

# Present habitat suitability for *Anopheles atroparvus* (Diptera, Culicidae) and its coincidence with former malaria areas in mainland Portugal

César Capinha<sup>1</sup>, Eduardo Gomes<sup>2</sup>, Eusébio Reis<sup>1</sup>, Jorge Rocha<sup>1</sup>, Carla A. Sousa<sup>3</sup>, V. E. do Rosário<sup>2</sup>, A. Paulo Almeida<sup>3</sup>

<sup>1</sup>Centro de Estudos Geográficos, Universidade de Lisboa, Alameda da Universidade, 1600-214 Lisboa, Portugal; <sup>2</sup>CMDT-LA, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, 96, 1349-008 Lisboa, Portugal; <sup>3</sup>Unidade de Entomologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, 96, 1349-008 Lisboa, Portugal

**Abstract.** Malaria was a major health problem in the first half of the 20th Century in mainland Portugal. Nowadays, although the disease is no longer endemic, there is still the risk of future endemic infections due to the continuous occurrence of imported cases and the possibility of transmission in the country by *Anopheles atroparvus* Van Thiel, 1927. Since vector abundance constitute one of the foremost factors in malaria transmission, we have created several habitat suitability models to describe this vector species' current distribution. Three different correlative models; namely (i) a multilayer perceptron artificial neural network (MLP-ANN); (ii) binary logistic regression (BLR); and (iii) Mahalanobis distance were used to combine the species records with a set of five environmental predictors. Kappa coefficient values from *k*-fold cross-validation records showed that binary logistic regression produced the best predictions, while the other two models also produced acceptable results. Therefore, in order to reduce uncertainty, the three suitability models were combined. The resulting model identified high suitability for *An. atroparvus* in the majority of the country with exception of the northern and central coastal areas. Malaria distribution during the last endemic period in the country was also compared with the combined suitability model, and a high degree of spatial agreement was obtained (kappa = 0.62). It was concluded that habitat suitability for malaria vectors can constitute valuable information on the assessment of several spatial attributes of the disease. In addition, the results suggest that the spatial distribution of *An. atroparvus* in the country remains very similar to the one known about seven decades ago.

**Keywords:** *Anopheles atroparvus*, habitat suitability, malaria, geographical information system.

---

## Introduction

Malaria is still one of the most devastating public health problems in the world. In 2006, there were an estimated 247 million cases that caused nearly a million deaths, mostly children under 5 years of age (WHO, 2008). The disease is caused by one of four

species of the genus *Plasmodium*: *Plasmodium falciparum* Welch, 1897, *Plasmodium vivax* Grassi and Feletti, 1890, *Plasmodium ovale* Stephens, 1922 and *Plasmodium malariae* Laveran, 1881. Infection occurs when the parasites are inoculated into humans by female mosquitoes of the genus *Anopheles*, designated as the vector of the disease.

Portugal, like several other European countries, had high incidence rates of this disease until the end of the first half of the 20th century. Although the disease has been eliminated, the former malaria vector *Anopheles atroparvus* Van Theil, 1927 (Cambournac, 1942) is still abundant all through the

Corresponding author:

César Capinha

Centro de Estudos Geográficos

Universidade de Lisboa

Alameda da Universidade, 1600-214 Lisboa, Portugal

Tel. +35 1 21 7940218; Fax +35 1 21 7938690

E-mail: cesarcapinha@hotmail.com

country (Ribeiro et al., 1988; Almeida et al., 2008). The distribution in Portugal of *An. atroparvus*, and of other members of the *Anopheles maculipennis* species complex, has been the object of several studies (e.g. Pires et al., 1982; Ribeiro et al., 1992). However, the continuous representation of this species distribution is currently unknown. Since vector control has been and continues to be a commonly used strategy in malaria control (Muturi et al., 2008), this kind of information can be very valuable in the fight against the disease, should it occur again.

In Portugal, where malaria transmission was mosquito-density dependent, several anti-vector measures were adopted such as the introduction of the mosquito fish (*Gambusia* sp.), intermittent rice field irrigation, improvement of the rice field workers' accommodations (appliance of protective nets in beds, doors and windows) and indoor residual spraying with DDT (Bruce-Chwatt and Zulueta, 1977; Borges, 2001). The implementation of these measures along with patients treatment resulted in a large reduction of the incidence of the disease, which was declared "eradicated" from the country by the WHO in 1973. Since then, despite an isolated case in 1975 (Antunes et al., 1987), the occurrence of the disease is exclusively in the form of imported cases. The highest number of these imported cases was reached with the end of the Portuguese colonial war in 1974 but much later, in the period of 1993 to 2003, Castro et al. (2004) still reported the occurrence of 50 to 85 imported cases per year. This situation, combined with the existence of the local *An. atroparvus* vector which can be infected with exotic strains of plasmodia such as *P. falciparum* (Sousa, 2008), raises the possibility of future endemic infections.

Climate change is considered a key issue for current modifications of the distribution and abundances observed for some mosquito species (e.g. Pascual et al., 2006; Minář et al., 2007). These changes may affect the transmission capability of the vector species and are, therefore, of major importance with regard to public health. Although the global surface temperature has been rising in the

last century, the growth rates are not homogeneously distributed (IPCC, 2007). Recent studies for Portugal show that the country has been experiencing temperature increases in the order of 0.5°C per decade since 1975 (Miranda et al., 2006). Due to the many implications that this environmental factor has on the biology of *An. atroparvus* (Cambournac and Hill, 1938) this modest increase can still already have contributed to change in its distribution.

Habitat suitability models, used to study the distribution of several species, were mainly developed in the fields of biogeography and conservation biology. These models are based on the quantification of relations between the species and several environmental factors considered prominent to its distribution (Guisan and Thuiller, 2005). Such quantifications enclose, in general, two possible approaches: the mechanistic model, based on the knowledge of the species physiology, and the correlative model (Guisan and Zimmerman, 2000). The obtained relations are then applied to spatial data using map algebra modules and a geographical information system (GIS). These approaches reduce the need for highly detailed field surveys and are one of the best ways to achieve spatially continuous representations of species distributions and several studies have adopted this thinking for modelling the distribution of disease vectors (e.g. Kuhn et al., 2002; Moffett et al., 2007). In the present context, the study of the propensity for vector infection through space analysis processes is one of the most potentially rewarding approaches. Therefore, the aims of this study were to:

- (i) obtain the first spatially continuous representation of *An. atroparvus* habitat suitability in the territory;
- (ii) identify the degree of agreement between the achieved distribution and former malaria areas;
- (iii) verify the validity of using vectors distribution models to assess the malaria transmission risk; and
- (iv) identify the existence of recent shifts on the distribution of suitable areas for *An. atroparvus*.

In order to achieve these objectives several habitat suitability models for *An. atroparvus* in mainland Portugal were computed. Three different correlative approaches were used relating records with the presence or absence of the species with several environmental factors. The obtained results were cross-validated and compared with the known malaria distribution in mainland Portugal during its last endemic period.

## Materials and methods

### *Abundance and distribution data of Anopheles*

To gather species distribution data several field surveys were carried out. The adult mosquitoes were collected in human and animal facilities using electrical aspirators. The surveys took place at several localities along mainland Portugal, between the years 2001 and 2003. Mosquito abundance was computed as the number of mosquitoes collected per hour and per collector. The obtained results were, however, inadequate to be used directly, due to (i) the high variability in the conditions of the surveyed facilities, and (ii) the lack of species-specific abundances for the different members of the *An. maculipennis* complex.

The distinct conditions of the surveyed facilities, namely the implementation of anti-mosquito devices, presence or absence of animal hosts and species as well as different degrees of isolation from the exterior environment, led to a high irregularity of abundance values, even for fairly close sites. In this context, it was not possible to consider the abundance values as such, i.e. a low level of abundance could not directly be taken as a representation of a lower suitability area. On the other hand, the opposite holds true, i.e. places where high abundances were registered indicate an environmental suitability to the species in question.

Due to the inter-species similarity of adult *Anopheles* mosquitoes, the survey results comprised abundance records of the whole *An. maculipennis* complex. In order to extract abundance

values considering *An. atroparvus* only, it was necessary to establish a prevalence value of this species relative to the other members of the complex for each locality. Ramos et al. (1978) identified *An. atroparvus* as the only species of the complex in Algarve, and the same was found for Alentejo (Pires et al., 1982). However, in the mountainous Montejunto-Estrela system and in northern areas, the complex is also represented by *An. maculipennis* Meigen, 1818 and some negligible records of *Anopheles melanoon* Hackett, 1934 (Sousa, 2008). For central Portugal, Ribeiro et al. (1992), report an abundance ratio of nine *An. atroparvus* for each *An. maculipennis*, while Ribeiro et al. (1999a) found the lower proportion of six to one for the Serra da Estrela Natural Park. None of the other studies provide exact proportion values, but they all report a majority abundance of *An. atroparvus* (e.g. Ribeiro et al., 1999b, 2002). Taking this information into consideration, a general abundance ratio of eight *An. atroparvus* for each *An. maculipennis* was established. The abundance records reported from locations in the Montejunto-Estrela system and the northern areas were then weighted using this proportion and the ones above the 6<sup>th</sup> decile were extracted from the resulting values. Totalling 76 records, these areas were considered as the ones with a higher suitability for *An. atroparvus*. However, taking into account that even these higher abundance values were also possibly biased due to the variability of the surveyed places, they were converted to “presence only” records, disregarding their abundance value.

Besides the presence records, absences were also attained. These are usually surveyed locations where the specimens were not found. In our study, and also due to the different conditions of the surveyed places, only those sites where other mosquitoes species were present but *An. maculipennis* complex showed null or extremely reduced values of abundance (<2%) were considered. This was done to remove false absences caused by anti-mosquito actions or other facility-specific effects. A total of 16 absence records were attained.

### *Environmental factors*

The gathering of the environmental data with the highest predictive potential, supported by prior knowledge about the prevailing factors in the studied species distribution, is foreseen as one of the major steps to be taken in species distribution modelling (Araújo and Guisan, 2006). In this sense it is crucial to make an appropriate review of existing knowledge concerning the ecological factors considered influential in the distribution of a species as this allows a better and safer choice of the independent variables eligible for integrating the predictive statistical models. Such choice is, nevertheless, many times postponed by data unavailability and/or difficulties in its spatial representation.

After compiling both the acquaintance of *An. atroparvus* ecology and the available data and accounting for the possibility that “overfitting” of the predictive models may occur when a large number of independent variables with reduced samples are used, five environmental factors were taken into account:

- (i) mean maximum temperature of the warmest trimester;
- (ii) mean minimum temperature of the coldest trimester;
- (iii) mean total annual precipitation;
- (iv) wetlands density and suitability index; and
- (v) agricultural density and suitability index.

In this approach, the temperature spatial variability models are included due to temperature’s direct influence on species physiology and behaviour (Cambournac and Hill, 1938) or depending on an indirect effect due to access to larval sites and variability of the depth of the water bodies. Based on the climatological records for the five decades from 1950 to 2000, two datasets were included: the mean maximum temperature of the warmest trimester and the mean minimum temperature of the coldest trimester. The use of two datasets simultaneously to represent one unique ecological factor prevents spatial and temporal homogenization resulting from the application of a unique annual mean model. All

information concerning these variables was directly acquired in raster structure from the WORLDCLIM project (<http://www.worldclim.org>) (Hijmans et al., 2005).

The use of precipitation as a predictor was considered due to the flushing effects on the productivity of the mosquito breeding sites, as already verified for other *Anopheles* species (Paaijmans et al., 2007). Moreover, it is strongly related to the availability and characteristics of *An. atroparvus*’ habitats in the aquatic phase of its life cycle, i.e. the depth of water bodies and the availability of temporary ponds. This information was also obtained from the WORLDCLIM project for the 1950-2000 period.

In order to include information about the profusion and proximity of wetlands, which are crucial for *An. artoparvus* due to the direct relationship to the aquatic life cycle (Becker et al., 2003), we employed the Corine Land Cover 2000 (CLC2000) spatial data sets for Portugal (scale 1:100,000) which is produced by the European Environment Agency (EEA) (<http://www.eea.europa.eu/themes/landuse/clc-download>). Since this information does not cover the terrain completely, a spatially continuous model based on wetland densities was elaborated by calculating wetland density within a circle with 10 km radius. Different weights were also attributed to each wetland type considering its known suitability for *An. atroparvus*. The differentiated areas were: rice fields, inland marshes, large inland water bodies, estuaries, coastal lagoons, intertidal flats, permanently irrigated areas, watercourses, and saltpans. Three different levels of weighting were applied based on expert knowledge regarding the suitability of each of the areas to the species larval development. The final model was obtained through the product of the density values and the factors used for weighting.

The density model based on agricultural land use and land cover was taken as indicator of livestock production. For example, places such as cattle farms have a high capability for the development of large populations of *An. atroparvus* due to the zoophilic nature of this mosquito species (Sousa, 2008). This approach is based on the broad pre-

sumption that areas characterized by a high agricultural intensity have a greater probability of containing cattle herds as the main sign of economic activity or being used for cattle breeding for domestic/private consumption. This model included the following classes derived from the CLC2000 classification: pastures, non-irrigated arable land, permanently irrigated land, rice fields, vineyards, fruit trees and berry plantations, olive groves, annual crops associated with permanent crops, complex cultivation patterns, land principally occupied by agriculture, with significant areas of natural vegetation and agro-forestry areas. The final model regarding this factor was obtained following the same method used for the wetlands model.

#### *Predictive models calibration*

When implementing predictive models, the distribution records are generally split into calibration and validation data sets. However, considering the relatively small amount of distribution records ( $n = 92$ ) in this study, a random multiple partitioning of the records was used. This method avoids excessive use of records for validation purposes, being more suitable than a single validation set, when only a small amount of species records is available (Hirzel et al., 2006). In this sense three calibration-validation sets were made, each validation fraction composed of 13 records (2 absences and 11 presences). This number of records corresponds to a random extraction of about 15% for each record type, which permitted the species prevalence to remain unaffected in the calibration sets. In addition, all spatial predictor models were re-sampled into a 1 km<sup>2</sup> grid system.

#### *Predictive systems*

Three distinct predictive models were implemented: Mahalanobis distance ( $D^2$ ); binary logistic regression (BLR); and a multilayer perceptron with back propagation artificial neural network (MLP-

ANN). The use of several predictive models is suggested as a method for reducing the uncertainty on species distribution modelling (Pearson et al., 2006).

#### *Mahalanobis distance calibration*

Mahalanobis distance ( $D^2$ ) corresponds to an  $n$ -dimensional distance to the centroid of the provided values. This measure, when applied to the habitat suitability modelling, calculates the similarity between the predictors multidimensional mean transmitted by presence records and each of the cells composing the study area. In this sense Mahalanobis distance requires only presence data, thus removes the influence of possible bias and errors present in absence sets. However, in general, presence-only methods have lower predictive performances than presence-absence models (Austin and Meyers, 1996). Even so, in this group of models Mahalanobis distance presents some of the best results (Tsoar et al., 2007). Besides, different measurement scales and correlations amongst the predictors do not affect this similarity index, so it can be applied to non-normally distributed variables and its outputs are easily interpreted. This measure was implemented directly over the spatial data using the Mahalanobis distances ArcView extension (Jenness, 2003). Due to the existence of three calibration sets the final model was obtained through the arithmetic mean of their resulting outputs.

#### *BLR calibration*

Briefly, BLR is a well-known method in habitat suitability modelling, belonging to the family of generalized linear models, which is used to estimate the occurrence probability of a certain event, given the values of the predictors. BLR is particularly well suited for species distribution modelling, since it works with a binary response like presence-absence and does not assume normal distributions. The resulting values are expressed in a 0 to 1 interval and interpreted as the predicted distribution of the

species under study. For each of the calibrated models, the Hosmer-Lemeshow tests (Hosmer and Lemeshow, 1989) were performed, and all showed significant values ( $P < 0.05$ ), indicating that the models adequately fitted the calibration data.

#### Artificial neural network calibration

Artificial neural networks are adaptive statistical models based on an analogy with how the brain is thought to function (Abdi, 2003). Simple units, the nodes, which are connected by weighted links, constitute the structure. The model calibration (which can also be called training) takes place by applying different weights to the signal passing these connections. In our study, an MLP-ANN composed by three interconnected layers was used. The nodes in the input layer receive the predictors values, while the second layer (the hidden layer) runs the learning back-propagation process before passing the information on to the output layer (presence-absence data) for prediction. The measured error between the expected and the actual output provides the basis for the training process which consists of running numerous cycles until the desired error level is achieved.

Our model included five nodes in the hidden layer, resulting in as many predictors. This was done to reduce the risk of “overfitting” of the model which is favoured by the use of an excessive number of parameters in non-parametric models (Liu et al., 2007). The error measure used was the mean square error and the networks were trained through 400,000 cycles.

The heuristic nature of artificial neural networks implies that different modelling episodes produce different results. In order to deal with this variability a common approach is to calculate the mean of a set of outputs. In this way the final model of each calibration set was represented by the arithmetic mean of three different calibration outputs. The final model corresponded to the arithmetic mean of these means.

#### Calibration calculations

The BLR and the MLP-ANN results come in a range from 0 to 1, while the Mahalanobis results, being standardized values of similarity, can vary from 0 (maximum similarity) to any positive number of maximum dissimilarity. In this study, the final Mahalanobis model was normalized into a 0 - 1

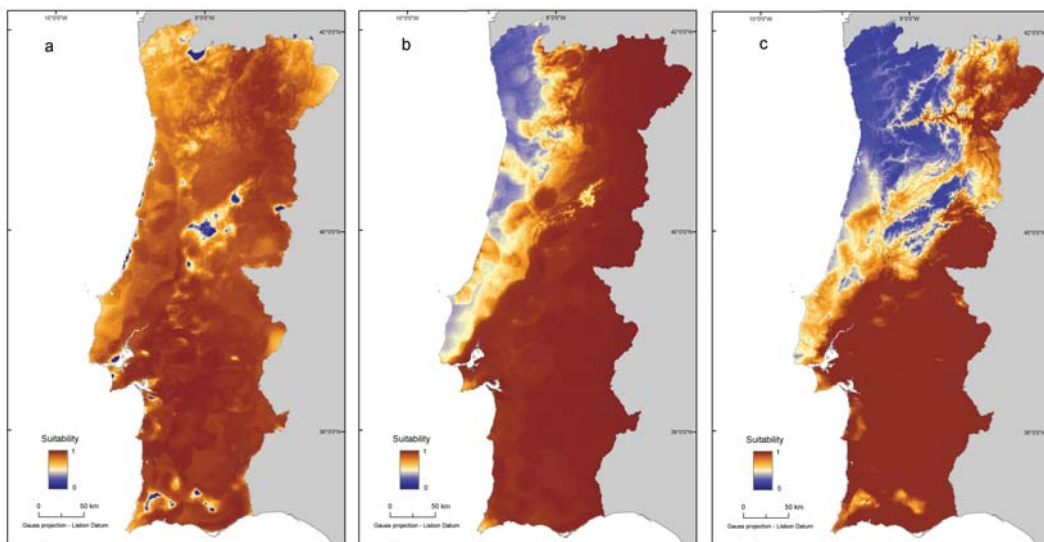


Fig. 1. Habitat suitability for *An. atroparvus* across Portugal obtained by the different correlative methods: a) Mahalanobis distance, b) logistic regression, and c) artificial neural network.

interval based on its mean and standard deviation values. These values were then inverted, since lower values of this measure correspond to higher similarity with the places where the species was recorded. The different habitat suitability models can be seen in Figure 1.

Before using the obtained models to effectively characterize the distribution of the suitability areas for *An. atroparvus*, it was necessary to evaluate their predictive performances. Since it is only possible to evaluate the appropriateness of the results for the intended purpose after this procedure has been carried out, it is considered a fundamental stage in species distribution modelling (Araújo and Guisan, 2006). In the context of the present work, the results evaluation also allowed us to differentiate the performances obtained by each of the modelling methods.

The kappa index (Cohen, 1960) was used as predictive evaluation measure. This index expresses the amount of agreement between the correct classifications achieved and the ones expected merely by chance. Its values generally vary from 0 (but can be negative), representing results identical to the ones expected by chance, and 1 corresponding to a perfect agreement between actual and obtained results. To proceed with the calculation of this index it is necessary to establish suitability threshold values. A commonly adopted approach is to use a range of thresholds and calculate the corresponding kappa value for each one of them (e.g. Elith et al., 2006) adopting the maximum value achieved as the predictive performance of the model. All three models were partitioned in 20 intervals of equal amplitude (0.05) and the kappa value calculated for each. The obtained results showed BLR to be the best performing model ( $\kappa = 0.77$ ). According to Landis and Koch (1977), this value corresponds to the existence of an excellent agreement between the model results and the validation records. The model obtained by MLP-ANN presented a fairly lower result ( $\kappa = 0.51$ ) which even so is considered by the same authors to be a good level of agreement. The lowest predictive performance value was registered with the Mahalanobis model with  $\kappa = 0.42$ , which is close to

the lower limit of the good-agreement range.

Considering that all three models achieved good to excellent agreement levels, it was decided to combine the output achieved, i.e. all the binary maps. This conjugation allowed us to reduce the uncertainty that would have resulted from the output of a single method (Pearson et al., 2006), which is similar what has been proposed by Araújo and New (2006) concerning uncertainty reduction for projecting species distributions in future climate scenarios. Figure 2 shows the resulting model amounting to the agreement between the three binary models.

In order to assess the spatial similarity between the former malaria areas and habitat suitability model, the kappa value was calculated. Taking into account the existence of four classes in the combined suitability model, different binary class combinations were created in order to be comparable with the binary map of malaria distribution. These outcomes are expressed in Table 1.

## Results

The obtained kappa values (Table 1) show different agreement levels between all of the binary suitability models and the former malaria distribution

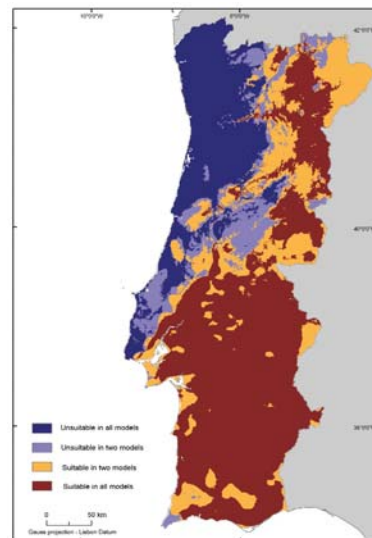


Fig. 2. Combination of binary habitat suitability models for *An. atroparvus* across Portugal.

Table 1. Kappa values of former malaria distribution and present habitat suitability models for *An. artoparvus* in Portugal.

Threshold	L1	L2	L3	L4	$D^2$	BLR	MLP-ANN
Kappa	0.32	0.43	0.45	0.62	0.36	0.38	0.42

L1 = unsuitable in all models *vs* remaining area; L2 = unsuitable for 2 models *vs* remaining area; L3 = suitable in all models *vs* remaining area; L4 = unsuitable in all models *vs* suitable in all models;  $D^2$  = Mahalanobis distance using best kappa threshold; BLR = binary logistic regression using mean best kappa threshold; MLP-ANN = multilayer perceptron artificial neural network using mean best kappa threshold.

in mainland Portugal. Amongst the single-model outputs, and in contrast to previous validation results, the MLP-ANN achieved the highest agreement ( $\kappa = 0.42$ ), followed by BLR ( $\kappa = 0.38$ ) and  $D^2$  ( $\kappa = 0.36$ ). In general, all of the outputs resulting from multi-model combination show higher predictive performance than single-model binary distributions. The exception is model L1, which presents the lowest value ( $\kappa = 0.32$ ). The agreement value is much higher when the disagreement areas between the models are removed (model L4,  $\kappa = 0.62$ ).

The resulting suitability maps for *An. atroparvus* reveal a clear spatial separation in the way the country fits its habitat preferences. The northwest and central coastal areas present lower suitability, while the south and interior central and northern areas comprise almost the entire distribution of higher suitability values.

## Discussion

The habitat suitability threshold where maximum kappa is achieved can be used to produce binary maps of suitable and unsuitable areas for the target species. Although information is lost in this conversion, a binary map based on the maximum kappa value allows a more precise spatial discrimination between areas of higher and lower suitability. This transformation is useful in the present case where the Mahalanobis model presents a clear concentration of values close to the maximum of suitability (Fig. 1a). This is also valid for both BLR and Mahalanobis models, where a direct interpretation of the continuous values can be biased since the cut values obtained by the maximum kappa (BLR = 0.65;  $0.7 - D^2 = 0.95$ ) are distinct from the expected theoretical value of 0.5. In this sense three bina-

ry models were built based on the suitability threshold where the maximum kappa was achieved or their mean value when more than one threshold occurred.

In the obtained combined model, a clear distribution pattern emerges. Unsuitability agreement covers approximately all the northwest of the country, while the south and the interior north are almost entirely occupied by the least uncertain suitable areas. This spatial pattern is fairly coincident with the distribution of precipitation and temperature variations along the country. Unsuitable areas match coarsely to the wettest places with mild temperatures in the summer, while dryer areas reaching higher temperatures are almost entirely suitable areas. Unsuitable areas are seen here not as places with complete absence of the species, but instead as least suitable places where lower abundances occur. This is in agreement with the distribution pattern achieved by several studies based on direct interpretation of field data. For example, Almeida et al. (2008), found that despite being present along the entire country *An. maculipennis* s.l. presented lower abundances in the districts of Braga, Leiria, Lisbon, Oporto and Viana do Castelo, which correspond closely with the unsuitable areas obtained, and higher abundances in the south and north-east of the country, also coinciding with the least uncertain suitable areas. Higher abundances of *An. atroparvus* have previously been recorded in Alentejo and Algarve (the two southernmost regions of the country) by Ramos et al. (1978) and Pires et al. (1982). However, this spatial pattern is somewhat divergent with the known distribution of the species outside the Iberian Peninsula. In fact, *An. atroparvus* is found mainly in northern countries with colder and wetter climate conditions such



as the Netherlands, United Kingdom, Germany or Poland, being absent from some of the warmer Mediterranean regions, such as south Italy, Greece and Turkey. Considering the higher suitability of this species in the southern areas of Portugal, its absence from many of the other Mediterranean areas can be due to competitive exclusion involving other species. In this sense, it is possible to infer that this species has reached, not only in Portugal but also in places along the Iberian Peninsula, some of the most highly suitable conditions for its existence. This would justify its former dominant role in malaria transmission in these areas.

The combined suitability model was also used for comparison with the known distribution of malaria during its last endemic period. The spatial distribution of the disease was documented by Cambournac (1942) who presented a map based on known values of spleen rates (i.e. the proportion of people of a given population with enlarged spleens) by municipality, totalling six distinct regions (Fig. 3). Although this author considers the existence of different spleen rates between the areas, he also refers that they were all related with the existence of high abundance levels for *An. atroparvus*.

With the exception of the L1 model, all the results

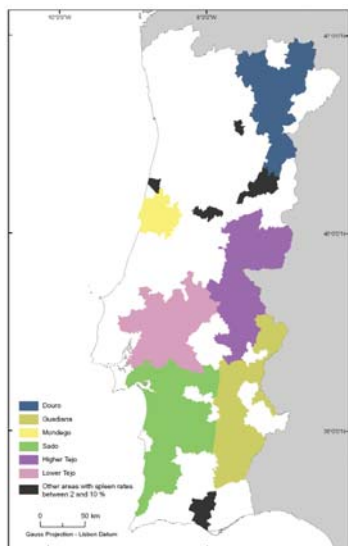


Fig. 3. Former malaria distribution in continental Portugal (Cambournac, 1942).

from the multi-model combination show higher predictive performance than single-model binary distributions. However, the agreement value is far higher when the disagreement areas between models are removed. The higher kappa value, achieved by the L4 binary model, points to two different assumptions, i.e. the combination of different suitability models can be a good possibility for achieving better predictive results and the *An. atroparvus* distribution pattern in mainland Portugal is still substantially similar to the one existing when the last endemic malaria episodes occurred. This suggests that despite the recent temperature increase in Portugal, the distribution of this species remains unaltered from the one registered nearly seven decades ago. This invariance is important when one considers that the country has registered as much as a mean temperature increase of  $0.5^{\circ}\text{C}/\text{decade}$  since 1975 (Miranda et al., 2006). Since *An. atroparvus* reaches some of the highest temperatures in its distribution range in the southernmost part of the Iberian Peninsula and survives there, it can be expected that northern populations will remain resilient to possible temperature increments in the next decades. It is also clear that vector habitat suitability models for vector species can act as one of the components of information in health risk analysis studies.

The option to apply more than one predictive model and their subsequent combination for achieving the interpreted habitat suitability model allowed a more precise spatial discrimination of its distribution. This achieved distribution presented higher agreement with the former malaria areas than individual models, which suggests higher appropriateness of this technique to achieve species habitat suitability models.

## Conclusions

The current *An. atroparvus* distribution is fairly similar to former malaria distribution in Portugal, suggesting that habitat suitability models of vectors can be a good surrogate in the spatial assessment of

malaria risk. Opposite to what usually happens in other malaria-endemic areas, the incidence of the disease in the country was dependent on vector density. Thus, to spatially model the current abundance of the vector is, indirectly, to model malaria transmission rates and therefore also the receptivity (and risk) for malaria re-emergence. This similarity suggests that *An. atroparvus* still presents the same preferential distribution as nearly seven decades ago and that the species is still not affected by the increasing temperatures recorded in the country for the last decades. This resilience of the species in one of the warmer and dryer area of its distribution indicates that northern populations may indeed not present, at least not in the short-term, large distributional variations due to a direct influence of potential temperature increases.

### Acknowledgements

This publication was partially funded by: (i) EU grant GOCE-2003-010284 EDEN and is catalogued by the EDEN Steering Committee as EDEN0144 ([www.eden-fp6project.net](http://www.eden-fp6project.net)), being the contents of this publication of the sole responsibility of the authors and do not necessarily reflect the views of the European Commission, and; (ii) "Arbovirus dos mosquitos de Portugal" (POCTI/35775/ESP/2000).

### References

- Abdi H, 2003. Neural networks. In: Encyclopedia for research methods for the social sciences. Sage, Thousand Oaks, pp. 725-728.
- Almeida A, Galão R, Sousa C, Novo M, Parreira R, Pinto J, Piedade J, Esteves A, 2008. Potential mosquito vectors of arboviruses in Portugal: species, distribution, abundance and West Nile infection. *Trans R Soc Trop Med Hyg* 102, 823-832.
- Antunes F, Forte M, Tavares L, Botas J, Carvalho C, Carmona H, Araújo F, 1987. Malaria in Portugal 1977-1986. *Trans R Soc Trop Med Hyg* 81, 561-562.
- Araújo M, Guisan A, 2006. Five (or so) challenges for species distribution modelling. *J Biogeogr* 33, 1677-1688.
- Araújo M, New M, 2006. Ensemble forecasting of species distributions. *Trends Ecol Evol* 22, 42-47.
- Austin M, Meyers J, 1996. Current approaches to modeling the environmental niche of *Eucalyptus*: implications for management of forest biodiversity. *Forest Ecol Manag* 85, 95-106.
- Becker N, Petric D, Zgomba M, Boase C, Dahl C, Lane J, Kaiser A, 2003. Mosquitoes and their control. Kluwer Academic/Plenum Publishers, 518 pp.
- Borges F, 2001. A malária no vale do rio Sado. In: Memórias do Instituto de Malariologia de Águas de Moura. Da luta anti-palúdica ao museu, Câmara Municipal de Palmela/Centro de estudos de vectores e doenças infecciosas, Palmela, pp. 59-64.
- Bruce-Chwatt LJ, Zulueta J, 1977. Malaria eradication in Portugal. *Trans R Soc Trop Med Hyg* 71, 232-240.
- Cambournac F, 1942. Sobre a epidemiologia do sezonismo em Portugal, Lisboa: Sociedade Industrial de Tipografia, 235 pp.
- Cambournac F, Hill R, 1938. The biology of *Anopheles atroparvus* in Portugal. Transactions of the Third International Congress of Tropical Medicine and Malaria 2, 178-184.
- Castro L, Cardoso A, Queirós L, Gonçalves G, 2004. Malária na Região Norte de Portugal (1993-2002). Caracterização epidemiológica. *Acta Med Port* 17, 291-298.
- Cohen J, 1960. A coefficient of agreement for nominal Scales. *Educ Psychol Meas* 20, 37-46.
- Elith J, Graham C, Anderson R, Dudik M, Ferrier S, Guisan A, Hijmans R, Huettmann F, Leathwick J, Lehmann A, Li J, Lohmann L, Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J, Peterson T, Phillips S, Richardson K, Scachetti-Pereira R, Schapire R, Soberon J, Williams S, Wisz M, Zimmermann N, 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29, 129-151.
- Guisan A, Thuiller W, 2005. Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 10, 993-1009.
- Guisan A, Zimmermann N, 2000. Predictive habitat distribution models in ecology. *Ecol Model* 135, 147-186.
- Hijmans R, Cameron S, Parra J, Jones P, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25, 1965-1978.

- Hirzel A, Lay G, Helfer V, Randin C, Guisan A, 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecol Model* 199, 142-152.
- Hosmer DW, Lemeshow S, 1989. *Applied logistic regression*, John Wiley & Sons Inc, New York, USA 397 pp.
- IPCC, 2007. *Climate change 2007: the physical science basis. Summary for policy makers working group I to the Intergovernmental Panel on Climate Change. Fourth Assessment Report (AR4)*.
- Jenness J, 2003. Mahalanobis distances (mahalanobis.avx) extension for ArcView 3.x. Jenness Enterprises. Available: <http://www.jennessent.com/arcview/mahalanobis.htm>.
- Kuhn K, Campbell-Lendrum D, Davies C, 2002. A continental risk map for malaria mosquito (Diptera: Culicidae) vectors in Europe. *J Med Entomol* 39, 621-630.
- Landis J, Koch G, 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.
- Liu Y, Starzyk JA, Zhu Z, 2007. Optimizing number of hidden neurons in neural networks. In: *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*. IASTED/ACTA Press, Innsbruck, Austria, pp. 138-143.
- Minář J, Halgoš J, Bartalová A, Jalili N, 2007. Current climatic change and its impact on mosquito fauna in conditions of Slovakia and Czech Republic. *Acta Zool Univ Comen* 47, 177-182.
- Miranda P, Valente A, Tomé AR, Trigo R, Coelho F, Aguiar A, Azevedo F, 2006. O clima em Portugal nos séculos XX e XXI. In: *Alterações climáticas em Portugal. Cenários, Impactos e Medidas de Adaptação, Projecto SIAM II*, Gradiva, Lisboa, Portugal, pp. 45-113.
- Moffett A, Shackelford N, Sarkar S, 2007. Malaria in Africa: vector species' niche models and relative risk maps. *PLOS One* 2, e824.
- Muturi E, Burgess P, Novak R, 2008. Malaria vector management: where have we come from and where are we headed? *Am J Trop Med Hyg* 78, 536-537.
- Paaijmans K, Wandago M, Githeko A, Takken W, 2007. Unexpected high losses of *Anopheles gambiae* larvae due to rainfall. *Plos One* 2, e1146.
- Pascual M, Ahumada J, Chaves L, Rodo X, Bouma M, 2006. Malaria resurgence in the East African highlands: temperature trends revisited. *P Natl Acad Sci USA* 103, 5829-5834.
- Pearson R, Thuiller W, Araújo M, Martinez-Meyer E, Brotons L, McClean C, Miles L, Segurado P, Dawson T, Lees D, 2006. Model-based uncertainty in species range prediction. *J Biogeogr* 33, 1704-1711
- Pires C, Ribeiro H, Capela R, Ramos H, 1982. Research on the mosquitoes of Portugal (Diptera, Culicidae) VI - The mosquitoes of Alentejo. *An Inst Hig Med Trop* 8, 79-101.
- Ramos H, Ribeiro H, Pires C, Capela R, 1978. Research on the mosquitoes of Portugal (Diptera, Culicidae) II - The mosquitoes of Algarve. *An Inst Hig Med Trop* 1, 79-101.
- Ribeiro H, Alves-Pires C, Ramos H, Capela R, 2002. Research on the mosquitoes of Portugal (Diptera, Culicidae) XII - The mosquitoes of Minho and Douro Litoral. *Garcia de Horta - Série Zoológica* 24, 31-50.
- Ribeiro H, Pires C, Capela R, 1988. An annotated checklist of the mosquitoes of continental Portugal (Diptera, Culicidae). *Actas do III Congresso Ibérico de Entomologia*, 233-253.
- Ribeiro H, Ramos H, Capela R, Pires C, 1992. Research on the mosquitoes of Portugal (Diptera, Culicidae) II - The mosquitoes of the Beiras. *Garcia de Horta - Série Zoológica* 16, 137-161.
- Ribeiro H, Ramos H, Pires C, 1999a. Os mosquitos do Parque Natural da Serra da Estrela (Insecta, Diptera, Culicidae). *Garcia de Horta - Série Zoológica* 23, 119-168.
- Ribeiro H, Ramos H, Pires C, 1999b. Os mosquitos do Parque Natural de Montesinho (Insecta, Diptera, Culicidae). *Garcia de Horta - Série Zoológica* 23, 23-67.
- Sousa CA, 2008. Malaria vectorial capacity and competence of *Anopheles atroparvus* Van Thiel, 1927 (Diptera: Culicidae): implications for the potential re-emergence of malaria in Portugal. PhD Thesis, New University of Lisbon, Portugal, 187 pp.
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R, 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Divers Distrib* 13, 397-405.
- WHO 2008. *World Malaria Report 2008*. World Health Organization, Geneva, Switzerland, 215 pp.