MAXIMUM ENTROPY NICHE MODELLING TO ESTIMATE THE POTENTIAL DISTRIBUTION OF *PHYTOPHTHORA MEGAKARYA* BRASIER & M. J. GRIFFIN (1979) IN TROPICAL REGIONS

Maxwell C. Obiakara[1] (ORCID: 0000-0002-0635-8068),
Peter M. Etaware[1] (ORCID: 0000-0002-9370-8029),
Kanayo S. Chukwuka[1] (ORCID: 0000-0002-8050-0552)

[1] *Plant Ecology Unit, Department of Botany, University of Ibadan, Ibadan, Nigeria*
[1] *Plant Pathology Unit, Department of Botany, University of Ibadan, Ibadan, Nigeria*

Abstract.

Background: Phytophthora megakarya *is an invasive pathogen endemic to Central and West Africa. This species causes the most devastating form of black pod disease of cacao (*Theobroma cacao*). Despite the deleterious impacts of this disease on cocoa production, there is no information on the geographic distribution of* P. megakarya.

Aim: In this study, we investigated the potential geographic distribution of P. megakarya *in cocoa-producing regions of the world using ecological niche modelling.*

Methods: Occurrence records of P. megakarya *in Central and West Africa were compiled from published studies. We selected relevant climate and soil variables in the indigenous range of this species to generate 14 datasets of climate-only, soil-only, and a combination of both data types. For each dataset, we calibrated 100 candidate MaxEnt models using 20 regularisation multiplier (0.1−1.0 at 0.1 interval, 2−4 at 0.5 interval, 4−8 at 1 interval, and 10) and five feature classes. The best model was selected from statistically significant candidates with an omission rate ≤ 5% and the lowest Akaike Information Criterion corrected for small sample sizes, and projected onto cocoa-producing regions in Southeast Asia, Central and South America. The risk of extrapolation in model transfer was measured using the mobility-oriented parity (MOP) metric.*

Results: We found an optimal goodness-of-fit and complexity for candidate models incorporating both climate and soil data. Predictions of the model with the best performance showed that nearly all of Central Africa, especially areas in Gabon, Equatorial Guinea, and southern Cameroon are at risk of black pod disease. In West Africa, suitable environments were observed along the Atlantic coast, from southern Nigeria to Gambia. Our analysis suggested that P. megakarya is capable of subsisting outside its native range, at least in terms of climatic and edaphic factors. Model projections identified likely suitable areas, especially in Brazil and Colombia, from southwestern Mexico down to Panama, and across the Caribbean islands in the Americas, and in Sri Lanka, Indonesia, Malaysia, and Papua New Guinea in Asia and adjacent areas

Conclusion: The outcomes of this study would be useful for developing measures aimed at preventing the spread of this pathogen in the tropics.

Key words: Black pod disease; Cocoa; MaxEnt; kuenm; West Africa; Central Africa.

1. Introduction

The oomycete genus *Phytophthora* essentially comprises pathogenic species of economic and ecological importance. These organisms have been responsible for devastating plant diseases worldwide, including potato blight, jarrah dieback and sudden oak death. This genus has gained an important phytopathological status in the last two decades, with more than 100 newly described species, classified on the basis of their invasiveness, either as cosmopolitan generalists or agriculture-related specialists (Scott et al., 2019). Although many of these species are pathogenic to cocoa (Table S1, Supplementary Material),

P. megakarya is the most economically important given its severe impacts and recent westward incursion in Africa from Cameroon (Akrofi, 2015; Bailey et al., 2016). The origin of this pathogen is not clear; its earliest occurrences were recorded in southwestern Cameroon in 1906, following an unusually high black pod disease incidence that caused cocoa yield losses about three times higher than those due to *P. palmivora*. Although *P. megakarya* has not been reported in cocoa-growing regions outside Africa, new introductions of this species cannot be overlooked especially in this era of increased global exchanges.

Phytophthora megakarya thrives in very humid climates. Although soil texture plays a central role in the epidemiology of some *Phytophthora* species (Corcobado et al., 2013), there is a dearth of information on the effect of edaphic factors on the persistence of *P. megakarya*. A sharp contrast exists between weather conditions required for zoospore production in laboratory and field conditions (Bailey et al., 2016). In the field, rainfall rather than temperature is critical for development of epidemics caused by *P. megakarya* (Deberdt et al., 2008). The amount of precipitation required for the survival of *P. megakarya* (1100–2000 mm) nearly matches that which is required for cocoa production (Akrofi, 2015). The minimum temperature for the survival of *P. megakarya* is 10–11º C, while the optimum temperature range for spore production and pathogenic activities is 24–26ºC (Bailey et al., 2016).

Under the conditions described above, the soil-borne spores of *P. megakarya* develop into sporangia which are dispersed from cocoa stems to pods, and eventually up the foliage mainly by rain splash (Bailey et al., 2016). Adaptations that make *P. megakarya* specifically harmful to cocoa include its spores' ability to survive in soil for extended periods and sustain multiple cycles of infestations. In addition, *P. megakarya* is more prolific in zoopore production, and readily forms sizeable necrotic lesions on cocoa pods compared to *P. palmivora* (Ali et al., 2016), and symptoms on pods (rot), branches and trunks (canker) are more conspicuous (Bailey et al., 2016). Lesions on pods as a result of *P. megakarya* infestation can drastically reduce cocoa bean yield, while canker considerably shortens its life span.

Predictive modelling to determine the distributional potential of species based on abiotic conditions is a widely used tool in ecology and biodiversity conservation (Franklin, 2009; Peterson et al., 2011). Advances in ecological niche modelling have provided opportunities to explore and predict the spatial extents of environmentally conducive habitats for disease-causing microorganisms (Peterson, 2014). In view of the serious environmental and economic impacts of *Phytophthora* species worldwide, diverse modelling approaches have been used to identify potentially suitable habitats for them (Burgess et al., 2017; Duque-Lazo et al., 2016; Hernández-Lambraño et al., 2018; La Manna et al., 2012; Vaclavık & Meentemeyer, 2011; Ramírez-Gil & Peterson, 2019). For example, Burgess et al., (2017) predicted large areas with ideal environmental conditions for *P. cinnamomi* in Sub-Saharan Africa, Southeast Asia, Europe, and America. However, the geographic coverage of most of these studies is limited, perhaps owing to the low dispersal ability of fungal and fungal-like species, or limited data availability. Considering the risk of a widespread outbreak of black pod disease in Africa as a result of increased East-West exchanges, we assessed the ecological niche of *P. megakarya* based on current climatic and edaphic factors in West and Central Africa, and provided information on its potential spatial distribution, more broadly across tropical regions, where cocoa is a major agricultural product.

2. Materials and Methods
2. 1 Study area, occurrence data cleaning and partitioning

The study area encompasses all cocoa-producing countries across Africa, Central and South America, and Southeast Asia, based on country-specific yield data from the Food and Agricultural Organisation of the United Nation database (http://www.fao.org/faostat/en/#data). Available geo-referenced occurrence records of *P. megakarya*, which were all confined to West and Central Africa were gathered from published studies. The raw occurrence dataset contained 753 records across seven countries: Cameroon, Gabon, Sao Tome & Principe, Nigeria, Togo, Ghana and Cote d'Ivoire. The bulk of these data (96.4%) were already compiled in a doctoral thesis (Mfegue, 2012) and the remaining were obtained from studies published prior to 2013 (Table S2, Supplementary Material).

Data cleaning was done manually; dubious and duplicate records were identified when visualized in relation to their provenances using ArcMap 10.4 (https://www.esri.com/). For instance, many records were listed in Cote d'Ivoire, Ghana, and Sao Tome and Principe but mapped in other countries and the Atlantic Ocean. Others mapped right on the coastline or very close to it (≤0.3 km), and were excluded to avoid ambiguity. Finally, to reduce spatial bias, occurrences were sub-sampled based on a minimal distance of 10 km between any two records using SDM Toolbox 2.4 (Brown, 2014). This cleaning and reduction process resulted in 92 useful occurrence records.

All occurrences of *P. megakarya* were restricted to tropical rainforest and savanna regions in Central and West Africa, which correspond to three subtypes

(Af, Am and Aw) of the Köppen-Geiger climate classification (Beck et al., 2018). To reduce uncertainties associated with choice of model calibration extent (Barve et al., 2011), we assumed that current populations of *P. megakarya* would be able to disperse across the geographical footprint within Af, Am and Aw in Africa (Fig. 1). Thus, model calibration was restricted to tropical rainforest and savanna climates in Africa. This spatial extent, represent the accessible area for *P. megakarya*, and it is referred to as "**M**" in the Biotic-Abiotic-Movement framework of species distributions proposed by Sobéron & Peterson (2005).

Spatial relatedness in occurrence data, as shown by the high density of records of *P. megakarya* in southern Cameroon (Fig. 1b) has been shown to affect model predictions, especially when projections are made in non-analogous habitats (Roberts et al., 2017). To ensure a better degree of independence in training and testing data, we divided occurrence data equally into four distinct spatial calibration sets (bins) using the *ENMeval* package (Muscarella et al., 2014). We

Figure 1: Study area. Cocoa-producing countries outside Africa with no known records of *P. megakarya* (CPCOA) are shown in grey (a); presence records and potentially accessible areas for *P. megakarya* in Africa, M are shown in green shades (b). GIS layers were obtained from Natural Earth (https://www.naturalearthdata.com/).

withheld 12 of the 92 records to serve as evaluation data and equally divided them among the four bins (Fig. S1, Supplementary Material).

2. 2 Environmental variables

Previous studies have shown that temperature and rainfall are the most important growth-limiting factors for *P. megakarya* (Bailey et al., 2016; Puig et al., 2018). This pathogen thrives in very humid habitats, where rainfall exceeds 2000 mm (Ndoumbè-Nkeng et al., 2009). Relatively elevated temperatures (>30° C) have been shown to inhibit its growth (Puig et al., 2018). Based on this information, we selected six climatic variables that may represent this species' climatic requirements. We downloaded these variables from the climatologies at high resolution for the earth's land areas, CHELSA (Karger et al., 2017). Although climate is the major driver of species' distributions at coarse scales, the utility of ecological niche models can be improved with the addition of edaphic variables (Velazco et al., 2017; Zuquim et al., 2019). Thus, we used topsoil (15 cm depth) physico-chemical properties and climatic variables both separately, and in combination to model the niche of *P. megakarya*. Seven topsoil (0 - 30 cm depth) physico-chemical properties were downloaded from the Harmonized World Soil Database (Wieder, 2014). All environmental layers were upscaled to a 5' (100 km^2) resolution using the arithmetic mean of values of higher resolution cells, and cropped to the study area as defined in section 2.1 using SDM Toolbox 2.4 (Brown, 2014). The complete list of predictor variables is provided in Table 1.

To reduce redundancy and dimensionality among predictor variables, the environmental space characterized by the selected predictor variables was summarized using principal component analysis. We considered the first three principal components, which cumulatively explained 91.99% and 77.48% of the original variation in the climatic and edaphic datasets, respectively, and retained the most important predictors (Fig. 2). This analysis was done using the *FactoMineR* (Lê et al., 2008) and *factoextra* (Kassambara & Mundt, 2020) packages in R (R Core Team 2020). We examined the information gain associated with both variable categories by grouping them incrementally to create 14 sets of predictor variables from a combination of the most important ones as shown in Table 2.
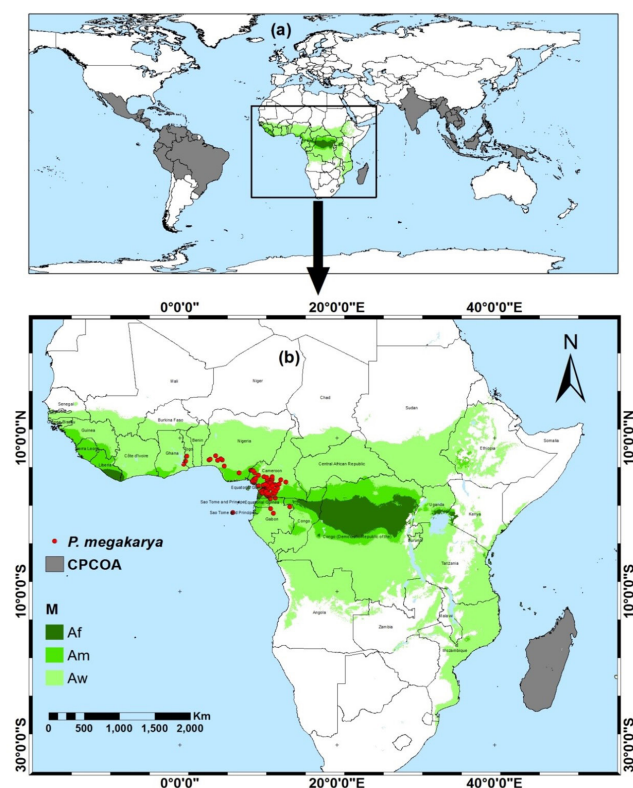
Table 1: List of bioclimatic variables used in this study

| Variable Source	Code	Description	Unit
Climate	Bio 01	Annual Mean Temperature	°C/10
	Bio 05	Maximum Temperature Warmest Month	°C/10
	Bio 06	Minimum Temperature Coldest Month	°C/10
	Bio 12	Annual Precipitation	mm/year
	Bio 13	Precipitation Wettest Month	mm/month
	Bio 14	Precipitation of Driest Month	mm/month
Soil	Bulk_Density	Bulk density	Kg/dm^3
	Gravel	Gravel	% volume
	Sand	Sand	% weight
	Silt	Silt	% weight
	Clay	Clay	% weight
	CEC	Cation Exchange Capacity	Cmol/kg
	OC	Organic Carbon	% weight
	pH	pH (in Water)	-log(H$^+$)
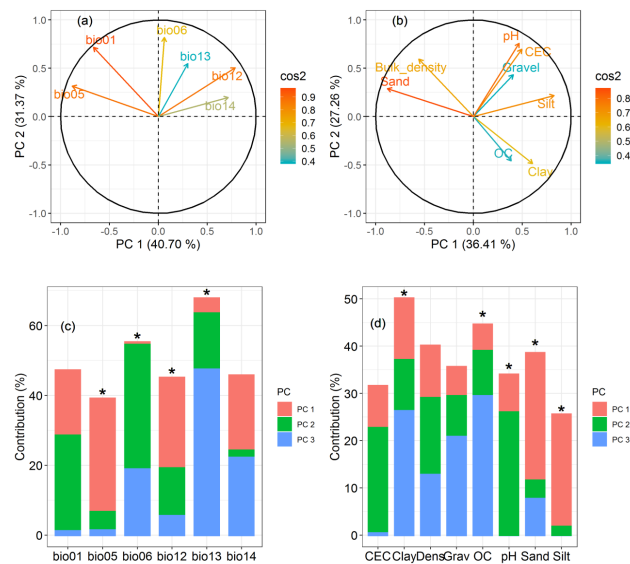
Figure 2: Principal components analysis of predictor variables. The importance of a variable is measured by the quality of its representation, cos2 (a & b), and its contribution (c & d) to each principal component. Warmer colours depict better representation (top row), while the most important variables are depicted with an asterisk (bottom row).

2.3 Maximum entropy ecological niche modelling

Ecological niche models were developed using the maximum entropy modelling algorithm, implemented in MaxEnt v. 3.4.0 (Phillips et al., 2006). Previous studies have demonstrated that spatial predictions from MaxEnt models tend to be not only more accurate than other comparable methods, but also reliable even with limited presence data (van Proosdij et al., 2016; West et al., 2016). We used kuenm (Cobos et al., 2019a), an R package designed to make MaxEnt ecological niche model calibration and evaluation as comprehensive and reproducible as possible. MaxEnt model parameterisation in kuenm is user-friendly and does not require high computational power, even with relatively large datasets. The utility of this platform has been demonstrated in its ability to perform model selection across all possible combinations of environmental data and parameters (Cobos et al., 2019b). Using the 14 datasets previously generated, we created a suite of candidate models with 70% of the presence data, which amounts to 56 records in total (14 per bin). These models were cali-

Table 2: List of the predictor variable combinations

Variable dataset code	Variable combination
Set 01:	Bio 05, Bio 06, Bio 13
Set 02:	Bio 05, Bio 06, Bio 13, Bio 12
Set 03:	Sand, pH, OC
Set 04:	Sand, pH, OC, Silt
Set 05:	Sand, pH, OC, Clay
Set 06:	Sand, pH, OC, Silt, Clay
Set 07:	Bio 05, Bio 06, Bio 13, Sand, pH, OC
Set 08	Bio 05, Bio 06, Bio 13, Sand, pH, OC, Silt
Set 09	Bio 05, Bio 06, Bio 13, Sand, pH, OC, Clay
Set 10	Bio 05, Bio 06, Bio 13, Sand, pH, OC, Silt, Clay
Set 11	Bio 05, Bio 06, Bio 13, Bio 12, Sand, pH, OC
Set 12	Bio 05, Bio 06, Bio 13, Bio 12, Sand, pH, OC, Silt
Set 13	Bio 05, Bio 06, Bio 13, Bio 12 Sand, pH, OC, Clay
Set 14	Bio 05, Bio 06, Bio 13, Bio 12 Sand, pH, OC, Silt, Clay

brated with a wide range of regularisation coefficients (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.5, 0.8, 0.9, 1, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10) and the most basic sets of feature classes (quadratic, linear-quadratic, linear-product, quadratic-product and linear-quadratic-product), to ensure simple relationships between presence data and predictor variables (Merow et al., 2013).

In kuenm, candidate models are evaluated based on their statistical significance first, which is assessed by means of the area under the curve of the receiver operating characteristic, pROC (Peterson et al., 2008). Then, evaluation proceeds with model performance taking precedence over complexity. Partial ROC was calculated by bootstrap with 100 iterations and 50% of the data. Internal validation was based on 30% of the records (i.e., 6 per bin). After we had selected statistically significant models using an omission rate of $\leq 5\%$ (Anderson et al., 2003), the most parsimonious ones were retained for further analyses based on the lowest Akaike Information Criterion corrected for small samples sizes, ΔAICc ≤ 2 (Warren and Seifert 2011).

The final models were calibrated based on the parameters of the best candidate models using 100 bootstrap replicate samples with all 80 *P. megakarya* occurrences. The probability of occurrence of this pathogen was estimated using the recommended complementary log-log (cloglog) transformed output (Phillips et al., 2017). Final model evaluation (in **M**) was based on the 12 records initially withheld from analyses. Model performance was further assessed using the Boyce index, a robust metric for presence-only models (Hirzel et al., 2006). Values of this metric close to +1 indicate reliable model predictions suggesting a model's ability to detect habitat suitability at test locations. We used the *ecospat* package (Di Cola et al., 2017) for this analysis. Binary suitability maps were generated by reclassifying the average of replicate models using the minimum and 10% training presence thresholds. Preference was given to the former considering the robustness of the processed occurrence dataset (Pearson et al., 2006). Uncertainties in the spatial distribution of *P. megakarya* in **M** were estimated using the standard deviation of replicate models. Projections in geographic space, outside **M** were created by extrapolation and clamping. Extrapolation risks associated with these projections were assessed using the mobility-oriented parity metric, MOP (Owens et al., 2013).

3. Results
3. 1. Models performance and variable contributions

We generated 1,400 candidate models for *P. megakarya* using a combination of 20 regularisation multipliers and five feature classes across 14 sets of environmental variables. All candidate model predictions were significantly better than null expectations, and 1152 had an omission rate $\leq 5\%$. Predictor variable source had a profound effect on model complexity (Fig. 3). Models based on both climatic and soil predictors provided the greatest parsimony as opposed to
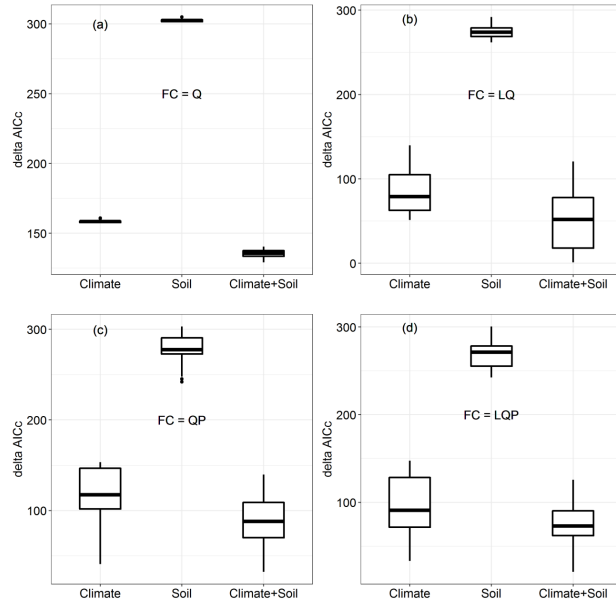
Figure 3: Summary of candidate model performance in relation to feature class and predictor variable source. Only statistically significant models meeting the omission rate criterion were considered. The performance of each predictor dataset is presented in the supplementary material (Fig. S2).

soil-only models. Of these, only four met the ΔAICc criterion (Table 3). Omission rate and Boyce index for the best model after final evaluation were 8.3% and 0.91, respectively.

The contribution of bioclimatic variables was higher than that of edaphic variables (Table 4). Thermal variables, namely Maximum Temperature of the Warmest Month and Minimum Temperature of the Coldest Month, were the most influential predictors,

as evidenced by their average contribution of $35.60 \pm 0.49\%$ and $22.01 \pm 0.44\%$, respectively. Precipitation of the Wettest Month was the third most important variable ($20.44 \pm 0.37\%$). Two other variables had about 4% importance in the best model: Clay content and Annual Precipitation. Of the four edaphic variables, sand content had the highest contribution ($9.51 \pm 0.24\%$), while soil carbon had the lowest contribution.

Model response curves suggested that predicted suitability was highest at values of Maximum Temperature of Warmest Month $\leq 26 \degree$ C and decreased exponentially thereafter (Fig. 4 a). Suitability exponentially increased with increasing Minimum Temperature of Coldest Month and plateaued from about 23°C (Fig. 4b). Precipitation of Wettest Month, Sand and Clay content followed a bell-shaped pattern with maximum suitability between 400-600 mm/month, 50-60% and 30-40% respectively (Fig. 4 c-e).

3. 2 Model spatial predictions in the calibration extent

The predicted distributions of *P. megakarya* based on the minimum and 10 percentile training presence thresholds were largely restricted between 10° S and 10° N (Fig. 5 c & d). The largest region with environmental conditions more likely to support the proliferation of *P. megakarya* was in Central Africa, encompassing São Tome and Principe, Gabon, Equatorial Guinea, and southern Cameroon. Habitat suitability was equally high in the Republic of Congo, including areas south of the Batéké Plateau to the northern limits of the Odzala National Park, and in the Democratic Republic of Congo, in the northern

Table 3: Model performance in relation to predictor variable combinations and MaxEnt parameters for the four best models identified in our analyses

Dataset	FC	RM	pROC	OR	AICc	ΔAICc	WAICc	BI	N
Set 14	LQ	0.40	0.00	0.17	1523.09	0.00	0.23	0.92	16
Set 11	LQ	0.40	0.00	0.17	1523.19	0.10	0.25	0.91	11
Set 13	LQ	0.50	0.00	0.08	1523.48	0.39	0.16	0.91	13
Set 09	LQ	0.40	0.00	0.17	1524.30	1.20	0.11	0.81	12

FC=feature class, RM= Regularisation multiplier, OR = Omission rate, BI: Boyce index, ΔAICc, Akaike Information Criterion corrected for small sample sizes, WAICc = weighted AIC.

Table 4: Variable contributions and permutation importance

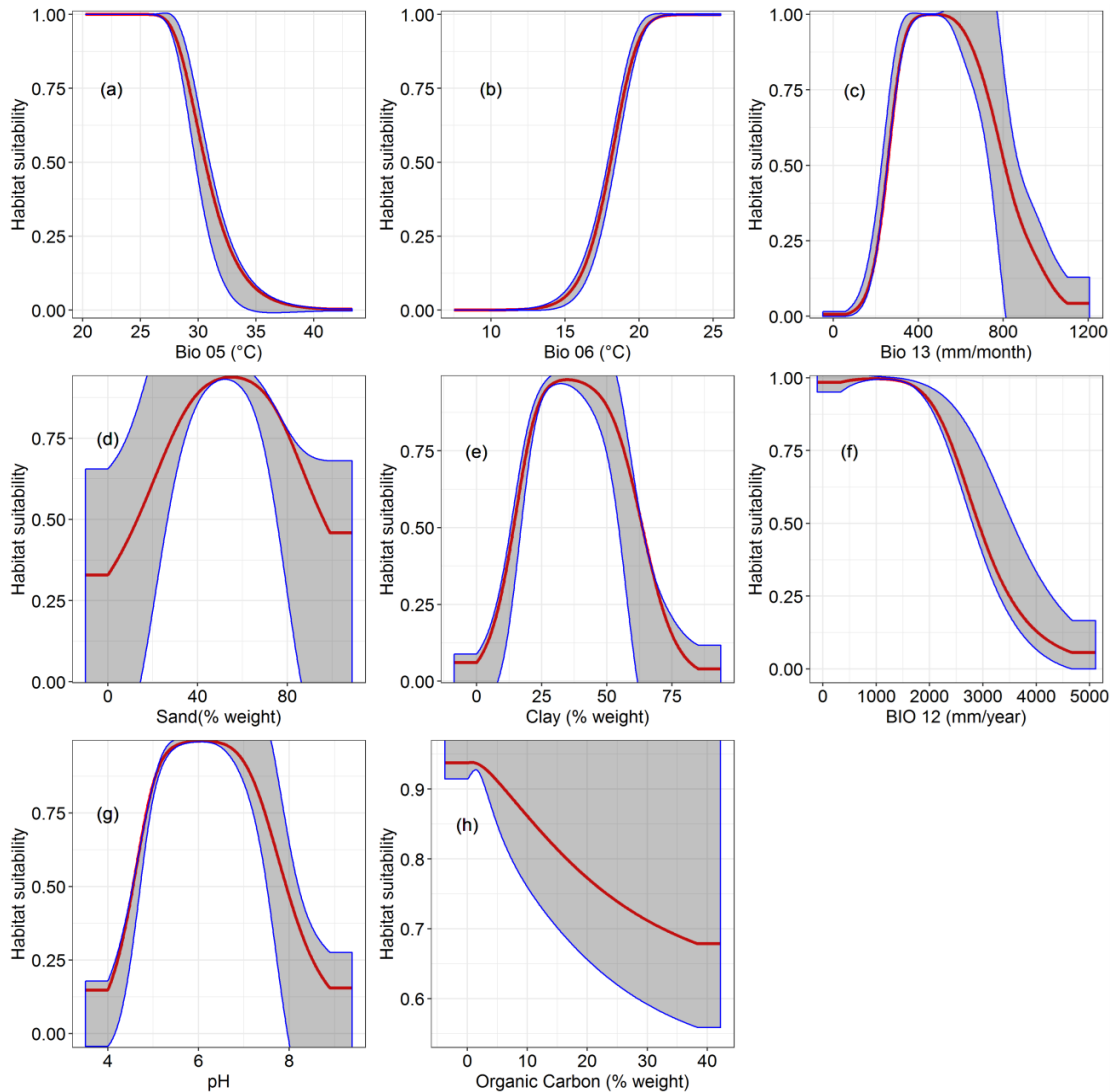| Variable	Contribution (%)	Permutation importance (%)
Max Temperature of Warmest Month	35.60 ± 0.49	31.60 ± 0.60
Min Temperature of Coldest Month	22.01 ± 0.44	29.84 ± 0.50
Precipitation of Wettest Month	20.44 ± 0.37	21.24 ± 0.50
Sand	9.51 ± 0.24	1.12 ± 0.16
Clay	4.78 ± 0.14	8.39 ± 0.34
Annual Precipitation	4.47 ± 0.25	1.82 ± 0.08
pH	3.12 ± 0.13	5.94 ± 0.25
OC	0.07 ± 0.02	0.06 ± 0.01

Figure 4. Response curves of the best model. Curves depict the relationship between predicted habitat suitability (cloglog) and each environmental variables, keeping all others at their mean value. The average response for the 100 replicates is shown in red and the standard deviation in grey. Variable codes are given in Table 1.

half of the Congo River Basin.

In west Africa, suitable areas were observed along the Atlantic coast, from southern Nigeria to Gambia. This area was interrupted by a discontinuity that coincides with the Dahomey Gap in Benin and Togo. Abiotic conditions over the entire Liberian territory and much of Sierra Leone, and in the major cocoa-producing regions in southern Cote d'Ivoire and Ghana were found to be conducive for *P. megakarya*. A broad swath of Guinea was predicted as unsuitable for this species as opposed to an isolated area ranging from the estuaries of Guinea-Bissau to north-western Gambia. In East Africa, our model also predicted an association of *P. megakarya* with coastal environments, in Kenya, Tanzania, Mozambique and Madagascar, and areas surrounding Lake Victoria. These predictions followed an overall similar pattern for both thresholds, and were associated with a moder-

ate level of uncertainty, resulting from the variability among replicate models (Fig. 5 b). MOP analysis showed that all areas predicted as conducive for the study species in Madagascar had conditions similar to those of the calibration extent.

3. 3 Model spatial predictions outside the calibration range

Model transfer to Central America predicted propitious environments in areas between southern Mexico and Panama, and across the Caribbean islands with a very low extrapolation risk (Fig. 6 a & b). The largest suitable area in South America was central Brazil; other such areas appeared scattered in Suriname, Venezuela, Peru and Colombia. We noted very low habitat suitability in an area broadly matching the meanders of the Amazon River (Fig. 6 a). However, predictions within this space were based
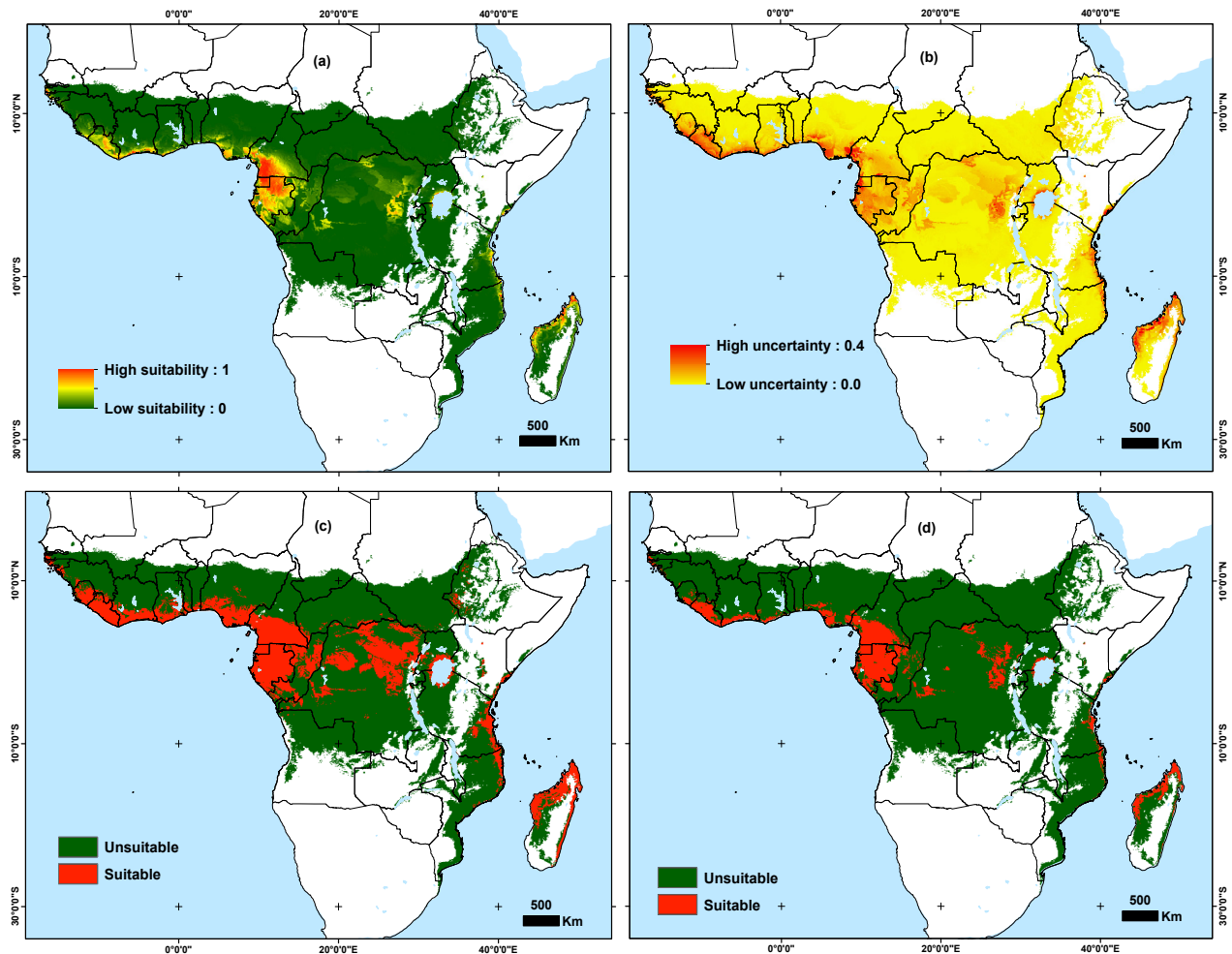

Figure 5. MaxEnt prediction of suitable areas for *P. megakarya* in Sub-Saharan Africa. Continuous suitability map based on the mean of 100 replicate models (a) and uncertainty (standard deviation) associated with this prediction (b). Green and red areas depict the unsuitable and suitable geographic space in the calibration extent after reclassification using the minimum training presence threshold (c) and 10 percentile training presence threshold (d).
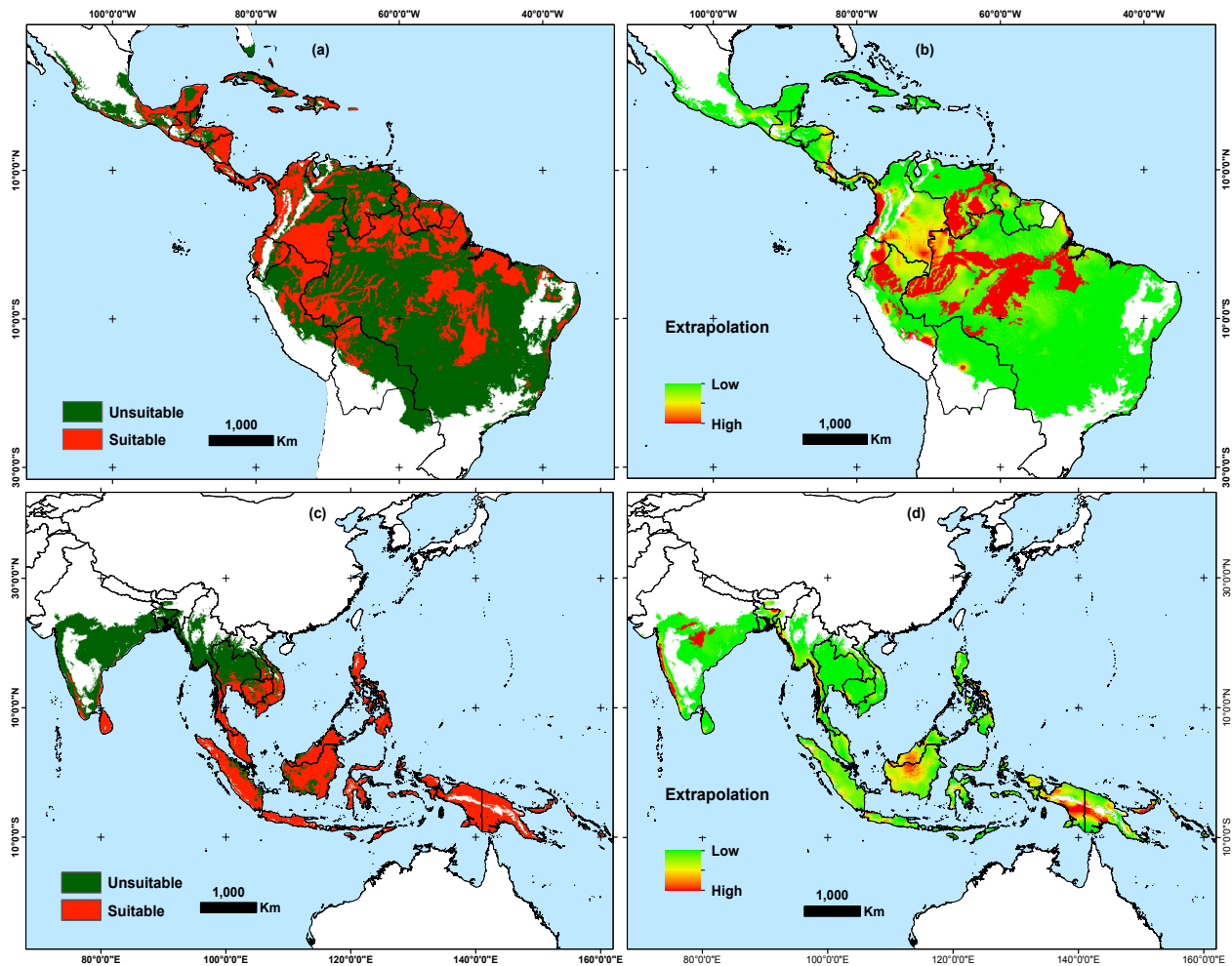
Figure 6 . MaxEnt prediction of suitable areas for *Phytophthora megakarya* in South America and Southeast Asia. Suitable environmental conditions (in red) in Central and South America (a) and South east Asia (c). MOP analysis comparing environmental conditions in the calibration range with those in Central and South America (b) and Southeast Asia (d). High model extrapolation in these predictions are depicted in red. Binary maps were made using the minimum training presence threshold.

on strict extrapolation according to MOP analysis as were those in southern Venezuela, northern Peru and along the Pacific coast of Colombia (Fig. 6 b).

The environment in many cocoa-growing countries in Southeast Asia was conducive for *P. megakarya* (Fig. 6 c & d). Strict model extrapolation was observed on the northern side of the Deccan Plateau and Western Ghats in India. Areas extending from the Eastern Ghats to Sri Lanka were particularly suitable for this species. Moderate extrapolation risk was associated with predictions in Central Indonesia, Malaysia and Papua New Guinea. Although India is a major player in the global cocoa production, model predictions were generally unfavourable for *P. megakarya* in many areas of this country.

4. Discussion

In view of the emergence of *P. megakarya* and the resulting increase in the incidence of black pod disease in West and Central Africa, we assessed the potential geographical distribution of this species using maximum entropy modelling. MaxEnt models were calibrated with 80 presence locations of *P. megakarya* in Central and West Africa, while evaluation was based on 12 records. The use of relevant climatic and edaphic predictor variables revealed areas with environmental conditions that are conducive for the proliferation of this species. Predicted distributional areas were fragmented in South America and more continuous in Southeast Asia, with more novel combinations of environmental factors in the former than the latter. The application of ecological niche

models in the field of plant pathology is relatively recent, and the *Phytophthora* genus has been the primary focus of several studies (Burgess et al., 2017; Duque-Lazo et al., 2016; Kluza et al., 2007; La Manna et al., 2012; Ramírez-Gil & Peterson, 2019).

The paucity of data on the presence of plant pathogens poses a major constraint in predicting their potential distributions. Georeferenced occurrence data of *P. megakarya* are rather poorly documented compared with its congeners outside the Afro-tropical region (Scott et al., 2019). The spatial distribution of other *Phytophthora* species has mostly been explored through ecological niche modelling at relatively limited geographical extents, usually within national boundaries. To date, *P. cinnamomi* is the only species that has been mapped across continents owing to its wide host range (Burgess et al., 2017). Many studies rely on primary data, collected during field sampling (Duque-Lazo et al., 2016; La Manna et al., 2012; Ramírez-Gil & Peterson, 2019). In contrast, our study exploited occurrences data gathered from a broader area, large enough to support a robust intercontinental niche modelling effort. However, we noted that many field-based studies on *P. megakarya* neglect to share detailed locality data for this species. We also noted that records of this species were conspicuously unavailable for Côte d'Ivoire even though it has been reported there (Pokou et al., 2008). Although, our results are based on presence records mostly from eastern Cameroon, where this species was first detected, the approach used in data partitioning has been shown to reduce effects of spatial structure in occurrence data and improve model performance (Roberts et al., 2017).

It is generally accepted that climate is the major driver of species' distributions at wide geographic scales; as such, niches have often been modelled with regard to climate predictors only. For example, an efficient climate-only epidemiological model has been developed to forecast black pod disease outbreaks in Nigeria (Etaware et al., 2020). Although these tools are relatively simple, reliable and easy to implement, given the direct effects of climatic factors on the pathogen's physiology, we found that climate is not sufficient to characterize the environmental requirements of *P. megakarya*. Our results showed that edaphic variables, though not very useful in themselves, played a supplemental role in model predictions. In other words, we obtained better models for *P. megakarya* as a result of combining climatic and edaphic variables. This result was perhaps unexpected as *Phytophthora* species are essentially soil-borne pathogens. For instance, results from previous studies highlighted the affinity of *P. cinnamomi* for fine-textured soils, which promote high water retention (Corcobado et al., 2013). In the same vein, clay content was also found to be among the factors that influence the distribution of this species in the Iberian Peninsula (Hernández-Lambraño et al., 2018). However, our findings support the hypothesis that both climate and soil related determine species' distributions (Velazco et al., 2017; Zuquim et al., 2019).

We found a trend in the probability of predicted environmental suitability that is consistent with the temperature and rainfall requirements of *P. megakarya*. Rainfall in West and Central Africa generally follows a bimodal distribution pattern, with the highest peak in September. Thus, the maximum suitability for precipitation of the wettest month, which was predicted between 400 and 500 mm, is in agreement with the findings of Efombagn et al., (2004), who measured yearly black pod disease progress in three cocoa clones in Cameroon between 1999 and 2001, and recorded the highest number of rotten pods in September (rainfall ≈ 300 mm). In addition, we obtained the highest habitat suitability for Minimum Temperature of the Coldest Month values above 22° C, and Maximum Temperature of the Warmest Month values below 28° C. This range matches the 24° C - 26° C optimal temperatures for *P. megakarya* (Bailey et al., 2016).

Biotic factors, parasitism in the case of *P. megakarya* can play a significant role in shaping its geographic distribution, and as such reinforce model predictions. However, such interactions are often more expressed at local scales and tend to fade at coarser resolutions, comparable with those at which our analyses were done (Ramírez-Gil & Peterson 2019). Hernández-Lambraño et al., (2018) used the presence of two primary hosts of *P. cinnamomi* as a predictor to estimate the potential geographic distribution of this pathogen in the Iberian Peninsula. A similar specific host-pathogen relationship exists between *P. megakarya* and cocoa (Bailey et al., 2016). Although this pathogen might develop a multi-host capacity in the future, considering the number of its alternative hosts (Akrofi et al., 2015; Opoku et al., 2002), its close interaction with cocoa in Africa might be helpful in small scale modelling efforts.

The size of the area that is potentially accessible to a species by means of natural dispersal (**M**) has profound influences on niche modelling outcomes (Barve et al., 2011). In this study, we adopted a relatively broad **M** for model calibration. Although there is evidence that the study species currently occupies a small fraction of **M** (Akrofi, 2015), its impending spread beyond this area, through contaminated planting material cannot be overlooked considering the economic value of its primary host. This approach of defining **M** has been used in predictive models of pathogens affecting Hass avocado in Colombia, including, *P. palmivora* and *P. cinnamomi* (Ramírez-Gil & Peterson, 2019).

Model predictions indicated that *P. megakarya* would be of less concern in Africa, except in Central and West Africa where we identified two distinct areas of high environmental suitability, separated by the Dahomey Gap, in Benin. This result might explain the existence of two genetically different groups of *P. megakarya*, one in each region (Nyassé et al., 1999). We hypothesize that the drier climate in the Dahomey Gap acts as an effective natural barrier to the convergence of these two genotypes. It is noteworthy that our models were robust enough to detect high habitat suitability in the main cocoa-producing regions, in southern Côte d'Ivoire (Pokou et al., 2008), even though presence data from this area were not available.

Though we assessed the performance of different MaxEnt models using a wide range of parameters and predictor variables, our use of a single modelling algorithm is one obvious limitation of this study. In fact, there is a multitude approaches in correlative species distribution modelling to choose from, and more calls have been made to explore model performance across a suite of algorithms (Qiao et al. 2015; Zhu & Peterson 2017). Thus, further studies are needed to better understand the distributional potential of *P. megakarya* using individual and ensemble models.

Conclusions

Only four countries in West and Central Africa contribute the largest proportion of global cocoa bean production. However, further spread of *P. megakarya* is likely with increasing global trade, especially in developing economies. Adoption of pre-emptive measures would be crucial for effective coordinated international control measures. Models for *P. mega-karya* incorporating soil and climate data can be used to predict regions at risk of black pod disease. This study underscored the potential use of correlative species distribution modelling for forecasting plant pathogen occurrences at regional and global scales. This understanding of the spatial pattern of occurrence of *P. megakarya* could help in controlling the spread of black pod disease in tropical regions.

Corresponding Author
mc.obiakara@gmail.com
+234 813 362 7222

Author contributions
Conceptualization: MCO; Methods: MCO; Data curation and analyses: MCO; Writing - original draft preparation: MCO and PME; Writing - review and editing: MCO, PME and SKC

Declarations of interest: none.

References
Akrofi, A. Y. (2015). *Phytophthora megakarya*: A Review On Its Status As A Pathogen On Cacao In West Africa. *African Crop Science Journal*, *23*(1), 21.

Akrofi, A. Y., Amoako-Atta, I., Assuah, M., & Asare, E. K. (2015). Black pod disease on cacao (Theobroma cacao, L) in Ghana: Spread of *Phytophthora megakarya* and role of economic plants in the disease epidemiology. *Crop Protection*, *72*, 66–75. https://doi.org/10.1016/j.cropro.2015.01.015

Ali, S. S., Amoako-Attah, I., Bailey, R. A., Strem, M. D., Schmidt, M., Akrofi, A. Y., Surujdeo-Maharaj, S., Kolawole, O. O., Begoude, B. A. D., ten Hoopen, G. M., Goss, E., Phillips-Mora, W., Meinhardt, L. W., & Bailey, B. A. (2016). PCR-based identification of cacao black pod causal agents and identification of biological factors possibly contributing to *Phytophthora megakarya* 's field dominance in West Africa. *Plant Pathology*, *65*(7), 1095–1108. https://doi.org/10.1111/ppa.12496

Anderson, R. P., Lew, D., & Peterson, A. T. (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling*, *162*(3), 211–232. https://doi.org/10.1016/S0304-3800(02)00349-6

Bailey, B. A., Ali, S. S., Akrofi, A. Y., & Meinhardt, L. W. (2016). *Phytophthora megakarya*, a Causal Agent of Black Pod Rot in Africa. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 267–303). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_8

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*(11), 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Brown, J. L. (2014). SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, *5*(7), 694–700. https://doi.org/10.1111/2041-210X.12200

Burgess, T. I., Scott, J. K., Mcdougall, K. L., Stukely, M. J. C., Crane, C., Dunstan, W. A., Brigg, F., Andjic, V., White, D., Rudman, T., Arentz, F., Ota, N., & Hardy, G. E. St. J. (2017). Current and projected global distribution of *Phytophthora cinnamomi* , one of the world's worst plant pathogens. *Global Change Biology*, *23*(4), 1661–1674. https://doi.org/10.1111/gcb.13492

Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019a). kuenm: An R package for detailed development of ecological niche models using Maxent. *PeerJ*, *7*, e6281. https://doi.org/10.7717/peerj.6281

Cobos, M. E., Peterson, A. T., Osorio-Olvera, L., & Jiménez-García, D. (2019b). An exhaustive analysis of heuristic methods for variable selection in ecological niche modeling and species distribution modeling. *Ecological Informatics*, *53*, 100983. https://doi.org/10.1016/j.ecoinf.2019.100983

Corcobado, T., Solla, A., Madeira, M. A., & Moreno, G. (2013). Combined effects of soil properties and *Phytophthora* cinnamomi infections on Quercus ilex decline. *Plant and Soil*, *373*(1–2), 403–413. https://doi.org/10.1007/s11104-013-1804-z

Deberdt, P., Mfegue, C. V., Tondje, P. R., Bon, M. C., Ducamp, M., Hurard, C., Begoude, B. A. D., Ndoumbe-Nkeng, M., Hebbar, P. K., & Cilas, C. (2008). Impact of environmental factors, chemical fungicide and biological control on cacao pod production dynamics and black pod disease (*Phytophthora megakarya*) in Cameroon. *Biological Control*, *44*(2), 149–159. https://doi.org/10.1016/j.biocontrol.2007.10.026

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*(6), 774–787. https://doi.org/10.1111/ecog.02671

Duque-Lazo, J., van Gils, H., Groen, T. A., & Navarro-Cerrillo, R. M. (2016). Transferability of species distribution models: The case of *Phytophthora* cinnamomi in Southwest Spain and Southwest Australia. *Ecological Modelling*, *320*, 62–70. https://doi.org/10.1016/j.ecolmodel.2015.09.019

Efombagn, M. I. B., Marelli, J. P., Ducamp, M., Cilas, C., Nyasse, S., & Vefonge, D. (2004). Effect of Fruiting Traits on the Field Resistance of Cocoa (Theobroma cacao L.) Clones to *Phytophthora megakarya*. *Journal of Phytopathology*, *152*(10), 557–562. https://doi.org/10.1111/j.1439-0434.2004.00895.x

Etaware, P. M., Adedeji, A. R., Osowole, O. I., & Odebode, A. C. (2020). ETAPOD: A forecast model for prediction of black pod disease outbreak in Nigeria. *PLOS ONE*, *15*(1), e0209306. https://doi.org/10.1371/journal.pone.0209306

Franklin, J. (2009). *Mapping Species Distributions: Spatial inference and Prediction*. 340.

Hernández-Lambraño, R. E., González-Moreno, P., & Sánchez-Agudo, J. Á. (2018). Environmental factors associated with the spatial distribution of invasive plant pathogens in the Iberian Peninsula: The case of *Phytophthora* cinnamomi Rands. *Forest Ecology and Management*, *419–420*, 101–109. https://doi.org/10.1016/j.foreco.2018.03.026

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*(2), 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder,

H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7*. https://CRAN.R-project.org/package=factoextra

Kluza, D. A., Vieglais, D. A., Andreasen, J. K., & Peterson, A. T. (2007). Sudden oak death: Geographic risk estimates and predictions of origins. *Plant Pathology*, *56*(4), 580–587. https://doi.org/10.1111/j.1365-3059.2007.01602.x

La Manna, L., Matteucci, S. D., & Kitzberger, T. (2012). Modelling *Phytophthora* disease risk in Austrocedrus chilensis forests of Patagonia. *European Journal of Forest Research*, *131*(2), 323–337. https://doi.org/10.1007/s10342-011-0503-7

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, *25*(1). https://doi.org/10.18637/jss.v025.i01

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Mfegue, C. V. (2012). *Origine et mécanismes de dispersion des populations de Phytophthora megakarya, pathogène du cacaoyer au Cameroun* [Montpellier SupAgro]. https://www.supagro.fr/theses/extranet/12-0015_MFEGUE.pdf

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198–1205. https://doi.org/10.1111/2041-210X.12261

Ndoumbè-Nkeng, M., Efombagn, M. I. B., Nyassé, S., Nyemb, E., Sache, I., & Cilas, C. (2009). Relationships between cocoa *Phytophthora* pod rot disease and climatic variables in Cameroon. *Canadian Journal of Plant Pathology*, *31*(3), 309–320. https://doi.org/10.1080/07060660909507605

Nyassé, S., Grivet, L., Risterucci, A. M., Blaha, G., Berry, D., Lanaud, C., & DesprÉAux, D. (1999). Diversity of *Phytophthora megakarya* in Central and West Africa revealed by isozyme and RAPD markers. *My-*

cological Research, *103*(10), 1225–1234. https://doi.org/10.1017/S0953756299008369

Opoku, I. Y., Akrofi, A. Y., & Appiah, A. A. (2002). Shade trees are alternative hosts of the cocoa pathogen *Phytophthora megakarya*. *Crop Protection*, *21*(8), 629–634. https://doi.org/10.1016/S0261-2194(02)00013-3

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, *263*, 10–18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, A. (2006). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar: Predicting species distributions with low sample sizes. *Journal of Biogeography*, *34*(1), 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Peterson, A. T. (2014). *Mapping disease transmission risk: Enriching models using biogeography and ecology*. Johns Hopkins University Press.

Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*(1), 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Peterson, A. T., Soberón Jorge, Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions*. Princeton University Press.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pokou, N. D., N'Goran, J. A. K., Kébé, I., Eskes, A., Tahi, M., & Sangaré, A. (2008). Levels of resistance to *Phytophthora* pod rot in cocoa accessions selected on-farm in Côte d'Ivoire. *Crop Protection*, *27*(3–5), 302–309. https://doi.org/10.1016/j.cropro.2007.07.012

Puig, A. S., Ali, S., Strem, M., Sicher, R., Gutierrez, O. A., & Bailey, B. A. (2018). The differential influence of temperature on *Phytophthora megakarya* and *Phytophthora* palmivora pod lesion expansion, mycelia

growth, gene expression, and metabolite profiles. *Physiological and Molecular Plant Pathology*, *102*, 95–112. https://doi.org/10.1016/j.pmpp.2017.12.002

Qiao H., Soberón J., Peterson A. T., (2015). No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution 6*, 1126–1136. https://doi.org/10.1111/2041-210X.12397

R Core Team (2020). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. URL https://www.R- project.org/.

Ramírez-Gil, J. G., & Peterson, A. T. (2019). Current and potential distributions of the most important diseases affecting Hass avocado in Antioquia Colombia. *Journal of Plant Protection Research*, *59*(2), 214–228. https://doi.org/10.24425/JPPR.2019.129288

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Scott, P., Bader, M. K.-F., Burgess, T., Hardy, G., & Williams, N. (2019). Global biogeography and invasion risk of the plant pathogen genus *Phytophthora*. *Environmental Science & Policy*, *101*, 175–182. https://doi.org/10.1016/j.envsci.2019.08.020

Vaclavık, T., & Meentemeyer, R. K. (2011). Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions*, 11.

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, *39*(6), 542–552. https://doi.org/10.1111/ecog.01509

Velazco, S. J. E., Galvão, F., Villalobos, F., & De Marco Júnior, P. (2017). Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. *PLOS ONE*, *12*(10), e0186025. https://doi.org/10.1371/journal.pone.0186025

West, A. M., Kumar, S., Brown, C. S., Stohlgren, T. J., & Bromberg, J. (2016). Field validation of an invasive species Maxent model. *Ecological Informatics*, *36*, 126–134. https://doi.org/10.1016/j.ecoinf.2016.11.001

Wieder, W. (2014). *Regridded Harmonized World Soil Database v1.2*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1247

Zhu G.-P., Peterson A. T., (2017). Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. *Biological Invasions 19*, 2519-2532. https://doi.org/10.1007/s10530-017- 1460-y

Zuquim, G., Costa, F. R. C., Tuomisto, H., Moulatlet, G. M., & Figueiredo, F. O. G. (2019). The importance of soils in predicting the future of plant habitat suitability in a tropical forest. *Plant and Soil*. https://doi.org/10.1007/s11104-018-03915-9

Supplementary material

Table S1: Major pathogenic *Phytophthora* species of cocoa.

Species	Countries affected	Authors
P. megakarya	West Africa	Brasier and Griffin (1979)
P. palmivora	All cocoa-growing regions	Brasier and Griffin (1979)
P. citrophthora	Brazil, Mexico, Indonesia, India	Lawrence et al. (1990); Appiah et al. (2004); Chowdappa and Chandramohanan (1996)
P. capsici	Americas, Caribbean, Central Africa	Erwin and Ribeiro (1996); Guest (2007)
P. katsurae	Côte d'Ivoire	Liyanage and Wheeler (1989)
P. botryos	Malaysia	Kroon et al. (2004)
P. megasperma	Venezuela	Zentmeyer (1988)

References

Appiah A, Flood J, Archer SA, Bridge PD (2004) Molecular analysis of the major *Phytophthora* species on cocoa. Plant Pathol 53:209-219

Brasier CM, Griffin MJ (1979) Taxonomy of *Phytophthora palmivora* on cocoa. Trans Br Mycol Soc 72:111-143.

Chowdappa P, Chandramohanan R (1996) Occurrence of *Phytophthora citrophthora* on cocoa in India. Trop Agric 73:158-160.

Erwin DC, Ribeiro OK (1996) *Phytophthora* diseases worldwide. APS Press.

Guest D (2007) Black pod: diverse pathogens with a global impact on cocoa yield. Phytopathol 97:1650-1653.

Kroon LPNM, Bakker FT, van den Bosch GBM, Bonants PJM, Flier WG (2004) Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclearDNA sequences. Fungal Genet Biol41: 766-782

Lawrence JS, Campelo AMFL, de Figueiredo JM (1990). Diseases of cacao. I-Fungal diseases which occur on the fruit (in Portuguese). Agrotropica 2:121-136.

Liyanage NIS, Wheeler BEJ (1989)*Phytophthora katsurae* from cocoa. Plant Pathol 38:627-29.

Zentmeyer GA (1988) Taxonomic relationships and distribution of *Phytophthora* causing black pod of cocoa. Proceedings of the 10th International Cocoa Research Conference,17th - 23rd May, Santo Domingo, Dominican Republic. pp. 391-395.

Table S2: Sources of occurrences of *P. megakarya* in West Africa

Author(s)	Count	Country
Mfegue	726	Cameroon, Gabon, Sao Tome & Principe, Nigeria, Togo, Ghana and Cote d'Ivoire
Mfegue et al. (2012)	8	Cameroon, Nigeria,
Ndoumbè-Nkeng et al. (2009)	3	Cameroon
Ndoungue et al. (2018)	1	Cameroon
Olu-Owolabi et al. (2013)	15	Nigeria
Total	753	

REFERENCES

Mfegue CV, Herail C, Adreit H, Mbenoun M, Techou Z, Ten Hoopen M, Tharreau D, & Ducamp M. (2012) Microsatellite Markers For Population Studies of *Phytophthora Megakarya* (Pythiaceae), A Cacao Pathogen In Africa. American Journal of Botany: e353–e356.

Mfegue, CV (2012) *Origine et mécanismes de dispersion des populations de Phytophthora megakarya, pathogène du cacaoyer au Cameroun* [Montpellier SupAgro]. https://www.supagro.fr/theses/extranet/12-0015_MFEGUE.pdf

Ndoumbè-Nkeng, M., Efombagn, M. I. B., Nyassé, S., Nyemb, E., Sache, I., & Cilas, C. (2009). Relationships between cocoa *Phytophthora* pod rot disease and climatic variables in Cameroon. *Canadian Journal of Plant Pathology*, *31*(3), 309–320. https://doi.org/10.1080/07060660909507605

Ndoungue M, Petchayo S, Techou Z, Nana WG, Nembot C, Fontem, D, & Ten Hoopen GM (2018) The impact of soil treatments on black pod rot (caused by *Phytophthora megakarya*) of cacao in Cameroon. *Biological Control.* https://doi.org/10.1016/j.biocontrol.2018.04.016

Olu-Owolabi BI, Agunbiade FO, Fagbayigbo BO & Adebowale KO (2013) Monitoring Copper Bioaccumulation in Cocoa from Copper-Based Pesticide–Treated Cocoa Farms Using Fuzzy Similarity Method *Bioremediation Journal*, 17(3):131–147,
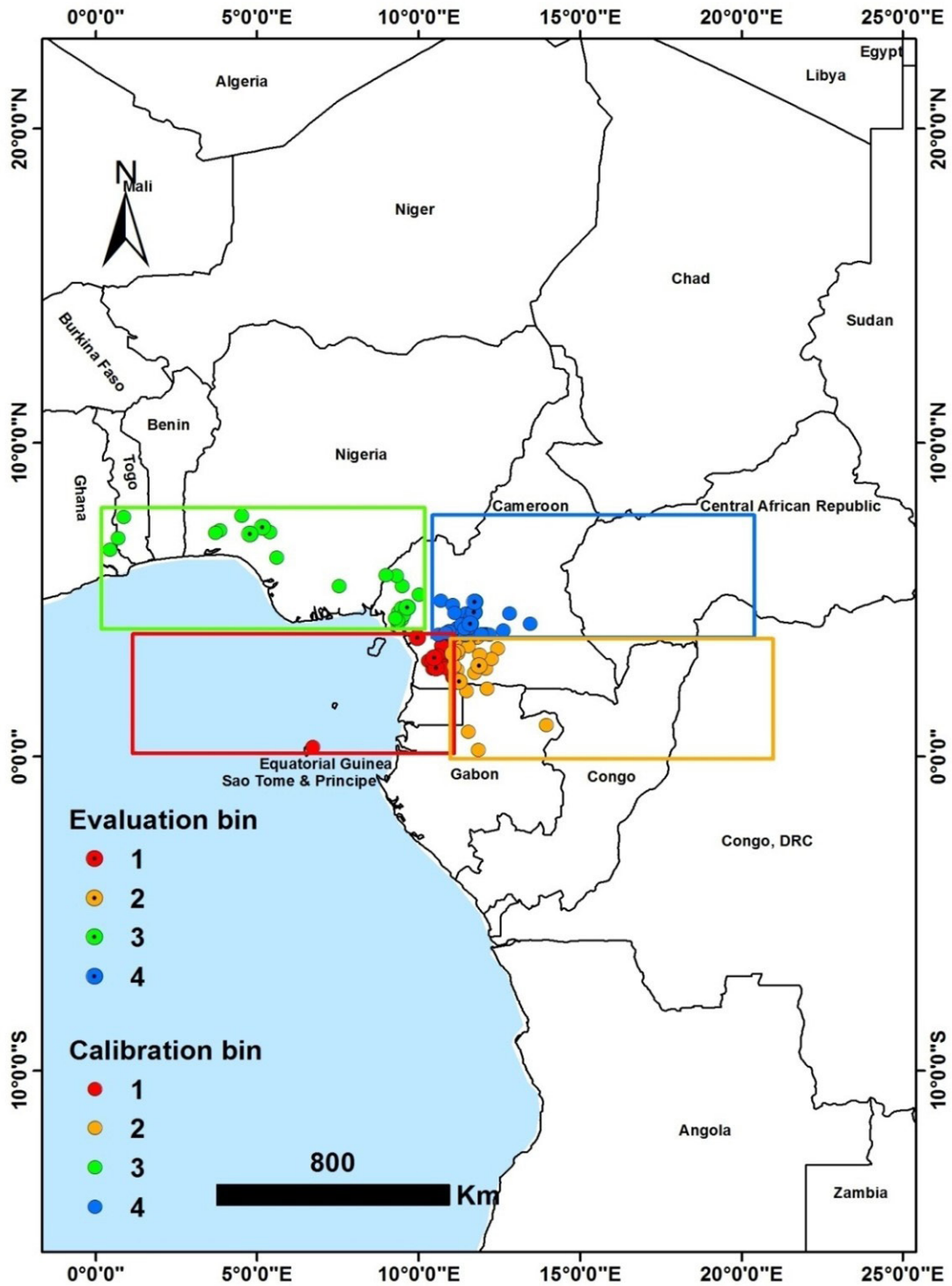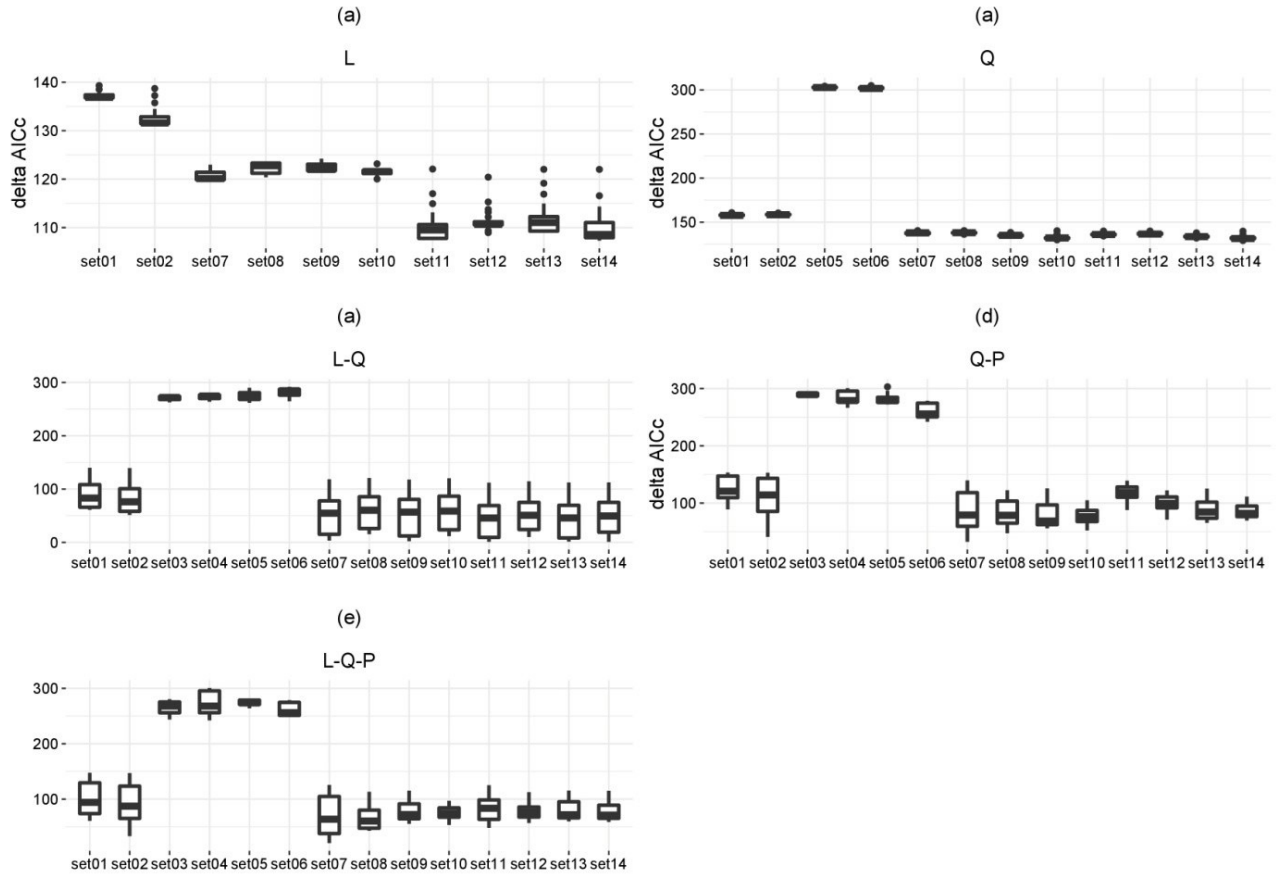
Figure S1: Occurrence data partitioning

Figure S2: Candidate model performance in relation to feature class and predictor variable category