

1 **Mapping suitability for Buruli ulcer at fine spatial scales across**  
2 **Africa: a modelling study**

3 Hope Simpson MSc<sup>1\*</sup>, Earnest Njih Tabah PhD<sup>2</sup>, Richard Phillips MD<sup>3</sup>, Michael Frimpong PhD<sup>3</sup>, Issaka  
4 Maman PhD<sup>4</sup>, Joseph Timothy PhD<sup>1</sup>, Paul Saunderson MD<sup>5</sup>, Rachel L Pullan PhD<sup>1</sup>, Jorge Cano PhD<sup>1</sup>

5 \*corresponding author: [hope.simpson@lshtm.ac.uk](mailto:hope.simpson@lshtm.ac.uk)

6 **Author Affiliations**

7 <sup>1</sup>London School of Hygiene and Tropical Medicine, United Kingdom

8 <sup>2</sup>National Yaws, Leishmaniasis, Leprosy and Buruli ulcer Control Programme, Cameroon

9 <sup>3</sup>School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Ghana

10 <sup>4</sup>National Reference Laboratory for Buruli Ulcer Disease in Togo; Ecole Supérieure des Techniques  
11 Biologiques et Alimentaires (ESTBA), Laboratoire des Sciences Biologiques et des Substances  
12 Bioactives, Université de Lomé, Lomé, Togo

13 <sup>5</sup>Accelerating Integrated Management (AIM) Initiative, Accra, Ghana.

14

15 **Abstract**

16 **Background**

17 Buruli ulcer (BU) is a disabling and stigmatising neglected tropical disease (NTD). Its distribution and  
18 burden are unknown because of underdiagnosis and underreporting. It is caused by *Mycobacterium*  
19 *ulcerans*, an environmental pathogen whose environmental niche and transmission routes are not  
20 fully understood. Active BU case searches can limit morbidity by identifying cases and linking them  
21 to treatment, but these are mostly restricted to well-known endemic areas. A better understanding  
22 of environmental suitability for environmental reservoirs of *M. ulcerans* and BU disease would  
23 advance understanding of the disease's ecology and burden, and could inform targeted surveillance.

24 **Methodology/Principal Findings**

25 We used previously compiled point-level datasets of BU and *M. ulcerans* occurrence, evidence for  
26 BU occurrence within national and sub-national areas, and diverse environmental datasets. We  
27 fitted relationships between BU and *M. ulcerans* occurrence and environmental predictors by  
28 applying regression and machine learning based algorithms, combined in an ensemble model to  
29 characterise the optimal ecological niche for the disease and bacterium across Africa at a resolution  
30 of 5km x 5km. Climate and atmospheric variables were the strongest predictors of both  
31 distributions, while indicators of human disturbance including damming and deforestation, drove  
32 local variation in suitability. We identified patchy foci of suitability throughout West and Central  
33 Africa, including areas with no previous evidence of the disease. Predicted suitability for *M. ulcerans*  
34 was wider but overlapping with that of BU. The estimated population living in areas predicted  
35 suitable for the bacterium and disease was 29.1 million.

36 **Conclusions/Significance**

37 These maps could be used to inform burden estimations and case searches which would generate a  
38 more complete understanding of the spatial distribution of BU in Africa, and may guide control  
39 programmes to identify cases beyond the well-known endemic areas.

40

41

42

43 **Author summary**

44 Like many neglected tropical diseases primarily affecting the rural poor, Buruli ulcer (BU) is under-  
45 detected and under-reported within routine health information systems. As such, the burden and  
46 distribution are not fully known, impeding appropriate targeting of health resources, control, and  
47 care for people affected. Having previously evaluated and mapped the existing evidence for BU and  
48 its causative agent *M. ulcerans*, we concluded that the disease was likely to occur beyond the range  
49 of known endemic areas. However, we were left with the question of where exactly these  
50 undetected cases might be occurring. Answering this question required a more fine-scale approach:  
51 BU is highly focal, presumably due to local variation in the environmental factors which determine  
52 suitability for *M. ulcerans* survival and transmission to humans. We used the compiled evidence and  
53 geographical datasets to build statistical models representing the relationship between  
54 environmental factors and previously reported cases. This allowed us to define the ecological niche  
55 of BU, and subsequently to identify areas across Africa where this niche was met, providing suitable  
56 conditions for the disease. We constructed separate models of suitability for *M. ulcerans*, using  
57 locations where its DNA had been detected in environmental sources. Unsurprisingly, suitability for  
58 *M. ulcerans* was predicted to be wider than, but geographically overlapping with that for BU. This  
59 implies that beyond the conditions necessary for survival of the bacterium, additional factors are  
60 required for transmission to humans. The high-resolution suitability maps we present are intended  
61 to guide case search activities which may identify endemic areas beyond the known endemic range.  
62 Data on the true prevalence of BU from targeted case searches within predicted-suitable areas will  
63 also allow us to validate and refine the models, and potentially predict the actual probability of cases  
64 occurring within predicted suitable areas.

65

66

## 67 **Introduction**

68 Buruli ulcer (BU) is a chronic necrotizing disease of the skin and soft tissue, which causes debilitating  
69 symptoms and sequelae, associated with a high burden of morbidity and stigma for patients and  
70 economic costs for affected households [1-3]. These impacts are felt particularly strongly in  
71 impoverished rural communities with poor access to health services [3, 4]. The infectious agent is  
72 *Mycobacterium ulcerans*, a slow-growing environmental bacterium which may be transmitted from  
73 aquatic environments to humans via inhalation or ingestion, or by penetration of the skin [1, 5, 6].  
74 The main control strategy is active case finding in endemic areas to promote early case detection  
75 and effective treatment, which limits disease progression [7, 8]. BU occurs mostly in tropical and  
76 subtropical areas of West and Central Africa, with smaller foci in parts of Asia, South America, the  
77 Western Pacific and Australasia [9]. However, the disease is recognised to be underdiagnosed and  
78 under-reported, and may occur undetected in other parts of the world [9-12].

79 The distribution of BU is presumably linked to environmental suitability for *M. ulcerans* survival and  
80 replication, as well as to human and environmental factors favouring transmission [13]. On a  
81 continental scale, BU appears to be limited by climatic factors: it is restricted to tropical and  
82 subtropical regions and absent from arid areas [14]. Within endemic areas, the disease shows a  
83 highly focal distribution [15-17], but reasons for this are not well understood, since the precise niche  
84 and transmission routes of *M. ulcerans* have been difficult to characterise [18]. The pathogen has  
85 only been cultured from environmental and animal samples a handful of times [19-21], although it  
86 has been detected by PCR in aquatic environments of endemic and non-endemic areas, and a in  
87 wide range of potential hosts including mammals, fish, amphibians, and aquatic and terrestrial  
88 insects [22-26]. Consistent with the ecology of an environmental pathogen, the distribution of *M.*  
89 *ulcerans* in the environment appears to be wider than that of BU, suggesting that factors beyond  
90 environmental suitability for *M. ulcerans* are required for transmission [13, 14, 27].

91 Our understanding of the pathways of BU infection is also limited, partly by its long and variable  
92 incubation period, which makes it difficult for patients and clinicians to attribute particular events or  
93 activities to disease acquisition [28]. Local spatial analysis has identified several environmental  
94 variables associated with increased BU incidence, primarily proximity to rivers, as well as  
95 environmental disturbance and land-use changes including deforestation, urbanisation,  
96 agriculturalization and mining [17, 29]. Case control studies have identified contact with unprotected  
97 waterbodies as a risk factor for disease [30], suggesting that activities which bring people into  
98 contact with water sources harbouring *M. ulcerans* increase the risk of disease acquisition [31-33].

99 Given the recognised scale of BU under-detection and under-reporting, it is likely that BU occurs  
100 beyond the known range of reported cases. A better understanding of potential suitability for the  
101 pathogen in the environment and the disease in humans would help to improve its surveillance and  
102 control in countries where it is known to be endemic. Furthermore, characterisation of the  
103 environmental factors linked to suitability for *M. ulcerans* and BU may reveal areas at risk for disease  
104 emergence, or places harbouring unrecognised cases.

105 In this investigation, we aim to identify environmental factors which characterise the environmental  
106 niche of *M. ulcerans* and BU disease in humans, and to model their respective relationships with BU.  
107 These analyses will be used to identify areas of continental Africa which may be suitable for *M.*  
108 *ulcerans* or BU based on their environmental characteristics.

## 109 **Methods**

### 110 **Data on Buruli ulcer and *M. ulcerans* distribution**

111 We used previously compiled spatial datasets of point locations of recorded occurrences of BU  
112 disease in humans, and of detection of *M. ulcerans* genetic material in biotic and abiotic  
113 environmental samples [9, 34]. Data for the final models was extracted from the database on  
114 03/01/2020.

115 BU occurrence locations were restricted to those where BU infection was confirmed by a positive  
116 result for PCR targeting IS2404, or histopathology consistent with BU disease. To explore the model's  
117 sensitivity to the case definition, we repeated the analysis using all locations where clinically  
118 diagnosed BU had been reported. We hereon refer to the two datasets as '*confirmed occurrences*'  
119 and '*all occurrences*' respectively.

120 The environmental dataset was restricted to locations where *M. ulcerans* DNA had been identified  
121 and distinguished from that of other mycobacteria: either by multiplex qPCR assays quantifying the  
122 relative copy numbers of IS2404, IS2606 and the KR-B domain [35]; by variable nucleotide tandem  
123 repeat (VNTR); or mycobacterial interspersed repetitive unit (MIRU) typing [36, 37]. We hereon refer  
124 to this dataset as '*environmental occurrences*'.

125 All records were restricted to locations with reliable geographical coordinates and deduplicated by  
126 geographical location. Human and environmental locations were weighted by the year and the  
127 specificity of confirmatory tests reported (S1 Text).

### 128 **Environmental datasets used in ecological modelling**

129 We assembled gridded datasets of 51 environmental variables considered relevant to the ecological  
130 niche of *M. ulcerans* [18]. All variables and their sources are shown in Table 1 and full details are  
131 provided in S3 Text.

132

133 **Variable selection**

134 We compiled the gridded predictor variables at a resolution of 5km x 5km within a rectangular area  
135 of West Africa from latitude -13.57195, longitude-4.11032, to lat. 16.67107, long. 14.493. This area  
136 contained 94% of all BU occurrence locations, 95% of confirmed BU occurrence locations, and all  
137 environmental occurrence locations. We extracted the values of predictor variables at the locations  
138 of BU cases (all occurrences) and environmental occurrences of *M. ulcerans* DNA.

139 We used principal components analysis (PCA) to identify the minimum set of variables that best  
140 characterized the environment at observations of BU and of *M. ulcerans* from the global  
141 environment. We undertook separate PCAs on the human (confirmed cases) and environmental  
142 datasets, selecting variables that contributed most strongly to the minimum number of principal  
143 components collectively accounting for at least 80% of the total dataset variance.

144 **Pseudoabsence and Background data**

145 The occurrence data used in the modelling framework were supplemented with systematically  
146 generated pseudo-absence and background data. The use of artificial absence data is a common  
147 approach in species distribution modelling, designed to account for geographically biased presence  
148 data and sparse absence data [38]. The terms background and pseudoabsence are often used  
149 interchangeably to describe artificial absence data, though Elith and Hijmans distinguish them on the  
150 basis that background data are intended to characterise the 'environmental domain' of the study-  
151 helping to account for geographical bias, while pseudoabsence data represent areas assumed to be  
152 unsuitable for the species and are intended to capture the environment in these areas [39]. We  
153 hereon use the terms in this sense.

154 Pseudoabsence and background points were both selected from a restricted geographical extent  
155 around occurrence points, defined by the spatial structure of the occurrence predictors. This has  
156 been recommended by previous authors as way to provide an ecologically meaningful definition of  
157 the study range [40]. Pseudoabsence points were sampled at higher density in areas of weaker  
158 evidence according to a systematic review of the geographical distribution of BU [41]. Background  
159 points were sampled at higher density around recorded occurrence points. More details on the  
160 generation of pseudoabsence and background points are provided in S2 Text.

161 Pseudoabsence and background weights were uniform within datasets and assigned so their sum  
162 was equal to the sum of occurrence weights in each model [42]. Human background points were  
163 restricted to a minimum distance of 10km from any occurrence location, and environmental  
164 background points were restricted to 10km from any human or environmental occurrence location.

165 The distribution of BU and *M. ulcerans* occurrences, pseudoabsence and background points are  
166 shown in S1 Figure and S2 Figure.

### 167 **Ensemble modelling**

168 The environmental factors selected through PCA were used as predictor variables. We used the  
169 *biomod2* package in R [43, 44] to implement seven algorithms: generalized linear models (GLM),  
170 generalized additive models (GAM), generalized boosted regression models (GBM), artificial neural  
171 networks (ANN), multiple adaptive regression splines (MARS), maximum entropy (MaxEnt) and  
172 random forest (RF).

173 Individual model algorithms were each run 20 times with a random sample of 80% of data points,  
174 and evaluated with the remaining 20%. For each algorithm we calculated the mean true skill statistic  
175 (TSS), the mean positive correctly classified (PCC) and the mean area under the curve (AUC) of the  
176 receiver operation characteristic (ROC) [45]. Models with mean AUC above 0.8 were integrated in an  
177 ensemble using committee averaging to attribute higher weight to better performing models.

178 We plotted the importance values representing each variable's contribution to the model and  
179 created marginal effect plots for the modelled covariates in the highest performing model ensemble.

### 180 **Estimating total population living in suitable areas**

181 We calculated the total area suitable for BU, *M. ulcerans*, and the total area suitable for both, and  
182 extracted estimates of the population living in each of these areas from a raster representing  
183 estimated number of people per 1km<sup>2</sup> grid square in 2020 [46].



## 184 **Results**

### 185 **Datasets of BU occurrence in humans and *M. ulcerans* DNA detection in the environment**

186 The modelled data included 2,183 unique point locations with reported cases of BU in Africa (Figure  
187 1A). BU was confirmed by PCR or histopathology at 738 unique locations. There were 91 unique  
188 locations where *M. ulcerans* DNA had been detected by MIRU, VNTR or qPCR (Figure 1B). The  
189 dataset of clinically diagnosed human cases represented 16 countries, mostly in West and Central  
190 Africa, with a few in East and southeast Africa. The confirmed cases were restricted to 13 countries.  
191 The time period of human case detection was from 1957 to 2019. The median year of diagnosis was  
192 2010. The 91 records of environmental detection of *M. ulcerans* represented three countries: Ghana,  
193 Cameroon and Benin, and covered the period from 2006 to 2018 with a median year of detection of  
194 2013.

195 **Figure 1A:** Selection of BU occurrence points from BU database

196 **Figure 1B:** Selection of model environmental occurrence points from global database of MU  
197 occurrences

### 198 **Principal components analysis**

199 Ten principal components (PCs) collectively contributed 86% of variation in the human BU  
200 occurrence locations. Seven PCs characterised the environmental locations. The variables selected  
201 for each model are shown in S3 Text. The mean contribution of the most important predictors of  
202 human BU occurrence dropped sharply at 150km, while the contribution of the predictors of  
203 environmental detection of *M. ulcerans* declined at 300km (S4 Text).

### 204 ***Environmental suitability for BU***

205 All individual distribution models performed well with ROC scores above 0.8 (S5 Text). Mean PCC  
206 scores were between 79.3 and 91.6% and mean TSS scores were between 0.57 and 0.79 (Figures A  
207 and B, S5 Text). RF performed best with a mean PCC of 91.6%, a mean TSS of 0.79 and mean ROC  
208 0.95. The final ensemble model showed an overall mean ROC of 0.96 with sensitivity of 87.1% and  
209 specificity of 92.9%. The mean TSS was 0.80 and the mean kappa score was 0.80 (Table 2).

210

211 Table 2: Validation metrics for ensemble models for BU and *M. ulcerans* suitability

		Weighted Mean	Lower CI	Upper CI
BU suitability	TSS	0.797	0.797	0.794
	ROC	0.963	0.961	0.963
	kappa	0.801	0.804	0.797
MU suitability	TSS	0.885	0.885	0.896
	ROC	0.96	0.96	0.961
	kappa	0.894	0.894	0.894

212

213 The minimum temperature of the coldest month was the strongest contributor to the RF models,  
214 followed by PET and distance to dams (Figure A in S6 Text). Environments with minimum  
215 temperature below 18°C were unsuitable, but suitability increased sharply at this temperature,  
216 decreased a little from 20-21°C, before increasing again, remaining high up to 24°C (Figure A in S7  
217 Text). Optimal values of PET were between 1,000 and 1,500mm per month, corresponding to  
218 tropical rainforest canopy cover. There was a gradual decline in suitability for BU with increasing  
219 distance to the nearest dam up to a distance of 80km, after which suitability dropped rapidly.

220 The overall distribution was constrained to humid tropical areas and local scale variation appeared  
221 to be driven by hydrological features and deforestation patterns (Figure 2). The total area predicted  
222 to be suitable for BU was 338,500 km<sup>2</sup>, and the total population living in areas predicted suitable was  
223 69.7 million (Table 3). Pockets of suitability for BU totalling >100 km<sup>2</sup> were predicted in 16 countries  
224 in Africa, including all 12 countries along the west-central African coastline from Liberia to Angola  
225 (S2 Table). Angola had the widest area predicted suitable, followed by the Democratic Republic of  
226 the Congo, although the patches of suitability predicted in these countries were associated with high  
227 uncertainty. Nigeria had the largest population at risk, with 17.8 million predicted to be living in  
228 areas suitable for BU, followed by the Democratic Republic of the Congo where 10.8 million were  
229 predicted to be living in suitable areas (S2 Table).

230 The model including all cases of BU (S3 Figure) gave similar results to the model including confirmed  
231 cases only. The Pearson coefficient of correlation between the two models was 0.86.

232 Figure 2: Predicted environmental suitability for the occurrence of BU disease in humans and  
233 associated error of prediction.

#### 234 ***Environmental suitability for M. ulcerans***

235 All models performed well with ROC above 0.8, apart from MAXENT Phillips which was excluded  
236 from the ensemble model. Mean PCC varied from 0.80- 0.89 between models and mean TSS was  
237 between 0.54 and 0.77 (Figures C and D, S5 Text). RF outperformed other algorithms in predicting

238 the occurrence of *M. ulcerans*. The final ensemble model had a mean TSS score of 0.87, with a  
239 sensitivity of 92.3 and specificity of 94.5% (Table 2). The ROC score was 0.99 and the kappa score was  
240 0.87.

241 The minimum temperature of the coldest month was the strongest predictor of *M. ulcerans*  
242 occurrence in the RF models, accounting for 24.6% of all variance in the model (Figure B in S6 Text).  
243 Distance to deforested areas was also a strong predictor, accounting for 23.7% of the variance.  
244 Suitability was low at coldest month minimum temperatures below 18°C, and increased sharply to a  
245 peak at 25°C. There was a strong response to the distance to deforested areas, with high suitability  
246 at close range to deforested areas, decreasing sharply at 25km (Figure B in S6 Text).

247 The overall distribution appeared to be restricted by suitability for minimum temperature and  
248 precipitation in the driest month, with local variation driven by deforestation (Figure 3). The total  
249 area predicted to be suitable for *M. ulcerans* was 833,975km<sup>2</sup>, and the total population living in  
250 areas predicted suitable was 71.2 million (Table 3). Pockets of suitability were predicted in 31  
251 countries (S2 Table). The DRC had the widest area predicted suitable (184,500 km<sup>2</sup>) followed by Côte  
252 d'Ivoire (120,600 km<sup>2</sup>). The highest population living in suitable areas was in Nigeria (26.7 million).

253 Figure 3: Predicted environmental suitability for the occurrence of *M. ulcerans* in the environment  
254 and associated error of prediction.

#### 255 **Overlap of suitability for BU and *M. ulcerans***

256 The total area predicted to be suitable for both BU and *M. ulcerans* was 126,775 km<sup>2</sup>, with 29.1  
257 million people predicted to be living in areas at risk. There were notable differences in the extents of  
258 the areas predicted suitable for BU disease and environmental *M. ulcerans* (Figure 4). There were  
259 wide areas in Sierra Leone and Cote d'Ivoire predicted suitable for *M. ulcerans* but not for BU  
260 disease, which was restricted to smaller pockets within these countries. Suitability for *M. ulcerans*  
261 was also predicted outside the predicted range of BU across central Africa and along the north-west  
262 coast of Mozambique, eastern Tanzania and coastal Kenya. In contrast, there were large patches in  
263 Central Africa predicted suitable for BU but not suitable for *M. ulcerans*. Ghana followed by  
264 Cameroon had the widest area predicted suitable for both BU and *M. ulcerans*. The highest  
265 populations living in areas predicted suitable for both BU and *M. ulcerans* were in Nigeria and  
266 Ghana, with 11.4 and 8.0 million respectively at risk.

267 Figure 4: Predicted overlap of environmental suitability for BU and of *M. ulcerans* occurrence.

268

269 Table 3: Total area predicted suitable and population in areas at risk for Buruli ulcer, *M. ulcerans*,  
270 and both, in Africa

	Total area suitable (km <sup>2</sup> )	<i>Lower bound</i>	<i>Upper bound</i>	Population in suitable areas	<i>Lower bound</i>	<i>Upper bound</i>
BU	338,500	283,950	425,600	69,740,630	62,573,091	83,050,754
MU	912,600	767,775	1,112,275	88,542,574	75,380,436	104,645,747
BU & MU	126,775	94,800	177,700	29,124,902	22,597,071	39,371,153

271

272 Suitability for BU and *M. ulcerans* is shown by country in S1 Supporting Maps.

## 273 **Discussion**

274 We have used ecological niche modelling to identify environmental factors associated with the  
275 occurrence of Buruli ulcer and its causative agent *M. ulcerans*, and to predict environmental  
276 suitability for the disease and bacterium across continental Africa. Incorporating existing data on BU  
277 distribution and a geo-environmental definition of the range of occurrences, the resulting maps  
278 represent evidence-based predictions within a relevant spatial context.

279 There was substantial overlap in the factors contributing to suitability for human cases and  
280 environmental occurrence. Both BU and *M. ulcerans* were constrained to particular bioclimatic  
281 zones by environmental predictors which varied over large areas, characterising the humid tropical  
282 realm where BU is endemic in Africa. Our finding that areas with minimum temperature less than  
283 18°C were unsuitable for BU supports evidence for a different epidemiology of the disease in Africa  
284 compared to endemic areas of temperate Australia and Japan [47]. Local-scale variation in factors  
285 including the distance to dams, deforestation and hydrology resulted in a patchy distribution of  
286 predicted suitability, consistent with our understanding of the epidemiology of BU, which is  
287 recognised to be highly focal in endemic settings [48].

288 We identified pockets of suitability for BU in patchy foci throughout the known-endemic range of  
289 the disease, particularly in the tropical zones of countries around the Gulf of Guinea. Throughout  
290 this range, suitability was predicted in regions not previously recognised as endemic. For example,  
291 two foci of suitability were predicted in Equatorial Guinea, which had no evidence of cases reported  
292 in peer-reviewed literature. However, these two areas correspond to the origins of a number of  
293 cases diagnosed by an expert in BU between 1995 and 2005 [49, 50].

294 Some locations in northern Cameroon with previous evidence of PCR confirmed BU were found to  
295 be unsuitable for the disease. This discordance may be due to the model's failure to identify suitable  
296 environments in areas of lower BU incidence. However, given the great volume of surveillance data  
297 collected by the well-established BU control programme in Cameroon, some patients are likely to

298 have been diagnosed outside the region where they acquired the disease [51], and we consider it  
299 plausible that some regions where BU has been recorded are not actually suitable for transmission.

300 The suitability maps provide a depiction of areas potentially at risk for BU beyond what is known  
301 from the distribution of reported cases, currently the basis for targeting of surveillance and control.  
302 Given the recognised scale of underreporting of BU [41], the current approach is likely to exclude  
303 cases outside of known disease foci, and we suggest that areas predicted suitable for BU could be  
304 considered as targets for case finding activities, with the aim of identifying unrecognised foci and  
305 patients not known to the health system. Based on the wide areas of suitability predicted by this  
306 work and existing evidence of under-reporting of BU [52], the south of Nigeria would be a key target  
307 for case finding activities. There were wide areas predicted suitable in the Republic of Congo, the  
308 DRC and Angola, although these predictions were associated with significant uncertainty, which  
309 should be considered in the design of any future surveys. Suitable areas of Equatorial Guinea with  
310 historical evidence of cases would also be targets for case finding, although in this case the predicted  
311 suitable areas are more restricted, potentially necessitating a more stratified approach. A  
312 comparable approach has been applied to target malaria elimination efforts to transmission  
313 hotspots predicted through geospatial risk mapping [53], employing environmental modelling to  
314 impute risk in the absence of full surveillance coverage.

315 The model predictions could also be used to inform the design of cross-sectional surveys for BU,  
316 combining exhaustive case searches with environmental modelling to achieve robust estimates of  
317 prevalence. In a nationwide survey for podocooniosis in Cameroon, the selection of survey  
318 communities was stratified according predicted suitability for the disease based on a model trained  
319 mainly using data from Ethiopia [54]. This survey identified higher rates of podocooniosis in  
320 communities that were predicted suitable, implying a benefit in terms of the cost per case identified,  
321 compared to a survey employing random selection of survey communities.

322 The validation metrics we calculated demonstrate the ability of the models to predict BU and *M.*  
323 *ulcerans* occurrence with high accuracy. However, these measures do not indicate the models'  
324 generalisability to areas beyond the range of known locations. Validation against external datasets  
325 would be required to assess this quality and is a target for future analysis.

326 The scale of analysis (grid cells at 5km x 5km) may have limited our ability to quantify the effect of  
327 predictors varying over small geographical scales. For example, the relatively small contribution of  
328 distance to waterways on suitability for BU was a surprising result, given that proximity to and  
329 contact with rivers have been identified as risk factors for BU disease [13, 18]. However, the  
330 variation we were able to capture in this predictor was limited, since most of the land area in  
331 tropical and sub-tropical Africa is within 10km of a river or stream, and much is within 5km. A  
332 previous analysis of land use and landcover and BU presence at large spatial scales in Benin found no  
333 association of BU at community level with proximity to water bodies [55]. The scale of analysis may  
334 also have limited our ability to capture fine scale variation in environmental suitability for BU. Our  
335 models predicted large contiguous areas of suitability in some areas with suitable bioclimatic  
336 conditions and within close proximity to stable night lights and deforested areas. Such areas may be  
337 suitable in reality, but exhibit an uneven distribution of disease due to factors not included in our  
338 models.

339 We have identified areas of high suitability for BU and *M. ulcerans* within known endemic-areas, and  
340 in areas not currently recognised as endemic, but with evidence of possible undiagnosed or  
341 misdiagnosed BU. The population at highest risk of BU is within areas where BU and *M. ulcerans*  
342 niches overlap, comprising almost 30 million people in 2020. The focal nature of BU distribution, the  
343 recognised scale of under-detection, and the impact of late diagnosis on disease severity strongly  
344 suggest a targeted approach to active case finding as a means to control this disease. The fine-scale,  
345 evidence-based predictions presented here could provide a tool to target such efforts, which will

346 help to increase the proportion of cases linked to treatment, and contribute further research to

347 establish the burden and distribution of this devastating disease.

348

349



350 **Acknowledgements**

351 The AIM Initiative was the sole funder of this work. We would like to recognise the contribution of  
352 all health workers, researchers and data managers who recorded cases of Buruli ulcer which were  
353 compiled into the global database of infections underpinning this study.

354 **Author contributions**

355 RP and JC acquired the funding that enabled this project. JC conceptualised the study. JC and HS  
356 together developed the investigation and methodology- including the software (code for the  
357 analysis), and produced the visualisations presented. HS curated the data analysed, some of which  
358 was originally collected by other authors: ET, RP, MF, IM and JT. HS undertook the formal analysis  
359 and prepared the original draft. All authors critically reviewed the draft.

360

361

362 **Supporting Information**

363 **S1 Text:** Weighting of occurrence points

364 **S2 Text:** Selection of background and pseudoabsence points

365 **S3 Text:** Environmental variables used in modelling

366 **S1 Figure:** Distribution and weights of confirmed BU occurrence locations, background and  
367 pseudoabsence points

368 **S2 Figure:** Distribution and weights of environmental *M. ulcerans* DNA occurrence locations,  
369 background and pseudoabsence points

370 **S3 Figure:** Predicted environmental suitability for the occurrence of BU disease in humans and  
371 associated error of prediction, including all clinically diagnosed cases of BU

372 **S4 Text:** Spatial structure of occurrence predictors

373 **S5 Text:** Individual model performance evaluation statistics

374 **S6 Text:** Variable importance plots for random forest models

375 **S7 Text:** Marginal effect plots for random forest models

376 **S1 Table:** Total area predicted suitable and population living in suitable areas for Buruli ulcer, *M.*  
377 *ulcerans*, and both, by country in African continent.

378 **References**

- 379 1. Walsh DS, Portaels F, Meyers WM. Buruli ulcer (Mycobacterium ulcerans infection).  
380 Transactions of the Royal Society of Tropical Medicine and Hygiene. 2008;102(10):969-78. Epub  
381 2008/07/29. doi: 10.1016/j.trstmh.2008.06.006. PubMed PMID: 18657836.
- 382 2. Roltgen K, Pluschke G. Epidemiology and disease burden of Buruli ulcer: a review. Research  
383 and Reports in Tropical Medicine. 2015;6:59-73. doi: 10.2147/rrtm.s62026. PubMed PMID:  
384 WOS:000219035500008.
- 385 3. Amoakoh HB, Aikins M. Household cost of out-patient treatment of Buruli ulcer in Ghana: a  
386 case study of Obom in Ga South Municipality. BMC health services research. 2013;13:507. Epub  
387 2013/12/10. doi: 10.1186/1472-6963-13-507. PubMed PMID: 24313975; PubMed Central PMCID:  
388 PMCPCmc4029088.
- 389 4. Chukwu JN, Meka AO, Nwafor CC, Oshi DC, Madichie NO, Ekeke N, et al. Financial burden of  
390 health care for Buruli ulcer patients in Nigeria: the patients' perspective. International health.  
391 2017;9(1):36-43. Epub 2016/12/18. doi: 10.1093/inthealth/ihw056. PubMed PMID: 27986841.
- 392 5. Duker AA, Portaels F, Hale M. Pathways of Mycobacterium ulcerans infection: a review.  
393 Environment international. 2006;32(4):567-73. Epub 2006/02/24. doi: 10.1016/j.envint.2006.01.002.  
394 PubMed PMID: 16492390.
- 395 6. Rodhain F. [Buruli ulcer: hypothetical modes of transmission of Mycobacterium ulcerans].  
396 Bulletin de l'Academie nationale de medecine. 2012;196(3):685-90; discussion 90-1. Epub  
397 2013/03/12. PubMed PMID: 23472356.
- 398 7. Organisation WH. Buruli ulcer (Mycobacterium ulcerans infection) 2019 [27/08/2019].  
399 Available from: [https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-\(mycobacterium-](https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection))  
400 [ulcerans-infection\)](https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection)).
- 401 8. Phillips RO, Robert J, Abass KM, Thompson W, Sarfo FS, Wilson T, et al. Rifampicin and  
402 clarithromycin (extended release) versus rifampicin and streptomycin for limited Buruli ulcer lesions:  
403 a randomised, open-label, non-inferiority phase 3 trial. The Lancet. 2020.
- 404 9. Simpson H, Deribe K, Tabah EN, Peters A, Maman I, Frimpong M, et al. Mapping the global  
405 distribution of Buruli ulcer: a systematic review with evidence consensus. Lancet Glob Health.

- 406 2019;7(7):e912-e22. Epub 2019/06/16. doi: 10.1016/S2214-109X(19)30171-8. PubMed PMID:  
407 31200890; PubMed Central PMCID: PMC6614043.
- 408 10. Kanga JM. Aspects épidémiologiques de l'ulcère de Buruli en Côte d'Ivoire : résultats d'une  
409 enquête nationale. Bull Soc Pathol Exot. 2001;94:46-51.
- 410 11. Amofah G, Bonsu F, Tetteh C, Okrah J, Asamoah K, Asiedu K, et al. Buruli ulcer in Ghana:  
411 results of a national case search. Emerging infectious diseases. 2002;8(2):167-70. Epub 2002/03/19.  
412 doi: 10.3201/eid0802.010119. PubMed PMID: 11897068; PubMed Central PMCID: PMC6614043.
- 413 12. Noeske J, Kuaban C, Rondini S, Sorlin P, Ciaffi L, Mbuagbaw J, et al. Buruli ulcer disease in  
414 Cameroon rediscovered. The American journal of tropical medicine and hygiene. 2004;70(5):520-6.  
415 Epub 2004/05/25. PubMed PMID: 15155984.
- 416 13. Garchitorena A, Roche B, Kamgang R, Ossomba J, Babonneau J, Landier J, et al.  
417 Mycobacterium ulcerans ecological dynamics and its association with freshwater ecosystems and  
418 aquatic communities: results from a 12-month environmental survey in Cameroon. PLoS neglected  
419 tropical diseases. 2014;8(5):e2879. doi: 10.1371/journal.pntd.0002879. PubMed PMID: 24831924;  
420 PubMed Central PMCID: PMC4022459.
- 421 14. Williamson HR, Benbow ME, Nguyen KD, Beachboard DC, Kimbirauskas RK, McIntosh MD, et  
422 al. Distribution of Mycobacterium ulcerans in buruli ulcer endemic and non-endemic aquatic sites in  
423 Ghana. PLoS neglected tropical diseases. 2008;2(3):e205. Epub 2008/03/28. doi:  
424 10.1371/journal.pntd.0000205. PubMed PMID: 18365034; PubMed Central PMCID:  
425 PMC6614043.
- 426 15. Campbell LP, Finley AO, Benbow ME, Gronseth J, Small P, Johnson RC, et al. Spatial Analysis  
427 of Anthropogenic Landscape Disturbance and Buruli Ulcer Disease in Benin. PLoS neglected tropical  
428 diseases. 2015;9(10):e0004123. Epub 2015/10/17. doi: 10.1371/journal.pntd.0004123. PubMed  
429 PMID: 26474482; PubMed Central PMCID: PMC6614043.
- 430 16. Wu UI, Holland SM. Host susceptibility to non-tuberculous mycobacterial infections. The  
431 Lancet Infectious diseases. 2015;15(8):968-80. Epub 2015/06/08. doi: 10.1016/s1473-  
432 3099(15)00089-4. PubMed PMID: 26049967.
- 433 17. Landier J, Gaudart J, Carolan K, Lo Seen D, Guegan JF, Eyangoh S, et al. Spatio-temporal  
434 patterns and landscape-associated risk of Buruli ulcer in Akonolinga, Cameroon. PLoS neglected

- 435 tropical diseases. 2014;8(9):e3123. Epub 2014/09/05. doi: 10.1371/journal.pntd.0003123. PubMed  
436 PMID: 25188464; PubMed Central PMCID: PMC4154661.
- 437 18. Merritt RW, Walker ED, Small PL, Wallace JR, Johnson PD, Benbow ME, et al. Ecology and  
438 transmission of Buruli ulcer disease: a systematic review. PLoS neglected tropical diseases.  
439 2010;4(12):e911. Epub 2010/12/24. doi: 10.1371/journal.pntd.0000911. PubMed PMID: 21179505;  
440 PubMed Central PMCID: PMC3001905.
- 441 19. Portaels F, Meyers WM, Ablordey A, Castro AG, Chemlal K, de Rijk P, et al. First cultivation  
442 and characterization of Mycobacterium ulcerans from the environment. PLoS neglected tropical  
443 diseases. 2008;2(3):e178. Epub 2008/03/28. doi: 10.1371/journal.pntd.0000178. PubMed PMID:  
444 18365032; PubMed Central PMCID: PMC2268003.
- 445 20. Aboagye SY, Danso E, Ampah KA, Nakobu Z, Asare P, Otchere ID, et al. Isolation of  
446 Nontuberculous Mycobacteria from the Environment of Ghanaian Communities Where Buruli Ulcer Is  
447 Endemic. Applied and environmental microbiology. 2016;82(14):4320-9. Epub 2016/05/22. doi:  
448 10.1128/aem.01002-16. PubMed PMID: 27208141; PubMed Central PMCID: PMC4959205.
- 449 21. Marsollier L, Aubry J, Milan G, Brodin P. Aquatic insects and transmission of Mycobacterium  
450 ulcerans. M S-Med Sci. 2007;23(6-7):572-5. doi: 10.1051/medsci/20072367572. PubMed PMID:  
451 WOS:000248183300005.
- 452 22. Willson SJ, Kaufman MG, Merritt RW, Williamson HR, Malakauskas DM, Benbow ME. Fish  
453 and amphibians as potential reservoirs of Mycobacterium ulcerans, the causative agent of Buruli  
454 ulcer disease. Infection ecology & epidemiology. 2013;3. Epub 2013/02/27. doi:  
455 10.3402/iee.v3i0.19946. PubMed PMID: 23440849; PubMed Central PMCID: PMC3580280.
- 456 23. Eric Benbow M, Kimbirauskas R, McIntosh MD, Williamson H, Quaye C, Boakye D, et al.  
457 Aquatic macroinvertebrate assemblages of Ghana, West Africa: understanding the ecology of a  
458 neglected tropical disease. EcoHealth. 2014;11(2):168-83. Epub 2013/12/07. doi: 10.1007/s10393-  
459 013-0886-7. PubMed PMID: 24306551.
- 460 24. Carson C, Lavender CJ, Handasyde KA, O'Brien CR, Hewitt N, Johnson PD, et al. Potential  
461 wildlife sentinels for monitoring the endemic spread of human buruli ulcer in South-East australia.  
462 PLoS neglected tropical diseases. 2014;8(1):e2668. Epub 2014/02/06. doi:  
463 10.1371/journal.pntd.0002668. PubMed PMID: 24498452; PubMed Central PMCID:  
464 PMC3907424.

- 465 25. McIntosh M, Williamson H, Benbow ME, Kimbirauskas R, Quaye C, Boakye D, et al.  
466 Associations between *Mycobacterium ulcerans* and aquatic plant communities of West Africa:  
467 implications for Buruli ulcer disease. *EcoHealth*. 2014;11(2):184-96. Epub 2014/01/21. doi:  
468 10.1007/s10393-013-0898-3. PubMed PMID: 24442959.
- 469 26. Marion E, Chauty A, Yeramian E, Babonneau J, Kempf M, Marsollier L. A case of guilt by  
470 association: Water bug bite incriminated in *M. ulcerans* infection. *International journal of*  
471 *mycobacteriology*. 2014;3(2):158-61. Epub 2014/06/01. doi: 10.1016/j.ijmyco.2014.01.004. PubMed  
472 PMID: 26786340.
- 473 27. Williamson HR, Benbow ME, Campbell LP, Johnson CR, Sopoh G, Barogui Y, et al. Detection  
474 of *Mycobacterium ulcerans* in the environment predicts prevalence of Buruli ulcer in Benin. *PLoS*  
475 *neglected tropical diseases*. 2012;6(1):e1506. Epub 2012/02/04. doi: 10.1371/journal.pntd.0001506.  
476 PubMed PMID: 22303498; PubMed Central PMCID: PMC269429.
- 477 28. Johnson PDR. Buruli Ulcer in Australia. In: Pluschke G, Röltgen K, editors. *Buruli Ulcer:*  
478 *Mycobacterium Ulcerans Disease*. Cham: Springer International Publishing; 2019. p. 61-76.
- 479 29. Wu J, Tschakert P, Klutse E, Ferring D, Ricciardi V, Hausermann H, et al. Buruli Ulcer Disease  
480 and Its Association with Land Cover in Southwestern Ghana. *PLoS neglected tropical diseases*.  
481 2015;9(6):e0003840. Epub 2015/06/20. doi: 10.1371/journal.pntd.0003840. PubMed PMID:  
482 26091265; PubMed Central PMCID: PMC4474842.
- 483 30. N'Krumah R T, Kone B, Tiembre I, Cisse G, Pluschke G, Tanner M, et al. Socio-Environmental  
484 Factors Associated with the Risk of Contracting Buruli Ulcer in Tiassale, South Cote d'Ivoire: A Case-  
485 Control Study. *PLoS neglected tropical diseases*. 2016;10(1):e0004327. Epub 2016/01/09. doi:  
486 10.1371/journal.pntd.0004327. PubMed PMID: 26745723; PubMed Central PMCID:  
487 PMC4712845.
- 488 31. Marston B, Diallo MO, Horsburgh CR, Diomande I. Emergence of Buruli ulcer Disease in the  
489 Daloa region of Côte d'Ivoire. *The American journal of tropical medicine and hygiene*. 1995;52:291-  
490 24.
- 491 32. Raghunathan PL, Whitney EA, Asamoah K, Stienstra Y, Taylor TH, Jr., Amofah GK, et al. Risk  
492 factors for Buruli ulcer disease (*Mycobacterium ulcerans* Infection): results from a case-control study  
493 in Ghana. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of*  
494 *America*. 2005;40(10):1445-53. Epub 2005/04/22. doi: 10.1086/429623. PubMed PMID: 15844067.

- 495 33. Aiga H, Amano T, Cairncross S, Adomako J, Nanas OK, Coleman S. Assessing water-related  
496 risk factors for Buruli ulcer: a case-control study in Ghana. *The American journal of tropical medicine*  
497 *and hygiene*. 2004;71(4):387-92. Epub 2004/11/02. PubMed PMID: 15516631.
- 498 34. Simpson H, Deribe K, Tabah EN, Maman I, Frimpong M, Ampadu E, Saunderson P, Pullan,  
499 RL and Cano, J. Occurrences of Buruli ulcer infections in humans and *Mycobacterium ulcerans* in  
500 animal and environmental samples. 2018.
- 501 35. Fyfe JA, Lavender CJ, Johnson PD, Globan M, Sievers A, Aзуolas J, et al. Development and  
502 application of two multiplex real-time PCR assays for the detection of *Mycobacterium ulcerans* in  
503 clinical and environmental samples. *Applied and environmental microbiology*. 2007;73(15):4733-40.  
504 Epub 2007/05/29. doi: 10.1128/aem.02971-06. PubMed PMID: 17526786; PubMed Central PMCID:  
505 PMCPMC1951036.
- 506 36. Lavender CJ, Stinear TP, Johnson PD, Aзуolas J, Benbow ME, Wallace JR, et al. Evaluation of  
507 VNTR typing for the identification of *Mycobacterium ulcerans* in environmental samples from  
508 Victoria, Australia. *FEMS microbiology letters*. 2008;287(2):250-5. Epub 2008/08/30. doi:  
509 10.1111/j.1574-6968.2008.01328.x. PubMed PMID: 18754785.
- 510 37. Stragier P, Ablordey A, Meyers WM, Portaels F. Genotyping *Mycobacterium ulcerans* and  
511 *Mycobacterium marinum* by using mycobacterial interspersed repetitive units. *J Bacteriol*.  
512 2005;187(5):1639-47. doi: 10.1128/jb.187.5.1639-1647.2005. PubMed PMID:  
513 WOS:000227191600011.
- 514 38. Elith\* J, H. Graham\* C, P. Anderson R, Dudík M, Ferrier S, Guisan A, et al. Novel methods  
515 improve prediction of species' distributions from occurrence data. *Ecography*. 2006;29(2):129-51.
- 516 39. Hijmans RJaE, J. Species distribution modeling with R. 2017.
- 517 40. Senay SD, Worner SP, Ikeda T. Novel three-step pseudo-absence selection technique for  
518 improved species distribution modelling. *PloS one*. 2013;8(8):e71218.
- 519 41. Simpson H, Deribe K, Tabah EN, Peters A, Maman I, Frimpong M, et al. Mapping the global  
520 distribution of Buruli ulcer: a systematic review with evidence consensus. *The Lancet Global Health*.  
521 2019;7(7):e912-e22.

- 522 42. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species  
523 distribution models: how, where and how many? *Methods in ecology and evolution*. 2012;3(2):327-  
524 38.
- 525 43. Stoffel V, Barthelme B, Chague F. [Tropical ecopathology: up hill and down dale Buruli ulcer].  
526 *Sante publique (Vandoeuvre-les-Nancy, France)*. 2005;17(2):191-7. Epub 2005/07/09. PubMed  
527 PMID: 16001561.
- 528 44. Wilfried Thuiller DG, Robin Engler and Frank Breiner. biomod2: Ensemble Platform for  
529 Species Distribution Modeling. R package version 3.3-15/r728. 2017.
- 530 45. Liu C, White M, Newell G. Measuring and comparing the accuracy of species distribution  
531 models with presence-absence data. *Ecography*. 2011;34(2):232-43.
- 532 46. School of Geography and Environmental Science UoS. WorldPop Africa Continental  
533 Population Datasets (2000- 2020). WorldPop 2018.
- 534 47. Yotsu RR, Suzuki K, Simmonds RE, Bedimo R, Ablordey A, Yeboah-Manu D, et al. Buruli Ulcer:  
535 a Review of the Current Knowledge. *Current Tropical Medicine Reports*. 2018:1-10.
- 536 48. Roltgen K, Qi W, Ruf MT, Mensah-Quainoo E, Pidot SJ, Seemann T, et al. Single nucleotide  
537 polymorphism typing of *Mycobacterium ulcerans* reveals focal transmission of buruli ulcer in a highly  
538 endemic region of Ghana. *PLoS neglected tropical diseases*. 2010;4(7):e751. Epub 2010/07/24. doi:  
539 10.1371/journal.pntd.0000751. PubMed PMID: 20652033; PubMed Central PMCID:  
540 PMCPMC2907412.
- 541 49. III IdSC. Buruli ulcer in Equatorial Guinea [presentation]. In press 2006.
- 542 50. Ham A. AFIP Mycobacteriology Chief continues to attack Buruli ulcer in Third World  
543 countries. *the AFIP Letter*. 2000;158(3):5.
- 544 51. Agbo I. The gendered impact of Buruli ulcer on the household production of health: why  
545 decentralization favours women. *Tropical Medicine & International Health*. 2015;20:298-. PubMed  
546 PMID: WOS:000360758801329.
- 547 52. Vincent QB, Ardant MF, Adeye A, Goundote A, Saint-Andre JP, Cottin J, et al. Clinical  
548 epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *The Lancet Global  
549 health*. 2014;2(7):e422-30. Epub 2014/09/05



550 2014/08/12. doi: 10.1371/journal.pntd.0003123

551 10.1016/s2214-109x(14)70223-2. PubMed PMID: 25103396.

552 53. Clements AC, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: how can  
553 geospatial science help to achieve malaria elimination? *The Lancet Infectious Diseases*.

554 2013;13(8):709-18. doi: 10.1016/S1473-3099(13)70140-3.

555 54. Deribe K, Beng AA, Cano J, Njouendo AJ, Fru-Cho J, Awah AR, et al. Mapping the  
556 geographical distribution of podoconiosis in Cameroon using parasitological, serological, and clinical  
557 evidence to exclude other causes of lymphedema. *PLoS neglected tropical diseases*. 2018;12(1).

558 55. Wagner T, Benbow ME, Burns M, Johnson RC, Merritt RW, Qi J, et al. A Landscape-based  
559 model for predicting *Mycobacterium ulcerans* infection (Buruli Ulcer disease) presence in Benin,  
560 West Africa. *EcoHealth*. 2008;5(1):69-79. Epub 2008/07/24. doi: 10.1007/s10393-007-0148-7.

561 PubMed PMID: 18648799.

562

563



6,027 locations  
in BU database



5,627 locations with recorded  
cases



3,533 point locations with reliable  
geo-coordinates



3,318 locations in continental  
Africa



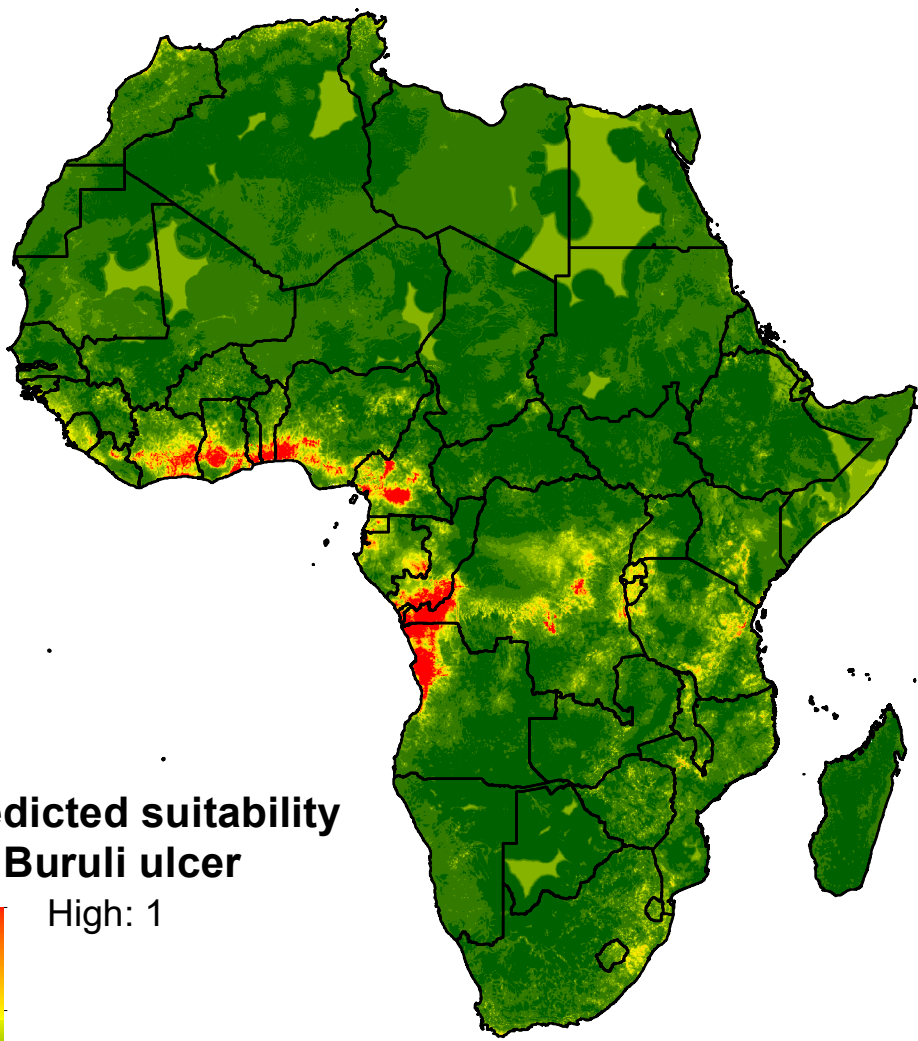
3,318 locations with clinically-  
diagnosed cases

1,336 locations with cases  
confirmed by PCR/histopathology

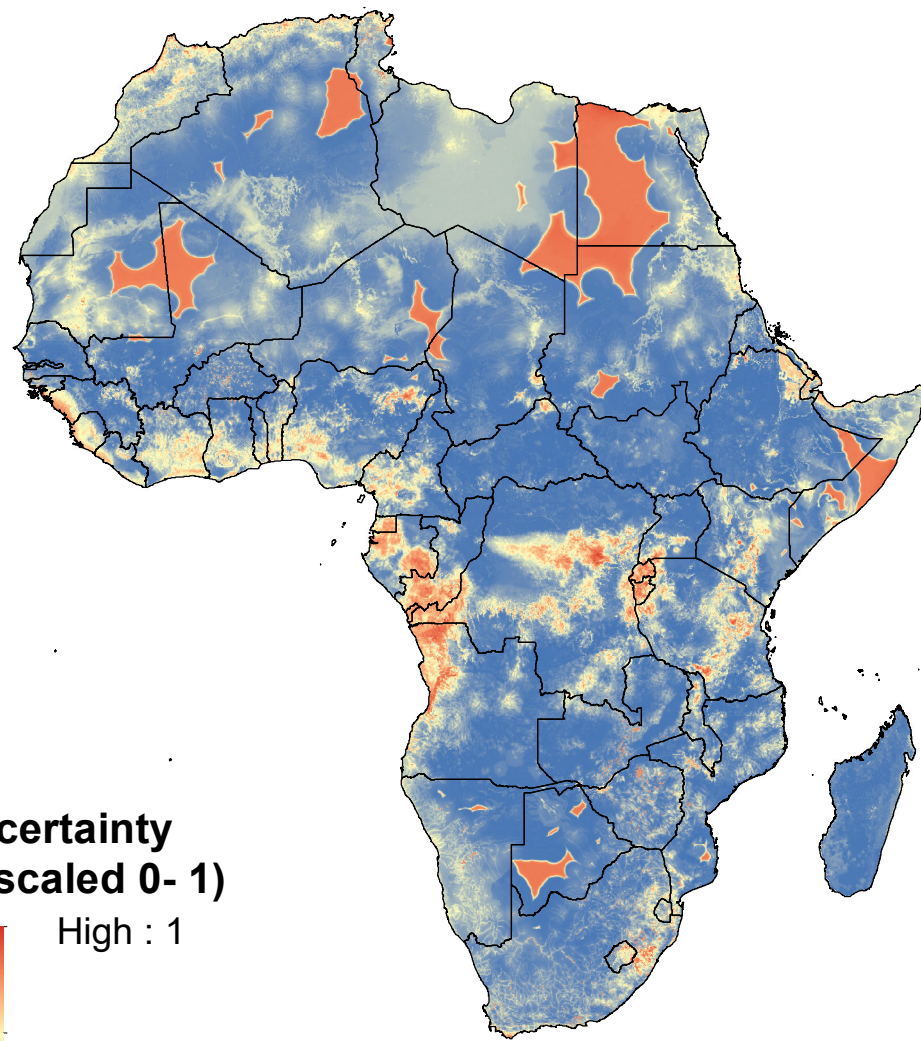
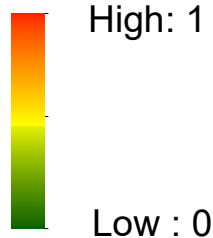


2,183 unique locations

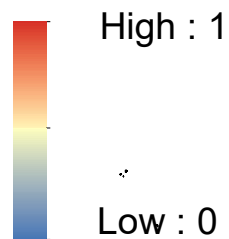
738 unique locations

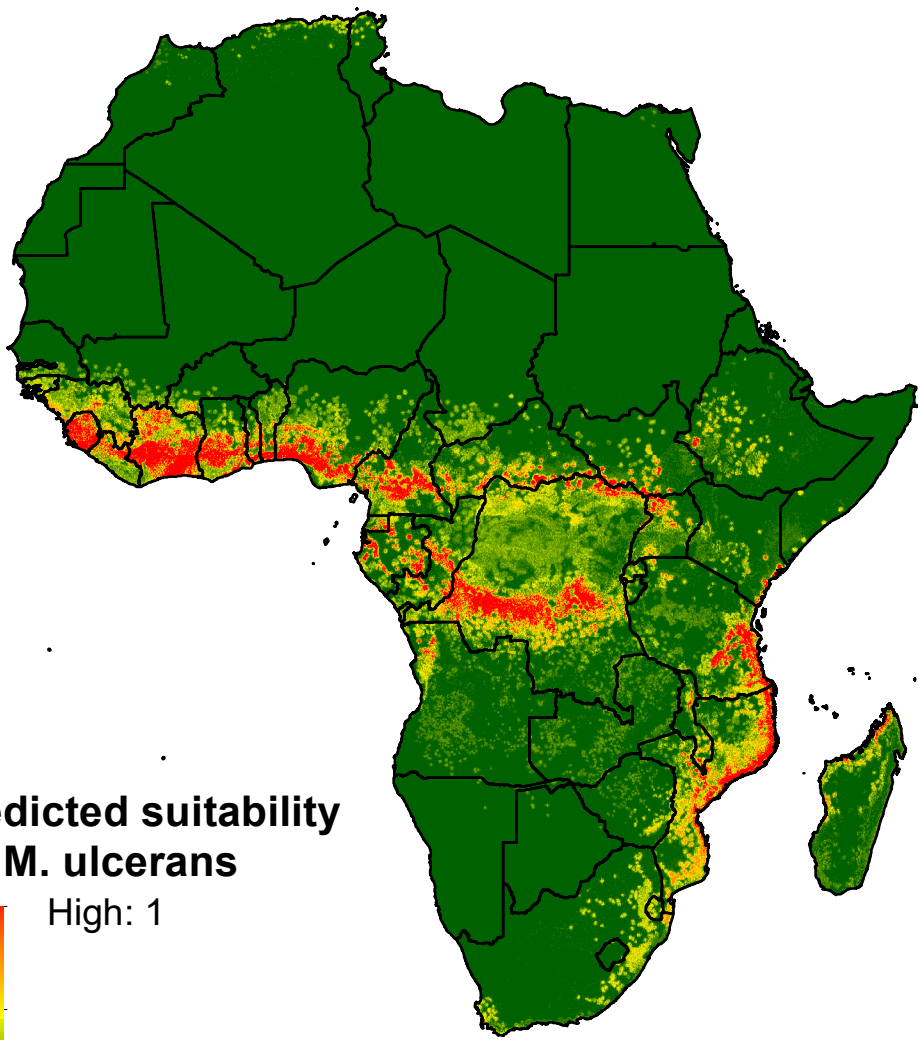


**Predicted suitability  
for Buruli ulcer**

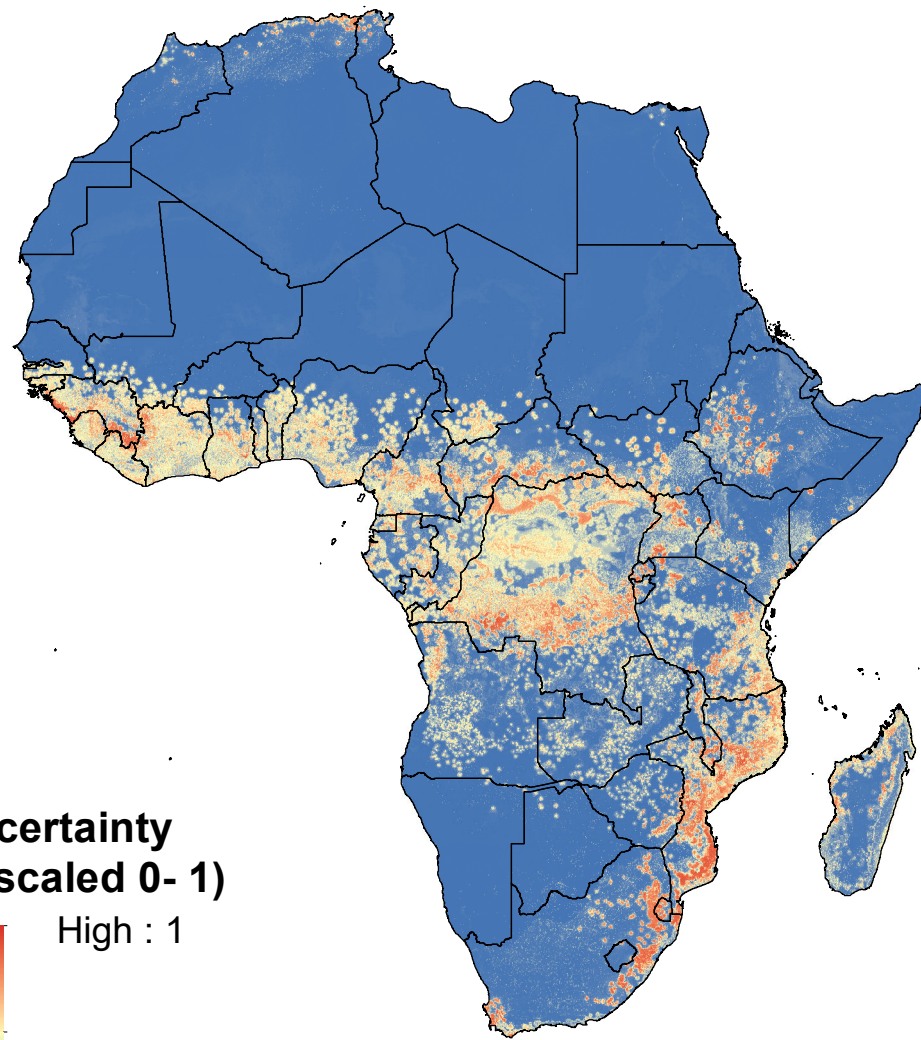
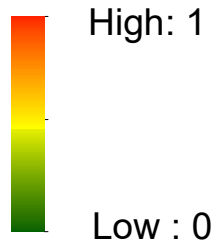


**Uncertainty  
(rescaled 0- 1)**





**Predicted suitability  
for *M. ulcerans***



**Uncertainty  
(rescaled 0- 1)**

