Modelling the spatial distribution of mycetoma in Sudan

Article (Accepted Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/98691/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk
Modelling the spatial distribution of mycetoma in Sudan.

Rowa Hassan\textsuperscript{1,2}, Hope Simpson\textsuperscript{3}, Jorge Cano\textsuperscript{3}, Sahar Bakhiet\textsuperscript{1}, Eltayeb Ganawa\textsuperscript{4}, Daniel Argaw\textsuperscript{5}, Melanie J Newport\textsuperscript{2}, Kebede Deribe\textsuperscript{2,6}, Ahmed Hassan Fahal\textsuperscript{1}

\textsuperscript{1} Mycetoma Research Centre, Soba University Hospital, University of Khartoum, Khartoum, Sudan
\textsuperscript{2} Department of Global Health and Infection, Brighton and Sussex Medical School, Falmer, Brighton, BN1 9PX, UK
\textsuperscript{3} Department of Disease Control, London School of Hygiene and Tropical Medicine, United Kingdom
\textsuperscript{4} Department of GIS and Cartography, Faculty of Geography and Environmental Sciences, University of Khartoum, Khartoum, Sudan
\textsuperscript{5} World Health Organization, Control of Neglected Tropical Diseases, Innovative and Intensified Disease Management, Geneva, Switzerland
\textsuperscript{6} School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding author:
Rowa Hassan
Email: Roaalbasha2016@outlook.com, R.hassan@bsms.ac.uk
Tele: +249129050542
Abstract

Mycetoma is a neglected tropical disease that is described worldwide, and Sudan has the highest reported number of mycetoma infections across the globe. The incidence, prevalence and burden of mycetoma globally are not precisely known, and its risk factors remain largely unelucidated. This study aimed to identify the environmental predictors of fungal and bacterial mycetoma in Sudan and to identify areas of the country where these niche predictors are met. Demographic and clinical data from confirmed mycetoma patients seen at the Mycetoma Research Centre between 1991 and 2018 were included in this study. Regression and machine learning techniques were used to model the relationships between mycetoma occurrence in Sudan and environmental predictors. The strongest predictors of mycetoma occurrence were aridity, proximity to water, low soil calcium and sodium concentrations and the distribution of various species of thorny trees. The models predicted the occurrence of eumycetoma and actinomycetoma in the central and southeastern states of Sudan, and along the Nile river valley and its tributaries. Our results showed that the risk of mycetoma in Sudan varies geographically, tied to identifiable environmental risk factors. Suitability maps are intended to guide health authorities, academic institutes and organizations involved in planning national scale surveys for early case detection and management leading to better patient treatment, prevention and control of mycetoma.

Key words: mycetoma; Sudan; environmental modeling; machine learning; ensemble models
Introduction

Mycetoma is a neglected tropical disease (NTD) that in common with other NTDs has numerous health and socioeconomic impacts on patients and communities. (1-3) It is a chronic infectious disease that leads to destructive granulomatous inflammation, eventually causing deformity to the affected body part limb, impeding mobility. (4, 5) Mycetoma is reported worldwide, but it is endemic in tropical and subtropical regions. (6, 7) Mycetoma-endemic areas are characterized by moderate aridity, low humidity, and a short rainy season. (8-10) Sudan is considered one of the endemic regions for mycetoma among a list of other countries. (11, 12) In a recent review mycetoma was noted to be present in 102 countries around the world and Sudan had the highest burden of cases. (13) More than 70 microorganisms have been incriminated in causing mycetoma, and according to the causative organism, it is classified into eumycetoma, caused by fungi, and actinomycetoma caused by bacteria. (14, 15) Eumycetoma is frequently reported in Africa and India, whereas actinomycetoma is more predominant in Asia, and Central and South America. (16, 17) In Sudan, *Madurella mycetomatis* is the leading causative agent of eumycetoma, while *Streptomyces somaliensis*, *Actinomadura madurae* and *Actinomadura pelletierii* are the most frequently reported actinomycetoma causative organisms. (18-21)

Mycetoma is a painless disease at the initial stages, of gradual onset and progress. Most affected patients are of low socioeconomic status and have poor levels of health education. (22) Furthermore, in mycetoma-endemic regions, the medical and health facilities are meager so most patients report late to medical care with advanced disease. (19, 20) For this reason, most of the reported patients are hospitalized rather than diagnosed at an earlier, more treatable stage of the disease. (21, 23) Late presentation of mycetoma patients leads to a reduced cure rate and a higher chance of recurrence, which in turn increases the rate of morbidity and disability. (24, 25) Despite the long history of mycetoma in Sudan and efforts to fight it, the incidence, prevalence and geographical distribution remain unclear. (9) Mycetoma patients in Sudan are often poor and illiterate and contracting mycetoma adds to their socio-economic burden.

Most of the epidemiological characteristics of mycetoma disease are uncertain. The route of infection, the incubation period, and the factors that determine susceptibility and resistance to disease are largely unknown. (22, 26) Likewise, its incidence,
prevalence, and burden have not been properly estimated at global or national levels. This lack of data is one reason for the absence of control and prevention programmes targeting the disease. (22, 27)

Although the infection pathway is not completely known, bacteria and fungi causing mycetoma are thought to penetrate through skin lesions and wounds. (28) It is hypothesized that thorns from trees and shrubs of the Acacia genus are a major cause of wounds through which mycetoma infectious agents can infect humans. In most mycetoma-endemic countries, thorny trees and shrubs such as Acacia species and other species from the pea family (Family: Fabaceae) are present. Many Acacia species bear sharp thorns that protect their leaves from herbivorous animals, and which on the ground, pose a threat to individuals walking barefoot or in open shoes such as sandals. (29) Other potential environmental risk factors that have been identified include hygiene practices and proximity to animals. (22) Many individuals living in rural areas have poor access to adequate quality water and sanitation facilities, which impedes good hygiene and could be one of the factors allowing the organism to establish infection, in the context of direct and sustained contact with soil, animals, and their dung. (30) Many mycetoma-causative organisms have been found in dung-enriched soil, hence dung could also provide a natural habitat for mycetoma-causative agents. (26)

The distribution of mycetoma has previously been modelled using Environmental Niche Modelling (ENM) techniques, which suggested central Sudan is suitable for mycetoma occurrence due to evident overlap of Acacia distribution and mycetoma cases, but only 44 mycetoma confirmed cases were used and they were distributed between 12°S and 19°N. (8) In this work, we aim to provide an update to these models by constructing models using a large occurrence dataset, categorised by type of mycetoma (actinomycetoma vs eumycetoma), using a wider set of environmental predictors, and with a range of modelling algorithms. The present study was conducted at the Mycetoma Research Centre (MRC), University of Khartoum, the only WHO collaborating centre on mycetoma globally. (31) The MRC was established in 1991 to tackle the burden of mycetoma in Sudan with well-trained dedicated staff for the diagnosis and management of mycetoma patients attending from all the states of
Sudan. To date, more than 9,500 patients have been seen and managed at the Centre.
Materials and methods

Data Source

The data included in this study were extracted from the patient database at the MRC. The study included only patients from Sudan who were seen at the centre during the period 1991-2018. The variables of interest included the patients' demographic characteristics, details of their clinical presentation and diagnosis, place of birth, their current home address if different and their address at the time of onset of symptoms. For most people, the place of birth and current residence were within the same vicinity but if the person had moved outside of their birth area the address where they were living at the onset of symptoms was used for the analysis. These locations were remotely geo-referenced using ArcGIS to obtain geographical coordinates defining the mycetoma occurrence locations. The diagnosis of mycetoma and its classification into eumycetoma and actinomycetoma were based on a careful clinical interview and examination, grain culture, cytological examination of aspirates from mycetoma lesions and histopathological examination of surgical biopsy samples. Histopathological methods have been used for mycetoma diagnosis since 1950s. Molecular diagnosis using conventional polymerase chain reaction (PCR) with pan-fungal/bacterial primers and species-specific primers was introduced in 2017.

Environmental predictors and variable selection

Candidate predictors

Since some determinants of mycetoma occurrence remain unelucidated, to avoid exclusion of relevant predictors, we started with a wide range of candidate predictors and used correlation analysis followed by principal components analysis (PCA), a technique used to reduce the dimensionality of large datasets.

Continuous gridded datasets (rasters) of 52 environmental variables considered relevant to the mycetoma ecological niche were assembled. These raster datasets included precipitation and temperature, soil composition and pH, livestock distributions, proximity to water sources, elevation and related topographical variables, the predicted distributions of Acacia species and other thorny vegetation, measures of atmospheric moisture availability which determine potential vegetative growth and an
index of vegetation coverage. Previous studies have indicated these variables to be associated with mycetoma occurrence in Sudan and elsewhere. \(^{(9, 33-35)}\)

Climate variables such as minimum, maximum and mean temperature and precipitation levels were obtained from the WorldClim database (version 2.0), a repository of climatic indicators based on long-term data collection from weather stations. \(^{(36)}\) In addition, we obtained from the Consortium for Spatial Information (CGIAR CSI) estimates of aridity (rainfall deficit) and potential evapo-transpiration (PET, an indicator of atmospheric demand for moisture), \(^{(37)}\) both modelled by using WorldClim datasets and a dataset of modelled elevation, based on data collected by the Shuttle Radar Topography Mission (SRTM). \(^{(38)}\)

We compiled soil composition and pH datasets from the International Soil Reference and Information Centre (ISRIC)-World Soil Information project. \(^{(39)}\) Raster datasets of predicted livestock distributions (cattle, chicken, sheep and goat), modelled by collaborators from the International Livestock Research Institute (ILRI), the Food and Agriculture Organization of the United Nations (FAO) and the Université Libre de Bruxelles (ULB-LUBIES), were obtained from the Gridded Livestock of the World (GLW) version 2 database. \(^{(40)}\) We downloaded spatial data on waterbodies and waterways from the OpenStreetMap project (OSM) through the Geofabrik platform and produced continuous surfaces of straight-line (Euclidean) distance in km to each. \(^{(41)}\) We obtained the Enhanced Vegetation Index (EVI) dataset, generated using satellite imagery data from the MODIS (Moderate Resolution Imaging Spectroradiometer) satellite, from the Earth Explorer NASA site. \(^{(42)}\) Full details of these data sources are shown in (Supplementary File 1, Table 1).

Finally, we modelled the distributions of 11 species of thorny trees which are considered potential vehicles for the inoculation of mycetoma-causative organisms using occurrence records downloaded from the Global Biodiversity Information Facility \(^{(43)}\) along with a suite of environmental variables. Full details of the modelling of these thorny plant species are given in a supplemental file (Supplementary File 2).

Each of these gridded datasets was resampled to 1 km\(^2\) resolution using bilinear interpolation and clipped and aligned to the outline of Sudan. Raster processing was
done using the raster package in R v3.3.2 and the final map layouts were created with ArcGIS 10.5 software (ESRI Inc., Redlands CA, USA).

**Variable Selection**

Initially, we calculated correlation coefficients between all candidate predictors across Sudan and identified groups of predictors that were correlated with a Pearson’s correlation coefficient >0.8. We then extracted values of each of the compiled raster datasets at each of the mycetoma occurrence locations to produce a matrix of covariates (a table showing the x and y coordinates of the occurrences, and at each of these locations, the values of all candidate covariates). We then ran PCA separately for actinomycetoma and eumycetoma to identify the most relevant predictors of the distribution.

In this process, the original variables are reduced into a set of “components” (or axes), which are linear combinations of variables, and are uncorrelated with each other. The components are ordered by the fraction of the total dataset variance they contribute, and the variance contribution of each component was summarised by a value termed an eigen-value. From the minimum set of components which collectively contributed at least 80% of the total variation in occurrence locations, we identified the original variable with the highest influence (measured via the factor loading). We ensured that we did not select any variables which were found to be correlated with another included variable. This process resulted in a selection of uncorrelated variables which collectively contributed the greatest proportion of variance to the occurrence locations. We used the prcomp package in R to implement this analysis.

**Ensemble modelling**

We built separate ensemble distribution models for actinomycetoma and eumycetoma, using the occurrence records and the suite of selected covariates for each. The ensembles were constructed from four algorithms from the BIOMOD package: Generalized Boosted Regression Model (GBM), Random Forest (RF), Generalized Linear Models (GLM) and Generalized Additive Models (GAM). For each model, we used the default parameters set by the biomod2 R package. These algorithms are all classified as presence-absence models, meaning that they require both occurrence and absence points to determine environmental suitability, absence points being locations confirmed by surveys to be clear of the outcome under
To account for the lack of absence points within the dataset, we generated “pseudoabsence” points, representing areas presumably unsuitable for mycetoma. The use of pseudoabsence points to represent areas presumed to be unsuitable for a species is a well-established approach in species distribution modelling. We implemented pseudoabsence selection within the BIOMOD framework, using the “surface range envelope” approach to define the area of assumed unsuitability. The envelope is estimated through a presence-only suitability model which identifies the range of locations at which the values of the chosen environmental covariates are within a specified range (here between the 5th and 95th percentiles) of the covariate values at the occurrence locations. A set of pseudoabsence samples (here a set of 5 samples), equal in size to the number of occurrence points, was randomly extracted from outside this envelope each time the model was run. Each algorithm was run 50 times with a random sample of 80% of the data points used for model training and the remaining 20% reserved for evaluation. The mean true skill statistic (TSS), the mean proportion correctly classified (PCC) which is equivalent to Kappa and the mean area under the curve (AUC) of the receiver operating characteristic (ROC) were used to evaluate the performance of each algorithm.

**Variables of final models**

For eumycetoma, aridity index, soil calcium concentration, wetness index, mean diurnal temperature, distance to the nearest river, presence of cattle, goats, and chickens, and predicted occurrence of *Acacia mellifera* and *Faidherbia albida* trees in the area were included. For actinomycetoma, aridity index, distance to the nearest river, distance to the nearest water body (pond or lake), wetness index, soil sodium and iron concentrations, presence of cattle and sheep in the area, mean diurnal temperature and mean temperature in the coldest year quarter were included.

All models with an AUC > 0.8 were compiled into the final ensemble distribution model. Within the ensemble, the predicted suitability value was the mean of the included models, weighted by their relative AUC. Upper and lower limits of suitability were estimated by calculating confidence intervals around the ensemble mean suitability value for each cell within the grid. Variable contribution plots were produced to show
the relative contribution of each variable to the model, along with marginal effect plots to show the response of both mycetoma types to changes in each modelled covariate.
Results

Dataset of mycetoma occurrence

The modelled data included 7812 unique points after the exclusion of patients (829) from outside Sudan and those lacking diagnostic or geographical information. The patients were seen at the MRC in the period 1991- 2018, and they came from all the states of Sudan. The study included 5513 patients (79%) with confirmed eumycetoma and 1470 patients (21%) with actinomycetoma (Figure 1). Most of the Mycetoma patients were from Al Jazirah State (34.4%) and Khartoum State (14.5%), (Table 1).

Environmental predictors of eumycetoma occurrence in Sudan

The PCA for eumycetoma occurrence in Sudan identified ten variables that characterized the environmental conditions at the occurrence locations. These were aridity index, soil calcium concentration, wetness index, mean diurnal temperature, distance to the nearest river, presence of cattle, goats, and chickens, and predicted occurrence of *Acacia mellifera* and *Faidherbia albida* trees in the area. Two variables - the predicted density of cattle and chickens - were excluded after the initial modelling step as each contributed less than 1% of the variability in the response variable (eumycetoma occurrence).

All models for eumycetoma occurrence performed well with an ensemble ROC score of 0.993, sensitivity of 96.475 % and specificity of 94.476%. The mean TSS was 0.909, and mean kappa score was 0.898 (Table 2).

Across the RF models, the distance to the nearest river was the strongest environmental predictor of suitability for eumycetoma occurrence, and the suitability decreased with increased distance to the nearest river. The diversity of thorny trees, represented by the number of species predicted to be present, was the second most important predictor of suitability for eumycetoma occurrence, with a greater diversity of thorny trees associated with a higher probability of occurrence of the disease. We found the suitability for eumycetoma to be higher in areas with a lower probability of occurrence of *Acacia mellifera* and arid areas. Soil calcium concentration suitable for eumycetoma occurrence was between 3000 and 11000 mg/kg while mean diurnal
temperature range and the probability of occurrence contributed to a lesser extent and goat density was the least contributor. (Supplementary File 1, figure 1).

Environmental predictors of actinomycetoma occurrence in Sudan

The PCA for actinomycetoma occurrence also revealed ten variables as potential environmental predictors. The variables were aridity index, distance to the nearest river, distance to the nearest water body (pond or lake), wetness index, soil sodium and iron concentrations, presence of cattle and sheep in the area, mean diurnal temperature and mean temperature in the coldest year quarter. After the primary analysis, four variables; distance to the nearest river, distance to the nearest water body (pond or lake), mean temperature in the coldest year quarter and variety in thorny trees were included.

All models for actinomycetoma occurrence performed well, with a ROC score of 0.995, sensitivity of 95.445% and specificity of 96.709%. The mean TSS was 0.921, and the mean kappa was 0.903 (Table 2).

Across the RF models, distance to the nearest water body was the major contributor to actinomycetoma occurrence, followed by the distance to the nearest river. Mean temperature during the coldest quarter in a range of 18-25°C contributed to the predicted suitability for actinomycetoma occurrence by 13%. Arid areas with a low concentration of soil sodium contributed to almost 11 %. The diversity of thorny trees in an area showed no difference between areas with no diversity of thorny trees and areas with diverse group of thorny trees. (Supplementary File 1, figure 3).

Predicted suitability for mycetoma occurrence in Sudan

Both eumycetoma and actinomycetoma were predicted to occur around the central and south-eastern parts of Sudan and along the Nile river valley and its tributaries. The central states of Sudan, Al Jazirah, White Nile, Khartoum, and Sennar State had the largest areas predicted suitable for eumycetoma. The predicted distribution of actinomycetoma was slightly wider than that of eumycetoma, with more patchy suitability predicted in the western states of Sudan (Supplementary File 1, figure 5 and
The overlapping areas for actinomycetoma and eumycetoma risk are illustrated in Figure 4.

**Discussion**

We have applied ecological niche modelling techniques to identify environmental variables associated with the occurrence of mycetoma in Sudan. Models were run separately for eumycetoma and actinomycetoma and gave similar results in terms of the predicted distribution. Both types were predicted along the River Nile and its tributaries, mainly around the central and south-eastern parts of Sudan covering Khartoum, Al Jazirah and White Nile States, the northern part of Sennar State and the eastern part of North Kordofan State which are known to have a high burden of cases. (48) The models also predicted the occurrence of eumycetoma in the north and eastern parts of the country, where cases have not previously been recorded. This might indicate underreporting or misdiagnosis of cases from these areas. Such areas must be considered as priorities for future surveillance activities and to offer targeted training of medical personnel for early detection and better diagnosis of mycetoma.

This study showed the suitability for mycetoma occurrence was widespread in Khartoum State, and this can be attributed to the high influx of people from rural endemic areas to the state, due to various socioeconomic difficulties, often secondary to having the disease. Also, Khartoum residents regularly visit their families in endemic regions and can contract the infection during such visits. Furthermore, much of the state beyond the city is undeveloped rural land that harbours the ecological and socio-demographic characteristics associated with mycetoma.

As previously mentioned, mycetoma is distributed globally in a belt form, and areas within the belt are dry with annual rainfall between 50 and 1000 mm and high temperatures. (49) Given the recognized influence of temperature on the survival of mycetoma-causative organisms, (50) it is not surprising that temperature-related variables had a large impact on the suitability for both types of mycetoma. Our results indicate that environments with mean temperatures in the coldest year quarter between 20°C and 25°C are most suitable for actinomycetoma-causative organisms. Daily variation in temperature was a strong predictor of suitability for eumycetoma. It has been reported that fungal organisms that cause eumycetoma cannot survive in
environments where the daily range in temperature is more than 15°C on average. This could be explained by the fact that the organisms do not live in regions that have extreme variation in temperature. According to our results, mycetoma seems to be predicted in arid areas, as mentioned previously in the literature, mycetoma prevails in reasonably arid areas with a short rainy season of 4–6 months.

Analysis of the MRC data showed that most of the patients reside in locations close to water sources, and the models predicted the same pattern. This observation is mostly likely explained by the fact that while mycetoma-causing species thrive in drier environments, people living in such environments settle near a water source to enable activities such as washing and growing crops. Thus distance to rivers correlates with mycetoma occurrence as a result of human behavior, rather than environmental conditions near rivers supporting organism survival.

To our knowledge, the influence of the soil mineral components on the growth of mycetoma-causative organisms has not been studied thoroughly. Our findings suggested that soil calcium concentration had a moderate effect on suitability for eumycetoma and sodium concentrations contribute to the disease occurrence suitability. In histopathological studies calcium, among other elements, was found in abundance in the cell wall of *Madurella mycetomatis*. It is an important element in the formation of the cement substance which is a hard-brown material containing melanin, heavy metal ions, proteins, and lipids. Cement is a fungal protective mechanism, acting as a sink for harmful unpaired electrons and provides the cell walls with structural rigidity as well as storing water and ions to prevent desiccation. This may indicate that calcium could be an element in the organism’s virulence by contributing to the myelinization and sclerosis of the grain. In a recent cross-sectional study conducted in Mexico, cases were prevalent among three soil types that had high level of calcium. This observation warrants further laboratory studies to explore the possible role of these soil elements in the pathogens’ survival.

We found that environments with a greater number of species of Acacia trees were more suitable for eumycetoma, supporting previous evidence that had suggested a role for skin penetration by thorns in the transmission of the disease. In our study, and in contrast to the study conducted by Samy and colleagues in 2014, not only was
the distribution of Acacia trees tested but also the role of diversity of these trees in mycetoma occurrence in both its types. (9) However, areas with either no thorny trees or areas with more than five species of thorny trees were both suitable for actinomycetoma occurrence. In our study eumycetoma, but not actinomycetoma, occurrence seems to be predicted by the diversity of thorny trees. This could be explained by the observation that eumycetoma-causative organisms live on the thorns. This study showed no strong association between the individual species types and mycetoma occurrence. These observations on the association between disease occurrence and individual tree species, their botanical properties and other parameters need further investigation. Such studies are important to gain more in-depth knowledge about disease transmission and the life cycle of the causative microorganisms to develop ways to interrupt them.

In conclusion, the studied models have indicated that arid areas proximal to water sources, soil with low concentrations of calcium and sodium and with a variety of thorny tree species provide the most suitable environment for the occurrence of mycetoma. This work included environmental probable risk factors that have not been studied previously which could be used to guide the implementation of preventive interventions and control strategies in the affected areas. Until there is robust evidence on disease transmission and the life cycle of the microorganisms involved, it may be advisable to limit the use of thorny trees in building animal enclosures and houses. However, such simple and effective interventions may not align with long-standing cultural practices and should be comprehensively discussed with all stakeholders and community members to ensure community acceptance. Protective measures such as the use of footwear should be encouraged in people who are in direct contact with these trees.

Our study also had considered the suitable environmental indicators across the whole of Sudan with updated data from the MRC database records using a wide range of environmental risk factors. The results could help guide planning for national and international scale surveys for mycetoma identification and to design targeted training for medical staff on early detection and diagnosis. Moreover, they could aid in the design of national and global mycetoma advocacy and awareness programmes.
leading to early active case detection with early appropriate patient management in areas that are highly endemic for mycetoma.

**Authors’ statements:**
RH, HS, JC, SB, EG, DA, MJN, KD and AHF conceived and designed the study; RH, HS, JC, SB, EG, DA, AHF, MJN and KD performed the study; RH, HS, JC, SB, EG, DA, MJN, KD and AHF analyzed the data; RH and KD wrote the first draft of the manuscript. RH, HS, JC, SB, EG, DA, MJN, KD and AHF revised the manuscript. All authors read and approved the final manuscript.

**Acknowledgement**
None

**Funding:**
This research was funded by the National Institute for Health Research (NIHR) Global Health Research Unit on NTDs at BSMS (16/136/29) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK government. KD is supported by the Wellcome Trust [grant number 201900/Z/16/Z], as part of his International Intermediate Fellowship.

**Competing interests**
The authors declare that they have no conflicts of interest.

**Data availability**
Data available upon request
**Ethical approval:**

This study was approved by the Mycetoma Research Centre, Khartoum, Sudan (IRB, No. SUH 11/12/2018) and from the BSMS Research Governance and Ethics Committee (ER/BSMS435/1). A written informed consent was obtained from each adult patient, and verbal assent was obtained from minors (age below 18 years) in addition to signed written informed consent from the parents/guardian.
References

47. Sui Z. Modeling tree species distribution and dynamics under a changing climate, natural disturbances, and harvest alternatives in the southern United States: Mississippi State University; 2015.
Figure Legends:

Figure 1. Records of mycetoma cases in Sudan (1991-2018), upper right map shows where Sudan stand in Africa, the blue dots stand for Actinomycetoma and yellow dots stand for Eumycetoma. Mycetoma cases are distributed along the River Nile with focus on central Sudan. Cases are distributed sporadically in western part of Sudan.

Figure 2. Predicted environmental suitability for eumycetoma in Sudan. Eumycetoma is predicted to occur along the River Nile and its tributaries with a hot spot on Central and south eastern part of Sudan.

Figure 3. Predicted environmental suitability for actinomycetoma in Sudan. Actinomycetoma is predicted to occur along the River Nile and its tributaries with a hot spot in central and south-eastern part of Sudan with sporadic occurrence in western part of Sudan.

Figure 4. Overlaps of actinomycetoma and eumycetoma environmental suitability in Sudan. Map shows central Sudan to be the most suitable area for the occurrence of both types of mycetoma.

Table Legends:

Table 1. Distribution by State of mycetoma patients from Sudan seen at the Mycetoma Research Centre in the period 1991 – 2018.

Table 2. Validation metrics for ensemble models for eumycetoma and actinomycetoma suitability

Supporting information:

S1 File: Covariates, marginal effect plots and predicted occurrences

S2 Text: Modelling the environmental suitability for the species of thorny vegetation present in Sudan.
Table 1.

<table>
<thead>
<tr>
<th>State</th>
<th>Eumycetoma</th>
<th>Actinomycetoma</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al Jazirah</td>
<td>2163</td>
<td>234</td>
<td>2397 (34.4%)</td>
</tr>
<tr>
<td>Khartoum</td>
<td>730</td>
<td>280</td>
<td>1010 (14.5%)</td>
</tr>
<tr>
<td>White Nile</td>
<td>753</td>
<td>101</td>
<td>854 (12.2%)</td>
</tr>
<tr>
<td>North Kordofan</td>
<td>479</td>
<td>341</td>
<td>820 (11.7%)</td>
</tr>
<tr>
<td>Sennar</td>
<td>562</td>
<td>69</td>
<td>631 (9%)</td>
</tr>
<tr>
<td>River Nile</td>
<td>128</td>
<td>59</td>
<td>187 (2.7%)</td>
</tr>
<tr>
<td>South Darfur</td>
<td>118</td>
<td>62</td>
<td>180 (2.6%)</td>
</tr>
<tr>
<td>North Darfur</td>
<td>97</td>
<td>73</td>
<td>170 (2.4%)</td>
</tr>
<tr>
<td>Al-Qadarif</td>
<td>108</td>
<td>42</td>
<td>150 (2.1%)</td>
</tr>
<tr>
<td>Kassala</td>
<td>95</td>
<td>40</td>
<td>135 (1.9%)</td>
</tr>
<tr>
<td>South Kordofan</td>
<td>80</td>
<td>46</td>
<td>126 (1.8%)</td>
</tr>
<tr>
<td>West Darfur</td>
<td>44</td>
<td>36</td>
<td>80 (1.1%)</td>
</tr>
<tr>
<td>Northern</td>
<td>45</td>
<td>29</td>
<td>74 (1.1%)</td>
</tr>
<tr>
<td>Blue Nile</td>
<td>38</td>
<td>10</td>
<td>48 (0.7%)</td>
</tr>
<tr>
<td>East Darfur</td>
<td>34</td>
<td>15</td>
<td>49 (0.7%)</td>
</tr>
<tr>
<td>West Kordofan</td>
<td>19</td>
<td>12</td>
<td>31 (0.4%)</td>
</tr>
<tr>
<td>Red Sea</td>
<td>13</td>
<td>13</td>
<td>26 (0.4%)</td>
</tr>
<tr>
<td>Central Darfur</td>
<td>7</td>
<td>8</td>
<td>15 (0.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>5513</td>
<td>1470</td>
<td>6983</td>
</tr>
</tbody>
</table>
### Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Weighted Mean</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eumycetoma</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSS</td>
<td>0.909</td>
<td>0.909</td>
<td>0.911</td>
</tr>
<tr>
<td>ROC</td>
<td>0.993</td>
<td>0.993</td>
<td>0.993</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.898</td>
<td>0.897</td>
<td>0.898</td>
</tr>
<tr>
<td><strong>Actinomycetoma</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSS</td>
<td>0.921</td>
<td>0.922</td>
<td>0.922</td>
</tr>
<tr>
<td>ROC</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.903</td>
<td>0.903</td>
<td>0.902</td>
</tr>
</tbody>
</table>