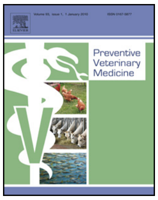
Preventive Veterinary Medicine 118 (2015) 8–21

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Species distribution modelling for *Rhipicephalus microplus* (Acari: Ixodidae) in Benin, West Africa: Comparing datasets and modelling algorithms

E.M. De Clercq[a,*], S. Leta[b], A. Estrada-Peña[c], M. Madder[d,e], S. Adehan[f], S.O. Vanwambeke[b]

[a] Research Fellow FNRS, Georges Lemaître Institute for Earth and Climate Research, Université catholique de Louvain, Place Louis Pasteur 3, 1348 Louvain-la-Neuve, Belgium
[b] Adami Tullu Agricultural Research Center, P.O.Box 35, Adami Tullu, Ethiopia
[c] Dept. of Parasitology, Veterinary Faculty, Miguel Servet 177, 50013 Zaragoza, Spain
[d] Vector Biology Unit, Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium
[e] Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort 0110, Pretoria, South Africa
[f] Institut National des Recherches Agricoles du Benin–INRAB, 01 BP 884 Cotonou, Benin

ARTICLE INFO

Article history:
Received 22 October 2013
Received in revised form 21 October 2014
Accepted 25 October 2014

Keywords:
Species distribution modelling
Invasive tick
Rhipicephalus microplus
Environmental data
WorldClim
MODIS

ABSTRACT

Rhipicephalus microplus is one of the most widely distributed and economically important ticks, transmitting *Babesia bigemina*, *B. bovis* and *Anaplasma marginale*. It was recently introduced to West Africa on live animals originating from Brazil. Knowing the precise environmental suitability for the tick would allow veterinary health officials to draft vector control strategies for different regions of the country. To test the performance of modelling algorithms and different sets of environmental explanatory variables, species distribution models for this tick species in Benin were developed using generalized linear models, linear discriminant analysis and random forests. The training data for these models were a dataset containing reported absence or presence in 104 farms, randomly selected across Benin. These farms were sampled at the end of the rainy season, which corresponds with an annual peak in tick abundance. Two environmental datasets for the country of Benin were compared: one based on interpolated climate data (WorldClim) and one based on remotely sensed images (MODIS). The pixel size for both environmental datasets was 1 km. Highly suitable areas occurred mainly along the warmer and humid coast extending northwards to central Benin. The northern hot and drier areas were found to be unsuitable. The models developed and tested on data from the entire country were generally found to perform well, having an AUC value greater than 0.92. Although statistically significant, only small differences in accuracy measures were found between the modelling algorithms, or between the environmental datasets. The resulting risk maps differed nonetheless. Models based on interpolated climate suggested gradual variations in habitat suitability, while those based on remotely sensed data indicated a sharper contrast between suitable and unsuitable areas, and a patchy distribution of the suitable areas. Remotely sensed data yielded more spatial detail in the predictions. When computing accuracy measures on a subset of data along the invasion front, the modelling technique Random Forest outperformed the other modelling approaches, and results with MODIS-derived variables were better than those using WorldClim data.

---

[*] Corresponding author. Tel.: +32 10479040; fax: +32 10472877.
E-mail address: eva.declercq@uclouvain.be (E.M. De Clercq).

http://dx.doi.org/10.1016/j.prevetmed.2014.10.015

The high environmental suitability for *R. microplus* in the southern half of Benin raises concern at the regional level for animal health, including its potential to substantially alter transmission risk of *Babesia bovis.* The northern part of Benin appeared overall of low environmental suitability. Continuous surveillance in the transition zone however remains relevant, in relation to important cattle movements in the region, and to the invasive character of *R. microplus*.

1. Introduction

The cattle tick *Rhipicephalus microplus* (Canestrini) is one of the most important disease vectors for livestock in tropical and sub-tropical countries. The largest challenge for control is its reported resistance to common acaricides. This tick transmits *Babesia bovis* and *B. bigemina*, as well as *Anaplasma marginale* (Madder et al., 2011). Bovine babesiosis is considered globally amongst the most significant tick-borne diseases of cattle (Hoch et al., 2012). The original distribution range of *R. microplus* covers southern and southeastern Asia. However, during the second half of the nineteenth century it was transported with livestock to eastern and southern Africa, the Comoro Islands, and Mascarene Islands. The approximate date and route of introduction and spread into Central and South America is unknown. This tick was also reported in northern and eastern Australia, New Caledonia and French Polynesia for about 100 years, but recent research has acknowledged this to be a different species, and renamed it *Rickettsia australis* Fuller, after its original name in Australia (Estrada-Peña et al., 2012). About a decade ago, the confirmed records of *R. microplus* in Africa were restricted to southern and south eastern Africa (Estrada-Peña et al., 2006). However, *R. microplus* has recently established viable populations in West African countries, namely Ivory Coast and Benin (De Clercq et al., 2012; Madder et al., 2012), Burkina Faso, Mali and Togo (Adakal et al., 2013), and Namibia (Nyangiwe et al., 2013). All four species of the *Rhipicephalus Boophilus* group (*Rhipicephalus annulatus*, *Rhipicephalus decoloratus*, *Rhipicephalus geigyi* and *R. microplus*) are thus currently present in the West-African region.

R. microplus is believed to have been introduced to Benin in 2004 from Brazil by imported Girolando cattle (Madder et al., 2012). The farm that first imported Girolando cattle, Kpinnou state farm, is situated in southern Benin and is assumed to be the original site of introduction.

Spatially explicit knowledge of environmental suitability can contribute to an efficient allocation of resources for control. In an attempt to predict the likely spread of *R. microplus*, a number of approaches using spatially explicit data are available to estimate the potential species distribution (Elith et al., 2006). The majority of these approaches compare the known distribution of a species to eco-climatic variables and derive statistical associations, which are then extended to areas with unknown species occurrence. Different techniques vary in how they model the distribution of the response variable, select relevant predictor variables, define functions for each variable, and predict geographic patterns of occurrence (Guisan and Zimmermann, 2000). Some studies use general linear models (GLM) to predict the distribution of disease and their vectors (Klingseisen et al., 2013) or linear discriminant analysis (LDA) (Ayala et al., 2009), while the number of papers using machine-learning methods (such as RandomForests – RF) is rising (Elith et al., 2011, 2006).

Not only the modelling approach, but also the environmental information sources used to explain species distribution is evolving. Eco-climatic variables are routinely used as predictor variables for tick distribution; like many other arthropods, ticks are sensitive to high temperatures and drought (Zeman and Lynen, 2010). The local climate can promote or inhibit several critical parts of their life cycle (Corson et al., 2004). Survival of the eggs and larvae depends largely on saturation deficit, regulated by temperature and relative humidity (Corson et al., 2004; Zeman and Lynen, 2010). Data layers describing vegetation characteristics can be a valuable addition to the predictor variables, since vegetation patches can locally provide sheltered habitats for ticks where climatic conditions are too harsh (Corson et al., 2004).

Two major approaches have been developed to monitor the factors creating favourable conditions for tick activity and survival, one involving the interpolation of weather station data, and another involving a diversity of remotely sensed data. While weather station data have been processed to produce biologically meaningful climate-related variables, the remotely sensed data have provided a broader diversity of environmental descriptors, including information on vegetation cover. Data layers describing vegetation characteristics (such as the Normalized Derived Vegetation Index, NDVI) can be a valuable addition, since a description of the vegetation could replace data on relative humidity, a key factor for tick survival. In the late nineties, pioneering work has been done on the biological link between disease vector distributions and NDVI as measured by satellite sensors (Randolph, 1997; Rogers et al., 1996). Distribution models for *Ixodes ricinus* have identified NDVI as the best performing single variable describing its habitat (Estrada-Peña, 1999). Data on climate have been collected and distributed by meteorological networks on the internet, and remotely sensed datasets have become available for download (Calistri et al., 2013; Klingseisen et al., 2013) and are being rapidly adopted by epidemiologists for characterising vector habitats (Kalluri et al., 2007).

Given the increasing use of species distribution modelling techniques and growing accessibility of environmental data layers of high spatial resolution, a clear need exists for comprehensive analyses of the predictive performance of species distribution modelling techniques and the effect of different environmental data layers on the

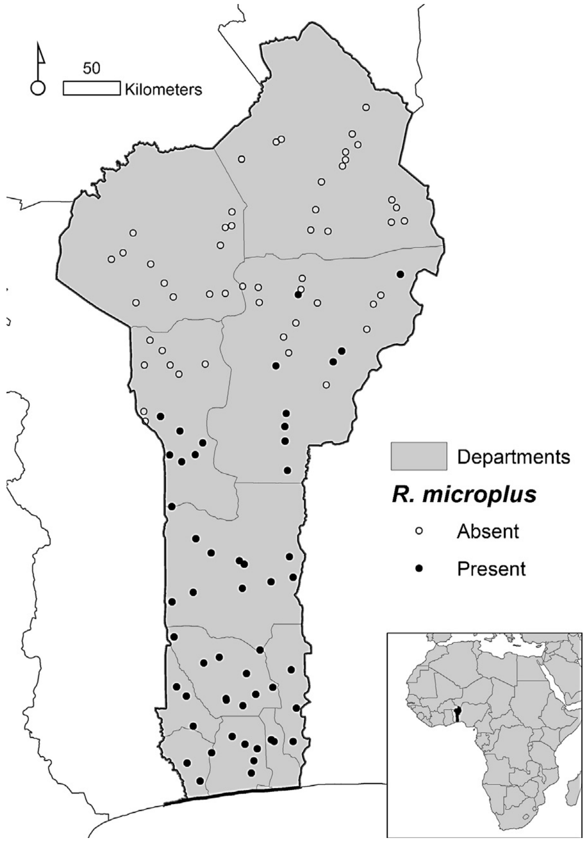


Fig. 1. (a): Situation of Benin in West Africa and (b) the location of the sampled points.

performance of these techniques (Zeman and Lynen, 2006). This study aims at evaluating the predictive performance of different modelling algorithms and different environmental datasets. The results produced also highlight the environmental suitability for *R. microplus* in Benin.

2. Materials and methods

2.1. Study area

The study was conducted in Benin (Fig. 1). Benin covers a land area of 112,622 km^2 and is located between 1°00′–3°51′E and 6°14′–12°24′N. The country is divided in two latitudinal climatic zones. The southern climatic zone has an equatorial type of climate with four seasons; both the dry and wet seasons occur twice a year. In the north, the wet and dry seasons occur only once a year. About 65% of the whole territory of Benin is covered by bushy vegetation suitable for grazing.

2.2. Data

Observed presence or absence of the cattle tick was collected in 104 randomly selected farms throughout the country. Each farm was visited once at the end of the rainy season in 2011 (De Clercq et al., 2012). Geographic coordinates on latitude and longitude in the WGS84 system were recorded using smart phones equipped with the EpiCollect application for the Android platform (Aanensen et al., 2009). Two animals were sampled at each farm, selecting those with the highest tick load. Ticks of the *R. Boophilus* genus were found at all farms, with an average 55 ticks per animal (median 42 ticks per animal), varying between 2 and 250 *R. microplus* ticks per infested animal. This upper limit was probably an underestimation since from heavily infested animals not all ticks were collected because of time constraints. Most cattle herders had applied tick control measures one month prior to sampling, and thus only presence/absence at the herd level was used in further studies. More information on sampling and tick identification protocols can be found in De Clercq et al. (2012). *R. microplus* was found in 50% of the sampled herds, which were mostly located in the southern half of the country, while it was largely absent from the northern half of the country (Fig. 1).

Table 1
Code and description of all used variables.

Code	Description	Source
bio_01	Annual mean temperature	Worldclim
bio_02	Mean diurnal range	Worldclim
bio_03	Isothermality	Worldclim
bio_04	Temperature seasonality	Worldclim
bio_05	Max temperature of warmest month	Worldclim
bio_06	Min temperature of coldest month	Worldclim
bio_07	Temperature annual range	Worldclim
bio_08	Mean temperature of wettest quarter	Worldclim
bio_09	Mean temperature of driest quarter	Worldclim
bio_10	Mean temperature of warmest quarter	Worldclim
bio_11	Mean temperature of coldest quarter	Worldclim
bio_12	Annual precipitation	Worldclim
bio_13	Precipitation of wettest month	Worldclim
bio_14	Precipitation of driest month	Worldclim
bio_15	Precipitation seasonality	Worldclim
bio_16	Precipitation of wettest quarter	Worldclim
bio_17	Precipitation of driest quarter	Worldclim
bio_18	Precipitation of warmest quarter	Worldclim
bio_19	Precipitation of coldest quarter	Worldclim
dLST_a0	Average day temperature	MODIS
dLST_a1	Annual variability in day temperature	MODIS
dLST_a2	Biannual variability in day temperature	MODIS
dLST_a3	Triannual variability in day temperature	MODIS
NDVI_a0	Average NDVI	MODIS
NDVI_a1	Annual variability in NDVI	MODIS
NDVI_a2	Biannual variability in NDVI	MODIS
NDVI_a3	Triannual variability in NDVI	MODIS

Environmental data layers were collected from different sources. A list of all variables is included in Table 1. The Worldclim and MODIS datasets are the most widely used datasets in the species modelling literature. Worldclim data are based on interpolated weather station data, while the data based on MODIS imagery are not interpolated, but derived from satellite observations. Bioclimatic data were obtained from the WorldClim database (http://www.worldclim.org/current). WorldClim data layers were generated through interpolation of average monthly climate data from weather stations at spatial resolution of 30 arc-second (Hijmans et al., 2005). Based on average monthly measurements from 1950 to 1990, this dataset contained layers describing the mean annual temperature, as well as the maximum and minimum temperature. Other layers included the minimum and maximum monthly precipitation, as well was the mean temperature during the rainy season. Since these data might be considered outdated, the result obtained using the WorldClim data were compared with a more timely remotely sensed alternative derived from the MODIS sensor. The MODIS constellation has been described in detail in Klingseisen et al. (2013). The products used in the present study were the daytime land surface temperature (dLST) and the Normalized Difference Vegetation Index (NDVI), both with a spatial resolution of 1 km^2. These MODIS products were downloaded for the year 2011 from the LP DAAC website (USGS, 2013); both the dLST as the NDVI images were subjected to a temporal spline interpolation to remove noise. The dLST-derived data were distributed with a time interval of 8 days, while NDVI had a time interval of 16 days. After spline interpolation, yearly averages and seasonal trends (referred to as harmonics) were extracted from the dataset using Fourier decomposition (Rogers and Randolph, 2006; Rogers et al., 1996). For the dLST variable, the first three harmonics explained 73% of the variance, whereas for the NDVI variables, the first three harmonics explained 79% of the variance. The mean values, as well as the amplitude of the annual, bi-annual and tri-annual cycle were used.

Both ArcGIS (ESRI, 2009) and QuantumGIS (Quantum GIS Development Team, 2013) were used to process spatial data, project the data in geographic coordinates (datum: WGS84), resample all raster layers to 1 km resolution and prepare the maps.

2.3. Variable preprocessing

Both datasets contained variables describing patterns in temperature and precipitation, which can be highly correlated. Including highly correlated variables in a model makes it difficult to determine exactly how each variable influences the species distribution. The correlation between variables was computed for the two sets of predictor variables by a Spearman rank correlation test. Correlation between variables was handled using Principal components analysis (PCA). PCA is a standard method to analyse the pattern of variation in a dataset and translates the data into independent axes—referred to as principal components (Cumming, 2000). The resulting principal components were screened for multivariate normality using QQ plots. Also, the distribution of the sampling points in the environmental feature space was assessed by calculating the Mahalanobis distance of each pixel to the multivariate mean for both the WorldClim and the MODIS dataset. The first five PCA components of the WorldClim and the MODIS datasets were used for modelling.

2.4. Modelling

As *R. microplus* has been present in this region for a little less than a decade, it may not yet have filled its entire ecological niche. The only way to assess this would be to monitor the realized tick distribution continuously over a long time period. However, this tick has an invasive character (Madder et al., 2012) and the extensive livestock movements taking place in Benin from the north to the southern part of the country and vice versa facilitate a rapid expansion of the tick, leading to the assumption that it is indeed possible that the tick has filled the majority of its ecological niche after a decade.

Three modelling techniques were used: generalized linear models (GLM), linear discriminant analysis (LDA) and random forests (RF). All analyses were implemented in R statistical package version 2.15.3 (R Development Core Team, 2013).

The observations were randomly divided into five parts; each model was run five times using respectively four parts to train the model (i.e. the training set) and the fifth part to validate the model (i.e. the validation set), each of the five parts becoming the validation set in turn. This procedure was repeated 100 times, resulting in 500 iterations for each of the models (GLM_WC, LDA_WC, RF_WC, GLM_MOD, LDA_MOD, RF_MOD). The predicted probabilities by each model run were referred to as *suitability maps*, indicating the degree to which the local environment is suitable for *R. microplus*. The 500 runs for each model and dataset were aggregated into mean and standard deviation suitability maps.

2.4.1. Generalized Linear Models (GLM)

Generalized Linear Models (GLM) describe the relationship between the response variable and a set of predictor variables by looking for the best fitting and most parsimonious model (Pohar et al., 2004). Models were fit using maximum likelihood and by allowing the linear model to be related to the response variable via its link function. We used the R package *dismo* (Hijmans et al., 2013), and the ‘binomial’ family with a ‘logit’ link.

2.4.2. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is commonly used to determine which variables better discriminate between two or more classes (here between presence and absence sites) by finding combinations of variables that make the groups as statistically distinct as possible. To discriminate between the presence and absence sites, a linear discriminant function that passes through the centroids of the two groups is used. The statistical computation of LDA is further discussed by Kachigan (Kachigan, 1991). For linear discriminant analysis, we used the R package MASS (Venables and Ripley, 2007).

2.4.3. Random Forests (RF)

Random Forests (RF) is a machine learning technique which generates multiple classifications or regression trees which are aggregated afterwards to compute a classification based on majority voting. The forest is grown by a procedure called bootstrap aggregation. Each tree in the forest is created using a random sample with replacement of the full data set. A randomly selected subset of predictor variables is used to split each node. More details on RF computation can be found in Peters et al. (2011). We implemented this modelling approach using the R package *randomForest* (Liaw and Wiener, 2002).

2.4.4. *Model evaluation*

All models were checked for residual spatial autocorrelation using Moran's I (Haining, 2003). A five-fold cross-validation was used to assess the performance of the models. As 500 training-validation runs were performed, we obtained 500 sets of accuracy measures for each of the six combinations of modelling approaches and datasets (GLM_WC, LDA_WC, RF_WC, GLM_MOD, LDA_MOD, RF_MOD). Five measures of model performance were used to assess the accuracy of the models: percentage correctly classified (PCC), sensitivity (Se), specificity (Sp), Kappa index of agreement (KIA), and Area under the Receiver Operating Characteristic (ROC) curve (AUC).

The first four measures of model performance (PCC, Se, Sp, KIA) require a threshold value in order to convert the probability of species occurrence into absence/presence classes. Jiménez-Valverde and Lobo (2007) pointed out the importance that prevalence be taken into consideration when selecting the criterion to set this threshold. This threshold was therefore set at 0.5, which was the prevalence of *R. microplus* in the dataset. PCC is the percentage of test observations that are correctly classified. Se and Sp express respectively the proportion of correctly predicted presences and the proportion of correctly predicted absences. KIA measures the proportion of correctly predicted sites after the probability of chance agreement has been removed (Moisen and Frescino, 2002). KIA values are negative when the agreement between prediction and observation is worse than expected by chance, and reaches 1 in case of perfect agreement. The ROC plot is obtained by plotting Se against 1-Sp at varying values of the threshold. The area under the ROC curve (AUC) measures the ability of the model to discriminate between presence and absence sites, regardless of a probability threshold. The AUC ranges from 0 to 1, a score of greater than 0.5 indicates a model better than random guessing (Peters et al., 2011).

Accuracy measures were analysed for two datasets:

1. Based on the validation set in the five-fold cross-validation. The entire set of 104 points was randomly divided into four parts (making a total of 84 points) for fitting the model, and one part (consisting of 20 points) for model validation. These accuracy measures provided information on the overall predictive performance of the model.
2. Considering the apparent north-south gradient in the presence or absence of the tick (Fig. 1), points in the extreme north or south of the country are expected to be easily predictable. In a second phase, the accuracy measures were therefore computed taking only the 33 points in the transition zone, between 9.26 and 10.42 degrees latitude. These values corresponded respectively with the most southerly absence and the most northerly presence of the cattle tick.

The accuracy measures of these six models (GLM_WC, GLM_MOD, LDA_WC, LDA_MOD, RF_WC, RF_MOD) were tested for significant differences. As these values did not follow a normal distribution, we used a Kruskall–Wallis test to detect a difference between models, and a

Table 2

Spearman Rank correlation for the WorldClim variables. Significant values are indicated by “*” ($\alpha = 0.05$).

	bio_01	bio_02	bio_03	bio_04	bio_05	bio_06	bio_07	bio_08	bio_09	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	bio_16	bio_17	bio_18	bio_19
bio_01	1																		
bio_02	−0.29*	1																	
bio_03	0.12	−0.8*	1																
bio_04	−0.1	0.81*	−0.9*	1															
bio_05	0.07	0.84*	−0.86*	0.88*	1														
bio_06	0.32*	−0.88*	0.9*	−0.88*	−0.81*	1													
bio_07	−0.14	0.91*	−0.94*	0.93*	0.93*	−0.96*	1												
bio_08	0.92*	−0.48*	0.32*	−0.36*	−0.22*	0.48*	−0.36*	1											
bio_09	0.41*	−0.78*	0.83*	−0.78*	−0.69*	0.93*	−0.85*	0.53*	1										
bio_10	0.81*	0.21*	−0.37*	0.43*	0.55*	−0.22*	0.38*	0.63*	−0.08	1									
bio_11	0.96*	−0.39*	0.26*	−0.27*	−0.09	0.44*	−0.29*	0.96*	0.52*	0.71*	1								
bio_12	−0.75*	−0.12	0.13	−0.14	−0.32*	0.07	−0.21*	−0.65*	−0.05	−0.72*	−0.69*	1							
bio_13	−0.23*	0.67*	−0.77*	0.79*	0.64*	−0.75*	0.74*	−0.35*	−0.71*	0.23*	−0.29*	0.08	1						
bio_14	0.12	−0.86*	0.77*	−0.85*	−0.81*	0.84*	−0.87*	0.31*	0.68*	−0.36*	0.24*	0.22*	−0.67*	1					
bio_15	−0.13	0.86*	−0.89*	0.93*	0.88*	−0.89*	0.93*	−0.36*	−0.79*	0.38*	−0.26*	−0.17	0.87*	−0.85*	1				
bio_16	−0.63*	0.68*	−0.6*	0.7*	0.51*	−0.67*	0.63*	−0.74*	−0.64*	−0.16	−0.69*	0.45*	0.8*	−0.65*	0.72*	1			
bio_17	0.17	−0.88*	0.87*	−0.93*	−0.88*	0.93*	−0.96*	0.4*	0.81*	−0.35*	0.32*	0.16	−0.74*	0.93*	−0.93*	−0.67*	1		
bio_18	0.44*	−0.66*	0.52*	−0.57*	−0.58*	0.55*	−0.59*	0.66*	0.48*	0.15	0.56*	−0.17	−0.46*	0.6*	−0.65*	−0.66*	0.64*	1	
bio_19	−0.73*	0.41*	−0.12	0.25*	0.13	−0.25*	0.19*	−0.78*	−0.24*	−0.47*	−0.71*	0.59*	0.27*	−0.3*	0.23*	0.71*	−0.24*	−0.5*	1

Mann–Whitney *U* test as a post-hoc test for a pairwise comparison of all models. The Bonferroni correction was used to allow for multiple comparisons.

3. Results

3.1. Variable preprocessing

3.1.1. Correlation in the Worldclim dataset

Since the bioclimatic variables within Worldclim described the climate in alternative manners, these variables were highly correlated (Table 2). Nearly all pairs of variables featured a significant correlation, and some had a correlation coefficient with an absolute value higher than 0.9, such as Bio_01 (mean annual temperature) and Bio_11 (mean temperature of the coldest quarter). The variable Bio_07 (temperature annual range) was highly correlated with no less than seven other variables: Bio_02 (mean diurnal range), Bio_03 (isothermality), Bio_04 (temperature seasonality), Bio_05 (maximum temperature of the warmest month), Bio_06 (minimum temperature of the coldest month), Bio_15 (precipitation seasonality), Bio_17 (precipitation of coldest quarter).

3.1.2. Correlation in the dataset containing the Fourier components of the MODIS data

Although Fourier components were assumed to be independent, the results in Table 3 show that they were correlated. Nearly all pairs had significant correlation, but the values were lower than with the Worldclim variables. The highest (absolute) value (−0.81) was observed between dLST_a0 (average yearly daytime temperature) and NDVI_a0 (average yearly vegetation greenness).

3.1.3. Principal components of the WordClim datasets

Because of the high correlation between WorldClim variables, we transformed this dataset into principal components. The first five components explained 96% of the total variance. The loadings of the original variables for these factors are listed in Table 4. The first component had its highest loading for Bio_04 (Temperature Seasonality), Bio_05 (Max Temperature of Warmest Month), Bio_07 (Temperature Annual Range) and Bio_15 (Precipitation Seasonality). The second component was largely determined by Bio_08 (Mean Temperature of Wettest Quarter). For component three, the most influencing variables were Bio_09 (Mean Temperature of Driest Quarter), Bio_13 (Precipitation of Wettest Month) and Bio_16 (Precipitation of Wettest Quarter). Bio_13 was an important variable for component four, and its loading (0.49) was opposed to that of Bio_19 (Precipitation of Coldest Quarter) (−0.40). The fifth component had its highest negative loadings for Bio_02 (Mean Diurnal Range) and Bio_14 (Precipitation of Driest Month). Overall, all features included in the WorldClim dataset regarding temperature and precipitation were withheld in the first five principal components. Fig. 2 shows that this dataset did not diverge significantly from a normal multivariate distribution.

3.1.4. Principal components of the MODIS-derived dataset

Although collinearity in this dataset was not as high as in the Worldclim dataset, PCA was also performed on this dataset. The first five components explained 91% of the total variance. The loadings of the original variables for the first five factors can be found in Table 5. The first component had its highest loading for the mean day temperature (dLST_a0). The second component was largely determined by the difference between the hottest and coolest temperatures (dLST_a1). The third component was determined mainly by tri-annual fluctuations in vegetation greenness (NDVI_a3). Tri-annual variability in temperature (dLST_a3) was the main contributing variable for the fourth component. Finally, the fifth component could be attributed to bi-annual variability in vegetation greenness. Fig. 2 shows that this dataset did not diverge significantly from a normal multivariate distribution.

3.2. Modelling

The models developed did not feature residual autocorrelation and were generally found to have good performance when the entire country was considered (Table 6). All models had a mean PCC and AUC value of greater than 0.92, which indicated a very good fit. The values for KIA were lower, but still higher than 0.83, which was very good. Overall, all three modelling techniques thus performed well and had a comparable modelling performance. Pairwise comparison showed significant differences for all accuracy measures (Table 7). The model LDA_WC had a significantly lower value for PCC, AUC and KIA than the other models, while the model RF_MOD scored significantly higher on these accuracy measures. The models on the WordClim dataset performed worse than those on the MODIS dataset when Se was considered, and better when Sp was considered.

No difference was found between GLM and RF (except for Sp), but when using LDA, Se was significantly lower, while Sp was significantly higher. The difference between LDA and the other modelling techniques (GLM & RF) was less pronounced when the MODIS variables were used as predictor variables. Regardless of their statistical significance, the difference in accuracy between models was small (<0.05).

In a second phase, we focussed on the accuracies in the transition zone. When doing this, the PCC and AUC measures dropped by more than 0.1 for the GLM and LDA models (Table 8). For Se and KIA, only the RF models yielded good results. When compared to GLM, LDA yielded significantly worse results for Se, Sp, Kappa and AUC (Table 9). The RF models on the contrary yielded significantly better results than the GLM and LDA models for all accuracy measures in the transition zone. The comparison between the two datasets was less straightforward. For the RF models, no significant differences were found between the WorldClim and the MODIS dataset in PCC, Sp and AUC values. For KIA and Se, the RF model on MODIS variables gave slightly better results. For the GLM and LDA models, the choice of predictor variables had a significant impact on all accuracy measures. The use of the MODIS dataset

Table 3
Spearman Rank correlation for the MODIS-derived variables. Significant values are indicated by “*” ($\alpha = 0.05$).

	dLST_a0	dLST_a1	dLST_a2	dLST_a3	NDVI_a0	NDVI_a1	NDVI_a2	NDVI_a3
dLST_a0	1							
dLST_a1	0.52*	1						
dLST_a2	0.77*	0.35*	1					
dLST_a3	0.32*	0.42*	0.19*	1				
NDVI_a0	−0.81*	−0.33*	−0.69*	−0.24*	1			
NDVI_a1	0.54*	0.79*	0.34*	0.39*	−0.33*	1		
NDVI_a2	−0.54*	−0.43*	−0.45*	−0.28*	0.52*	−0.57*	1	
NDVI_a3	−0.32*	0	−0.31*	0.06*	0.3*	−0.09*	0.17*	1

Table 4
Loadings for the first five principal components for the dataset containing the WorldClim variables. The largest loadings for each component are printed in **bold**.

Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Cumulative Variance Explained (%)	57	79	88	94	96
bio_01	−0.22	−0.30	0.23	0.02	−0.08
bio_02	−0.26	0.18	−0.05	−0.20	**−0.45**
bio_03	0.28	-0.04	0.02	−0.23	−0.01
bio_04	**−0.29**	−0.01	−0.07	0.16	0.02
bio_05	**−0.29**	−0.02	0.07	−0.09	−0.29
bio_06	0.23	−0.26	0.20	−0.08	0.24
bio_07	**−0.29**	0.12	−0.06	−0.01	−0.29
bio_08	−0.10	**−0.43**	0.08	0.20	−0.02
bio_09	0.11	−0.24	**0.51**	−0.31	0.01
bio_10	−0.28	−0.17	0.13	0.09	−0.07
bio_11	−0.15	−0.34	0.34	0.03	−0.17
bio_12	0.26	0.18	0.18	0.12	−0.21
bio_13	−0.03	0.31	**0.41**	**0.49**	0.14
bio_14	0.25	−0.14	−0.11	0.33	**−0.43**
bio_15	**−0.29**	0.06	0.03	0.11	0.13
bio_16	0.00	0.37	**0.43**	0.25	0.02
bio_17	0.26	−0.17	−0.09	0.28	−0.31
bio_18	0.26	−0.10	−0.06	0.23	−0.24
bio_19	0.16	0.27	0.29	**−0.40**	−0.32

Table 5
Loadings for the first five principal components for the dataset containing the MODIS derived variables. The largest loadings for each component are printed in **bold**.

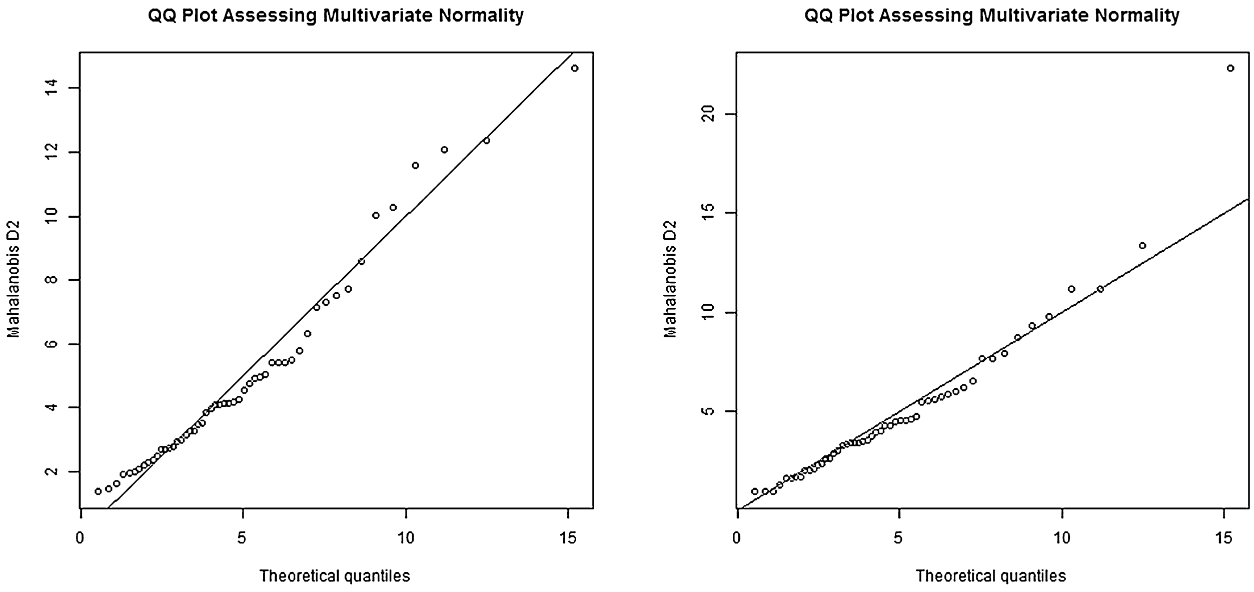
Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Cumulative Variance Explained (%)	46	65	75	84	91
dLST_a0	**−0.46**	0.14	0.09	0.00	−0.28
dLST_a1	−0.36	**−0.43**	−0.22	0.14	−0.38
dLST_a2	−0.38	0.36	0.22	0.05	−0.35
dLST_a3	−0.22	−0.41	0.37	**−0.79**	0.12
NDVI_a0	0.36	−0.39	−0.36	−0.05	−0.15
NDVI_a1	−0.38	−0.40	−0.33	0.19	0.00
NDVI_a2	0.39	0.06	0.00	−0.27	**−0.78**
NDVI_a3	0.18	−0.42	**0.72**	0.49	−0.11

improved Se, KIA, AUC for both GLM and LDA models, but decreased Sp. For PCC, the use of remotely sensed data improved for the LDA approach, but decreased for the GLM approach.

The predicted distribution of *R. microplus* for the different models is visualised in Figs. 3 and 5. As these are the mean predicted values for 500 iterations, the standard deviation around the mean is shown in Figs. 4 and 6.

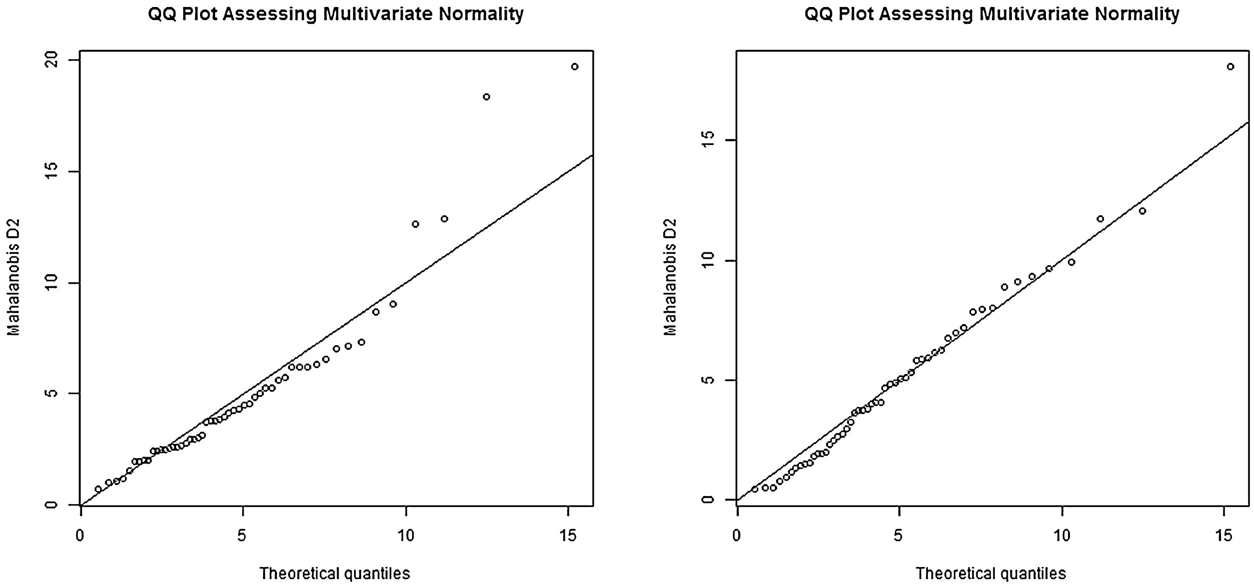
Table 6
Accuracy statistics of the different models. All accuracy measures were determined on the fifth fold, which was not used for the fitting of the model. Average values and standard deviation were computed for 500 iterations.

Approach	Variables	PCC		Se		Sp		KIA		AUC	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
GLM	WC	0.93	0.05	0.90	0.09	0.95	0.08	0.85	0.11	0.93	0.05
LDA	WC	0.92	0.05	0.85	0.10	0.98	0.04	0.83	0.11	0.92	0.05
RF	WC	0.93	0.05	0.89	0.09	0.96	0.07	0.85	0.11	0.93	0.05
GLM	MOD	0.93	0.06	0.92	0.09	0.93	0.08	0.85	0.12	0.93	0.06
LDA	MOD	0.93	0.05	0.90	0.09	0.96	0.06	0.86	0.10	0.93	0.05
RF	MOD	0.94	0.05	0.92	0.08	0.95	0.06	0.87	0.10	0.94	0.05



WorldClim Absences

WorldClim Presences



MODIS Absences

MODIS Presences

Fig. 2. QQ plots assessing multivariate normality of the respective datasets.

The general spatial trends in these six maps were similar. The southern and central parts of the country, subject to relatively moderate temperature and higher precipitation values, were considered suitable for the establishment of *R. microplus*. The northern hot and dry areas were less suitable. When using the Worldclim dataset (Fig. 3), the choice of the modelling algorithm slightly changed the shape of the transition zone between suitable and non-suitable areas. The transition zone from GLM and LDA modelling was smooth and gradual. Using RF, the transition zone was wider; a yellow strip indicating medium suitability descended almost to the coast along the mountain range in the west of Togo. When looking at the maps on variability between model runs (Fig. 4), GLM yielded relatively high variability along the transition zone, as well as in three other spots: around the middle of Togo, eastwards along the coast, and towards the north of Benin. Most of these spots were outside or on the country border of Benin. As is illustrated by the map of the Mahalanobis distances in feature space (Fig. 7), these areas did not

Table 7
Pairwise comparison of the accuracy measures determined on the fifth fold, which was not used for the fitting of the model. Significant differences at a 0.05 level are indicated by '*'. Combinations not tested are indicated by blanks.

PCC	GLM_WC	LDA_WC	RF_WC	GLM_MOD	LDA_MOD	RF_MOD
GLM_WC						
LDA_WC	*					
RF_WC	–	*				
GLM_MOD	–	*	–			
LDA_MOD	–	*	–	–		
RF_MOD	*	*	*	–	–	
SE						
GLM_WC						
LDA_WC	*					
RF_WC	–	*				
GLM_MOD	*	*	*			
LDA_MOD	–	*	–	*		
RF_MOD	*	*	*	–	*	
SP						
GLM_WC						
LDA_WC	*					
RF_WC	–	*				
GLM_MOD	*	*	*			
LDA_MOD	–	*	–	*		
RF_MOD	–	*	–	*	–	
KIA						
GLM_WC						
LDA_WC	*					
RF_WC	–	*				
GLM_MOD	–	*	–			
LDA_MOD	–	*	–	–		
RF_MOD	*	*	*	–	–	
AUC						
GLM_WC						
LDA_WC	*					
RF_WC	–	*				
GLM_MOD	–	*	–			
LDA_MOD	–	*	–	–		
RF_MOD	*	*	*	–	–	

have environmental conditions (temperature and vegetation patterns) similar to the locations sampled during the field sampling. The results for LDA appeared to be more stable, with lower standard deviations throughout, and variability only along the transition zone and eastwards along the coast. Variability in the RF model was generally low, and distributed in a patchy manner over the study area.

Using the MODIS dataset (Fig. 5), the prediction maps seemed to better fit the invasion front. It was difficult to see a difference between the maps obtained from the different modelling approaches. There was a small difference in width of the suitable strip along the Niger River in the north of the country, and the area indicated as 'medium suitability' (in yellow) occupied a larger part for the RF approach. The variability around the predicted mean suitability was generally lower than with the WorldClim dataset (Fig. 6). GLM results showed most variability, LDA and RF appeared to be more robust techniques. Most variability was observed in the transition zone, and along the Niger River along the northern border of Benin. This feature was also present in the map of the Mahalanobis distances in Fig. 7.

4. Discussion

The inference of the abiotic niche of the tick *R. microplus* and its projection over territories at country scale are

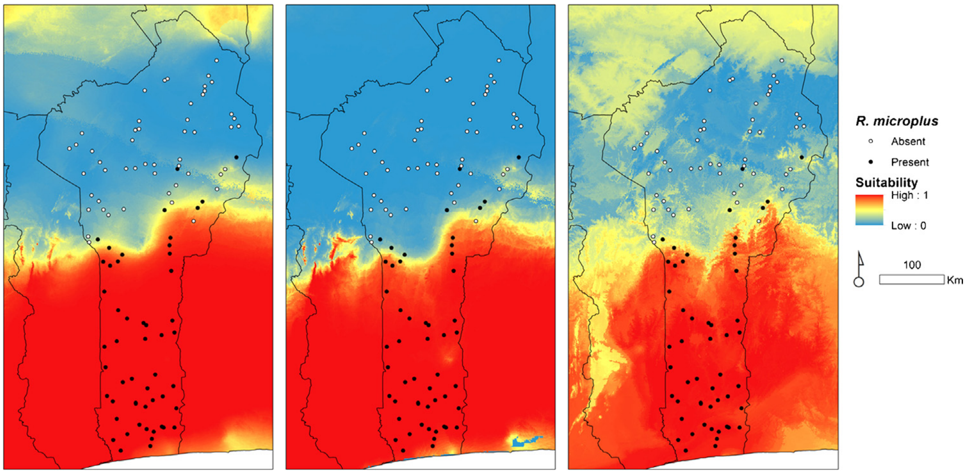
Table 8
Accuracy statistics of the different models. These accuracy measures were determined on the points between 9.26 and 10.42 degrees latitude. Average values and standard deviation were computed for 500 iterations for each model.

Approach	Variables	PCC		Se		Sp		KIA		AUC	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
GLM	WC	0.82	0.02	0.38	0.13	0.94	0.03	0.37	0.11	0.73	0.02
LDA	WC	0.80	0.01	0.14	0.01	0.98	0.02	0.17	0.03	0.74	0.01
RF	WC	0.95	0.03	0.85	0.13	0.98	0.03	0.86	0.11	0.94	0.06
GLM	MOD	0.81	0.03	0.51	0.11	0.90	0.02	0.42	0.1	0.72	0.06
LDA	MOD	0.81	0.03	0.36	0.14	0.93	0.01	0.33	0.13	0.66	0.02
RF	MOD	0.96	0.03	0.88	0.11	0.98	0.03	0.88	0.09	0.95	0.05

Table 9
Pairwise comparison of the accuracy measures, determined on the points between 9.26 and 10.42 degrees latitude. Significant differences at a 0.05 level are indicated by ‘*’. Combinations not tested are indicated by blanks.

PCC	GLM_WC	LDA_WC	RF_WC	GLM_MOD	LDA_MOD	RF_MOD
GLM_WC						
LDA_WC	*					
RF_WC	*	*				
GLM_MOD	*	*	*			
LDA_MOD	*	*	*	–		
RF_MOD	*	*	–	*	*	
SE						
GLM_WC						
LDA_WC	*					
RF_WC	*	*				
GLM_MOD	*	*	*			
LDA_MOD	–	*	*	*		
RF_MOD	*	*	*	*	*	
SP						
GLM_WC						
LDA_WC	*					
RF_WC	*	*				
GLM_MOD	*	*	*			
LDA_MOD	*	*	*	*		
RF_MOD	*	*	–	*	*	
KIA						
GLM_WC						
LDA_WC	*					
RF_WC	*	*				
GLM_MOD	*	*	*			
LDA_MOD	*	*	*	*		
RF_MOD	*	*	*	*	*	
AUC						
GLM_WC						
LDA_WC	*					
RF_WC	*	*				
GLM_MOD	*	*	*			
LDA_MOD	*	*	*	*		
RF_MOD	*	*	–	*	*	

necessary steps to evaluate the vulnerability of those territories to the invasive spread of the tick. An effective control strategy at the national level would spatially divide the country in a surveillance zone where the environment is unsuitable, a zone along the invasion front where spread prevention is attempted, and a control zone in the areas with high environmental suitability. If the survival of *R. microplus* is localized, eradication or control of this tick



(a) (b) (c)

Fig. 3. Maps of mean predicted abiotic suitability for *R. microplus* developed using the WorldClim data set and different modelling techniques: (a) GLM; (b) LDA; (c) RF.

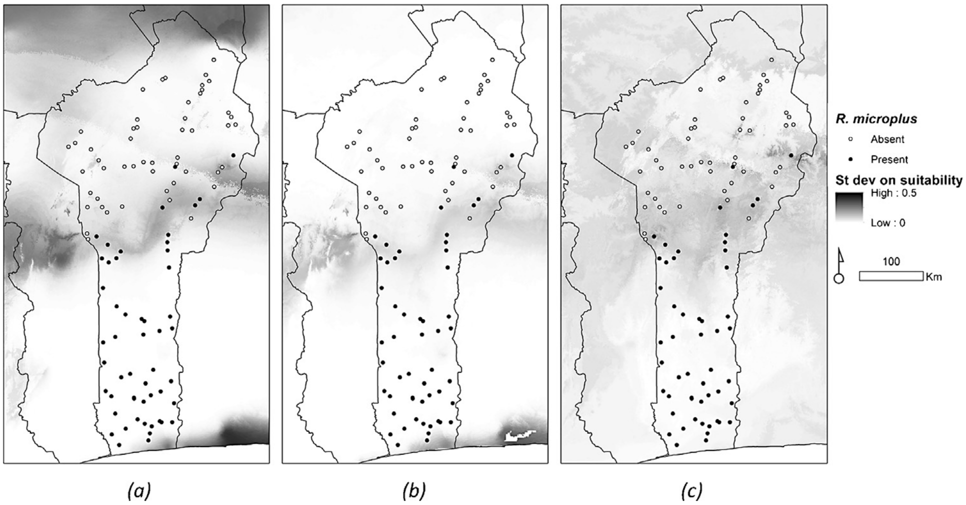


Fig. 4. Maps of standard deviation around the mean predicted abiotic suitability for *R. microplus* developed using the WorldClim data set and different modelling techniques: (a) GLM; (b) LDA; (c) RF.

becomes feasible. Predictive modelling can provide such cartographic information.

In this study we assessed the influence of modelling techniques and the type of environmental data used on the prediction of environmental suitability and its accuracy measures. Overall, when predicting disease risks, and especially the potential risk for invasive species, presence data are more reliable than absence data (Dicko et al., 2014). Sites where the tick was not observed might represent suitable areas which were not yet invaded or where the sampling method failed to catch an individual of the present population. It thus makes more sense to prioritize Se (i.e. true positive rate) over Sp (true negative rate).

During the preliminary study where different assumptions were verified, it was quite surprising to observe a high correlation between the MODIS-derived variables, since it is generally assumed that Fourier decomposition results in independent variables. In ecological suitability modelling as in this study, not the coefficients of the Fourier harmonics, but rather the derived amplitude and phase were used since they allow for a better biological interpretation. Also from a biological point of view, it is evident that vegetation greenness would be correlated with temperature. We would therefore like to stress the importance of carefully checking the correlation between the explanatory variables before modelling.

The performance of the modelling techniques and datasets tested over the entire country was comparable; significant differences were found, but they were relatively small (<0.05). Differences in accuracy between the models became apparent when zooming in to the transition zone, the area where the frontier of tick spread may change from year to year, and where the uncertainty observed in our results suggests large changes in the suitability for the tick. It is important to note that this transition zone is actually the main area of interest, both from a

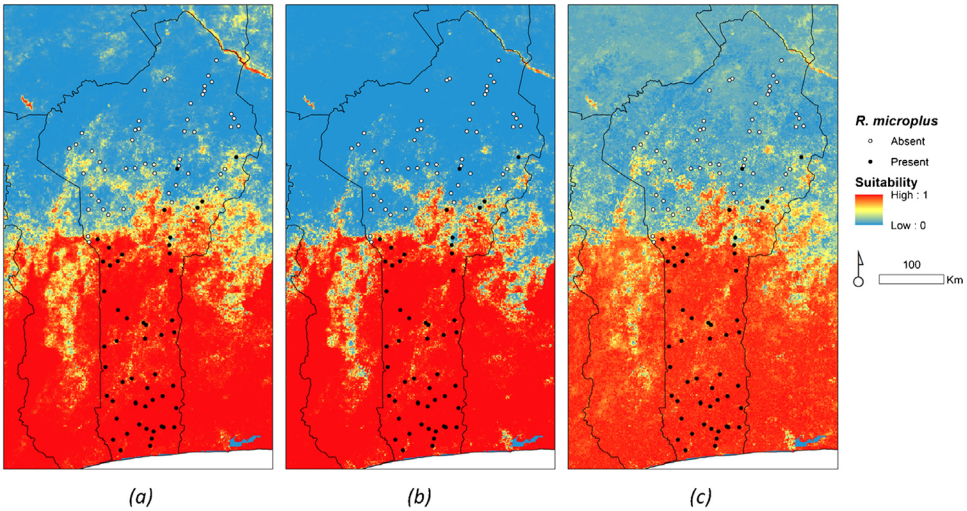


Fig. 5. Maps of the mean predicted abiotic suitability for *R. microplus* developed using the MODIS data set and different modelling techniques: (a) GLM; (b) LDA; (c) RF.

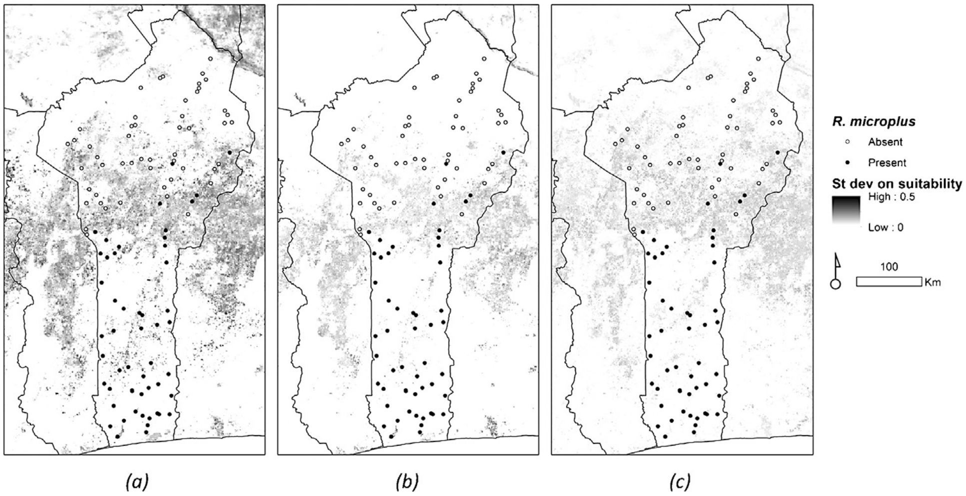


Fig. 6. Maps of standard deviation around the mean predicted abiotic suitability for *R. microplus* developed using the MODIS data set and different modelling techniques: (a) GLM; (b) LDA; (c) RF.

biological and from a disease management point of view. Here, RF came out as the best performing modelling technique with the highest values for various accuracy measures, and the use of remotely sensed variables yielded better results.

The areas of high abiotic suitability predicted by the models in this study were in the southern part of the country featuring an equatorial climate. This corresponds with the environmental requirements of *R. microplus* as reported in other studies (Estrada-Peña et al., 2006). While the differences in the maps produced by different modelling approaches were mainly seen in local spatial details, the differences in the maps obtained by using two different types of explanatory variables were more striking.

No gradual transition between suitable and non-suitable was seen in the MODIS-based maps, but rather a varying density of suitable areas. This clearly reflects the nature of the input data; the remotely sensed data gave a more fragmented result simply because they contain more spatial detail, since they were not based on interpolated data but yielded an explicit measurement for each pixel. MODIS imagery can detect thus smaller land cover elements that are not suitable for tick presence, such as water bodies. This makes MODIS imagery better suited for capturing the invasion front. Our results draw attention to the fact that visual patterns may differ widely between models and input data, even if the values for the accuracy measure do not reflect this.

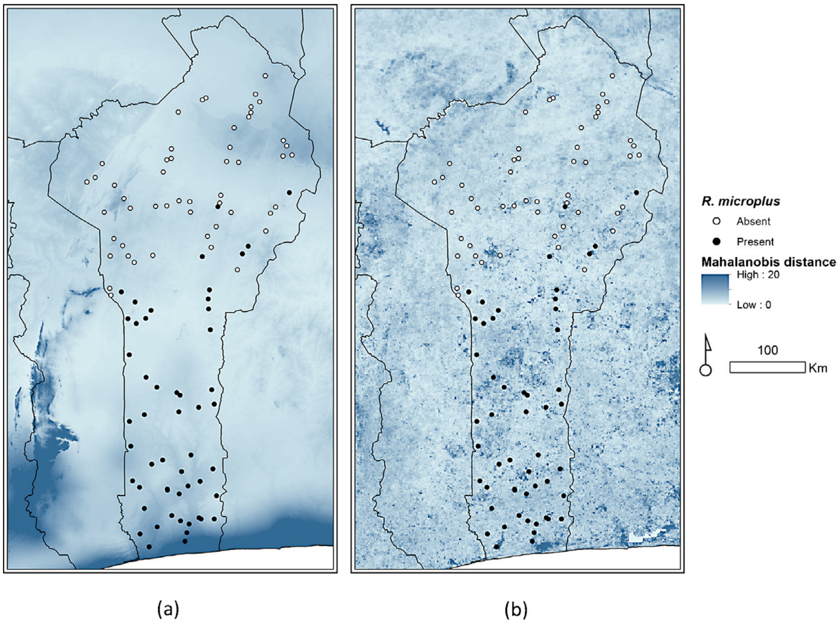


Fig. 7. Maps of the Mahalanobis distances in the dataset containing (a): WorldClim variables, and (b) MODIS-derived variables.

A second advantage is that the timing of the MODIS data fitted more closely to the sampling period. If data on tick abundance would be collected longitudinally and over different years, this could be put in relation with environmental data collected during the respective sampling period, and provide insight in intra- and inter-annual variability in tick presence and tick abundance.

We must consider the possibility that other datasets might have given yet better results. For instance, the Worldclim data give climate averages over the period 1950-1990, which does not coincide with the sampling period (September–December 2011). The reason for not using weather station data from the year 2011 is mainly that they are not readily available in a format suitable for modellers. Also, the value of specific conditions on a particular year is limited for spatial prediction purposes. Secondly, temperatures have risen by about 0.3 °C since the second half of the 20th century (Hulme et al., 2001; Sarr, 2012). No evidence has been found indicating significant increase or decrease in rainfall. Given the high accuracy of the obtained models, it is unlikely that more recent climate data would have improved the predictions. Finally, predictive models would probably improve when including information about movements of cattle and details of the life cycle processes of the tick.

5. Conclusions

We compared the ability of several modelling algorithms and datasets of explanatory variables (either climate interpolated or remotely sensed information) to describe the potential distribution of the tick *R. microplus* in parts of West Africa. In this study the overall performance of the modelling techniques and datasets tested was comparable. Predicted probabilities differed most in the transition zone, which is by definition the most difficult area to predict. In the transition zone, the models using remotely sensing data (MODIS) did better in capturing the invasion front. All models and dataset concurred in the high environmental suitability of the southern half of the Beninese territory for *R. microplus*. This raises concern at the regional level for animal health, including in regard to its potential to substantially alter transmission risk of *B. bovis*. The northern part of Benin appeared overall of low environmental suitability; however, continuous surveillance in the transition zone is highly relevant, in relation to important cattle movements in the region, and to the invasive character of *R. microplus*.

Acknowledgements

This work was funded by the Belgian Science Policy Programs (Belspo, SR/00/144). We would like to thank the anonymous reviewers for their constructive remarks, which have substantially improved the original manuscript.

References

Aanensen, D.M., Huntley, D.M., Feil, E.J., Al-Own, F., Spratt, B.G., 2009. EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. PLoS One, 4.

Adakal, H., Biguezoton, A., Zoungrana, S., Courtin, F., De Clercq, E.M., Madder, M., 2013. Alarming spread of the Asian cattle tick *Rhipicephalus microplus* in West Africa—another three countries are affected: Burkina Faso, Mali and Togo. Exp. Appl. Acarol., http://dx.doi.org/10.1007/s10493-013-9706-6.

Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.-P., Awono-Ambene, P., Fontenille, D., Simard, F., 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar. J. 8, http://dx.doi.org/10.1186/1475-2875-8-307.

Calistri, P., Ippoliti, C., Candeloro, L., Benjelloun, A., El Harrak, M., Bouchra, B., Danzetta, M.L., Di Sabatino, D., Conte, A., 2013. Analysis of climatic and environmental variables associated with the occurrence of West Nile virus in Morocco. Prev. Vet. Med. 110, 549–553, http://dx.doi.org/10.1016/j.prevetmed.2013.02.011.

Corson, M.S., Teel, P.D., Grant, W.E., 2004. Microclimate influence in a physiological model of cattle-fever tick (*Boophilus spp.*) population dynamics. Ecol. Model. 180, 487–514.

Cumming, G.S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. J. Biogeogr. 27, 441–455.

De Clercq, E.M., Vanwambeke, S., Sungirai, M., Adehan, S., Lokossou, R., Madder, M., 2012. Geographic distribution of the invasive cattle tick *Rhipicephalus microplus*, a country-wide survey in Benin. Exp. Appl. Acarol. 58, 441–452, http://dx.doi.org/10.1007/s10493-012-9587-0.

Dicko, A.H., Lancelot, R., Seck, M.T., Guerrini, L., Sall, B., Lo, M., Vreysen, M.J.B., Lefrançois, T., Fonta, W.M., Peck, S.L., Bouyer, J., 2014. Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. Proc. Natl. Acad. Sci. USA 111, 10149–10154, http://dx.doi.org/10.1073/pnas.1407773111.

Elith, J., Graham, R.P.A.C.H., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.Y., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

ESRI, 2009. ArcGIS 9.

Estrada-Peña, A., 1999. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Ixodidae) in South America. Vet. Parasitol. 81, 73–82, http://dx.doi.org/10.1016/S0304-4017(98)00238-6.

Estrada-Peña, A., Bouattour, A., Camicas, J.-L., Guglielmone, A., Horak, I., Jongejan, F., Latif, A., Pegram, R., Walker, A.R., 2006. The known distribution and ecological preferences of the tick subgenus *Boophilus* (Acari: Ixodidae) in Africa and Latin America. Exp. Appl. Acarol. 38, 219–235, http://dx.doi.org/10.1007/s10493-006-0003-5.

Estrada-Peña, A., Venzal, J.M., Nava, S., Mangold, A., Guglielmone, A.A., Labruna, M.B., De La Fuente, J., 2012. Reinstatement of *Rhipicephalus (Boophilus) australis* (Acari: Ixodidae) with redescription of the adult and larval stages. J. Med. Entomol. 49, 794–802, http://dx.doi.org/10.1603/ME11223.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186, http://dx.doi.org/10.1016/S0304-3800(00)00354-9.

Haining, R., 2003. Spatial Data Analysis: Theory and Practice, 1st edition. Cambridge University Press, Cambridge, UK; New York.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Hijmans, R.J., Phillips, S.J., Leathwick, J., Elith, J., 2013. Dismo: Species distribution modeling. R package.

Hoch, T., Goebel, J., Agoulon, A., Malandrin, L., 2012. Modelling bovine babesiosis: a tool to simulate scenarios for pathogen spread and to test control measures for the disease. Prev. Vet. Med. 106, 136–142, http://dx.doi.org/10.1016/j.prevetmed.2012.01.018.

Hulme, M., Doherty, R., Ngara, T., New, M., Lister, D., 2001. African climate change: 1900–2100. Clim. Res. 17, 145–168.

Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta Oecologica 31 (3), 361–369, http://dx.doi.org/10.1016/j.actao.2007.02.001.

Kachigan, S., 1991. Multivariate Statistical Analysis: A Conceptual Introduction [WWW Document]. Barnes Noble. URL ⟨http://

www.barnesandnoble.com/w/multivariate-statistical-analysis-samkash-kachigan/1101643729⟩ (accessed 8.1.13).

Kalluri, S., Gilruth, P., Rogers, D., Szczur, M., 2007. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. PLoS Pathog. 3, 1361–1371.

Klingseisen, B., Stevenson, M., Corner, R., 2013. Prediction of Bluetongue virus seropositivity on pastoral properties in northern Australia using remotely sensed bioclimatic variables. Prev. Vet. Med. 110, 159–168, http://dx.doi.org/10.1016/j.prevetmed.2012.12.001.

Liaw, A., Wiener, M., 2002. Classification and Regression by Random Forest.

Madder, M., Adehan, S., De Deken, R., Adehan, R., Lokossou, R., 2012. New foci of *Rhipicephalus microplus* in West Africa. Exp. Appl. Acarol. 56, 385–390.

Madder, M., Thys, E., Achi, L., Touré, A., De Deken, R., 2011. *Rhipicephalus (Boophilus) microplus:* a most successful invasive tick species in West-Africa. Exp. Appl. Acarol. 53, 139–145, http://dx.doi.org/10.1007/s10493-012-9522-4.

Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Model. 157, 209–225, http://dx.doi.org/10.1016/S0304-3800(02)00197-7.

Nyangiwe, N., Matthee, C., Horak, I., Matthee, S., 2013. First Record of the Pantropical Blue Tick *Rhipicephalus microplus* in Namibia.

Peters, J., De Baets, B., Van doninck, J., Calvete, C., Lucientes, J., De Clercq, E.M., Ducheyne, E., Verhoest, N.E.C., 2011. Absence reduction in entomological surveillance data to improve niche-based distribution models for *Culicoides imicola*. Prev. Vet. Med. 100, 15–28.

Pohar, M., Blas, M., Turk, S., 2004. Comparison of logistic regression and linear discriminant analysis: a simulation study. Metodol. Zv. 1, 143–161.

Quantum GIS Development Team, 2013. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. ⟨http://qgis.osgeo.org⟩.

Randolph, S.E., 1997. Abiotic and biotic determinants of the seasonal dynamics of the tick *Rhipicephalus appendiculatus* in South Africa. Med. Vet. Entomol. 11, 25–37.

R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogers, D.J., Hay, S.I., Packer, M.J., 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Ann. Trop. Med. Parasitol. 90, 225–241.

Rogers, D.J., Randolph, S.E., 2006. Climate change and vector-borne diseases. Adv. Parasitol. 62, 345–381, http://dx.doi.org/10.1016/S0065-308X(05)62010-6.

Sarr, B., 2012. Present and future climate change in the semi-arid region of West Africa: a crucial input for practical adaptation in agriculture. Atmos. Sci. Lett. 13, 108–112, http://dx.doi.org/10.1002/asl.368.

USGS, 2013. MODIS data products, version 05. ⟨http://edcdaac.usgs.gov/modis/dataproducts.asp⟩.

Venables, W.N., Ripley, B.D., 2007. Modern applied statistics with S. Springer, New York, NY [u.a].

Zeman, P., Lynen, G., 2006. Evaluation of four modelling techniques to predict the potential distribution of ticks using indigenous cattle infestations as calibration data. Exp. Appl. Acarol. 39, 163–176.

Zeman, P., Lynen, G., 2010. Conditions for stable parapatric coexistence between *Boophilus decoloratus* and *B. microplus* ticks: a simulation study using the competitive Lotka-Volterra model. Exp. Appl. Acarol. 52, 409–426.