Distribution of the Habitat Suitability of the Main Malaria Vector in French Guiana Using Maximum Entropy Modeling

Yi Moua, Emmanuel Roux, Romain Girod, Isabelle Dusfour, Benoit de Thoisy, F. Seyler, Sébastien Briolant

Journal of Medical Entomology

Distribution of the habitat suitability of the main malaria vector in French Guiana using Maximum Entropy modeling

Yi Moua[1], Emmanuel Roux[2], Romain Girod[3], Isabelle Dusfour[3], Benoit de Thoisy[4], Frédérique Seyler[2], Sébastien Briolant[3, 5, 6, 7]

[1] Université de Guyane, ESPACE-DEV, UMR 228 (IRD, UM, UR, UA, UG), Cayenne, French Guiana

[2] Institut de Recherche pour le Développement, ESPACE-DEV, UMR 228 (IRD, UM, UR, UA, UG), Montpellier, France

[3] Unité d'Entomologie Médicale, Institut Pasteur de la Guyane, Cayenne, French Guiana

[4] Laboratoire des Interactions Virus Hôtes, Institut Pasteur de la Guyane, Cayenne, French Guiana

[5] Direction Interarmées du Service de Santé en Guyane, Cayenne, French Guiana

[6] Institut de Recherche Biomédicales des Armées, Unité de Parasitologie et d'Entomologie Médicale, Marseille, France

[7] Unité de Recherche en Maladies Infectieuses Tropicales Emergentes, UMR 63, CNRS 7278, IRD 198, INSERM 1095, Faculté de Médecine La Timone, Marseille, France

Abstract

Malaria is an important health issue in French Guiana. Its principal mosquito vector in this region is *Anopheles darlingi*. Knowledge of the spatial distribution of this species is still very incomplete due to the extent of French Guiana and the difficulty to access most of the territory.

Species Distribution Modeling based on the maximal entropy procedure was used to predict the spatial distribution of *An. darlingi* using 39 presence sites.

The resulting model provided significantly high prediction performances (mean 10-fold cross-validated partial AUC and continuous Boyce index equal to, respectively, 1.11 – with a level of omission error of 20 % – and 0.42). The model also provided a habitat suitability map and environmental response curves in accordance with the known entomological situation.

Several environmental characteristics that had a positive correlation with the presence of *An. darlingi* were highlighted: non-permanent anthropogenic changes of the natural environment; the presence of roads and tracks; opening of the forest. Some geomorphological landforms and high altitude landscapes appear to be unsuitable for *An. darlingi*.

The Species Distribution Modeling was able to reliably predict the distribution of suitable habitats for *An. darlingi* in French Guiana. Results allowed completion of the knowledge of the spatial distribution of the principal malaria vector in this Amazonian region, and identification of the main factors that favor its presence. They should contribute to the definition of a necessary targeted vector control strategy in a malaria pre-elimination stage, and allow extrapolation of the acquired knowledge to other Amazonian or malaria-endemic contexts.

Keywords: Maxent, species distribution model, presence-only, *Anopheles darlingi*, sampling bias

Malaria is a public health issue in the Amazonian region, with major transmission foci depending on specific local characteristics associated with changing environmental and socio-demographic contexts. French Guiana is a French overseas territory with ~260,000 inhabitants. It remains one of the major malaria foci in the region, despite an improving epidemiological situation during the past ten years. The number of reported clinical cases has significantly dropped from 4,479 in 2005 to 434 in 2015 (Petit-Sinturel et al. 2016), and now corresponds to an incidence rate of two cases for 1,000 inhabitants for the whole territory, making it possible to target the pre-elimination of the disease in 2018 (Agence Régionale de Santé Guyane 2015). *Plasmodium vivax* is at present predominant and this species was responsible for 67% of the diagnosed cases of malaria in the territory in 2014, the others being mainly due to *Plasmodium falciparum* (Musset et al. 2014, Ardillon et al. 2015). However, this epidemiological situation is heterogeneous in space and time. In particular, a recrudescence of malaria cases is currently observed in the inland region (Saül, Cacao, and Régina) and eastern French Guiana (municipalities of Camopi and Saint-Georges-de-l'Oyapock), with a general incidence rate reaching 55.2 cases per 1,000 inhabitants in 2013 (Musset et al. 2014), likely due to the emergence and/or persistence of local foci of high malaria transmission (Berger et al. 2012, Musset et al. 2014). This Amazonian region, especially near the international borders, includes vulnerable populations. Some are hard-to-reach and have poor access to health services and treatment-seeking behaviors that may favor the development of resistance to antimalarial drugs (Musset et al. 2014, Wangdi et al. 2015). Uncontrolled areas of malaria transmission are also prevalent in illegal gold mining areas (Pommier de Santi et al. 2016a, Pommier de Santi et al. 2016b). The epidemiological situation remains quite unstable, and pre-elimination of malaria, corresponding to an incidence rate below one case for 1,000 inhabitants in any locality of French Guiana, remains a major challenge.

In this context, public health authorities must maintain control efforts while targeting them more precisely and objectively in time and space (Alimi et al. 2015). A map of malaria risk in French

Guiana is updated regularly by the regional unit of the French National Public Health Agency, based on the number of cases reported per locality and the data available on movements of human populations at risk, especially due to gold mining activities. This map is validated by the local expert committee of epidemic diseases (Comité d'Experts des Maladies à Caractère Épidémique, CEMCE), which brings together different experts of the disease in the region (from the Health Surveillance Agency, the Pasteur Institute of French Guiana, the Regional Unit of the National Public Health Agency, vector control services, hospitals, and other diagnosis and care centers, and the Defense Health Service in French Guiana). The lack of objective knowledge of several key factors, especially the spatiotemporal distribution of the main malaria vectors and human populations infected by *Plasmodium* and/or carrying gametocytes, makes such a map highly approximate.

Anopheles (Nyssorhynchus) darlingi Root (Diptera: Culicidae) is one of the most efficient malaria vectors in South America and is considered to be the primary malaria vector in French Guiana because of its anthropophilic behavior, natural infectability, high density, and sensitivity to *P. falciparum* (Girod et al. 2008, Hiwat et al. 2010, Fouque et al. 2010). Used entomological data collection for the entire territory, for the mapping of entomological risk indicators at the regional scale, is not feasible. French Guiana occupies a large territory (84,000 km²) which is mostly covered by rain forest (more than 80%) and highly inaccessible. Knowledge of the recent geographical distribution of *An. darlingi* is thus restricted to coastal areas, some villages along the international border rivers, and some illegal gold mining sites (Figure 1).

Species Distribution Modeling (SDM) offers an efficient solution to geographically extrapolate such knowledge to the entire territory (Pearson et al. 2007). Species Distribution Modeling produces maps of species habitat suitability by using known presence locations of the species and relevant environmental data. The use of SDM is thus encouraged to “improve and facilitate the development of alternative vector control strategies” (Alimi et al. 2015). Numerous SDM approaches are

proposed in the literature. Some of them, such as Maximum Entropy (Maxent; Phillips et al. 2006), Genetic Algorithm for Rule-Set Prediction (GARP; Stockwell 1999), Boosted Regression Trees (BRT; Friedman et al. 2000), Generalized linear and additive models (GLM and GAM; Guisan, et al. 2002), and Multivariate adaptive regression splines (MARS; Leathwick et al. 2005), exploit only species presence information, offering a significant advantage over methods that also require absence data. Indeed, absence data are often difficult to obtain. According to Peterson (2007) and Hirzel et al. (2002), absence can result from (1) the non-detection of the species in a suitable habitat, even if it is present, (2) the actual absence of the species for historical reasons, whereas the habitat is suitable, and (3) the true absence of the species and the unsuitability of the habitat. Comparative studies (Elith et al. 2006, Tognelli et al. 2009, Pearson et al. 2007, Hernandez et al. 2006, Wisz et al. 2008) show that Maxent is able to fit complex functions between habitat suitability and predictor variables, is the least sensitive to the size of the presence dataset, and tends to outperform other comparable methods when the dataset is small.

In this study, the mapping of the habitat suitability of *An. darlingi* at the scale of all of French Guiana was performed using the Maxent SDM approach. This work aims to provide reliable maps for improving malaria transmission risk mapping in French Guiana, and to identify the environmental factors and associated mechanisms that favor the presence of *An. darlingi*.

Materials and Methods

Study area

French Guiana (84,000 km²), a French overseas region located in South America, is separated from Suriname by the Maroni River and from Brazil by the Oyapock River and the Tumuc-Humac mountains. More than 80% of the territory is covered by rain forest. The country has an equatorial climate characterized by two annual dry seasons, from mid-August to mid-November and in March, and two wet seasons, from mid-April to mid-August and mid-November to February. The average annual rainfall reaches 4,000 mm and 2,000 mm in the wettest (north-east) and driest (north-west)

areas, respectively (Hammond 2005). The average monthly rainfall is >100 mm for the entire territory throughout the year, except for the three driest months: September, October, and November (Héritier, 2011). The average humidity is between 80% and 90%. The temperature is homogeneous over the entire territory throughout the year, with an average annual temperature of 26°C. The difference between the minimum and maximum daily temperature is more important than the annual variations. For example, in Maripasoula (on the border with Surinam) and Camopi (on the border with Brazil), the annual ranges of the minimum and maximum temperatures were, 4.3°C and 9.6°C (averages over the period 2001-2008), respectively, whereas the mean daily thermal amplitude was 9.8°C (average over the period 2001-2008; Météo-France, 2016). The population of ~260,000 inhabitants is unequally distributed throughout the territory. Approximately 90% of the population lives in the coastal area and most of the rest lives along the Maroni and the Oyapock rivers (Amerindians and Bush-Negroes). However many people live and/or transit through inland and remote areas of the territory (forestry workers, gold miners, and soldiers). According to many studies (Berger et al. 2012, Verret et al. 2006, Queyriaux et al. 2011, Hustache et al. 2007, Stefani et al. 2011, Pommier de Santi et al. 2016a), Amerindians, gold miners, and soldiers may be highly infected by malaria, whereas the areas in which they live and/or transit are those with the poorest knowledge of the presence and density of malaria vectors. It is thus of potential interest to consider the malaria risk, study the distribution of malaria vectors, and implement prevention and control actions over the entire territory of French Guiana.

Species Records

Presence sites of *An. darlingi* were provided by surveys of the Medical Entomology Unit of the Pasteur Institute of French Guiana and the Defense Health Service in French Guiana. *Culicidae* collections were performed using either human landing catches or traps (light traps or odor baited traps). Human landing catches consisted of exposing collector's lower leg and collecting landing mosquito with a mouth aspirator. Collectors were members of the Pasteur Institute or Defense

Health Services, they were aware of the risks associated with the method and had given their free consent. Malaria prophylaxis was proposed and information on the medication was explained. Light trap catches were performed with Center for Disease Control and Prevention (CDC) light traps, and odor baited catches were performed with Mosquito Magnet ® traps (Woodstream Corporation, Lititz, PA) baited with Octenol, a combination considered to be the best candidate for *Anopheles* surveillance in the region (Vezenegho et al., 2015).

Anopheles species were morphologically identified using taxonomic keys specific for the region (Floch and Abonnenc 1951, Faran and Linthicum 1981, Forattini 1962). Only *Culicidae* collections performed since the year 2000 were precisely geolocated by GPS coordinates and were used for the study (Figure 1). These data correspond to 74 capture sites for the family C*ulicidae*, and to 48 presence sites for the species *An. darlingi*.

The difficulty in accessing most of the French Guiana territory, and the priority given to the areas at risk of malaria transmission where many people live, led to a significant sampling bias with oversampling of the anthropized region of the territory, notably those easily accessible by roads (Figure 1).

Ecological knowledge and hypotheses

The presence of *An. darlingi* is linked to compositional and configurational features of the land cover and land use, as they partially determine breeding, feeding, and resting sites of the vector (Stefani et al. 2013). The natural environment for this vector in the Amazonian region includes floodable savanna, swamps (Girod et al. 2011, Zeilhofer et al. 2007), and flooded forest (Rozendaal 1992). Larvae are found along river edges, on flooded riverbanks, creeks, and pools formed near river-beds (Rozendaal 1992, Hiwat et al. 2010). Breeding sites are generally situated at low altitude (Mouchet 2004) and solely in freshwater, as *An. darlingi* is sensitive to salinity (Deane et al. 1948). Hydrological and geomorphological factors are responsible for the formation and destruction of *Anopheles* breeding sites (Smith et al. 2013).

Human activities, comprising deforestation and fish farming, also contribute to the creation of active breeding sites (Patz et al. 2000, Richard 1987, Stefani et al. 2013, Takken et al. 2005, Terrazas et al. 2015, Vittor et al. 2006, Vittor et al. 2009). Unpaved roads, tracks, and culverts form ideal breeding sites for *An. darlingi* in the Amazon region (Singer and Castro 2001). The presence of *An. darlingi* is also maintained by regular human presence due to its strong anthropophilic behavior. However, the presence and density of *An. darlingi* can either be favored or restricted depending on the type and intensity of the anthropogenic impacts. Stefani et al. (2013) systematically reviewed the literature and showed that all the studies describe the same mechanisms linking deforestation, land use, and the degree of urbanization with malaria transmission risk in the Amazonian region: opening the forest and maintaining a high degree of interaction between forested and deforested areas decreases the distance between feeding, breeding, and resting sites of *An. darlingi,* favoring the presence and high density of the vector (as well as a high probability of contact between humans and vectors); in contrast, intensifying deforestation and creating large urbanized and/or cultivated surfaces tends to decrease suitable habitat for *An. darlingi.* These two antagonistic consequences of human activities were considered in the SDM described here, by explicitly separating favorable and unfavorable factors in the environmental characterization.

The optimum temperature range for *An. darlingi* is between 20 and 30°C with a humidity of above 60% (Martens et al. 1995). Several studies established a minimal monthly rainfall threshold to designate suitable breeding habitats for *Anopheles* (reviewed in Smith et al. 2013). These values vary between 10 and 80 mm and need to be maintained for three or four months.

Environmental Variables

Environmental variables chosen as SDM inputs must characterize the ecological factors that influence the presence of *An. darlingi*, previously described. These factors are separated into three types: 1) natural environment features, associated with land cover, land use, and geomorphology for which the impact on the presence of *An. darlingi* depends on specific values or categorical classes;

2) anthropogenic activities that non-permanently alter the natural environment on a highly local scale and favor the presence of *An. darlingi*; 3) urbanization, corresponding to human presence and activities that permanently alter the natural environment over large areas and hinders the presence of the vector. Meteorological variables were not included in the model, because the temperature, rainfall, and humidity fall within the optimal ranges for presence of the species in French Guiana. Thus, these variables cannot significantly explain differences in the time average habitat suitability distribution over the year (this point is extensively discussed in the Discussion section).

Raw Geographic Data. Variables chosen as SDM inputs were derived from the following raw geographic data:

- Geomorphological landscape (*GLS*) and Geomorphological landforms (*GLF*) from the French Forest Office (ONF) (Guitet et al. 2013);
- Landscape types (*LS*) from the French Agricultural Research Centre for International Development (CIRAD) (Gond et al. 2011). This provides the distribution of landscape types in French Guiana, most being forested landscapes;
- Altitude (*ATL*) derived from the Digital Elevation Model provided by the Shuttle Radar Topography Mission (SRTM, spatial resolution: 30 meters) of the United States Aeronautics and Space Administration (NASA);
- Human footprint (*HFP*): An integrated human activity index that gives a general measure of the extent of expected threats on biodiversity, by assigning a score depending on the nature of the disturbance. It combines sublayers spatializing human population density, urban areas, legal and illegal mining sites, agriculture, forest settlements and camps, tourist camps, logged areas (forest activities), and potential hunting areas corresponding to a zone of two kilometers around roads, tracks and rivers, likely to be used by humans. The total disturbance score is the sum of all human activity scores (de Thoisy et al. 2010);
- Roads and tracks from the BD TOPO® database of the French Institute of Geographical

and Forestry Information (IGN).

Table 1 summarizes the main features of these raw geographic data.

Definition of Environmental Variables Used as Inputs for SDM. Several variables were extracted from the previously described raw data to better reflect the ecological knowledge and hypotheses mentioned above. The reference spatial resolution (pixel size) permitting the integration of all environmental layers was set to 1 by 1 km, *i.e.*, the coarsest resolution of the available layers, associated with the *LS* map.

The length of roads and tracks outside of urban areas (*ROADS*) was computed in the 1 km-cell grid from the BD TOPO® database.

The sublayers composing the *HFP* were first rasterized into 30-m grid cells, the smaller polygon of the *HFP* having a size of approximately 40 by 10 m. Distinct attributes were then extracted:

- The percentage of urbanization (*PER_URB*) within the 1 km grid cells;
- The percentage of urbanization within the eight neighbor cells of each urban cell (*PER_URB_NEIGH*), which permits distinguishing small from large urban areas. This layer was obtained for each 1 km-cell considered to be urban (*i.e.*, with *PER_URB* $\geq$ 50%), by averaging the *PER_URB* values for the eight (1 km side) neighbor pixels;
- The human activities which non-permanently alter the natural environment (*HA*), by first summing the scores of the following sublayers: tourist and forest camps, mining activities and logged areas, hunting areas nearby rivers, and then, by computing the minimum, median, and maximum values within the 1-km grid cells.

The agriculture sublayer from *HFP* was not used because it covers only the coastal area. The population density sublayer was also excluded because it did not have sufficient level of detail. The sublayer of potential hunting areas near roads and tracks were not used to avoid duplication of the length of roads and tracks outside of urban areas computed previously.

For each 1-km grid cell, the majority class of the categorical variables *GLS* and *GLF*, and the

minimum, median, and maximum altitude (*ALT*) values were computed.

Eventually, some corrections of the *LS* layer were performed as it did not identify urban areas and did not distinguish flooded forests associated with freshwater from those of the coastal strip associated with brackish water (mangroves): *LS* cells with an *PER_URB* value greater than or equal to 50% were reclassified into a new *LS* class referred to as *Urban*; *LS* cells classified as *Flooded forest* and corresponding to mangroves according to the coastal land use map provided by the ONF (Office National des Forêts Direction Régionale de Guyane, 2013) were reclassified as *Mangrove*. The variable *PER_URB* was excluded from the input SDM variables, as the urban areas were mapped, and their extent quantified, by the corrected *LS* and *PER_URB_NEIGH* layers, respectively.

Table 1 lists and describes the environmental variables used to build the model.

Maxent Model Principle

Maxent is an SDM which requires environmental variables and species presence-only data. It is based on the principle of maximum entropy to estimate an (*a priori*) unknown probability distribution over the entire study area. This probability distribution assigns a value that is proportional to the probability of the presence of the species to each pixel of the study area. It is therefore interpreted as a habitat suitability index (HSI) across the study area (Phillips et al. 2006). The Maximum Entropy principle consists of approximating the unknown probability distribution by finding the one that maximizes entropy and satisfies the constraints imposed by the environmental features at the known sites of presence. Environmental features are a set of input environmental variables chosen according to their expected relevance for the studied taxon (Phillips et al. 2006, Elith et al. 2011). The constraints ensure that the environmental values expected under the approximated probability distribution are consistent with environmental information observed at the presence points.

In practice, the Maxent distribution is defined on a set of points called background points. These

points should reflect the available environmental conditions of the study area and are chosen by uniform random sampling. This approach assumes that the presence data are not biased and that environmental conditions are uniformly sampled (Yackulic et al. 2013). However, in practice, some areas are more intensively sampled than others, and environmental conditions are not uniformly distributed and may imply a strong sampling bias. Phillips et al. (2009) proposed selecting the background points with the same environmental bias as the presence dataset to correct the effect of this sampling bias.

Model Building and Evaluation

Eleven environmental variables and 48 *An. darlingi* presence points (their coordinates were in the table in supplementary material S1) are used as inputs for Maxent. Only one presence site was selected to build the model when more than one occurred in the same pixel. As a result, only 39 presence sites were actually used for building the model. Hinge and categorical features were selected for the environmental variables. A hinge feature provides a good compromise between simplicity and the quality of the approximation of the species response curves (Elith et al. 2011, Phillips and Dudík 2008).

In this study, the distribution of the background points was biased so that the selection bias corresponds to that of the sampling. The sampling bias was defined as the relative sampling effort in the environmental space, and was estimated by considering the capture locations of *Culicidae*, obtained using the same capture techniques and supposed to be subjected to the same sampling bias as the *An. darlingi* species. The details of the method to create the relative sampling effort map are described in supplementary material S2.

The model was computed using version 3.3.3k of Maxent. The recommended values derived from Phillips and Dudík (2008) concerning the regularization parameters and the background set size, were applied. Regularization parameter values were set to 0.25 and 0.5 for categorical and hinge features, respectively, and the size of the background was set to 10,000. The extrapolation option

was not selected to avoid making predictions in environmental domains in which the model was not trained. The model was fitted using the full data set and evaluated using a 10-fold cross-validation procedure. The Receiver Operating Characteristic (ROC) curves and the associated Areas Under the ROC Curve (AUC) were computed. This was completed by computing the mean partial AUC ratios (Peterson et al., 2008), consisting of the ratios of the partial AUCs of the model over the null AUC (corresponding to random prediction), for omission errors (*E*) of 20, 10, and 5%. The Continuous Boyce Index (CBI), considered to better adapted to presence-only models than the AUC (Hirzel et al., 2006), was also computed. The gain (regularized training gain) was also used to evaluate the performance of the model prediction. It is a measure of the likelihood of the sample, and indicates how much better the estimated distribution fits the presence points than the uniform distribution, which corresponds to a null gain (Yost et al. 2008).

The importance of each variable was estimated using two methods, a heuristic method and the jackknife test. The heuristic method computes the percentage contribution of each variable to the model. During the training process, the increase of the gain is due to the adjustment of the feature weights and this increase is assigned to the environmental variable that the feature depends on. The sum of these increases in gain indicates the percentage contribution of each environmental variable. The jackknife test evaluates the individual contribution of each variable to the model by estimating the difference of the gain when removing each variable, one by one, and when considering the given variable alone to build the model.

Results

The mean AUC was 0.93, and the mean partial AUC ratios were 1.08, 1.03, and 1.01 for maximum omission errors sets to 20, 10, and 5% respectively. The mean CBI was 0.356 and the mean gain was 3.14. Three variables cumulatively contributed >80% (Table 2): the length of roads and tracks outside of urban areas (*ROADS*), the percentage of urbanization of neighboring pixels (*PER_URB_NEIGH*), and landscape (*LS*). The maximum value of the human activities which non-

297 permanently alter the natural environment (*HA_MAX*), geomorphological landscape (*GLS*),
298 minimum altitude (*ALT_MIN*), and geomorphological landform (*GLF)* contributed moderately to
299 the model, with contributions of 6.84%, 5.35%, 1.34%, and 1.19%, respectively. The following
300 input variables contributed very little to the model: minimum and median values of human activities
301 which non-permanently alter the natural environment (*HA_MIN* and *HA_MED*; 0.35 and 0.24%,
302 respectively); and median and maximum values of altitude (*ALT_MED* and *ATL_MAX*; 0.69 and
303 0.06%, respectively).
304 The results of the Jackknife test confirmed the non-significant contribution of the input variables
305 *HA_MIN*, *HA_MED*, *ALT_MED*, and *ALT_MAX* (Table 2).
306 A second model was built using only the most highly contributing environmental variables:
307 *ROADS*, *LS*, *PER_URB_NEIGH*, *HA_MAX*, *GLS*, *GLF*, and *ALT_MIN*. The overall performance of
308 this simpler model was very similar to the previous one, with the mean AUC and partial AUC ratios
309 equal to 0.93 and 1.11, 1.05, and 1.03, respectively. The mean gain was equal to 3.19 and the mean
310 CBI was 0.421. Relative contributions of the input variables were also very similar (Table 3).
311 The response curves of the environmental variables are represented in Figures 2 and 3. They show
312 that the HSI is maximal when the *PER_URB_NEIGH* is below 8%. Above this value, the HSI
313 decreases progressively towards 0. The HSI increases as *ROADS* increases up to 7,000 meters,
314 reaches a plateau value, and then tends to decrease above 10,000 meters. Among all *LS* classes,
315 *Woodland savanna/dry forest* and *Open forest* contribute the most to the high HSI values. The
316 geomorphological landscape classes *Coastal flat plain* and *Plain with residual relief* and the
317 geomorphological landform classes *Small-size and flat wet land*, *Small-size rounded hill,* and
318 *Lowered half-orange relief* – a tropical relief type corresponding to a hill with convex flanks giving
319 to it a roughly hemispherical shape (George, 1972) and usually linked to flat or swampy lowlands
320 drained by streams with meanders – are also associated with high HSI values. The HSI is maximal
321 when *ALT* is ~0, with a rapid decrease as altitudes increase. The *HA_MAX* response curve presents a

more complicated profile. The HSI increases for *HA_MAX* values between 0 and 8, decreases until *HA_MAX* reaches 24, and then again increases as values continue to climb above 24.

The map of habitat suitability for *An. darlingi*, based on all the presence data for modeling, shows six main areas (A – F) with a high HSI and a seventh area (G) corresponding to an epidemiological interest area (see Figure 4). A qualitative analysis was performed to determine the characteristics of the environmental variables of the areas with high HSI values (Table 4).

In the coastal area (A), where 90% of the Guyanese population lives, the HSI tends to be higher along the main road representing the main traffic route in French Guiana. Focusing on the main urban areas, represented in Figure 5, the HSI values within the highly urbanized districts of Cayenne and Kourou (rectangles in Figure 5) are lower than those of the surrounding pixels that are not considered to be highly urbanized. A very high HSI was predicted within the urban area of Saint-Laurent-du-Maroni. However, none of the pixels characterizing this city has a *PER_URB_NEIGH* value higher than or equal to 50%. The high HSI values in areas B, D, E, and G are characterized by the environmental variables *ROADS*, *HA_MAX*, the classes *Open forest* and *Mixed high and open forest*, and flat or moderately hilly terrain. The high HSI in areas C and F is essentially linked to *Open forest* and flat terrain.

The areas for which the model did not predict the HSI, due to the choice to not extrapolate to environmental domains not used to train the model, correspond to areas with an altitude higher than 400 meters. They represent a small number of pixels of the study area.

Discussion

The prediction performances of the model are excellent and significantly greater than those of the null model. The following discussion focuses on the ecological interpretation of the results and the methodological choices and alternatives.

Environmental Factors Explaining the Habitat Suitability

The geographic distribution of habitat suitability is consistent with existing knowledge of the

entomological situation despite the small number of presence points. The high HSI values can be explained by different environmental contexts depending on the geographical locations. In most areas (A, B, D, and E), the HSI values depend on human presence and activities, characterized by the environmental variables *HA_MAX* and *ROADS* (in areas D, E, and B, most roads are not paved and correspond mostly to tracks). The significantly positive correlation between the variable *ROADS* and the HSI confirms that road and track opening, accompanied by deforestation and pooling of rainwater at the roadside, may favor breeding sites (Singer and Castro 2001). The response curve for the variable *ROADS* (Figure 3) reaches a plateau above 7,000 meters of road per square kilometer and decreases thereafter. The decrease of the HSI at values above 7,000 meters suggests that the density of the road network leads to an improvement of the road quality (paved road eliminating culverts, adding sidewalks), thus limiting the availability of breeding and/or resting sites, in the same way as urbanization. Indeed, the response curve of the *PER_URB_NEIGH* variable confirms that highly urbanized areas provide a poorly suitable habitat for *An. darlingi* (Figure 3). Intensive urbanization implies concrete paving, the decrease or removal of green areas and forests, and consequently, the destruction of breeding and resting sites for *An. darlingi* (Stefani et al. 2013). This phenomenon is observed in the highly urbanized areas of Cayenne and Kourou (Figure 5). In contrast, Saint-Laurent-du-Maroni, the second largest urban area of French Guiana in terms of urbanization size and density, has high HSI values. In fact, unlike Cayenne and Kourou, this area is not considered to be highly urbanized using the criterion of this study (*PER_URB_NEIGH* $\geq$ 50%). However, the result for Saint-Laurent-du-Maroni seems unlikely because the presence of *An. darlingi* has not yet been reported in an urban area. Further field works could confirm the presence of this species in the city. The sensitivity of the model for the criterion that defines a highly urbanized area may also merit further study.

The values of *HA_MAX* in areas D and E were essentially associated with mining activity. In French Guiana, this activity is responsible for forest loss reaching 2,000 hectares per year (Office National

372 des Forêts Direction Régionale de Guyane, 2014). Between 2001 and 2013, Alvarez-Berríos and
373 Aide (2015) estimated that the largest forest loss due to gold mining in the tropical and subtropical
374 moist forest in South America was situated in the Guianan region including French Guiana. This
375 suggests that this activity, resulting in deforestation and creating sources of standing water such as
376 mining pits, combined with the presence of a large number of people, creates suitable conditions for
377 *An. darlingi*. The high HSI in these two areas is also explained by the *Mixed high and open forest*
378 landscape which is associated with human disturbance (Gond et al. 2011). Indeed, this landscape is
379 described as a forest environment linked to young or unstable vegetation mostly due to first stages
380 of anthropization. These results confirm the important role of human presence in the creation of
381 suitable habitats for *An. darlingi*, which is also consistent with the strong anthropophilic behavior of
382 this vector.
383 Some landscape types which are not directly associated with human presence or activities were also
384 associated with a HSI. The *Woodland savanna/dry forest* class appears to highly contribute to high
385 HSI values (Figure 2). It corresponds to the driest landscape in French Guiana (Gond et al. 2011),
386 but can be seasonally inundated due to its poor drainage, creating breeding sites (Rosa-Freitas et al.
387 2007). The high HSI values in this area are in accordance with previous studies (Vezenegho et al.
388 2015, Dusfour et al. 2013), which reported finding *An. darlingi* in the coastal savanna environments
389 of French Guiana. In uninhabited areas (zones F and C in Figure 4), a high HSI is associated with
390 the *Open forest* class (LS layer) and flat terrain. This LS class can be associated with different land
391 cover types in French Guiana (Gond et al. 2011) depending on the geographical location.
392 Consequently, this *LS* class may differentially affect *An. darlingi* habitat suitability. The *Open forest*
393 in area C mainly corresponds to wetlands (classified as *Flooded forest* according to the coastal land
394 use map provided by the Office National des Forêts Direction Régionale de Guyane, 2013), whereas
395 in area F, it corresponds to *Large surfaces of bamboo thicket and forbs*. *Anopheles darlingi* was
396 found in flooded forest; however, to our knowledge, no information is available concerning its

397 presence in large areas of bamboo thicket and forb*s*. The prediction in these areas should be taken
398 with precaution as a more precise description of the habitats within *Open forest* class is required.
399 Overall, this information highlights that natural environment could form highly suitable habitats
400 despite the high anthropophily of *An. darlingi*.

401 **Meteorological Variables**

402 In this study, meteorological variables were not used to build the model. Temperatures fall within
403 the optimal range for the species presence, and were considered to be geographically and
404 temporally too homogeneous to explain differences in the spatial distribution of habitat suitability.
405 Such a hypothesis is common in the Amazonian context. Olson et al. (2009) report that in their
406 study region (Amazon basin), "monthly temperatures were between 24.6°C and 29.4°C (well within
407 the range for optimal malaria transmission) for 95% of the observations," and consequently did not
408 include temperatures in their model. In French Guiana, several studies also used rainfall data to
409 study the intra-annual variations in *An. darlingi* density (Hiwat et al., 2010, Girod et al., 2011). The
410 exclusion of rainfall data is more debatable, as rainfall clearly influenced the intra-annual density of
411 *An. darlingi* in the study region (Hiwat et al., 2010, Girod et al., 2011, Vezenegho et al., 2015) even
412 if the relationship was not systematically observed (Girod et al., 2011). The evidence for this impact
413 on densities is that *An. darlingi* habitat suitability varies at an intra-annual scale, due to the
414 alternation of dry and wet seasons. However, the entire study area is subject to this alternation.
415 Moreover, given the high density of the French Guiana hydrological network and that the driest area
416 (north-west) still receives 2,000 mm a year, it can be reasonably assumed that *An. darlingi* can find
417 suitable conditions within the entire territory throughout most of the year. In French Guiana, the
418 geomorphological landscape highly influences the availability of breeding sites, and therefore their
419 spatial distribution, whereas the rainfall quantities influence the intra-annual variations of *An.*
420 *darlingi* densities. As a consequence, on an average over the year, we assume that the significant
421 factor influencing the distribution of habitat suitability is not the quantity of rainfall, but the

capacity of the landscape to provide suitable breeding sites when it rains.

Model Parametrization

The model was run by using the regularization parameter values and background set size recommended by Phillips and Dudík (2008), instead of those determined from specific experiments, as suggested by Merow et al. (2013). Phillips and Dudík (2008) tested a set of regularization parameter values with 48 species datasets that contained 11 to 13 environmental variables and a small number of categorical variables (1-3, as they considered discrete ordinal variables to be categorical). Nine of these datasets contained between 30 and 60 occurrences. The characteristics of the dataset exploited in our study (39 occurrence records; 13 and seven environmental variables including three categorical ones) are assumed to be quite similar of those of the datasets used by Phillips and Dudík (2008). We thus assumed that the pseudo-optimal parameters proposed by Phillips and Dudík (2008) could be confidently used in our study. Similarly, the background size was set to 10,000 based on the tests realized by Phillips and Dudík (2008), with 226 species and a median number of 57 presence sites. Better prediction performance may have been obtained by tuning the regularization values and background size and adding input environmental variables and features. However, the risk would have been to favor overfitting to the detriment of the bio-ecological interpretation of the model (see for example Merow et al., 2013). According to the entomologists who participated in the study, the model appears to be a good compromise between overfitting (that would have predicted suitable areas near occurrence points only) and being too general (that would have predicted suitable areas in too many environmental contexts for which the specialists have no species presence evidence).

Correction of the sampling bias effect

In this study, the effect of sampling bias was corrected by selecting background points with the same environmental bias as the sampled points. This approach appeared to be useful when applied to *An. darlingi* in French Guiana. Without a bias effect correction, the model predicted very high

447 HSI values in highly urbanized areas whereas these areas are known to be unsuitable for this vector
448 (see above). The biased background set is more concentrated around the sampled points (in the
449 environmental space) than the uniform random background, and is not likely to include
450 environmental conditions that are highly dissimilar to those encountered at the sampled points. As a
451 result, environmental conditions highly dissimilar to those of the sampled points can be subjected to
452 extrapolation, which may lead to erroneous habitat suitability predictions and bio-ecological
453 interpretations. This justifies not using the extrapolation option for modeling. The predicted HSI
454 map from the model with a biased background contains several excluded areas, whereas that of the
455 model with a uniform random background does not. Excluded areas correspond to high altitude
456 areas which are unsuitable for *An. darlingi* (Mouchet 2004).
457 When using a uniform random background, the three most contributive variables (cumulative
458 contribution equal to 85.5%) were all directly linked to human presence and territory accessibility
459 (*ROADS*: 38.1%, *PER_URB_NEIGH*: 34.9%, and *HA_MAX*: 12.5%). Thus, apart from urban areas,
460 high HSI values were associated with high *HA_MAX* and *ROADS* values. However, when
461 correcting the sampling bias effect, the *Landscape* (*LS*) variable was the second most contributive
462 variable (14.1%), the ROADS variable contribution increased to 62.6%, and the
463 PER_URB_NEIGH variable contribution decreased to 11.1% (see Table 3).
464 From a quantitative point of view, the two approaches (with and without applying the correction of
465 the sampling bias effect) resulted in identical AUC and partial AUC ratios. However, the regularized
466 gain and the CBI were lower without correction, with values equal to 2.81 (vs. 3.18) and 0.284 (vs.
467 0.421), respectively.
468 Thus, correction of the sampling bias effect gave better results: both more consistent with
469 knowledge from the field and more accurate in terms of prediction. The fact that the contributions
470 of the *LS* and *HA_MAX* variables respectively increased and decreased with the use of the biased
471 background, tends to show that the correction method actually manages to counterbalance the over-

representation of inhabited areas (cities, villages, and gold mining areas) in the sampled data. However, the very high contribution of the variable *ROADS* may be a residual effect of sampling bias, as sampling is essentially performed in the vicinity of accessible roads and tracks. Further studies are necessary to objectively and quantitatively assess the actual performance of the proposed methodology for correcting the effect of sampling bias.

Habitat Suitability and Malaria in French Guiana

Alimi et al. (2015) highlighted the utility of SDMs for gaining a better understanding of the geographical range and distribution of vectors for eliminating malaria and preventing outbreaks. The coastal strip in French Guiana is generally malaria free, although some cases resulting from local transmission are regularly diagnosed (Ardillon et al. 2015). This study, as well as that of Vezenegho et al. (2015), shows that the savanna in French Guiana may be highly suitable for *An. darlingi*. In the forest, Pommier de Santi et al. (2016c) found a link between mining, malaria cases, and the presence of *An. darlingi*. Indeed, >74% of malaria cases in French army soldiers were associated with operations to counteract illegal gold mining (Pommier de Santi et al. 2016a). According to the results of the present study, some areas associated with intense gold mining activity, known to be malaria transmission foci, are not necessarily associated with very high HSI values. In the village of Camopi, the annual malaria prevalence was 70% for children younger than seven years of age between 2000 and 2002 (Carme et al. 2005), reaching 100% in 2006 (Hustache et al. 2007). However, only some pixels on the border of the Camopi and Oyapock rivers have high values on the HSI map (area G in Figure 4). This is consistent with the study of Girod et al. (2011), which showed that the number of *An. darlingi* caught in this village was very low relative to the incidence of malaria cases. These findings collectively highlight two important points. First, the HSI map shown in Figure 4 does not correspond to a map of malaria transmission risk. Transmission risk depends on many factors that were not taken into account here, such as the parasitic charge and immunological status of the local population, compositional and

497 configurational features of the landscape (Stefani et al, 2013; Li et al, 2016), and behavioral factors.
498 Second, this highlights that malaria transmission can occur in areas where there is a very low
499 density of *An. darlingi.* This may be due to the presence of other *Anopheles* species such as *An.*
500 *(Nys.) nuneztovar*i Galbaldón*, An. (Nys.) oswaldo*i Peryassú*, An. (Nys.) intermedi*us Peryassú, *An.*
501 *(Nys.) marajoara* Galvão and Damasceno, or *An. (Nys.) ininii* Sénevet and Abonnenc (Diptera:
502 Culicidae), already known to be naturally infected with *Plasmodium* species and/or described as
503 efficient secondary malaria vectors (Dusfour et al. 2012, Pommier de Santi et al, 2016c).

504 **Environmental Characterization**

505 A significant limitation of this study was the spatial resolution of the environmental data. Capture
506 campaigns are generally carried out at a local scale (villages or camps; Vezenegho et al. 2015,
507 Dusfour et al. 2013). The spatial resolution of the study was not sufficient to take into account the
508 heterogeneity of the environment at the capture scale. The use of environmental data with higher
509 spatial resolution, such as the canopy height estimation from Fayad et al. (2014) or finer
510 characterization of the land cover could improve future studies. However, these data are not
511 consistently available across the entire territory.

512

513 In conclusion, the results of this study help to complete our knowledge on the spatial distribution of
514 the principal malaria vector in this Amazonian region, and to identify the main factors that favor its
515 presence. These results can be exploited to define the necessary targeted vector control strategies in
516 a malaria pre-elimination context, and to extrapolate the acquired knowledge to other Amazonian
517 contexts. They also suggest areas that need to be targeted to complete the field knowledge, validate
518 the prediction and strengthen the model. Eventually, these proposed methodological developments
519 can be applied to other species, including other disease vectors.

520 **Acknowledgements**

521 This study was funded by the Fonds Social Européen (FSE), Centre National d'Etudes Spatiales

(CNES), and Collectivité Territoriale de Guyane. Financial support was partially provided by the "Investissement d'Avenir" grants managed by the Agence Nationale de la Recherche (Center for the study of Biodiversity in Amazonia, ANR-10-LABX-0025) and by the GAPAM-Sentinela project of the Franco-brazilian scientific and academic cooperation program Guyamazon (funds: IRD, CIRAD, French Embassy in Brazil, Territorial Collectivity of French Guiana, Brazilian State-level research agencies of Amapá, Amazonas and Maranhão).

References cited

Agence Régionale de Santé Guyane. 2015. Plan de lutte contre le paludisme en Guyane 2015-2018. Agence Régionale de Santé Guyane, Cayenne, France.

Alimi, T.O., D. O. Fuller, M. L. Quinones, R-D. Xue, S. V. Herrera, M. Arevalo-Herrera, J. N. Ulrich, W. A. Qualls, and J. C. Beier. 2015. Prospects and recommendations for risk mapping to improve strategies for effective malaria vector control interventions in Latin America. Malar. J. 14:519.

Alvarez-Berríos, N. L., and T. M. Aide. 2015. Corrigendum: Global demand for gold is another threat for tropical forests (2014 Environ. Res. Lett. 10: 014006). Environ. Res. Lett. 10: 29501.

Ardillon, V., L. Carvalho, C. Prince, P. Abboud, and F. Djossou. 2015. Bilans 2013 et 2014 de la situation du paludisme en Guyane. Bulletin de Veille Sanitaire Antilles-Guyane. 1: 16-20.

Berger, F., C. Flamand, L. Musset, F. Djossou, J. Rosine, M.-A. Sanquer, I. Dusfour, E. Legrand, V. Ardillon, P. Rabarison, C. Grenier and R. Girod. 2012. Investigation of a sudden malaria outbreak in the isolated Amazonian village of Saul, French Guiana, January-April 2009. Am. J. Trop. Med. Hyg. 86: 591–597.

Carme, B., J. Lecat, and P. Lefebvre. 2005. [Malaria in an outbreak zone in Oyapock (French Guiana): incidence of malaria attacks in the American Indian population of Camopi] (in French) Médecine tropicale 65: 149–154.

Deane, L. M., O. R. Causey, and M. P. Deane. 1948. Notas sobre a distribuição ea biologia dos

547 anofelinos das regiões nordestina e Amazônica do Brasil. Revista do Serviço Especial de Saúde
548 Pública 1: 827–965.
549 **de Thoisy, B., C. Richard-Hansen, B. Goguillon, P. Joubert, J. Obstancias, P. Winterton, and**
550 **S. Brosse. 2010.** Rapid evaluation of threats to biodiversity: human footprint score and large
551 vertebrate species responses in French Guiana. Biodivers. Conserv. 19: 1567–84.
552 **Dusfour, I., J. Issaly, R. Carinci, P. Gaborit, and R. Girod. 2012.** Incrimination of *Anopheles*
553 *(Anopheles) intermedius* Peryassú, *An. (Nyssorhynchus) nuneztovari* Gabaldón, *An. (Nys.)*
554 *oswaldoi* Peryassú as natural vectors of *Plasmodium falciparum* in French Guiana. Memórias Do
555 Instituto Oswaldo Cruz 107: 429–432.
556 **Dusfour, I., R. Carinci, J. Issaly, P. Gaborit, and R. Girod. 2013.** A survey of adult *Anophelines*
557 in French Guiana: enhanced descriptions of species distribution and biting responses. J. Vector
558 Ecol. 38: 203–209.
559 **Elith, J., C. H. Graham, R. P Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F.**
560 **Huettmann, J. R. Leathwick, A. Lehmann, et al. 2006.** Novel methods improve prediction of
561 species' distributions from occurrence data. Ecography 29: 129–151.
562 **Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011**. A statistical
563 explanation of MaxEnt for ecologists: statistical explanation of MaxEnt. Divers. Distrib. 17: 43–
564 57.
565 **Faran, M. E., and K. J. Linthicum. 1981.** A handbook of the Amazonian species of *Anopheles*
566 *(Nyssorhynchus)* (Diptera: Culicidae). Mosq. Syst. 13: 1-81
567 **Fayad, I., N. Baghdadi, J-S. Bailly, N. Barbier, V. Gond, M. Hajj, F. Fabre, and B. Bourgine.**
568 **2014.** Canopy height estimation in French Guiana with LiDAR ICESat/GLAS data using
569 principal component analysis and Random Forest regressions. Rem. Sens. 6: 11883–11914.
570 **Floch, H., and E. Abonnenc. 1951.** Anophèles de la Guyane française. Archives de l'Institut
571 Pasteur de la Guyane et du territoire de l'Inini. 236:1-92

572 **Forattini, O. P. 1962.** Entomologia Médica : Vol. 1. Parte Geral, Diptera, Anophelini. Faculdade de
573 Higiene e Saúde Pública, Departamento de Parasitologia, Universidade de São Paulo, São Paulo,
574 Brazil.
575 **Fouque, F., P. Gaborit, R. Carinci, J. Issaly, and R. Girod. 2010.** Annual variations in the
576 number of malaria cases related to two different patterns of *Anopheles darlingi* transmission
577 potential in the Maroni area of French Guiana. Malar. J. 9: 80.
578 **Friedman, J., T. Hastie, and R. Tibshirani. 2000.** Additive Logistic Regression: a Statistical View
579 of Boosting. Ann. Stat. 28: 337–407.
580 **George, P. 1972.** Dictionnaire de la géographie. Presse Universitaire de France, Paris, France.
581 **Girod, R., P. Gaborit, R. Carinci, J. Issaly, and F. Fouque. 2008.** *Anopheles darlingi* bionomics
582 and transmission of *Plasmodium falciparum*, *Plasmodium vivax* and *Plasmodium malariae* in
583 Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana. Memórias Do
584 Instituto Oswaldo Cruz 103: 702–710.
585 **Girod, R., E. Roux, F. Berger, A. Stefani, P. Gaborit, R. Carinci, J. Issaly, B. Carme, and I.**
586 **Dusfour. 2011.** Unravelling the relationships between *Anopheles darlingi* (Diptera: Culicidae)
587 densities, environmental factors and malaria incidence: understanding the variable patterns of
588 malarial transmission in French Guiana (South America). Ann. Trop. Med. Parasitol. 105: 107–
589 122.
590 **Gond, V., V. Freycon, J-F. Molino, O. Brunaux, F. Ingrassia, P. Joubert, J-F. Pekel, M-F.**
591 **Prévost, V. Thierron, P-J. Trombe, and D. Sabatier. 2011.** Broad-scale spatial pattern of forest
592 landscape types in the Guiana Shield. Int. J. Appl. Earth Obs. Geoinf. 13: 357–367.
593 **Guisan, A., T. C. Edwards, and T. Hastie. 2002.** Generalized linear and generalized additive
594 models in studies of species distributions: setting the scene. Ecol. Model. 157: 89–100.
595 **Guitet, S., J-F. Cornu, O. Brunaux, J. Betbeder, J-M. Carozza, and C. Richard-Hansen. 2013.**
596 Landform and landscape mapping, French Guiana (South America). J. Maps. 9: 325–35.

Hammond, D. S. 2005. Tropical forests of the Guiana Shield: ancient forests in a modern world. CABI publishing, Wallingford, England.

Héritier, P. 2011. Le climat guyanais ; petit atlas climatique de la Guyane française. Météo France, Cayenne, France.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773–785. **Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002.** Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83: 2027–2036.

Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Model. 199: 142–152.

Hiwat, H., J. Issaly, P. Gaborit, A. Somai, A. Samjhawan, P. Sardjoe, T. Soekhoe, and R. Girod. 2010. Behavioral heterogeneity of *Anopheles darlingi* (Diptera: Culicidae) and malaria transmission dynamics along the Maroni River, Suriname, French Guiana. Trans. R. Soc. Trop. Med. Hyg. 104: 207–213.

Hustache, S., M. Nacher, F. Djossou, and B. Carme. 2007. Malaria risk factors in Amerindian children in French Guiana. Am. J. Trop. Med. Hyg. 76: 619–625.

Leathwick, J. R., D. Rowe, J. Richardson, J. Elith, and T. Hastie. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshw. Biol. 50: 2034–2052.

Li, Z., E. Roux, N. Dessay, R. Girod, A. Stefani, M. Nacher, A. Moiret, and F. Seyler. 2016. Mapping a knowledge-based malaria hazard index related to landscape using remote sensing: application to the cross-border area between French Guiana and Brazil. Remote Sens. 8: 1-22.

Martens, W. J., L. W. Niessen, J. Rotmans, T. H. Jetten, and A. J. McMichael. 1995. Potential impact of global climate change on malaria risk. Environ. Health Perspect. 103: 458–64.

622 **Météo-France. 2016**. Données pluviométriques disponibles au 01/01/2016
623 (http://pluiesextremes.meteo.fr/guyane/IMG/sipex_pdf/carte_reseau_dep973.pdf). Last accessed
624 the 7 Oct. 2016.
625 **Merow, C., M. J. Smith, and J. A. Silander. 2013.** A Practical guide to MaxEnt for modeling
626 species' distributions: what it does, and why inputs and settings matter. Ecography 36: 1058–
627 1069.
628 **Musset, L., S. Pelleau, R. Girod, V. Ardillon, L. Carvalho, I. Dusfour, M. SM Gomes, F.**
629 **Djossou, and E. Legrand. 2014.** Malaria on the Guiana Shield: a review of the situation in
630 French Guiana. Memórias Do Instituto Oswaldo Cruz 109: 525–33.
631 **Mouchet, J. 2004.** Biodiversité du paludisme dans le monde. John Libbey Eurotext, Montrouge,
632 France
633 **Office National des Forêts Direction Régionale de Guyane. 2013.** PROJET "EXPERTISE
634 LITTORAL 2011" Occupation du sol et sa dynamique sur la bande côtière de la Guyane de 2005
635 à 2011. Office National des Forêts et le Ministère de l'Agriculture, de l'Agroalimentaire et de la
636 Forêt, Cayenne, France.
637 **Office National des Forêts Direction Régionale de Guyane. 2014.** Rapport d'activité 2013.
638 Office National des Forêts et le Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt,
639 Cayenne, France.
640 **Olson, S. H., R. Gangnon, E. Elguero, L. Durieux, J. F. Guégan, J. A. Foley, and J. A. Patz.**
641 **2009.** Links between climate, malaria, and wetlands in the Amazon Basin. Emerg. Infect. Dis.15:
642 659-662.
643 **Patz, J. A., T. K. Graczyk, N. Geller, and A. Y. Vittor. 2000.** Effects of environmental change on
644 emerging parasitic diseases. Int. J. Parasitol. 30: 1395–1405.
645 **Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007.** Predicting
646 species distributions from small numbers of occurrence records: a test case using cryptic geckos

in Madagascar: Predicting species distributions with low sample sizes. J. Biogeogr. 34: 102–117.

Peterson, A. T. 2007. Ecological niche modelling and understanding the geography of disease transmission. Vet. Ital. 43: 393–400.

Peterson, A. T., M. Papeş, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol. Model. 213: 63–72.

Petit-Sinturel, M., L. Carvalho, A. Andrieu, C. Prince, P. Abboud, F. Djossou, and V. Ardillon. 2016. Situation du paludisme en Guyane française en 2015. Bulletin de Veille Sanitaire Antilles-Guyane. 2: 6-10.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190: 231–259.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161–175.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19: 181–197.

Pommier de Santi, V., A. Dia, A. Adde, G. Hyvert, J. Galant, M. Mazevet, C. Nguyen, S. B. Vezenegho, I. Dusfour, R. Girod, and S. Briolant. 2016a. Malaria in French Guiana Linked to Illegal Gold Mining. Emerging Infect. Dis. 22: 344–346.

Pommier de Santi, V., F. Djossou, N. Barthes, H. Bogreau, G. Hyvert, C. Nguyen, S. Pelleau, E. Legrand, L. Musset, M. Nacher, and S. Briolant. 2016b. Malaria hyperendemicity and risk for Artemisinin resistance among illegal gold miners, French Guiana. Emerging Infect. Dis. 22: 903–906

Pommier de Santi, V., R. Girod, M. Mura, A. Dia, S. Briolant, F. Djossou, I. Dusfour, A. Mendebil, F. Simon, X. Deparis, and F. Pagès. 2016c. Epidemiological and entomological studies of a malaria outbreak among French armed forces deployed at illegal gold mining sites

reveal new aspects of the disease's transmission in French Guiana. Malar. J. 15: 35.

Queyriaux, B., G. Texier, L. Ollivier, L. Galoisy-Guibal, R. Michel, and J. B. Meynard. 2011. *Plasmodium vivax* malaria among military personnel, French Guiana, 1998–2008. Emerging Infect. Dis. 17 : 1280–1282.

Richard, A. 1987. Le Paludisme en forêt, pp. 249–250. *In* Connaissance Du Milieu Amazonien, 15-16 October 1985, Paris, France. ORSTOM, Paris, France. **Rosa-Freitas, M. G., P. Tsouris, A. T. Peterson, N. A. Honório, F. S. M. de Barros, D. B. de Aguiar, H. C. Gurgel, M. de Arruda, S. D. Vasconcelos, and J. F. Luitgards-Moura. 2007.** An ecoregional classification for the state of Roraima, Brazil. The importance of landscape in malaria biology. Memórias Do Instituto Oswaldo Cruz. 102: 349–358.

Rozendaal, J. A. 1992. Relations between *Anopheles darlingi* breeding habitats, rainfall, river level and malaria transmission rates in the rain forest of Suriname. Med. Vet. Entomol. 6: 16–22.

Singer, B. H., and M. C. Castro. 2001. Agricultural colonization and malaria on the Amazon frontier. Ann. N. Y. Acad. Sci. 954: 184–222.

Smith, M. W., M. G. Macklin, and C. J. Thomas. 2013. Hydrological and geomorphological controls of malaria transmission. Earth-Sci. Rev. 116: 109–127.

Stefani, A., E. Roux, J. M. Fotsing, and B. Carme. 2011. Studying relationships between environment and malaria incidence in Camopi (French Guiana) through the objective selection of buffer-based landscape characterizations. Int. J. Health Geogr. 10: 1-13.

Stefani, A., I. Dusfour, A. P. S. A. Corrêa, M. C. B. Cruz, N. Dessay, A. K. R. Galardo, C. D. Galardo, R. Girod, M. S. M. Gomes, and H. Gurgel. 2013. Land cover, land use and malaria in the Amazon: a systematic literature review of studies using remotely sensed data. Malar. J. 12: 1–8.

Stockwell, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13: 143–158.

Takken, W., P. D. R. Vilarinhos, P. Schneider, and F. Dos Santos. 2005. Effects of environmental change on malaria in the Amazon region of Brazil. Frontis 9: 113–123.

Terrazas, W., V. Sampaio, D. de Castro, R. C. Pinto, B. C. de Albuquerque, M. Sadahiro, R. dos Passos, and J. U. Braga. 2015. Deforestation, drainage network, indigenous status, and geographical differences of malaria in the state of Amazonas. Malar. J. 14: 379.

Tognelli, M. F., S. A. Roig-Juñent, A. E. Marvaldi, G. E. Flores, and J. M. Lobo. 2009. An evaluation of methods for modelling distribution of Patagonian insects. Revista Chilena de Historia Natural 82: 347-360.

Verret, C., B. Cabianca, R. Haus-Cheymol, J-J. Lafille, G. Loran-Haranqui, and A. Spiegel. 2006. Malaria outbreak in troops returning from French Guiana. Emerg. Infect. Dis. 12: 1794–1795.

Vezenegho S.B., R. Carinci, P. Gaborit, J. Issaly, I. Dusfour, S. Briolant and R. Girod. 2015. *Anopheles darlingi* (Diptera: Culicidae) dynamics in relation to meteorological data in a cattle farm located in the coastal region of French Guiana: advantage of Mosquito Magnet trap. Environ. Entomol. 44: 454-462.

Vittor, A. Y., R. H. Gilman, J. Tielsch, G. Glass, T. I. M. Shields, W. Lozano, V. Pinedo-Cancino, and J. A. Patz. 2006. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. Am. J. Trop. Med. Hyg. 74: 3–11.

Vittor, A. Y., W. Pan, R. H. Gilman, J. Tielsch, G. Glass, T. Shields, W. Sánchez-Lozano, V. V. Pinedo, E. Salas-Cobos, and S. Flores. 2009. Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*. Am. J. Trop. Med. Hyg. 81: 5-12.

Wangdi, K., M. L. Gatton, G. C. Kelly, and A. C. A. Clements. 2015. Cross-Border Malaria: A Major Obstacle for Malaria Elimination. Adv. Parasitol., 89:79–107.

722 **Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, and NCEAS**
723 **Predicting Species Distributions Working Group. 2008.** Effects of sample size on the
724 performance of species distribution models. Divers. Distrib. 14: 763–773.
725 **Yackulic, C. B., R. Chandler, E. F. Zipkin, J. A. Royle, J. D. Nichols, E. H. Campbell Grant,**
726 **and S. Veran. 2013.** Presence-only modelling using MAXENT: when can we trust the
727 inferences? Methods Ecol. Evol. 4: 236–243.
728 **Yost, A. C., S. L. Petersen, M. Gregg, and R. Miller. 2008.** Predictive modeling and mapping
729 sage grouse (*Centrocercus Urophasianus*) nesting habitat using Maximum Entropy and a long-
730 term dataset from Southern Oregon. Ecol. Inform. 3: 375-386.
731 **Zeilhofer, P., E. Santos, A. L. M. Ribeiro, R. D. Miyazaki, and M. Santos. 2007.** Habitat
732 suitability mapping of *Anopheles darlingi* in the surroundings of the Manso hydropower plant
733 reservoir, Mato Grosso, Central Brazil. Int. J. Health Geogr. 6: 1-14.

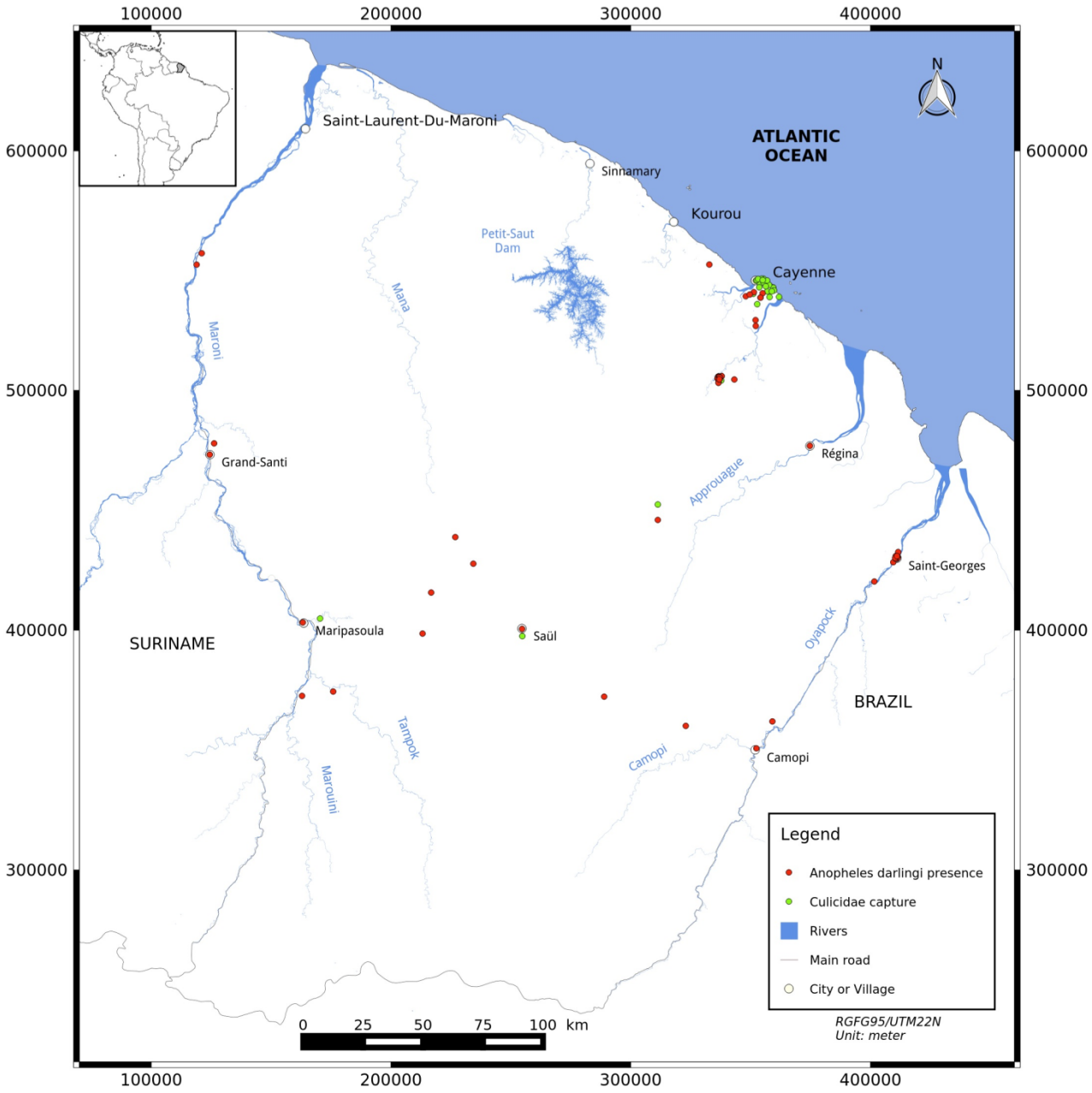

Figure 1. *Culicidae* capture points and *Anopheles darlingi* presence points (from 2000 to 2013).

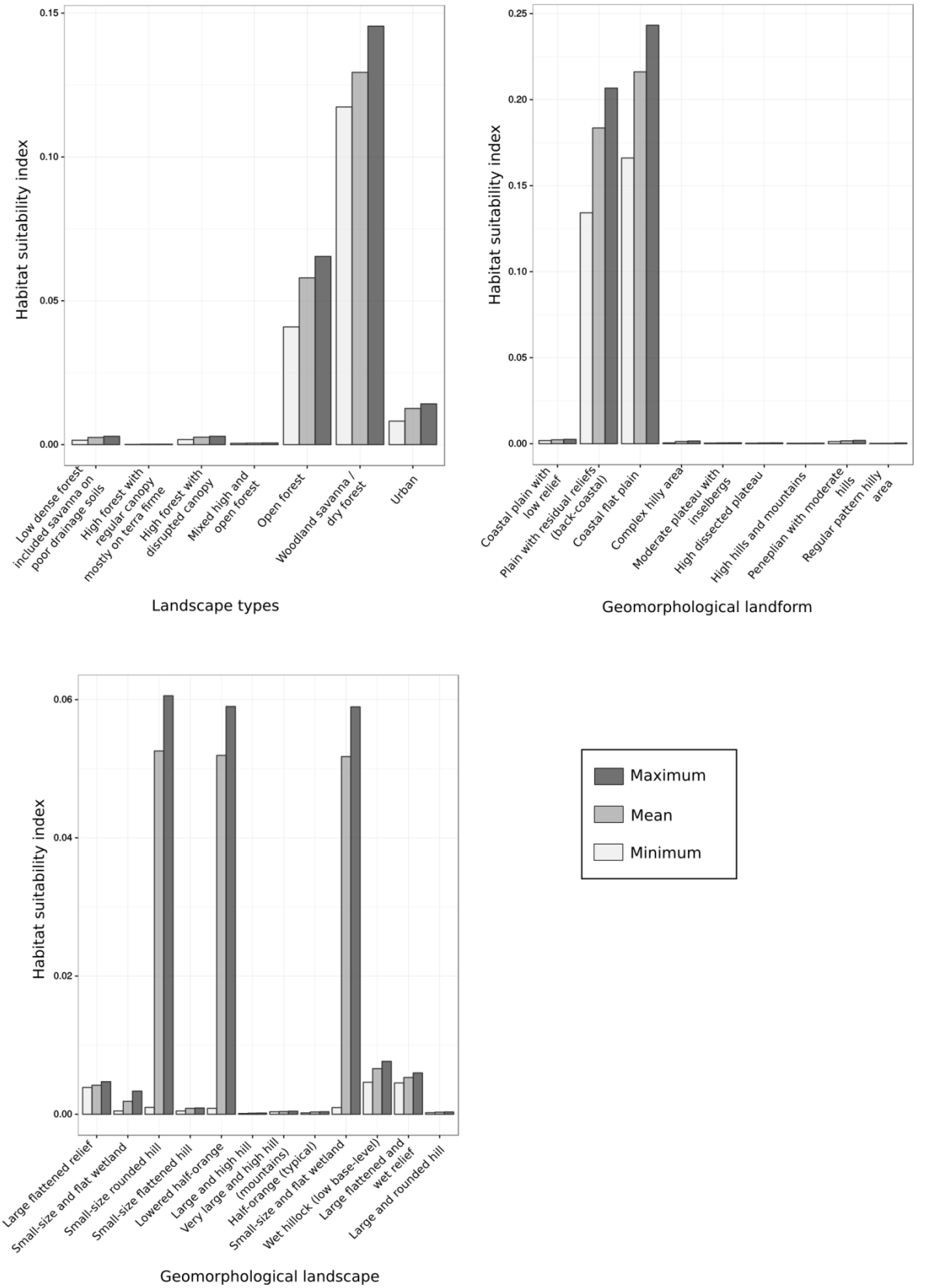

Figure 2. Response curves of categorical environmental variables.

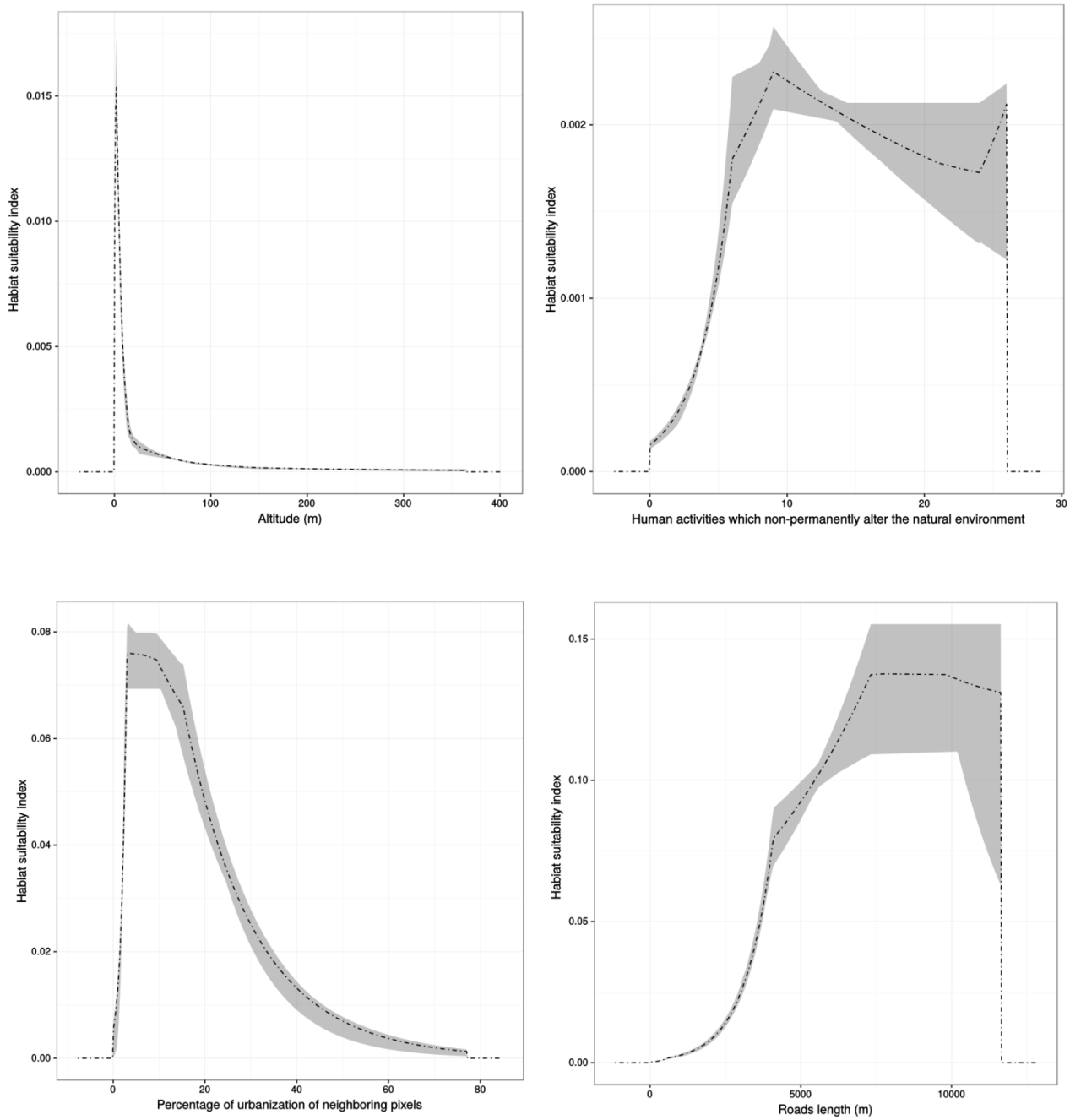

Figure 3. Response curves of numerical environmental variables. Dashed lines show the mean values and the grey regions represent the interval between the maximum and minimum values.

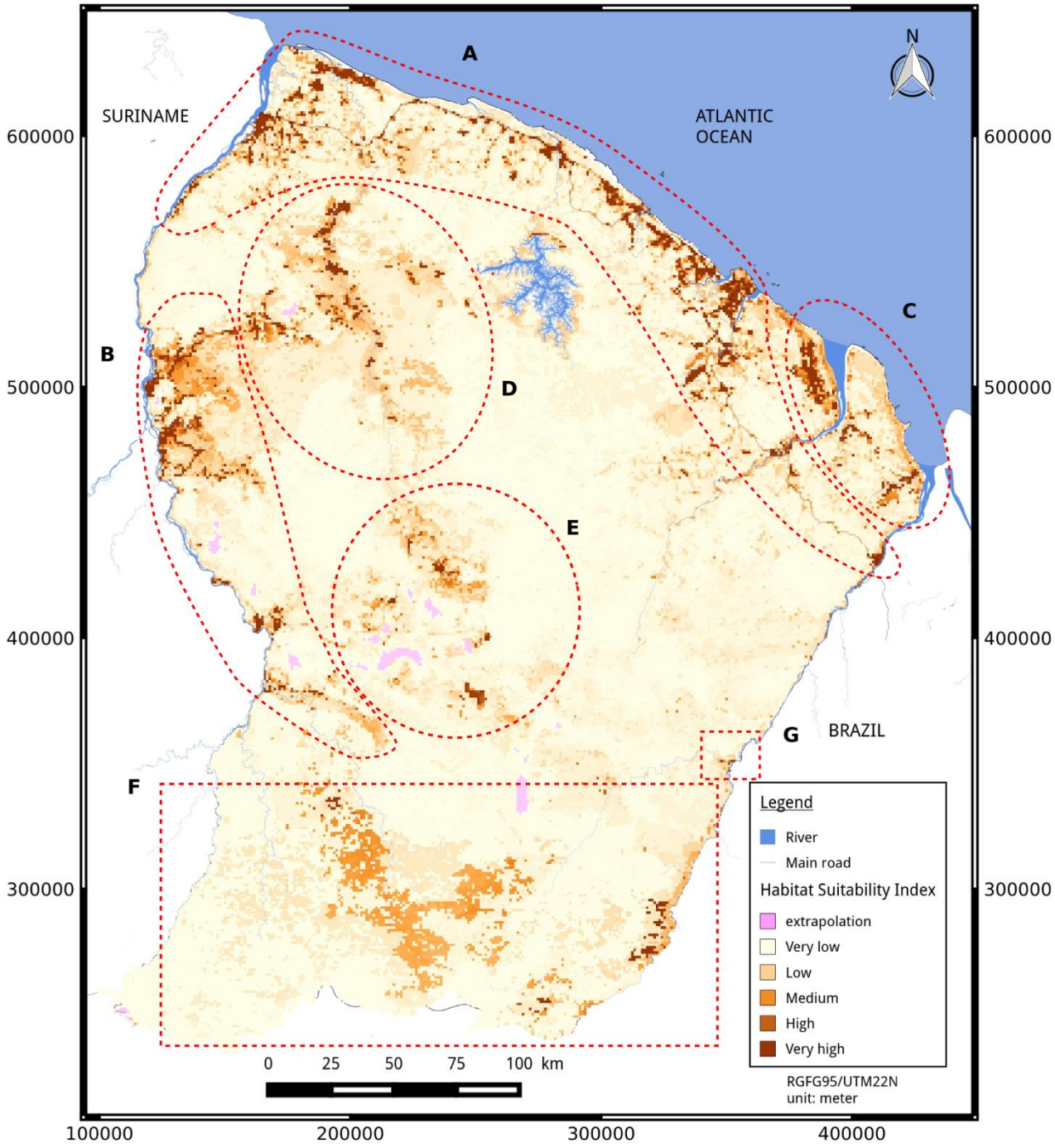

Figure 4. Habitat suitability index map. Six main areas with a high habitat suitability index (A to F) and Camopi village (G) are circumscribed by the red circles and rectangles.

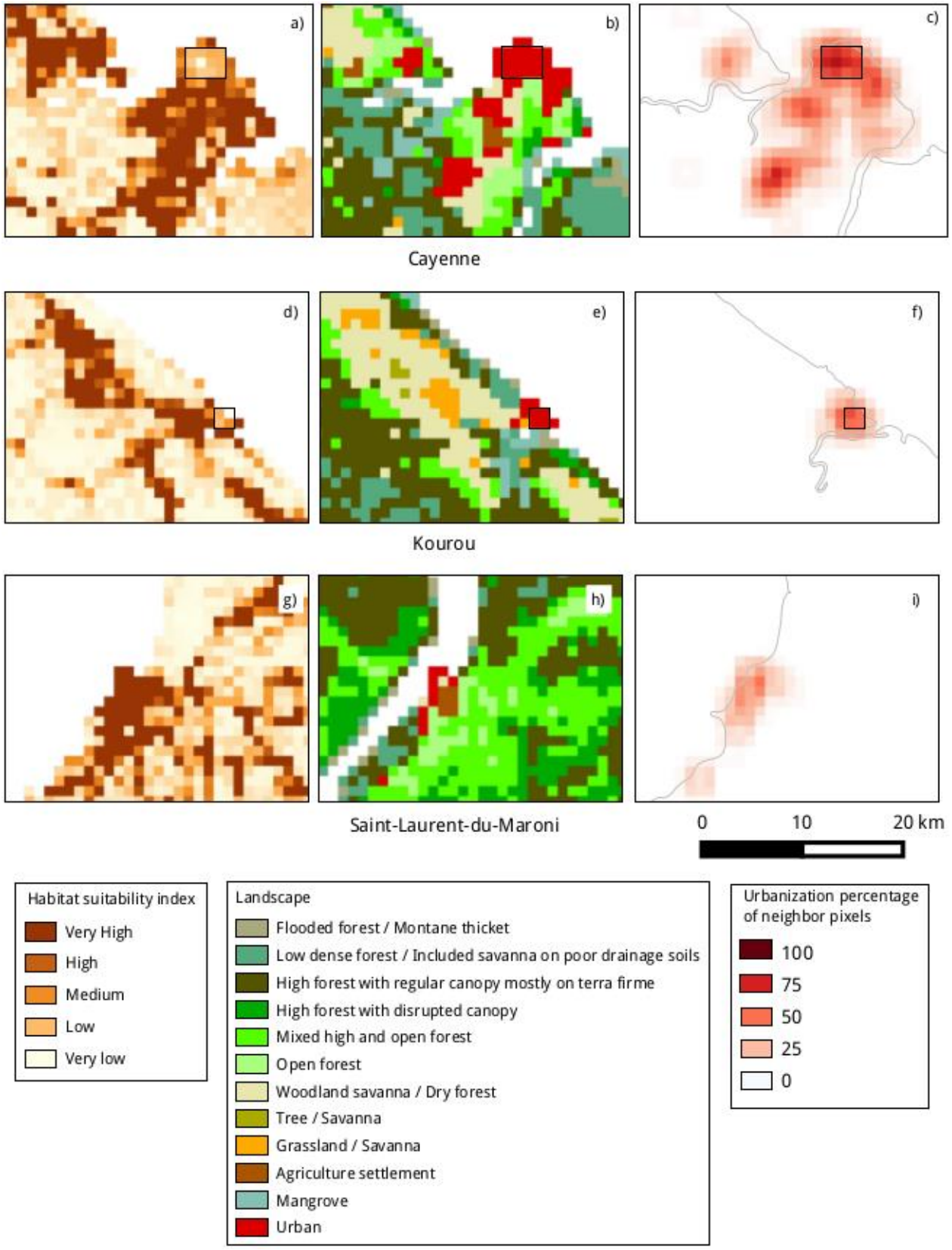

Figure 5. Zoom of urban areas. a, d, and g: habitat suitability index maps. b, e, and h: landscape type. c, f, and i: percentage urbanization of neighbor pixels. Rectangles correspond to highly urbanized areas (*LS* class is *Urban* and *PER_URB_NEIGH* $\geq$ 50%). Cayenne and Kourou include highly urbanized areas, but Saint-Laurent-du-Maroni does not.

Table 1. Raw environmental data and derived variables used to build the model.

Number of input variable	Producer, reference	Raw environmental data	Derived from (information source)	Date(s)	Original spatial resolution or interpretation scale	Derived SDM input variable(s)	Type of feature extraction for each 1x1 km pixel	Classes or range of values and units	Environment types [2]	*A priori* effect on An. darlingi presence [3] and bibliographic references	Input variable(s) type
1	Forest National Office (ONF), (Guitet et al. 2013)	Geomorphological landscape (*GLS*)	SRTM	2000	≥ 5000 m	Geomorphological landscape (*GLS*)	Majority class	12 classes	Natural environment	(/) Smith et al. (2013)	Categorical
2	Forest National Office (ONF), (Guitet et al. 2013)	Geomorphological landform (*GLF*)	SRTM	2000	≥ 200 m	Geomorphological landform (*GLF*)	Majority class	15 classes	Natural environment	(/) Smith et al. (2013)	Categorical
3	Agricultural Research Centre for International Development (CIRAD), (Gond et al. 2011)	Landscape types (*LS*)	Spot-Vegetation	2000	1000 m	Landscape types (*LS*)	Correction of pixels corresponding to urban areas and mangroves	14 classes	Natural environment and urbanization	(/) Stefani et al. (2013) Girod et al. (2011) Zeilhofer et al. (2007) Rozendaal (1992) Hiwat et al. (2010) Vittor et al. (2006) Vittor et al. (2009)	Categorical
4, 5, 6	National Aeronautics and Space Administration (NASA)	Altitude (*ALT*)	SRTM	2000	30 m	Altitude - minimum (*ALT_MIN*), - maximum (*ALT_MAX*) - median (*ALT_MED*)	Statistical computation	0 – 832 m	Natural environment	(-) Zeilhofer et al. (2007)	Continuous

Number of input variable	Producer, reference	Raw environmental data	Derived from (information source)	Date(s)	Original spatial resolution or interpretation scale	Derived SDM input variable(s)	Type of feature extraction for each 1x1 km pixel	Classes or range of values and units	Environment types [2]	*A priori* effect on An. darlingi presence [3] and bibliographic references	Input variable(s) type
7	National Institute of Geographic and Forestry Information (IGN)	Road and track network	BD TOPO®	2011	≥ 1000 m	Length of roads and tracks outside of urban areas (*ROADS*)	Computation of road/track lengths	0 – 12545 m	Non-permanent anthropogenic changes	(+) Singer and Castro (2001)	Continuous
8	Association Kwata 'Study and Conservation of French Guianan Wildlife ' (de Thoisy et al. 2010)	Human footprint (HFP)	From various sources [a]	2005	≥ 1000 m	Percentage of urbanization of neighboring pixels (*PER_URB_NEIGH*)	Percentage of urbanization within the eight neighbor cells	0-100%	Urbanization	(-) Stefani et al. (2013)	Continuous
9, 10, 11	Association Kwata 'Study and Conservation of French Guianan Wildlife ' (de Thoisy et al. 2010)	Human footprint (HFP)	From various sources [a]	2005	≥ 1000m	Human activities which non-permanently alter natural environment (*HA*) - minimum (*HA_MIN*) - maximum (*HA_MAX*) - median (*HA_MED*)	Statistical computation	0-30	Non-permanent anthropogenic changes	(+) Vittor et al. (2009)	Continuous

[a] French Institute for Statistical and Economic studies (INSEE); Regional Departments for Food, Agriculture and the Forest (DAAF); ONF; Regional Equipment, Habitat and Planning Authority (DDE) and Hammond et al. (2007).

[b] See the section on environmental variables in Materials and Methods.

[c] *A priori* effect on *An. darlingi* presence: (+) favorable; (-) unfavorable; (/) depends on categorical variable values.

Table 2. Mean contributions and jackknife results of the eleven input environmental variables.

Environmental variables	Contribution (%)	Cumulative contribution (%)	Gain with the variable only	Decrease of the gain without the variable (%)
ROADS	51.45	51.45	2.20	-7.98
PER_URB_NEIGH	17.17	68.62	1.86	-0.41
LS	15.32	83.94	2.23	-4.67
HA	7.43 (min: 0.35; median: 0.24; max: 6.84)	91.37	min: 0.02 median: 0.15 max: 0.43	min:-0.06 median: -0.22 max: -2.10
GLS	5.35	96.72	1.40	-2.59
ALT	2.09 (min: 1.34; median: 0.69; max: 0.06)	98.81	min: 1.12 median: 1.04 max: 0.76	min: - 1.04 median: -0.39 max: -0.03
GLF	1.19	100	0.80	-0.21

Table 3. Mean contributions and jackknife results of the seven input environmental variables of the simpler model.

Environmental variables	Contribution (%)	Cumulative contribution (%)	Gain with the variable only	Decrease of the gain without the variable (%)
ROADS	62.61	62.61	2.31	-8.61
LS	14.10	76.71	2.35	-6.23
PER_URB_NEIGH	11.15	87.86	2.05	-0.58
HA_MAX	5.39	93.25	0.37	-1.74
GLS	3.84	97.09	1.44	-1.90
GLF	2.1	99.19	1.01	-0.32
ALT_MIN	0.88	100	1.27	-1.29

Table 4. Characterization of areas with a high HSI

ns. signifies that the high HSI of the concerned area was not driven by that environmental variable, (+) signifies that when the value of the variable increases, the HSI also increases also, (-) signifies that when the value of the variable decreases, the HSI increases, and cells with classes name signifies that the presence of the given class implies a high HSI.

Area	*ROADS*	*LS* classes	*PER_URB_NEIGH*	*HA_MAX*	*GLS* classes	*GLF* classes	*ALT*
A	(+)	- Woodland savanna / Dry forest - Mixed high and open forest	(-)	(+)	- Coastal plain with low relief - Plain with residual reliefs (back coastal)	- Small size and flat wetland - Large flattened and wet relief - Wet hillock (low base-level)	(-)
B	(+)	- Open forest - Mixed high and open forest	*ns.*	(+)	- Peneplain with moderate hills	-Wet hillock (low base-level) - Large flattened relief	(-)
C	*ns.*	- Open forest	*ns.*	*ns.*	- Coastal flat plain	- Large flattened and wet relief	(-)
D	(+)	- Mixed high and open forest	*ns.*	(+)	*ns.*	*ns.*	(-)
E	(+)	- Mixed high and open forest	*ns.*	(+)	- Peneplain with moderate hills	- Large flattened relief	(-)
F	*ns.*	- Open forest	*ns.*	*ns.*	- Peneplain with moderate hills	- Lowered half-orange	(-)
G	(+)	Mixed high and open forest	*ns.*	(+)	*ns.*	*ns.*	(-)

S1. Coordinates of *Anopheles darlingi* presence sites (Coordinate system: RGFG95/UTM22N)

Number of sites	Locality	Longitude	Latitude
1	Cayodé	175866.785122426	374453.1653833
2	Taluène	162884.684665956	372648.888139721
3	Cacao	336895.086300153	505521.800163598
4	Midenangalanti	121041.015986301	557342.72581239
5	Grand Santi	124374.105534813	473261.338110987
6	Bois Martin	118927.138507594	552587.496776593
7	Flavien Campou	126212.611562883	477960.065899793
8	Régina	374698.048396995	476928.730177523
9	Camopi	352395.046055802	350764.216480451
10	Alikéné	322929.978284572	360135.527835386
11	Mine Boulanger	343189.582357019	504590.723307373
12	Carbet Légion crique Sikini	359046.553479	361996.097871
13	Cogneau - 23. Lot. Aquavilla	354091.053713437	538837.801956188
14	41. rue des Ixoras - Lot. Cogneau Larivot	351322.960529913	541116.10235073
15	Attila Cabassou	354987.967015231	540678.87835953
16	La Chaumière	347961.540051952	539402.864452834
17	1228. Ch. de La Chaumière	349626.610158034	540136.658859129
18	Saint-Georges	411052.118458134	429833.606330522
19	Quartier Espérance - Saint-Georges	411113.960068634	430709.665437749
20	Village Martin - Saint-Georges	411519.804866871	432628.481492214
21	Boulangerie – Saül	254644.262333224	400558.641871421
22	Chemin Mogès	352032.329517247	529456.717653458
23	Dorlin	216742.863044	415732.862195
24	Maripasoula	163199.637191727	403271.422431979
25	Repentir	234327.380239	427769.677349

Number of sites	Locality	Longitude	Latitude
26	Stoupan	352127.057631364	527017.037636427
27	Camp Pararé – Nouragues	311267.546406115	446010.858172732
28	Village Blondin - Saint-Georges	409492.529993713	428392.815435875
29	Quartier Adimo - Saint-Georges	410525.149504269	430989.662548913
30	Camp Bernet/ Légion étrangère	410437.343147834	429860.701524782
31	La ferme de Lait-Quateur	332757.212761399	552628.404024885
32	Grand Usine	288898.531939378	372330.073251345
33	Dagobert	226797.35968	438864.200286
34	Cacao	336515.199831583	505801.969900718
35	Saut-Maripa – Camp militaire	401579.145107468	420354.782186144
36	Eau-Claire	213145.824694	398626.838699
37	Cacao	336956.199791048	504651.969979654
38	Cacao	337909.199733996	506008.969874801
39	Cacao	336576.199808207	503413.970071365
40	Cacao	336555.199807874	503184.970087929
41	Impasse de la rafinerie – Cogneau	354091.053713437	538837.801956188
42	Quartier Bambou	411277.041477492	430178.275876772
43	Quartier Maripa - Saint-Georges	410252.974346512	430188.463058511
44	Quartier Savane - Saint-Georges	411024.737954334	430940.054473617
45	Cacao	336826.199809151	505764.969900894
46	Cacao	337091.199787723	505441.969921949
47	Cacao	336186.199848992	505049.969957266
48	Quartier Onozo- Saint-Georges	411339.040785181	430641.896115415

S2. Creation of the relative sampling effort map

Capture data of *Culicidae* (74 capture sites) were used to estimate the sampling effort of *An. darlingi*. The collection methods were identical and the sampling bias for the family was assumed to be representative of that for the focal species.

The sampling bias was defined as the relative sampling effort in the environmental space. For a pixel *i*, it corresponds to the ratio of the number of sampled pixels over the total number of pixels, within the *environmental neighborhood* of *i*.

First, all pixels of the study area were represented in the environmental variable space. This was accomplished by performing a Factorial Analysis of Mixed Data (FAMD) (Pagès, 2004). This analysis jointly takes into account numerical and categorical variables and makes it possible to represent the pixels within an Euclidean, orthonormal space defined from the whole set of environmental variables.

The membership degree of a pixel *j* to the neighborhood of pixel *i*, denoted w_{ij}, was defined by a Gaussian-like membership function:

$$w_{ij} = 0.5^{\left(d_{ij}/D_{min}\right)^2} \qquad (1)$$

with d_{ij} the euclidean distance between *i* and *j* in the factorial space, and D_{min} the threshold distance over which *j* does not significantly belong to the environmental neighborhood of *i*, *i.e.* over which $w_{ij} < 0.5$. The membership degree w_{ij} has the following properties:

- $w_{ij} \in]0,1]$;

- $w_{ij} = 1$ if $d_{ij} = 0$;

- $w_{ij} < 0.5$ if $d_{ij} > D_{\min}$.

The parameter D_{min} was set from *a priori* knowledge of *An. darlingi* bio-ecology. As highly urbanized areas are not suitable for *An. darlingi* (see § I.3), we stated that a pixel associated with *An. darlingi* presence cannot belong to a highly urbanized pixel. Reciprocally, a pixel considered to be highly urbanized cannot belong to the environmental neighborhood of a pixel where *An. darlingi* was observed.

Consequently, given *P*, the set of pixels where the species was observed and *U*, the set of pixels belonging to highly urbanized areas, D_{min} was defined as follows:

$$D_{min} = min\left(d_{pu}\right)_{p\in P, u\in U} \qquad (2)$$

A pixel is considered to be highly urbanized if it belongs to the *LC* class *Urban* and if its eight neighboring pixels present an average urbanization percentage (*PER_URB_NEIGH*) higher than or equal to 50%.

The concepts of environmental space and neighborhood, as well as the key method parameters are schematically represented in Figure S1.

Given *X*, the set of pixels of the study area, and $c = \{c_i\}_{i\in X}$, a vector such that $c_i = 1$ if *i* is sampled and $c_i = 0$ otherwise, the relative sampling effort at pixel *i*, z_i, is then defined as:

$$z_i = \sum_{j\in X} w_{ij} \cdot c \,/ \sum_{j\in X} w_{ij} \qquad (3)$$

The relative sampling effort was computed for each pixel of the study area. The resulting map was used to bias the random selection of background points. Consequently, for a given pixel, the greater the relative sampling effort, the higher the chance of selecting the pixel as a background point.

Reference

Pagès, J. 2014. Multiple Factor Analysis by Example Using R. Chapman & Hall, CRC Press.

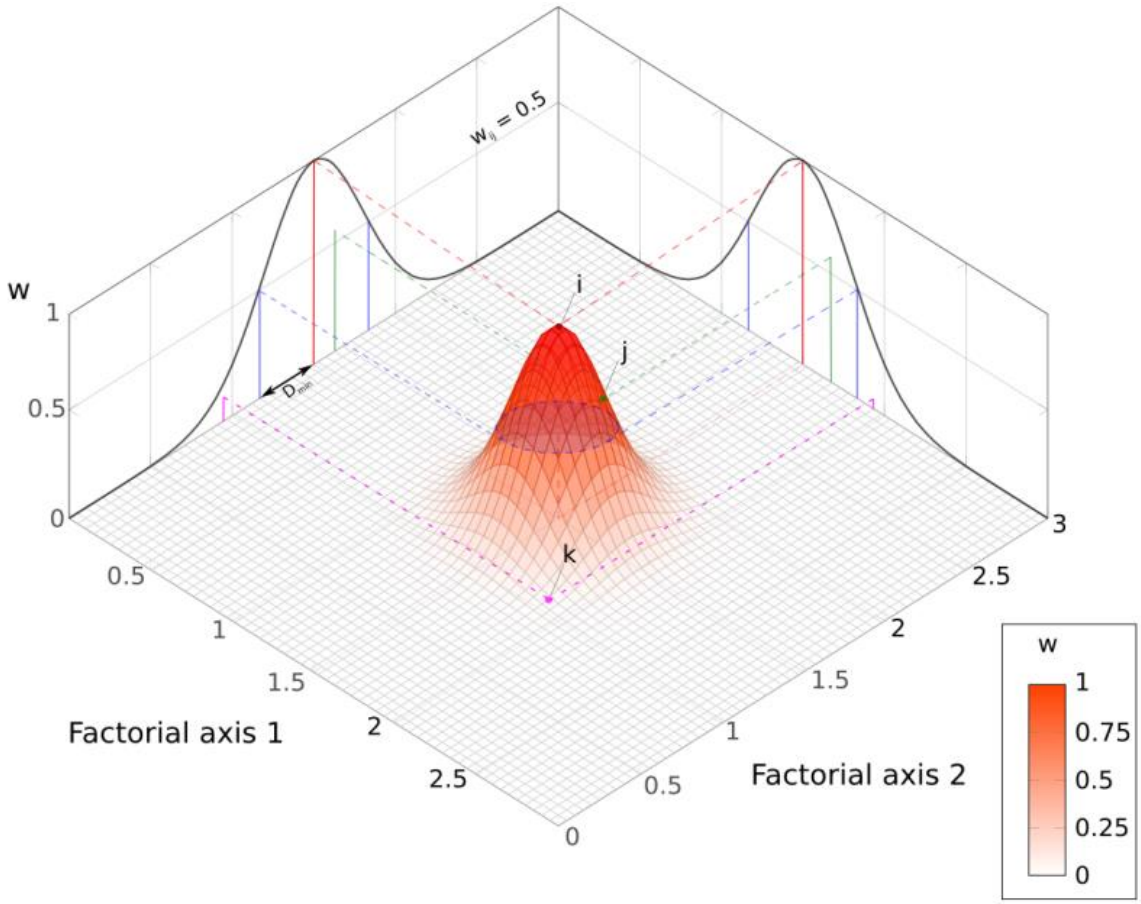

Figure S1. Neighborhood of a pixel *i* in the environmental space represented by the first and second factorial axes. The environmental neighborhood of point *i* is represented by the Gaussian function. The blue lines define the limit of the neighborhood of *i*. Only point *j* is situated above these lines. Thus *j* is in the neighborhood of *i* in the first factorial plane.