



OPEN

The potential distribution of *Bacillus anthracis* suitability across Uganda using INLA

V. A. Ndolo^{1✉}, D. Redding², M. A. Deka⁵, J. S. Salzer⁵, A. R. Vieira⁵, H. Onyuth³, M. Ocaido³, R. Tweyongyere³, R. Azuba³, F. Monje⁴, A. R. Ario⁴, S. Kabwama⁴, E. Kisaakye⁴, L. Bulage⁴, B. Kwesiga⁴, V. Ntono⁴, J. Harris⁵, J. L. N. Wood¹ & A. J. K. Conlan¹

To reduce the veterinary, public health, environmental, and economic burden associated with anthrax outbreaks, it is vital to identify the spatial distribution of areas suitable for *Bacillus anthracis*, the causative agent of the disease. Bayesian approaches have previously been applied to estimate uncertainty around detected areas of *B. anthracis* suitability. However, conventional simulation-based techniques are often computationally demanding. To solve this computational problem, we use Integrated Nested Laplace Approximation (INLA) which can adjust for spatially structured random effects, to predict the suitability of *B. anthracis* across Uganda. We apply a Generalized Additive Model (GAM) within the INLA Bayesian framework to quantify the relationships between *B. anthracis* occurrence and the environment. We consolidate a national database of wildlife, livestock, and human anthrax case records across Uganda built across multiple sectors bridging human and animal partners using a One Health approach. The INLA framework successfully identified known areas of species suitability in Uganda, as well as suggested unknown hotspots across Northern, Eastern, and Central Uganda, which have not been previously identified by other niche models. The major risk factors for *B. anthracis* suitability were proximity to water bodies (0–0.3 km), increasing soil calcium (between 10 and 25 cmolc/kg), and elevation of 140–190 m. The sensitivity of the final model against the withheld evaluation dataset was 90% (181 out of 202 = 89.6%; rounded up to 90%). The prediction maps generated using this model can guide future anthrax prevention and surveillance plans by the relevant stakeholders in Uganda.

Anthrax is a zoonotic disease caused by *Bacillus anthracis*, a Gram-positive, soil-borne, and spore-forming bacterium. In some environmental conditions, *B. anthracis* spores can survive in the soil for years or even decades^{1,2}, making long-term control of the disease very difficult. The disease primarily affects both domestic and wild animals^{3,4}, mostly herbivores, and can be transmitted to humans when they touch, ingest, or inhale bacterial spores from infected animals, carcasses, or animal by-products⁵. Anthrax is estimated to cause about 20,000–100,000 human infections annually across the world⁶. Approximately 1.1 billion livestock live in areas predicted to be at risk for anthrax globally⁶.

Outbreaks can reduce the economic productivity of the agricultural sector, including dairy production, harm biodiversity, and wildlife conservation efforts, and threaten the food security for communities that consume meat⁷. Effective control of anthrax disease in livestock needs robust surveillance efforts, annual livestock vaccination, and timely outbreak response involving ring vaccination, safe disposal of carcasses, and awareness creation campaigns³. There is currently no policy in Uganda concerning the routine annual livestock vaccination against anthrax since the benefits of vaccination against the disease are considered a “private good,” so farmers have to buy vaccines and prophylactic antibiotics privately for their animals, usually during ongoing anthrax outbreaks³.

The use of Ecological Niche Model (ENMs) in the fields of conservation biology, epidemiology, and environmental health has increased in recent years. ENMs seek to estimate the association between species presence or absence with environmental variables to predict the probability of the species suitability in non-sampled

¹Disease Dynamics Unit, Department of Veterinary Medicine, University of Cambridge, Madingley Rd, Cambridge, Cambridgeshire, UK. ²Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK. ³College of Veterinary Medicine Animal Resources and Biosecurity, Makerere University, Kampala, Uganda. ⁴Uganda National Institute of Public Health, Ministry of Health, Kampala, Uganda. ⁵US Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA, USA. ✉email: valentinendolo@gmail.com

locations or time periods⁸. ENMs have been used in the past to identify “hotspots” or priority areas likely to be inhabited by disease-causing micro-organisms or vectors^{9–11}. It is vital to identify these areas correctly to institute appropriate interventions. Several methods have been created over the past years to build ENMs for *B. anthracis* suitability; these include regression models such as Generalized Linear Models (GLM), Climate Envelope Models (e.g., BIOCLIM), Maximum Entropy Models (e.g., MAXENT), Artificial Neural networks such as SPECIES, and Classification and Regression Trees (e.g., BIOMOD)^{12–16}. However, ecological datasets are complex, often requiring sophisticated statistical techniques to analyse them¹⁷. It is now standard practice to require explicit spatial dependence structures when developing models for complex and non-linear associations between the species and environmental variables and to estimate the sources of uncertainty due to the sampling methods, reporting biases, input data, and other analytical errors¹⁷. Conventional approaches are limited in their ability to accommodate complex hierarchical dependency structures needed to adjust for spatial autocorrelation within ecological data¹⁸.

Hierarchical Bayesian models can account for spatial autocorrelation and include spatial random effects that might capture spatial distributions¹⁹. Posterior predictive distributions obtained from Bayesian models usually require numerical techniques to approach them. Simulation-based methods such as Markov Chain Monte Carlo (MCMC) may be used, but are computationally intensive^{20,21} while the novel Integrated-Nested Laplace Approximation (INLA) approach²² offers a better alternative. INLA has gained popularity as a framework for modelling spatial and temporal data of disease-causing organisms (e.g., Rift-Valley Fever and Lassa Fever)^{23,24}, as it can produce improved predictions compared to the conventional ENM approaches¹⁸. However, INLA appears yet to be used to model the spatial distribution of *Bacillus anthracis*.

This study aims to predict the geographical distribution of the species suitability of *B. anthracis* across Uganda using wildlife, livestock, and human data collected from 2004 to 2018. We use a Generalized Additive Model (GAM) using the INLA Bayesian framework to quantify the relationships between the occurrence of *B. anthracis* infection and the environment. Our dataset consolidates a national database of anthrax case records across Uganda built across multiple sectors bridging human and animal partners using a One Health approach.

Methods

Study setting. Uganda covers 241,037 km² (4° N, 2° S, 29° E, 35° E) and averages 1100 m above sea level²⁵. Uganda has a total population of 34.6 million people, with 7.4 and 27.2 million living in urban and rural areas, respectively²⁶. It borders Lake Victoria and has an equatorial climate. The study area is mainly plateau, with a few mountains. About 20 percent of Uganda’s area is covered by swamps and water bodies, including the four Great Lakes of East Africa (Lake Edward, Lake Victoria, Lake Albert, and Lake Kyoga). The country has ten national parks housing a high diversity of wildlife and endangered species²⁶. By 2009, agriculture ranked as the second leading contributor to the country’s Gross Domestic Product²⁷.

Surveillance data of *Bacillus anthracis* infection. Surveillance data of livestock and human cases from 2018 were provided by the Ministry of Health (MOH) in Uganda through the Field Epidemiology and Laboratory Training Program and the Uganda Ministry of Agriculture, Animal Industry and Fisheries (MAAIF). Geographical coordinates of anthrax cases among wildlife (from 2004 to 2010) in Queen Elizabeth National Park were obtained from a recently published study⁴. These cases were mapped to show the distribution of anthrax across Uganda (Fig. 1). It is only recently that Uganda has mandated systematic anthrax surveillance across the country following the outbreak that started in 2018. GPS coordinates gathered by field personnel during outbreak responses were used to map outbreak locations. The human anthrax cases were defined based on the CDC’s clinical criteria (signs and symptoms), presumptive laboratory diagnosis (Gram staining), and confirmatory laboratory diagnosis (bacterial culture, immunohistochemistry, ELISA, and PCR)²⁸. The animal anthrax cases were also defined based on the clinical presentation, presumptive laboratory diagnosis, and confirmatory laboratory diagnosis. All cases were classified as either ‘probable’ or ‘confirmed,’ with probable defined as cases that met both the clinical and presumptive laboratory diagnostic criteria and confirmed defined as cases that met the clinical and confirmatory laboratory diagnostic criteria. A total of 497 livestock (n = 171), humans (n = 32), and wildlife cases (n = 294), both confirmed (n = 32) and probable (n = 465), occurring from 2004 to 2018 were compiled. All methods were performed in accordance with the relevant guidelines and regulations.

Environmental variable processing. Correlative studies of environmental risk factors for anthrax outbreaks suggest that temperature^{6,29–37}, precipitation^{6,29–39}, elevation^{6,29,31,32,34,35,37–39}, soil (type, calcium concentration, pH, carbon content, and moisture)^{6,30,31,33,34,36,37,39–43}, vegetation^{6,29,31,34,36–40,42,43}, and hydrology^{37,44} are some of the major drivers of *B. anthracis* suitability. A total of 26 environmental predictors (Fig. 2) were selected for this study based on these known variables. These comprised 19 bioclimatic variables (the mean for the years 1970–2000) collected from the WorldClim database version 2 (<https://www.worldclim.org/data/worldclim21.html>) at a resolution of 30 s (~1 km²)⁴⁵. Four soil variables, including exchangeable calcium at a depth of 0–20 cm, soil water availability, soil pH (10×) in H₂O at a depth of 0 cm, and soil organic carbon at a depth of 0–5 cm, were retrieved from the International Soil Reference and Information Centre (ISRIC) data hub at a resolution of 250 m (<https://data.isric.org/geonetwork/srv/eng/catalog.search#/home>). Distance to permanent water bodies was derived from a global hydrology map provided by ArcGIS online version 10.6.1⁴⁶. Elevation data of 1 km² in resolution was obtained from the Global Multi-resolution Terrain Elevation Data (GMTED2010) dataset available from the United States Geological Service. Finally, the monthly Enhanced Vegetation Index (EVI) data for the years 2004, 2005, and 2010 (36 tiles in total) were obtained from The Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MYD13A3 v.6) at a spatial resolution of 1 km (<https://lpdaac.usgs.gov/products/myd13a3v006/>). The single variable, mean EVI, was calculated in QGIS by averaging

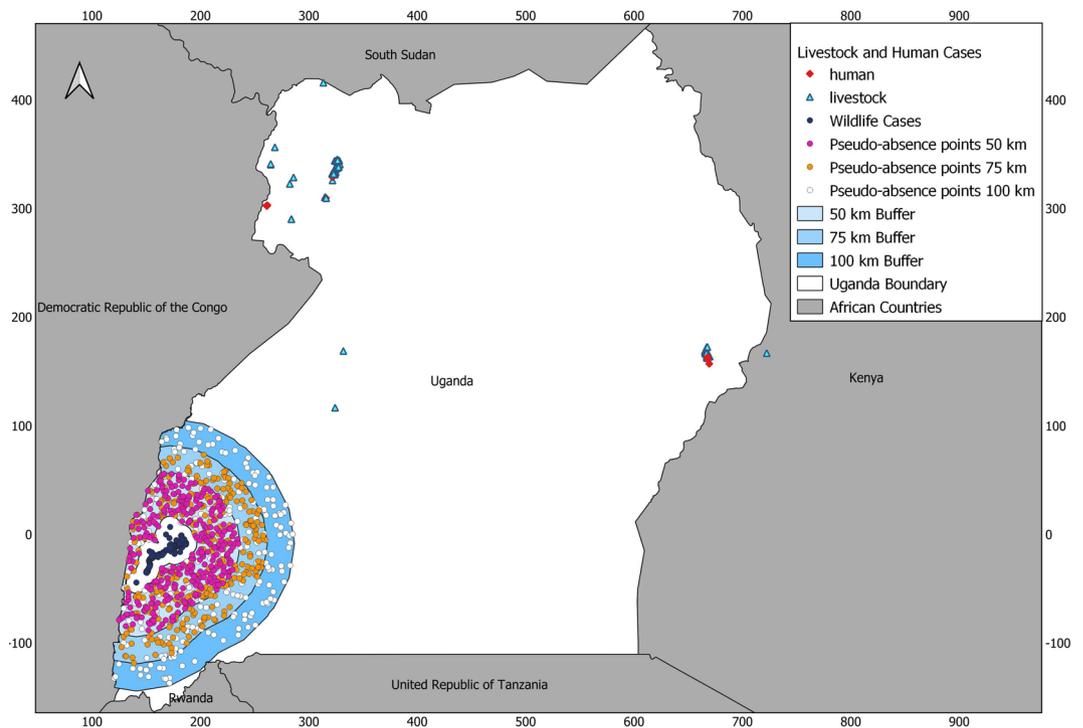


Figure 1. Distribution of anthrax presence and pseudo-absence locations across Uganda. The navy-blue circles show wildlife cases ($n = 294$) used for model training, the blue triangles show livestock cases ($n = 171$), and the red diamonds represent human cases ($n = 32$). The blue polygons show the locations of the 50 km, 75 km, and 100 km buffers which were constructed around the wildlife cases with a distance of 10 km between the buffers and the presence locations. The pink dots show the pseudo-absence points selected within the 50 km buffer, the orange dots show the pseudo-absence points selected within the 75 km buffer, and the white dots show the pseudo-absence points selected within the 100 km buffer. Prediction maps were developed using the Quantum Geographic Information System software (QGIS). URL: <https://qgis.osgeo.org> (2020).

all 36 tiles. The EVI minimizes variations in the canopy background and maintains precision over conditions with dense vegetation.

All environmental variables were resampled using the R package ‘Resample’⁴⁷ to a resolution of 1 km and clipped to the extent of Uganda. Since data sampling for wildlife data (used in model training) was not done systematically across the study area, target background buffers (polygon buffers created at certain radii from the training points and used for the random selection of pseudo-absences) were created for model calibration to reduce sampling bias. As there was no information on the sampling radius, a sensitivity analysis was conducted by creating target backgrounds using circular buffers of radius 50 km, 75 km, and 100 km around the presence points used for model training, leaving 10 km between the presence points and the various buffers (Fig. 1). Quantum Global Information System (QGIS) version 3.16 (<https://qgis.org>) software was then used to add 294 random pseudo-absence points within each buffer polygon giving a ratio of 1:1 for the training presences to pseudo-absences. A recent study explored how four approaches of pseudo-absence creation affect the performance of models across different species and three model types by building both terrestrial and marine models using boosted regression trees, generalised additive mixed models, and generalised linear mixed models⁴⁸. They then tested four methods for generating pseudoabsences across all the different model types: (1) correlated random walks (RW); (2) reverse correlated RW; (3) sampling pseudoabsences within a buffer area surrounding the presence points; (4) background sampling⁴⁸. The findings of the study suggested that the separation or distance in the environmental space between the presence locations and the pseudoabsences was the most significant driver of the model predictive ability and explanatory power, and thus finding was consistent across the three model types (boosted regression trees, generalised additive mixed models, and generalised linear mixed models) and both the terrestrial and marine habitats⁴⁸.

The values of the environmental variables were then extracted for each presence and pseudo-absence location using the raster package in R. With these, we did an initial data exploration to check for outliers within the covariates, collinearity, and to explore the relationships between the covariates and the response variables (presence or absence of anthrax). Cleveland dot plots were used to check for possible outliers. Following the outlier checks, variance inflation factors (VIF), pairwise plots, and Pearson correlation coefficients with correction for multiple comparisons were used to measure the statistically significant correlation between the covariates (Fig. 2). For variables that were highly correlated (correlation greater than 0.6) or those with high variance inflation ($VIF > 3$), only one was used in the modelling process. Five variables were selected following this analysis: Temperature seasonality (BIO4), elevation, distance to water, soil calcium, and soil water (Table 1).

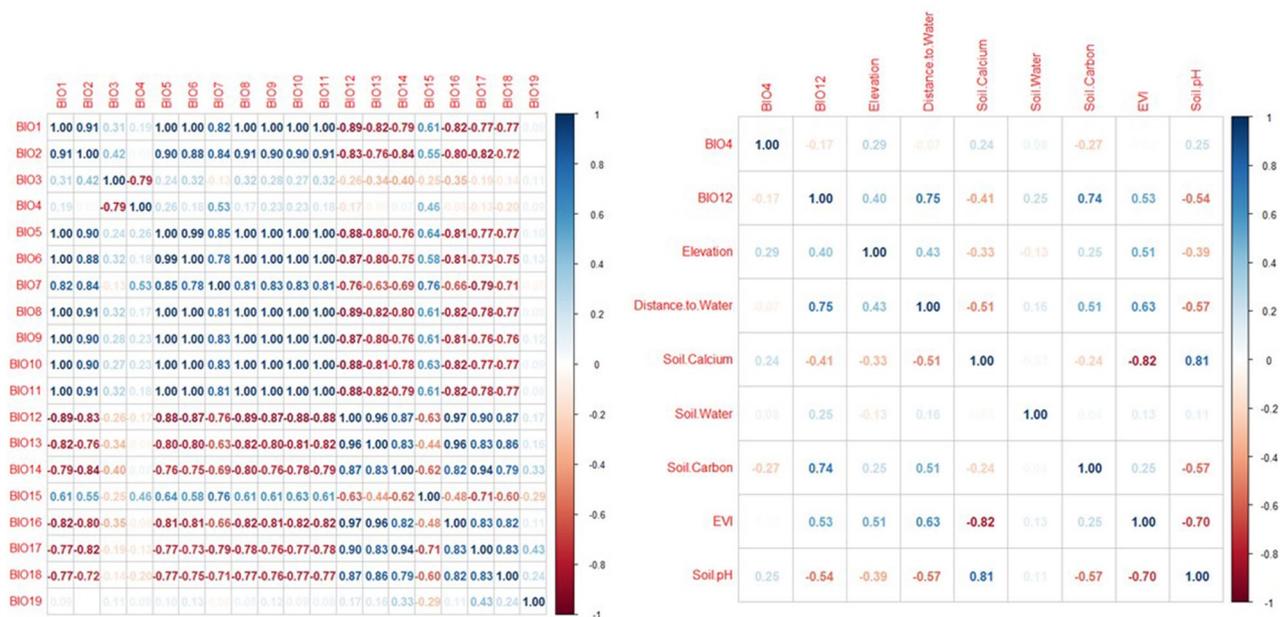


Figure 2. Results of correlation between covariates using Pearson's correlation test. Correlation between covariates was shown by red numbers (negative correlation) and blue numbers (positive correlation). BIO1 = Annual Mean Temperature, BIO2 = Mean Diurnal Range, BIO3 = Isothermality, BIO4 = Temperature Seasonality, BIO5 = Max Temperature of Warmest Month, BIO6 = Min Temperature of Coldest Month, BIO7 = Temperature Annual Range, BIO8 = Mean Temperature of Wettest Quarter, BIO9 = Mean Temperature of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter, BIO11 = Mean Temperature of Coldest Quarter, BIO12 = Annual Precipitation, BIO13 = Precipitation of Wettest Month, BIO14 = Precipitation of Driest Month, BIO15 = Precipitation Seasonality, BIO16 = Precipitation of Wettest Quarter, BIO17 = Precipitation of Driest Quarter, BIO18 = Precipitation of Warmest Quarter, BIO19 = Precipitation of Coldest Quarter.

Modelling anthrax suitability across Uganda. QGIS v.3.16 (<https://qgis.org>) and the R statistical package version 4.1.0⁴⁹ were used to conduct data visualization, cleaning, and model analysis (R code used available in a Github repository: https://github.com/valentinandolo/Uganda-Spatial_Model/tree/master). The wildlife cases (n = 294) were used for model training and testing. The remaining human (n = 32) and livestock (n = 171) cases were used for model evaluation. Since the wildlife case locations were recorded from 2004 to 2010, while the livestock and human cases were recorded in 2018, the latter locations were both spatially and temporally distinct from the wildlife cases, making an excellent basis for block cross-validation of the final model performance⁵⁰. Random partitioning of the data into training and testing sets can inflate the performance of a model and underestimate the error in the spatial prediction evaluation⁵⁰. Block cross-validation uses spatial blocks that separate the testing and training datasets; thus, the method has been suggested to be a good technique for error estimation and a robust approach for measuring a model's predictive performance⁵⁰.

The INLA package in R was applied to model the suitability of *B. anthracis* across Uganda. INLA calculates the spatial interaction effects using a Stochastic Partial Differential Equation (SPDE) method, which estimates a continuous Gaussian Markov Random Field (GMRF) where the correlation between two locations in space is specified by the Matérn correlation which is explained in more detail elsewhere⁵¹. The initial step in fitting an SPDE model is the creation of a Constrained Refined Delaunay Triangulation or a mesh to illustrate the spatial process⁵¹. R-INLA uses a function called 'inla.mesh.2d()' that applies a variety of arguments to build the mesh, these include: *loc*, *loc.domain*, *boundary*, *max.edge*, and *cutoff*⁵¹. The *loc* argument contains the point locations which provide information about the area of study and are used to create the triangulation nodes. Alternatively, a polygon of the study area can be used to identify the extent of the domain via *loc.domain*. We applied the point locations using the *loc* argument. We then specified the boundary of the mesh as a convex hull. We used the *max.edge* argument to specify the maximum edge length for the inner mesh domain/triangles and the outer triangles. We did this by first studying the distribution of distances between the point locations for the training presences and pseudo-absences. Most points were within a distance of about 90–100 km away from each other, thus, a possible guess for the range at which spatial autocorrelation persists was 100 km. We used a distance of 20 km as a range guess to create a finer mesh which has been shown to produce more precise models. We specified the maximum edge length for the inner triangles as 20 km divided by 5 (4 km) and the maximum edge length for the outer triangles as 20 km. Finally, the *cutoff* argument sets the minimum distance allowed between point locations. We divided the maximum edge length for the inner triangles by 5 (4 km divided by 5 = 0.8 km), meaning that points that were closer in distance than 0.8 km were replaced by one vertex to avoid the occurrence of small triangles.

The spatial effect, which is a numeric vector, then links each observation within the data to a spatial location, thus, accounting for region-specific variation that cannot be accounted for by the covariates. Following the

Variable name	Units	Spatial resolution	Source	Year
BIO1 = Annual Mean Temperature,	°C	30 arc-s	WorldClim database version 2 (https://www.worldclim.org/data/worldclim21.html) (36)	The average for 1970–2000
BIO2 = Mean Diurnal Range,				
BIO3 = Isothermality,				
BIO4 = Temperature Seasonality				
BIO5 = Max Temperature of Warmest Month				
BIO6 = Min Temperature of Coldest Month				
BIO7 = Temperature Annual Range				
BIO8 = Mean Temperature of Wettest Quarter				
BIO9 = Mean Temperature of Driest Quarter				
BIO10 = Mean Temperature of Warmest Quarter				
BIO11 = Mean Temperature of Coldest Quarter				
BIO12 = Annual Precipitation	ml	30 arc-s	WorldClim database version 2 (https://www.worldclim.org/data/worldclim21.html) (36)	The average for 1970–2000
BIO13 = Precipitation of Wettest Month				
BIO14 = Precipitation of Driest Month				
BIO15 = Precipitation Seasonality				
BIO16 = Precipitation of Wettest Quarter				
BIO17 = Precipitation of Driest Quarter				
BIO18 = Precipitation of Warmest Quarter				
BIO19 = Precipitation of Coldest Quarter				
Elevation (m)	m	1 km	Global Multi-resolution Terrain Elevation Data (GMTED2010) dataset available from the United States Geological Service	2010
Distance to permanent water bodies (km)	km	1 km	Derived from a global hydrology map provided by ArcGIS online version 10.6.1	2019
Mean Enhanced Vegetation index (EVI)	Units	1 km	The Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MYD13A3 v.6) (https://lpdaac.usgs.gov/products/myd13a3v006/)	Average of 36 tiles from the years 2004, 2005, and 2010
Soil exchangeable calcium (depth of 0–20 cm)	cmolc/kg	250 m	International Soil Reference and Information Centre (ISRIC) data hub (https://data.isric.org/geonetwork/srv/eng/catalog.search#/home)	
Soil water availability	v%			
Soil pH ($\times 10$) in H ₂ O (depth of 0 cm)	Units (1–14)			
Soil organic carbon (depth of 0–5 cm)	Dg/kg			

Table 1. Summary of the environmental variables used.

recommendation by Lindgren and Rue⁵², multivariate Gaussian distributions with means of zero and a spatially defined covariance matrix were used to model the spatial effect. Several versions of Bayesian hierarchical additive models were created by estimating a Bernoulli generalized additive model (GAM) with and without spatially correlated random effects. The Bernoulli GAM is defined as shown in Eqs. (1) and (2)

$$C_i \sim \text{Bernoulli}(p_i), \quad (1)$$

$$\text{logit}(p_i) = \alpha + \sum_{j=1}^m \beta_j(X_{j,i}) + \sum_{k=1}^l f_k(X_{k,i}) + \mu_i, \quad (2)$$

where C_i denotes the observed value, such that: *B. anthracis* presence or absence at a given location i ($i = 1, \dots, n$; $n = 588$) is given as C_i , where $C_i = 1$ if *B. anthracis* was present, and $C_i = 0$ if absent. Logit is the link function for binomial family, p_i is the expected value of the response variable (the probability of *B. anthracis* suitability) at location i , α is the intercept, $X_{j,i}$ and $X_{k,i}$ are the j th and the k th covariates at a location i , β_j are the beta coefficients, f_k are the smooth functions (cubic regression splines) for k th covariates, and μ_i is the spatial random effect at the location i ⁵³. The number of variables in linear term (m) and the non-linear term (l) are different because the variables employed in each term are different. We estimated both linear and non-linear effects for the covariates. Our overall database had 294 *B. anthracis* pseudo-absences generated randomly across the target background buffers to match the number of species presences recorded. Because we had no prior information, a Gaussian prior distribution with a mean of zero (default no effect prior unless data is informative) was applied for all the model parameters. The posterior mean, standard deviation, and 95% credible intervals were estimated for all the parameters.

Model selection. Several different candidate models were examined. First, a baseline model was built using only the intercept. A second baseline model was then built, which included the intercept and spatial random

effects only. Covariates were added to the second baseline model (intercept plus spatial effects model) without any smoothing function (i.e., only linear effects). The contribution of the spatial random effect was then re-examined by taking it out from the model. Smoothing functions were then added to all covariates, and the same procedure was repeated. Model selection was done using this forward stepwise approach. The final model was run using the three different target background buffers to identify the buffer distance with the lowest Deviance Information Criterion (DIC). The various options were assessed using the DIC⁵⁴, Watanabe-Akaike information criterion (WAIC), and the Conditional Predictive Ordinate (CPO). The DIC and WAIC were chosen because they are commonly used to assess model performance by measuring the compromise between the goodness of fit and complexity. For CPO, the logarithmic score ($-\text{mean}(\log(\text{CPO}))$) (LCPO) was calculated and used⁵⁵. CPO can also be used to conduct internal cross-validation of models using a leave-one-out approach to evaluate the predictive performance of the model. Lower LCPO, DIC, and WAIC estimates suggest superior model performance. Thus, the favoured model had the lowest values across the 3 metrics.

Model validation and evaluation. Model validation for the favoured model was conducted using an independent evaluation dataset comprising of livestock and human outbreaks occurring at different spatial locations and 8 years after the training data. The omission rate, which indicates the proportion of positive test locations that end up in pixels predicted to be unsuitable for *B. anthracis*⁵⁶, was used to validate the model. A low omission rate indicates good model performance. The threshold for suitability was the probability threshold that maximized the sensitivity and specificity of the model. The sensitivity was then derived by calculating the proportion of positive test locations that end up in pixels predicted to be suitable for *B. anthracis*⁵⁶. A high sensitivity indicates good model performance.

Model prediction. The favoured model selected using the criteria described above was used to generate countrywide prediction maps showing the posterior mean values, standard deviation, and the 95% credible intervals of the probability of *B. anthracis* suitability. The raster package in R was used to make the prediction maps. Bayesian kriging was done by treating all model parameters as random variables to include uncertainty in the prediction⁵⁷. This kriging is built into the INLA framework via the SPDE, which allows a Delaunay triangulation to be constructed around the presence and absence locations within the sampling frame⁵². INLA then conducts model inference and prediction at the same time by considering the prediction points as locations missing the response variable (set to NA)⁵¹. Following successful model prediction, additional linear interpolation functions then generate the output for the whole study area scaled from 0 to 1.

Ethical approval. Ethical approval for this study was provided by the Human Biology Research Ethics Committee, University of Cambridge, UK (Ref: HBREC.2019.02) the School of Veterinary Medicine and Animal Resources Institutional Animal Care and Use Committee, Makerere University, Uganda (Ref: SVAREC/21/2019); and the Uganda National Council for Science and Technology. Informed consent was obtained from all subjects and/or their legal guardian(s). All methods were performed in accordance with the relevant guidelines and regulations.

Results

A total of 497 anthrax cases comprising livestock ($n = 171$), humans ($n = 32$), and wildlife cases ($n = 294$), occurring from 2004 to 2018 were compiled (Fig. 1). Of these, 32 cases were classified as confirmed and 465 cases were probable. The wildlife cases were used to train and calibrate the model. Table 2 shows the results of the candidate models created across the different target background buffers. The 75 km background buffer radius gave the best results in terms of lower DIC values, so it was selected as the final background extent for model calibration. The baseline prediction model that incorporated the spatial random effect had a much lower DIC than the baseline model without it (Table 2) and addition of linear covariates reduced it further. The chosen best model (option 17 in Table 2) had the lowest DIC, WAIC and LCPO compared to the rest of the models. This final model applied a non-linear effect on elevation and linear effects on distance to water, soil calcium, and soil water (Table 2). The spatial random effect was not included in the model.

The posterior predicted mean probability of *B. anthracis* suitability and the 95% credible intervals for the parameters of the fixed effects used in the final INLA model are shown in Fig. 3. The results demonstrated a negative relationship between distance to water and the presence of *B. anthracis* between 0 and 0.3 km. However, soil calcium had a positive association with the occurrence of *B. anthracis* between 10 and 25 cmolc/kg. Similarly, the results showed that higher occurrences of *B. anthracis* are expected in areas with soil water availability of between 20 and 30 v% (the volumetric soil water content defined as the volume of water for each unit volume of soil) and elevation of 140–190 m (Fig. 3). The omission rate of the final model against the withheld evaluation dataset was 10%, indicating a sensitivity of 90% (181 out of 202 = 89.6%; rounded up to 90%). Only 21 locations out of the 202 positive test locations were omitted (predicted to be unsuitable) by the model projection, however, these locations were all within 50 km of the nearest predicted high-risk area. The recent study of the environmental factors influencing the distribution of anthrax across Queen Elizabeth National Park in Uganda had 5.2% omission rate, correctly predicting 94.8% of the test points applied for model validation³³. The other recent study of the distribution of *B. anthracis* across Africa had omission rates of 3.4%, 10%, and 5.9% when no thinning, 30 km thinning, and 50 km thinning was applied to the dataset⁵⁸.

The prediction maps from the INLA model identified several districts across Western, Central, Eastern, and Northern Uganda to be suitable for the occurrence of *B. anthracis* as shown in Table 3. These areas were identified to be regions that are likely to be suitable for the occurrence and persistent of *B. anthracis* (Fig. 4). The spatial effect which shows the intrinsic spatial variability of the observed data after excluding the environmental

Buffer radius	Model option	Type	Variables	DIC	WAIC	CPO
50 km	1	Linear	Baseline	817	817	0.69
	2	Linear	Baseline + μ	187	181	0.15
	3	Linear	BIO4 + Distw + Elev + Ca + Swater + μ	85	85	0.07
	4	Linear	BIO4 + Distw + Elev + Ca + Swater	83	84	0.07
	5	Linear	Distw + Elev + Ca + Swater	81	81	0.07
	6	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater})$	78	118	1.75
	7	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater}) + \mu$	120	263	5.29
	8	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater	60	61	0.05
	9	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater + μ	63	64	0.06
75 km	10	Linear	Baseline	817	817	0.69
	11	Linear	Baseline + μ	187	182	0.15
	12	Linear	BIO4 + Distw + Elev + Ca + Swater + μ	60	75	1.93
	13	Linear	BIO4 + Distw + Elev + Ca + Swater	50	60	0.08
	14	Linear	Distw + Elev + Ca + Swater	49	53	0.07
	15	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater})$	108	301	2.20
	16	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater}) + \mu$	184	604	9.79
	17	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater	44	47	0.05
	18	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater + μ	53	63	2.95
100 km	19	Linear	Baseline	817	817	0.69
	20	Linear	Baseline + μ	176	170	0.14
	21	Linear	BIO4 + Distw + Elev + Ca + Swater + μ	95	96	0.08
	22	Linear	BIO4 + Distw + Elev + Ca + Swater	95	96	0.08
	23	Linear	Distw + Elev + Ca + Swater	92	92	0.08
	24	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater})$	77	161	5.16
	25	Non-linear	$f(\text{Distw}) + f(\text{Elev}) + f(\text{Ca}) + f(\text{Swater}) + \mu$	145	613	9.50
	26	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater	85	85	0.07
	27	Non-linear	Distw + $f(\text{Elev})$ + Ca + Swater + μ	86	86	0.08

Table 2. Model comparison of the fitted INLA models using linear and non-linear associations across three background buffer radii. The criteria for comparison are DIC Deviance Information Criterion, WAIC Watanabe-Akaike Information Criterion, LCPO Conditional Predictive Ordinate, Variables acronyms are *Distw* Distance to water, *Ca* Soil calcium, *Elev* Elevation, *Swater* Soil water, *BIO4* Temperature Seasonality, μ Spatial effect.

predictors, showed a low (almost negligible) effect and was, therefore, excluded from the final model, implying that the variability of the species data for *B. anthracis* appears to be explained by the chosen variables used in the final model. Figure 5 is a binary map created using the threshold for maximum sensitivity and specificity. It shows areas that are likely to be suitable for anthrax and those that are not.

Discussion

This study uses a Bayesian framework to model the suitability of *B. anthracis* using wildlife data from Queen Elizabeth National Park, as well as livestock and human data from the MAAIF and MOH in Uganda. We demonstrate the INLA Bayesian framework to be a method that can predict hotspots of disease-causing organisms and inform prevention and surveillance strategies. ENMs have gained rapid popularity as powerful disease epidemiology tools^{23,24}. A major preliminary step to reducing anthrax outbreaks is to identify and prioritize areas that are suitable for the occurrence of *B. anthracis*³³. The correct detection of such areas can enhance effective management strategies for anthrax prevention. However, this requires a detailed understanding of the spatial distribution of *B. anthracis*, given that misidentification of anthrax “hotspots” can cause erroneous outbreak prevention practices leading to considerable economic losses.

The Bayesian model results show that the suitability of *B. anthracis* is influenced by soil properties, topography, as well as proximity to permanent water bodies. The non-linear relationship between elevation and species distribution demonstrated by the models implies that *B. anthracis* may occur in areas with distinct environmental properties but displaying higher preferences for areas with lower elevation (around 140–190 m). The linear relationships suggest that *B. anthracis* may prefer areas closer to water bodies (between 0 and 0.3 km). These results concur with the findings from a recent study on *B. anthracis* suitability conducted in Kruger National Park in South Africa which showed that low altitude (between 225 and 280 m) and close proximity to a water bodies were essential for survival and persistence of anthrax³⁷. Soil water availability had a weak positive effect on *B. anthracis* suitability. This variable, sometimes referred to as soil moisture has been shown to be an important predictor of *B. anthracis* suitability in Southern Kenya³⁶.

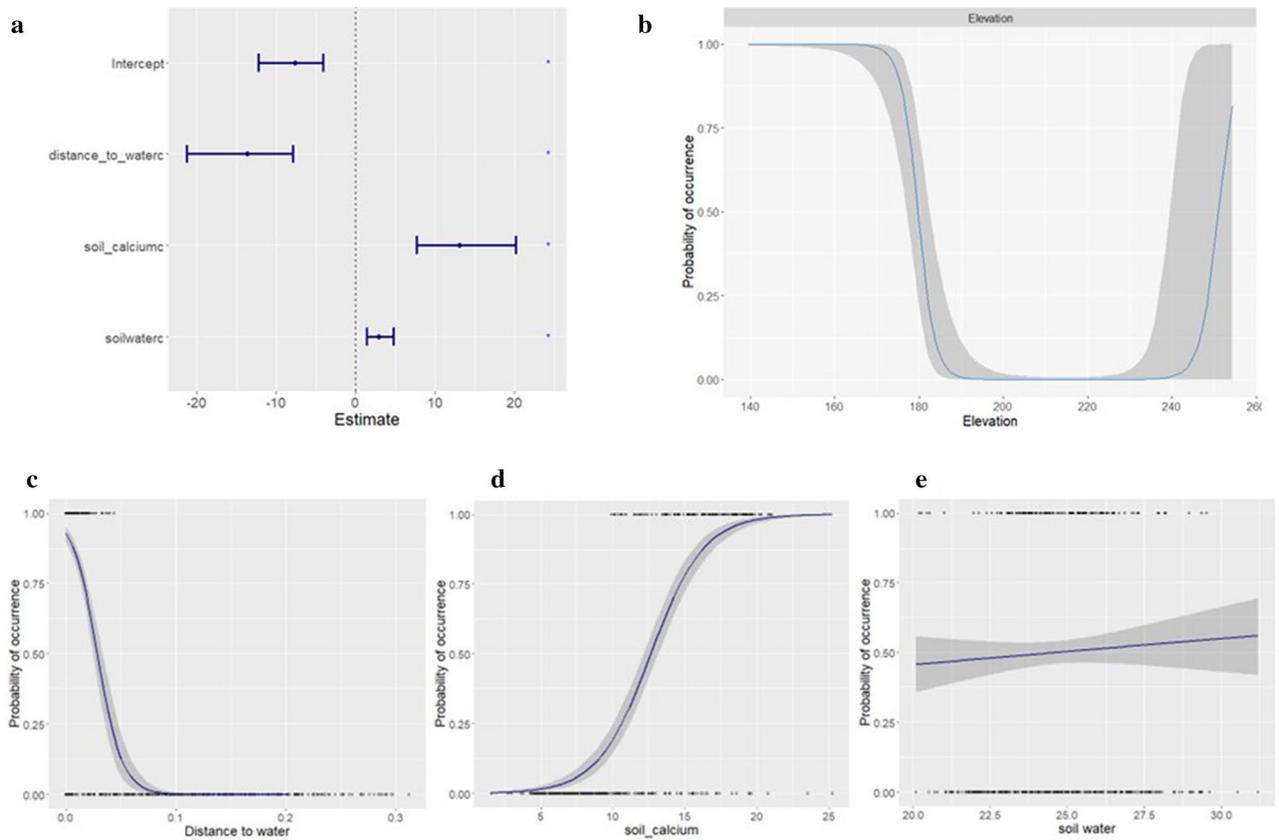


Figure 3. (A) The fixed effect and credible intervals for the linear covariates, (B) the smoothed fits of elevation (in m in x-axis) and linear fits of: (C) distance to water (in km in x-axis), (D) soil calcium (in cmolc/kg in x-axis), (E) soil water (in v% in x-axis) variables. The y-axis shows the estimated probability of presence. The shaded grey polygons represent 95% credible intervals.

Region	Districts
Western	Kasese, Rubirizi, Kitagwenda, Rukungiri, Kanungu, Kamwenge, Bunyangabu, Kabarole, Kazo, Kiruhura, Isingiro, Mbarara, Kabale, Kisoro, Kyegegwa, Bundibugyo, Kyenjojo, Ntoroko, Kagadi, Kakumiro, Kikuube, Hioma, Masindi, Buliisa, Kiryandongo
Central	Kyotera, Rakai, Lwengo, Lyantonde, Ssembabule, Mubende, Gomba, Kalungu, Kassanda, Mityana, Mpigi, Wakiso, Mukono, Kampala, Kayunga, Luwero, Nakaseke, Kiboga, Nakasongola, Kyankwanzi
Eastern	Kapchorwa, Kween, Bulambuli, Jinja, Kamuli, Buyende, Iganga, Luuka, Kaliro, Budaka, Kibuku, Pallisa, Ngora, Katakwi, Kapelebyong, Soroti, Serere, Kalaki, Kaberamaido
Northern	Kitgum, Kotido, Abim, Agago, Napak, Moroto, Nabilatuk, Nakapiripirit, Omoro, Oyam, Kole, Apak, Lira, Kwania, Dokolo, Abletong, Nwoya, Pakwach, Nebbi, Arua, Yumbe, Madi, Okollo, Obongi, Adjumani, Moyo, Amuru, and Lamwo

Table 3. Districts across Uganda with moderate to high *B. anthracis* suitability by region.

High soil calcium concentrations (10–25 cmolc/kg) also had a significant positive effect on the suitability of and area to *B. anthracis*. A recent study of *B. anthracis* suitability in Queen Elizabeth National Park in Uganda reported exchangeable calcium to be an important contributor in the creation of a species survival model for the bacteria³³. Steenkamp et al.³⁷ also reported that high calcium concentration was a significant driver of anthrax occurrence in South Africa. A growing body of evidence supports the fact that calcium has an important role in the preservation of anthrax spores in the soil⁵⁹. Studies have shown that calcium can be absorbed from the environment by the bacteria during spore formation and incorporated into the layers of the bacterial spore, particularly the core^{60,61}. Calcium then combines with dipicolinic acid to form a salt lattice that helps to stabilize the enzymes and DNA in the spore, hence, helping the bacteria to maintain dormancy and develop thermoresistance properties^{60–62}.

The INLA framework also allows the use of Delaunay triangulation as opposed to the regular grid structures that are commonly applied in conventional ENMs. This approach gathers more information from the areas with more observed data; thus, the density of triangulation is higher across these regions contributing to more accurate predictions. This approach considers the boundary effect by creating a mesh that transitions smoothly

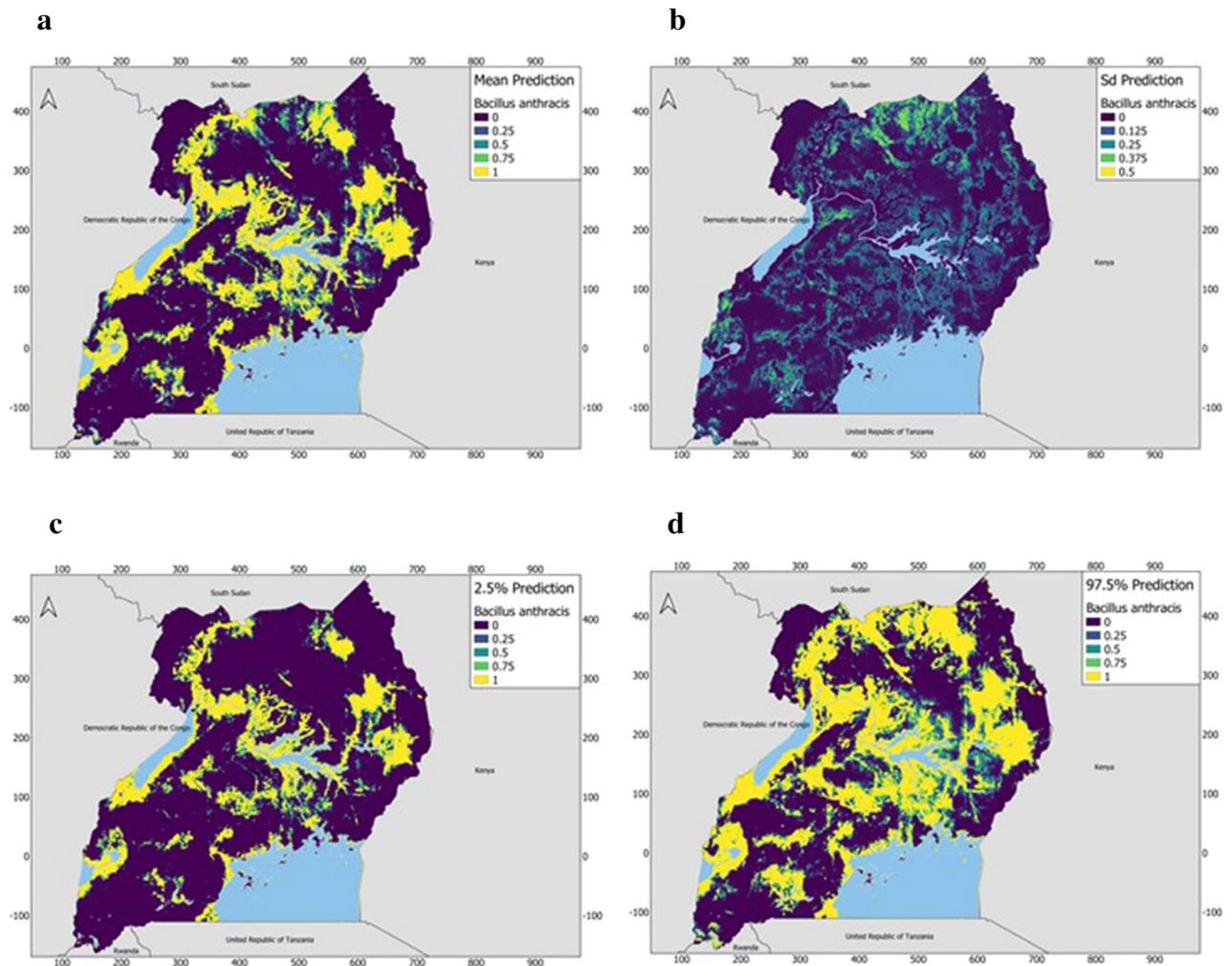


Figure 4. (A) The posterior predictive mean (B) posterior standard deviation (C) posterior lower credible interval (2.5% quantile) and (D) posterior credible upper interval (97.5% quantile) of the probability of *B. anthracis* suitability across Uganda from wildlife data (2004–2010). Prediction maps were developed using the Quantum Geographic Information System software (QGIS). URL: <https://qgis.osgeo.org> (2020).

from regions dominated by the smaller triangles (domain of interest in the model) to the regions with the large triangles (areas outside the domain of interest that help to avoid boundary effects within the model).

The areas of Western Uganda within Queen Elizabeth National Park, Northern Uganda around Arua, and Kween district in Eastern Uganda were predicted to be highly suitable for *B. anthracis* occurrence, confirming the results of previous studies with the accurate detection of the most suitable areas for the occurrence of the species^{33,58}. Compared to the most recent attempt to model the distribution of anthrax in Uganda³³, our model was different with regards to the parameters included and some of the covariates that were selected. For instance, the inclusion of the spatial random effect during model selection allowed the consideration of the effect of spatial autocorrelation before and after the addition of the covariates. The spatial effect was eventually excluded from the final model because the covariates used were able to account for most of the variation observed within the observed data. However, it is still important for future studies to test for any residual spatial variation and, if present, account for this and INLA provides a computationally efficient way of achieving this.

The INLA Bayesian framework applies probability distributions for the posterior estimates to model uncertainty within the parameter values²³. As such, a point estimate of the posterior probability of suitability as well as the uncertainty surrounding it can be obtained and assessed²⁴. Estimating this uncertainty via spatial maps is vital to offering end-users a realistic species suitability distribution to identify prevention and surveillance options. INLA models can not only deal with smoothing techniques such as GAMs, but they can also conduct inference and model prediction simultaneously, handle missing data, and incorporate biases within the data as spatial random effects, e.g., survey effort. One key limitation of this study is the lack of systematically collected and detailed anthrax case information. Therefore, it was impossible to account for structural dependencies such as temporal autocorrelation and analytical biases like survey effort or surveillance efficiency. Future studies should try to consider the incorporation of survey effort as well as temporal autocorrelation as model offsets or as other explanatory covariates within the model particularly when modelling incidence.

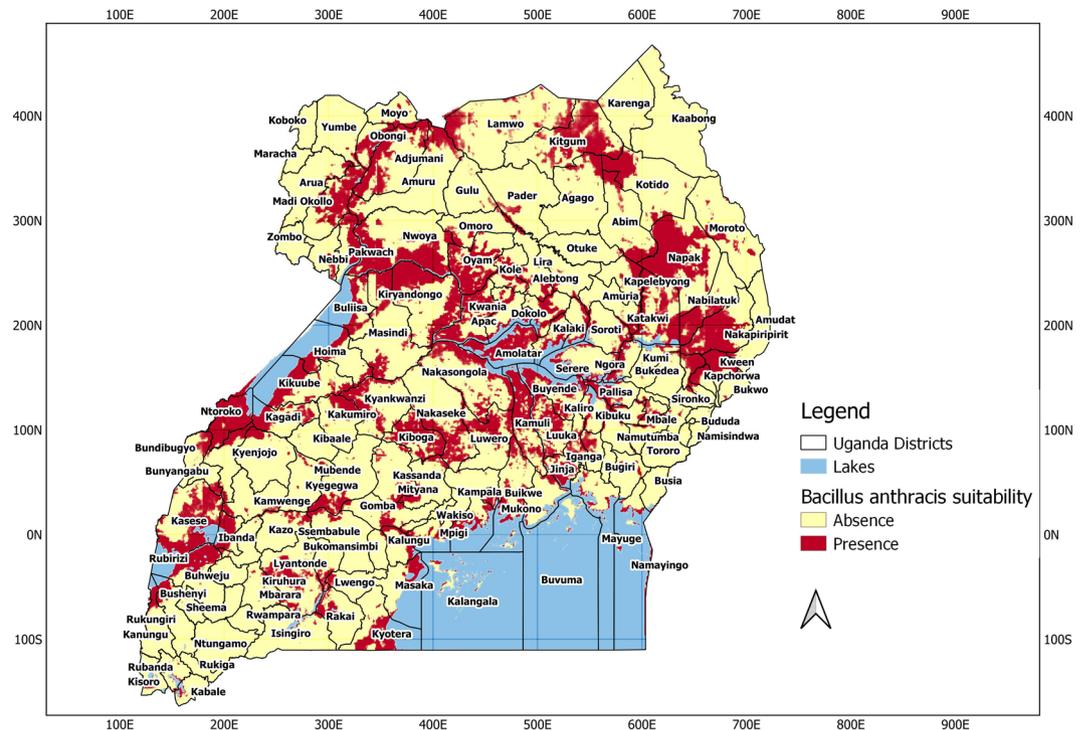


Figure 5. A binary map of the posterior predictive mean of the probability of *B. anthracis* suitability across Uganda using the threshold for maximum sensitivity and specificity (0.52). Prediction maps were developed using the Quantum Geographic Information System software (QGIS). URL: <https://qgis.osgeo.org> (2020).

This work employs a Bayesian GAM via INLA to examine the spatial suitability of *B. anthracis* using data from MAAIF, MOH, and published studies in Uganda. R-INLA has been shown to generate model predictions of species occurrence, which are better than those generated by Maxent and Boosted Regression Trees especially when the occurrence dataset is clumped or restricted¹⁸. Our model was able to pick up known locations of anthrax outbreaks occurring 8 years after the dataset used to calibrate the model with a sensitivity of 90%. The 10% of the test locations that were omitted by the model projection were all within 50 km of the nearest predicted high-risk area. There is a possibility that the animals could have acquired the infection within a high-risk area and died at a different location which was then georeferenced. These results demonstrate a high level of precision, making the prediction maps useful enough to be incorporated into future anthrax prevention and surveillance plans by the relevant stakeholders in Uganda. The model used in the past study to map the distribution of anthrax across Queen Elizabeth National Park in Uganda had a lower omission rate (5.2%) and slightly higher sensitivity of 94.8% against the test points³³. However, the occurrence data points used in the study were split randomly by a ratio of 3:1, with 75% of the data applied in model calibration, and the remaining 25% in model validation³³. The other recent continental study of the distribution of *B. anthracis* across Africa also had lower omission rates when no thinning (3.4%) and 50 km thinning (5.9%) were applied to the datasets⁵⁸. However, like the Queen Elizabeth study, the models were created by randomly sampling 50% of the occurrence dataset for model calibration and the rest (50%) for model validation⁵⁸. Random partitioning of the data into training and testing sets can inflate the performance of a model and underestimate the error in the spatial prediction evaluation⁵⁰.

The overall achievement of this study was to generate novel and highly relevant predictions of the spatial distribution of *B. anthracis* suitability across Uganda. Anthrax emerged as the top disease of highest priority for human and animal health sectors during the Ugandan One Health Zoonotic Disease Prioritization Workshop in 2017²⁴. The key recommendations from this workshop included an emphasis on building laboratory capacity, enhancing surveillance, improving outbreak response, and focusing on prevention and control, particularly using a multi-sectoral, One Health approach. The suitability maps we have generated in this study will help policy-makers in Uganda to effectively estimate the cost of implementing targeted anthrax prevention and surveillance campaigns and plan their management decisions effectively. This study also provides a basis for targeted studies to validate and improve predictions.

Conclusion

This work used a Bayesian framework to model the suitability of *B. anthracis* using data from MAAIF, MOH, and published studies in Uganda. The model identified known locations of anthrax outbreaks in Kiruhura, Arua, Pakwach, and Kween Districts, occurring 8 years later with excellent sensitivity. The prediction maps generated here can guide future anthrax prevention and surveillance plans by relevant stakeholders in Uganda. INLA methods can account for the analytical biases and structural dependencies within the input data, but more

effort still needs to be put in place to allow for detailed and systematic data surveillance to improve the quality of observation data available.

Data availability

The datasets and R code supporting the conclusions of this article are available in the GitHub repository, https://github.com/spatialmodels/Uganda-Spatial_Model.

Received: 20 August 2022; Accepted: 14 November 2022

Published online: 19 November 2022

References

- Gainer, R. Yamal and anthrax. *Can. Vet. J.* **57**, 985–987 (2016).
- Liskova, E. A. *et al.* Reindeer anthrax in the Russian arctic, 2016: Climatic determinants of the outbreak and vaccination effectiveness. *Front. Vet. Sci.* **8**, 1–9 (2021).
- Monje, F. *et al.* Anthrax Outbreaks among Domestic Ruminants Associated with Butchering Infected Livestock and Improper Carcass Disposal in Three Districts of Uganda, 2016–2018. (2020) <https://doi.org/10.21203/rs.2.20910/v1>.
- Driciru, M. *et al.* Spatio-temporal epidemiology of anthrax in Hippopotamus amphibious in Queen Elizabeth protected area, Uganda. *PLoS One* **13**, 1–21 (2018).
- World Health Organization. *Anthrax in Humans and Animals*. (WHO Press, 2008).
- Carlson, C. J. *et al.* The global distribution of *Bacillus anthracis* and associated anthrax risk to humans, livestock and wildlife. *Nat. Microbiol.* **4**, 1337–1343 (2019).
- Misgic, F., Atnaf, A. & Surafel, K. A review on anthrax and its public health and economic importance. *Acad. J. Anim. Dis.* **4**, 196–204 (2015).
- Peterson, A. T. & Soberón, J. Species distribution modeling and ecological niche modeling: Getting the concepts right. *Nat. Conserv.* **10**, 102–107 (2012).
- Neerinckx, S. B., Peterson, A. T., Gulinck, H. & Deckers, J. L. H. Geographic distribution and ecological niche of plague in sub-Saharan Africa. *Int. J. Health Geogr.* **54**, 1–12 (2008).
- Peterson, A. T., Sánchez-Cordero, V. & Beard, C. B. R. J. Ecologic niche modelling and potential reservoirs for Chagas disease, Mexico. *Emerg. Infect. Dis.* **8**, 662–667 (2002).
- Williams, R. & Fasina, F. O. P. A. Predictable ecology and geography of avian influenza (H5N1) transmission in Nigeria and West Africa. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 471–479 (2008).
- Thuiller, W. Editorial commentary on ‘BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change’. *Glob. Change Biol.* **20**, 3591–3592 (2014).
- Guisan, A., Edward, T. C. Jr. & Hastie, T. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecol. Modell.* **157**, 89–100 (2016).
- Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Divers. Distrib.* **20**, 1–9 (2014).
- Zhang, W. J., Zhong, X. Q. & Liu, G. H. Recognizing spatial distribution patterns of grassland insects: Neural network approaches. *Stoch. Environ. Res. Risk Assess.* **22**, 207–216 (2008).
- Phillips, S. J., Dudik, M. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* **190**, 231–259 (2006).
- Martínez-Minaya, J., Cameletti, M., Conesa, D. & Pennino, M. G. Species distribution modeling: A statistical review with focus in spatio-temporal issues. *Stoch. Environ. Res. Risk Assess.* **32**, 3227–3244 (2018).
- Redding, D. W., Lucas, T. C. D., Blackburn, T. & Jones, K. E. Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data. *PLoS One* **12**, 1–13 (2017).
- Paradinas, I., Conesa, D., López-Quílez, A. & Bellido, J. M. Spatio-Temporal model structures with shared components for semi-continuous species distribution modelling. *Spat. Stat.* **22**, 434–450 (2017).
- Poggio, L., Gimona, A., Spezia, L. & Brewer, M. J. Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. *Geoderma* **277**, 69–82 (2016).
- Banerjee, S., Gelfand, A. E., Finley, A. O. & Sang, H. Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70**, 825–848 (2008).
- Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71**, 319–392 (2009).
- Redding, D. W. *et al.* Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nat. Commun.* **12**, 1–10 (2021).
- Redding, D. W., Tiedt, S., Lo Iacono, G., Bett, B. & Jones, K. E. Spatial, seasonal and climatic predictive models of rift valley fever disease across Africa. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 1–9 (2017).
- World Atlas. Uganda Geography. <https://www.worldatlas.com/webimage/countrys/africa/uganda/ugland.htm> (2018).
- Ministry of Local Government. *Ministry of Local Government fact sheet*. <https://en.wikipedia.org/wiki/Uganda#Tourism> (2018).
- World Bank. *Uganda At a Glance*. https://web.archive.org/web/20090902211345/http://devdata.worldbank.org:80/AAG/uga_aag.pdf (2009).
- Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System (NNDSS). Anthrax (*Bacillus anthracis*) 2018 Case Definition. <https://www.cdc.gov/nndss/conditions/anthrax/case-definition/2018/> (2008).
- Abdrakhmanov, S. K. *et al.* Maximum entropy modeling risk of anthrax in the Republic of Kazakhstan. *Prev. Vet. Med.* **144**, 149–157 (2017).
- Assefa, A., Bihon, A. & Tibebu, A. Anthrax in the Amhara regional state of Ethiopia; spatiotemporal analysis and environmental suitability modeling with an ensemble approach. *Prev. Vet. Med.* **184**, 105155 (2020).
- Barro, A. S. *et al.* Redefining the Australian Anthrax Belt: Modeling the ecological niche and predicting the geographic distribution of *Bacillus anthracis*. *PLoS Negl. Trop. Dis.* **10**, 1–16 (2016).
- Chikerema, S. M., Murwira, A., Matope, G. & Pfukenyi, D. M. Spatial modelling of *Bacillus anthracis* ecological niche in Zimbabwe. *Prev. Vet. Med.* **111**, 25–30 (2013).
- Driciru, M. *et al.* Environmental determinants influencing anthrax distribution in Queen Elizabeth Protected Area, Western Uganda. *PLoS ONE* **15**, 1–21 (2020).
- Kracalik, I. T. *et al.* Modeling the environmental suitability of anthrax in Ghana and estimating populations at risk: Implications for vaccination and control. *PLoS Negl. Trop. Dis.* **11**, 1–17 (2017).
- Otieno, F. T. *et al.* Modeling the potential future distribution of anthrax outbreaks under multiple climate change scenarios for Kenya. *Int. J. Environ. Res. Public Health* **18**, 1–15 (2021).
- Otieno, F. T. *et al.* Modeling the spatial distribution of anthrax in southern Kenya. *PLoS Negl. Trop. Dis.* **15**, 1–16 (2021).

37. Steenkamp, P. J., van Heerden, H. & van Schalkwyk, O. L. Ecological suitability modelling for anthrax in the Kruger National Park, South Africa. *PLoS ONE* **13**, 1–13 (2018).
38. Mullins, J. C. *et al.* Ecological niche modeling of *Bacillus anthracis* on three continents: Evidence for genetic-ecological divergence?. *PLoS ONE* **8**, 1–8 (2013).
39. Mullins, J. *et al.* Ecological niche modelling of the *Bacillus anthracis* A1. a sub-lineage in Kazakhstan. *BMC Ecol.* **11**, 1–14 (2011).
40. Romero-Alvarez, D. *et al.* Potential distributions of *Bacillus anthracis* and *Bacillus cereus* biovar anthracis causing anthrax in Africa. *PLoS Negl. Trop. Dis.* **14**, 1–20 (2020).
41. Nath, S. & Dere, A. Applied geochemistry soil geochemical parameters in influencing the spatial distribution of anthrax in Northwest Minnesota, USA. *Appl. Geochem.* **74**, 144–156 (2016).
42. Blackburn, J. K. *et al.* *Bacillus anthracis* diversity and geographic potential across Nigeria, Cameroon and Chad: Further support of a Novel West African Lineage. *PLoS Negl. Trop. Dis.* **9**, 1–14 (2015).
43. Chen, W. *et al.* Mapping the distribution of anthrax in mainland China, 2005–2013. *PLoS Negl. Trop. Dis.* **10**, e0004637 (2016).
44. Morris, L. R., Proffitt, K. M., Asher, V. & Blackburn, J. K. Elk resource selection and implications for anthrax management in Montana. *J. Wildl. Manag.* **80**, 235–244 (2016).
45. Fick, S. E. & Hijmans, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
46. ESRI. ArcGIS Desktop: Release 10. (2011).
47. Hijmans, R. J. *et al.* Package 'raster': Geographic Data Analysis and Modeling. (2022).
48. Hazen, E. L. *et al.* Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. *Mov. Ecol.* **9**, 1–13 (2021).
49. R Core Team. R: A language and environment for statistical computing. (2019).
50. Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography (Cop.)* **40**, 913–929 (2017).
51. Zuur, A. F., Ieno, E. N. & Saveliev, A. A. *Beginner's Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA* (Highland Statistics Ltd., 2017).
52. Lindgren, F. & Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63**, 1–25 (2015).
53. Guisan, A., Edwards, T. C. & Hastie, T. Effect of boundary layer conductance on the response of stomata to humidity. *Ecol. Modell.* **157**, 89–100 (2002).
54. Berg, A., Meyer, R. & Yu, J. Deviance information criterion for comparing stochastic volatility models. *J. Bus. Econ. Stat.* **22**, 107–120 (2004).
55. Roos, M. & Held, L. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Anal.* **6**, 259–278 (2011).
56. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* <https://doi.org/10.1016/j.ecolmodel.2005.03.026> (2006).
57. Pennino, M. G., Muñoz, F., Conesa, D., López-Qúlez, A. & Bellido, J. M. Modeling sensitive elasmobranch habitats. *J. Sea Res.* **83**, 209–218 (2013).
58. Romero-álvarez, D. *et al.* Potential distributions of *Bacillus anthracis* and *Bacillus cereus* biovar anthracis causing anthrax in Africa. *PLoS Negl. Trop. Dis.* **14**, e0008131 (2020).
59. Himsworth, C. G. The danger of lime use in agricultural anthrax disinfection procedures: The potential role of calcium in the preservation of anthrax spores. *Can. Vet. J.* **49**, 1208–1210 (2008).
60. Dragon, D. C. & Rennie, R. P. The ecology of anthrax spores: Tough but not invincible. *Can. Vet. J.* **36**, 295–301 (1995).
61. Foerster, H. F. & Foster, J. W. Endotrophic calcium, strontium, and barium spores of *Bacillus megaterium* and *Bacillus cereus*. *J. Bacteriol.* **91**, 1333–1345 (1966).
62. Rode, L. J. & Foster, J. W. Quantitative aspects of exchangeable calcium in spores of *Bacillus megaterium*. *J. Bacteriol.* **91**, 1589–1593 (1966).

Acknowledgements

The authors wish to acknowledge the support received from the Ministry of Agriculture, Animal Industry, and Fisheries (MAAIF) Uganda. Finally, this work would also not have been possible without support from Dr. Alain Zuur and Dr. Elena Ieno, who provided some input on the model design and development.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Author contributions

V.A.N., J.L.N.W., and A.J.K.C. conceptualized the research. V.A.N. drafted the manuscript with input and revisions from J.L.N.W., A.J.K.C., D.R., H.O., M.O., R.T., R.A., F.M., A.R.A., S.K., L.B., E.K., B.K., V.N., J.H., J.S.S., A.V., M.D. All approved the final text.

Funding

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [Grant Number: OPP1144]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. Funding was also provided by the Dudley Stamp Memorial Award from the Royal Geographical Society. The Alborada Trust provides support for James Wood.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.A.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022