 to 2014 (yellow indicates absence and red indicates presence), a) CW, b) NZ and c) SZ Zones.

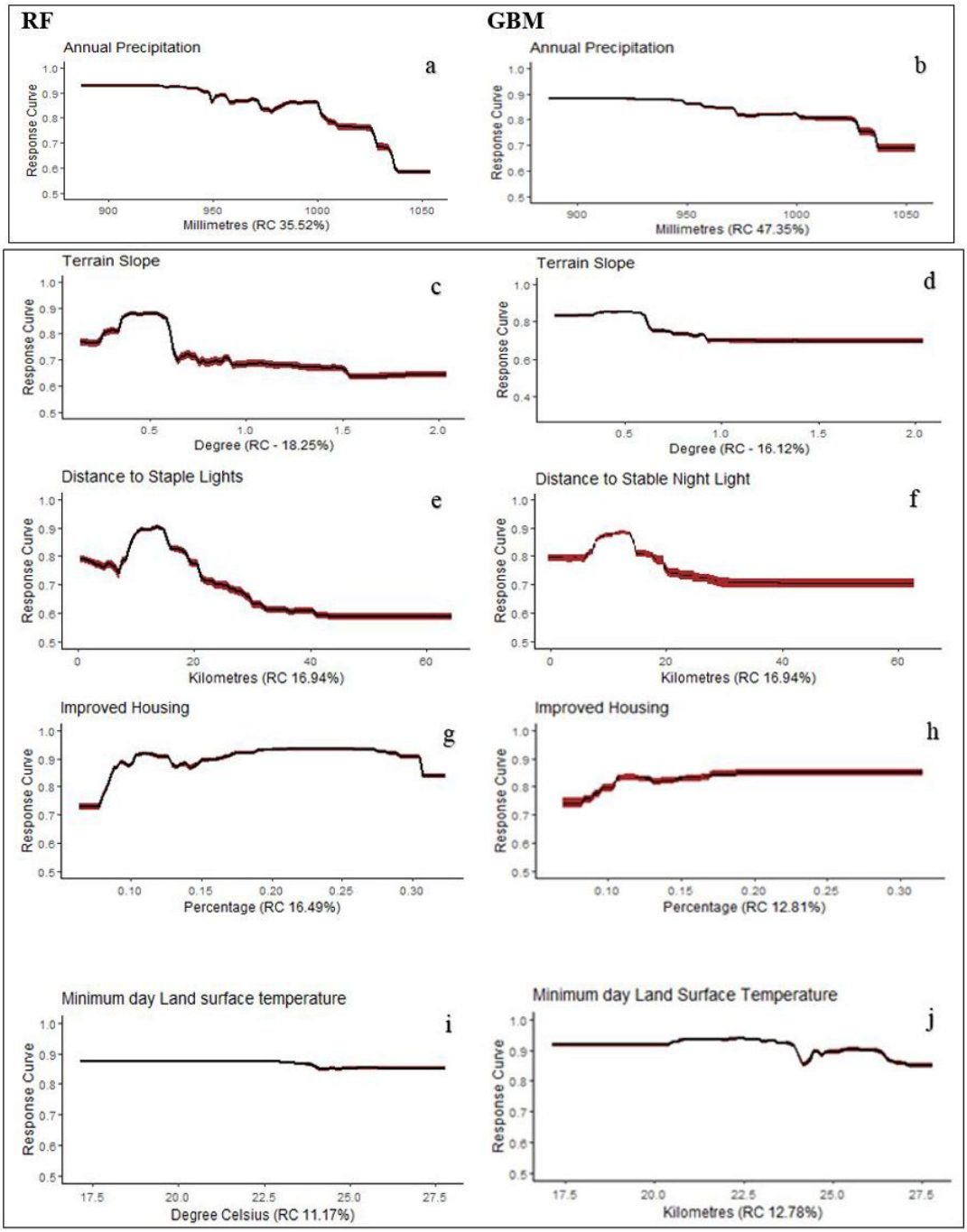

Figure 3

Response Curve for NZ using RF and GBM Models

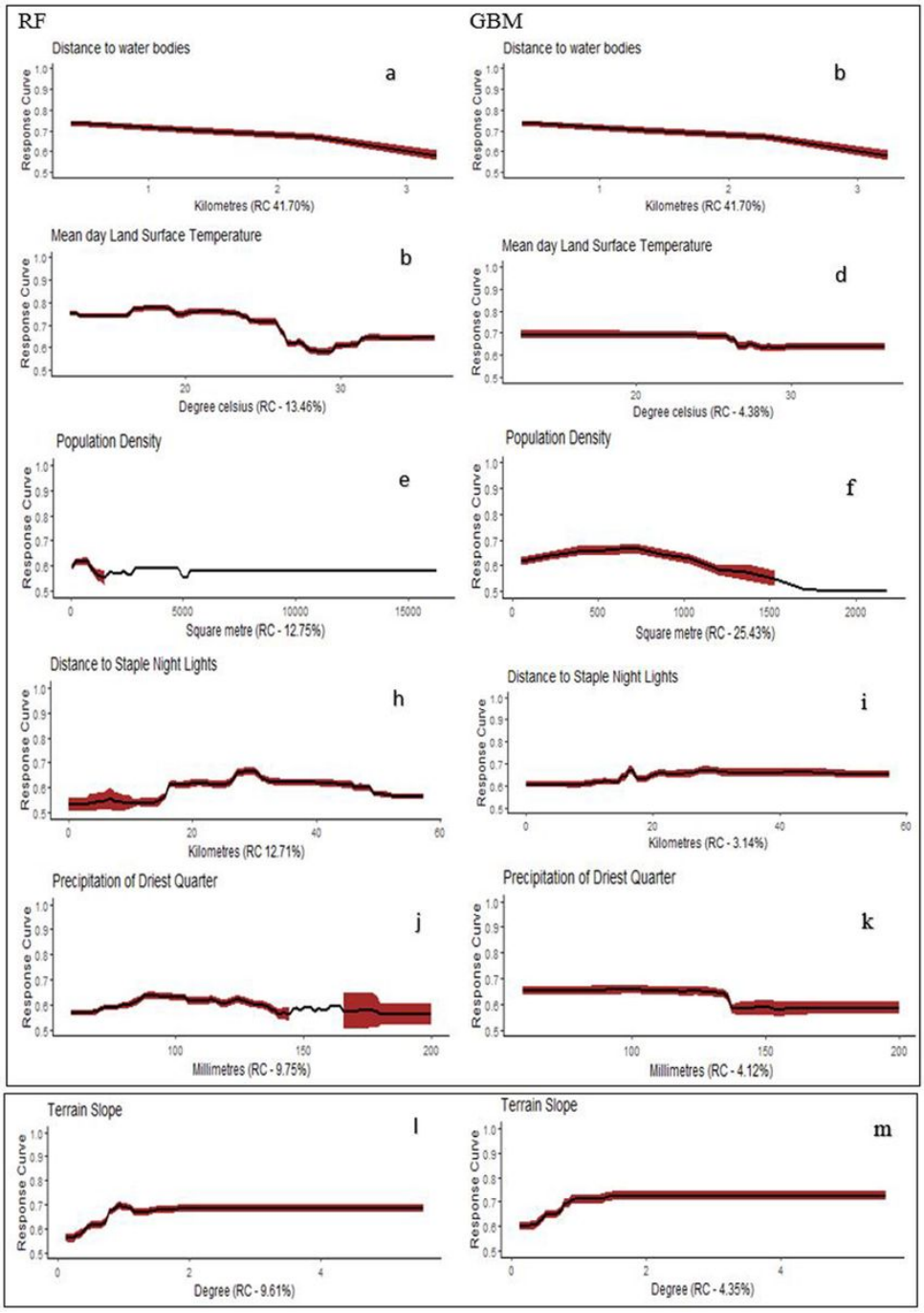

Figure 4

Response Curves for SZ using RF and GBM Models

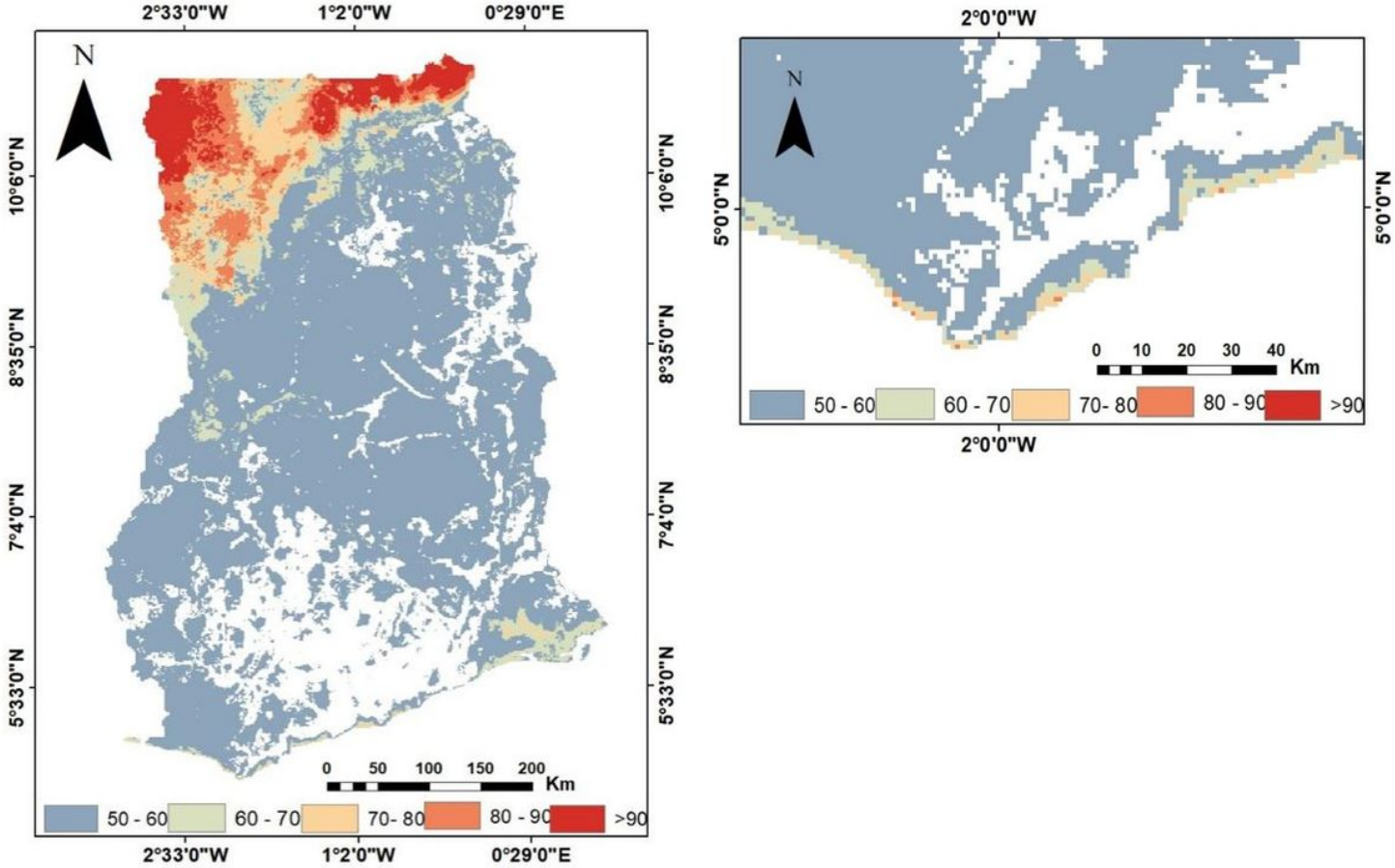

Figure 5

Probability maps of LF occurrence (%) generated with ensemble species distribution models, a) CW and b) coastline exaggerated for cartographic puposes. Only the areas with a probability ≥0.5 are presented. Probabilities of ≥0.8 were highlighted in red shades and represent most likely transmission areas.

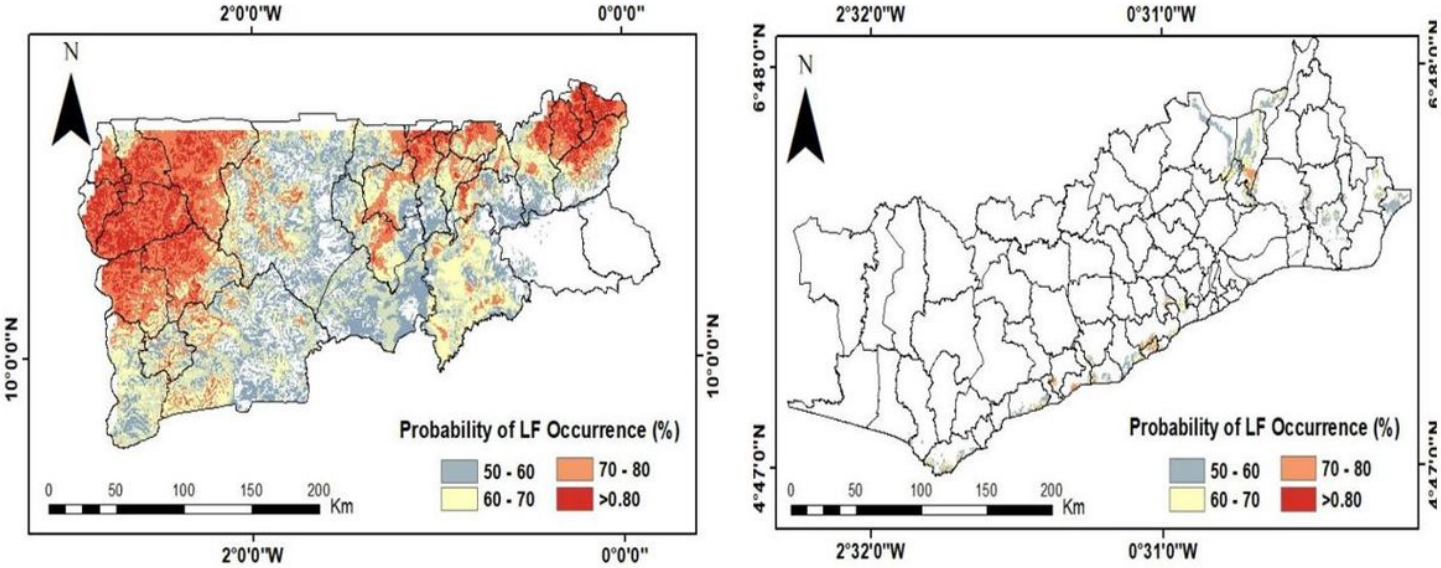

Figure 6

Probability maps of LF occurrence generated with ensemble species distribution models a) NZ and b) SZ. Only the areas with a probability ≥0.5 are presented. Probabilities of ≥0.8 were highlighted in red shades and represent most likely transmission areas.