

METHODOLOGY

Open Access



# Exploring fine-scale urban landscapes using satellite data to predict the distribution of *Aedes* mosquito breeding sites

Claire Teillet<sup>1\*</sup>, Rodolphe Devillers<sup>1</sup>, Annelise Tran<sup>2,3</sup>, Thibault Catry<sup>1</sup>, Renaud Marti<sup>3</sup>, Nadine Dessay<sup>1</sup>, Joseph Rwagitinywa<sup>4</sup>, Johana Restrepo<sup>4</sup> and Emmanuel Roux<sup>1,5\*</sup>

## Abstract

**Background** The spread of mosquito-transmitted diseases such as dengue is a major public health issue worldwide. The *Aedes aegypti* mosquito, a primary vector for dengue, thrives in urban environments and breeds mainly in artificial or natural water containers. While the relationship between urban landscapes and potential breeding sites remains poorly understood, such a knowledge could help mitigate the risks associated with these diseases. This study aimed to analyze the relationships between urban landscape characteristics and potential breeding site abundance and type in cities of French Guiana (South America), and to evaluate the potential of such variables to be used in predictive models.

**Methods** We use Multifactorial Analysis to explore the relationship between urban landscape characteristics derived from very high resolution satellite imagery, and potential breeding sites recorded from *in-situ* surveys. We then applied Random Forest models with different sets of urban variables to predict the number of potential breeding sites where entomological data are not available.

**Results** Landscape analyses applied to satellite images showed that urban types can be clearly identified using texture indices. The Multiple Factor Analysis helped identify variables related to the distribution of potential breeding sites, such as buildings class area, landscape shape index, building number, and the first component of texture indices. Models predicting the number of potential breeding sites using the entire dataset provided an  $R^2$  of 0.90, possibly influenced by overfitting, but allowing the prediction over all the study sites. Predictions of potential breeding sites varied highly depending on their type, with better results on breeding sites types commonly found in urban landscapes, such as containers of less than 200 L, large volumes and barrels. The study also outlined the limitation offered by the entomological data, whose sampling was not specifically designed for this study. Model outputs could be used as input to a mosquito dynamics model when no accurate field data are available.

\*Correspondence:

Claire Teillet  
claire.teillet@ird.fr  
Emmanuel Roux  
emmanuel.roux@ird.fr

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion** This study offers a first use of routinely collected data on potential breeding sites in a research study. It highlights the potential benefits of including satellite-based characterizations of the urban environment to improve vector control strategies.

**Keywords** Remote sensing, *Aedes aegypti*, Urban areas, Prediction, Landscape ecology, Vector control, Arbovirus diseases

## Introduction

According to the World Health Organization, mosquito-transmitted diseases, such as dengue, chikungunya, and Zika, have intensified during the past decades and are responsible for over 700,000 deaths each year. One of its main vectors, *Aedes aegypti*, is strongly adaptable to urban environments and thrives in urban settlements [1], favoring human dwellings and feeding almost exclusively on human blood [2]. Adult females get infected by feeding on the blood from an infected human host and can in turn transmit the disease to another host after pathogen proliferation in the vector (extrinsic incubation period) [3]. After a blood meal, females lay eggs in containers of stagnant water and larvae develop in the water before becoming adult mosquitoes. *Ae. aegypti* breeds mainly in small artificial or natural water containers, such as water storage containers, used tires, plastic containers, clogged gutters, and ornamental plants [4]. The configuration of the urban landscape is known to impact the availability and distribution of such containers, and therefore of potential *Ae. aegypti* breeding sites [5–7].

If the risk of introducing *Ae. albopictus* remains high in French Guiana, *Ae. aegypti* is currently the sole arboviruses vector for dengue, chikungunya, and zika viruses [8]. *Ae. aegypti* is present in nearly all inhabited regions, including smaller human settlements, and even in some wild areas [8]. The abundance of adult mosquitoes is strongly associated with building densities and is therefore higher in densely populated urban areas, in particular but not exclusively in the main cities, such as Saint-Laurent du Maroni, Kourou, Cayenne, Remire Montjoly, and Matoury [9]. In French Guiana, vector control mainly relies on indoor and outdoor insecticide spraying where human cases of any *Aedes*-borne disease have been identified, to control adult mosquitoes, and on routine mechanical or chemical elimination of larval breeding sites in urban areas (see paragraph 2.2 *Entomological Data* and [10]). Since the 1940s, the use of various insecticides for vector control in French Guiana has led to the development of resistance in *Ae. aegypti* populations, resulting in reduced efficacy in the territory [10, 11]. Vector control strategies are therefore exploring alternative methods that could target adult mosquitoes, acknowledging the need for more precise larval control in time and space [11]. Several epidemics linked to arboviruses have occurred in recent decades (i.e., 2006, 2009–2010, 2013, and 2020–2021). Since April 2023,

French Guiana has been experiencing new dengue fever epidemics, mainly due to serotypes DEN-2 and DEN-3 [12]. Unlike other Latin American countries (Brazil in particular), there are not many studies on dengue fever or its associated vector in French Guiana [13].

Reducing the risks associated with *Aedes*-borne diseases can be achieved by a better understanding of the influence of urban environment factors on the distribution of *Ae. aegypti* breeding sites. If such a knowledge can be achieved by *in-situ* entomological data collection, especially data on potential breeding sites, such a task can be very time-consuming and challenging, given the complexities of obtaining exhaustive data in complex urban environments and is not done in many places.

Remote sensing proved its ability to help characterize urban environmental and climatic variables associated with the mosquito life cycle, including breeding sites [14, 15]. Such an approach can hence provide an effective method to characterize urban landscapes, offering cost-effective and reproducible approaches that can be used at different spatial and temporal resolutions [14, 16]. A wide range of information contributing to the urban landscape characterization can be derived from remote sensing data, such as: land use or land cover [17, 18] and the landscape metrics that can be calculated from it [19, 20], spectral, thermal and texture indices [21–24], and elevation models [25].

Most studies using remote sensing data for modeling vector populations used entomological data associated with presence and/or abundance of adult mosquitoes (BG-Sentinel trap) or larvae (Ovitrap) [26–28]. Another approach consists of using remote sensing data to predict the distribution of potential mosquito breeding sites (i.e., recipients with larvae, with water but without larvae, or without water). Studies employing such approaches are rare, primarily due to the time and resources required for identifying and counting domestic and peridomestic containers by vector control efforts [29–33].

In these studies, input data and variables are usually satellite-derived land surface temperature, land use or land cover (LULC) classifications with few broad classes, in addition to radiometric indices like the normalized difference vegetation index (NDVI). Those are commonly used as predictors in models predicting the presence of *Aedes* breeding sites. However, urban features at fine spatial resolution have rarely been analyzed to inform those models. Arduino et al. [33] studied the impact of

micro-environments on positive breeding sites (positive to larvae) using very high resolution Ikonos satellite images, without considering the urban configuration in detail. Different approaches used Haralick texture indices [28] and landscape metrics (Shannon diversity index [34]) for urban space zoning, but not as direct predictors of the number and distribution of breeding sites. As far as we know, urban landscape characteristics at a fine spatial scale have rarely been considered as input variables in mosquito-borne disease models [35], especially when modeling potential breeding sites of *Aedes* mosquitoes.

Various methodological approaches have been used for modeling the distribution of *Aedes* mosquitoes and their breeding sites, such as linear regression [29], species distribution models [30] and more recently Boosted Regression Tree [31]. Machine learning models, such as Random Forest (RF), have already shown good results to study the relationship between ovitraps or adult mosquitoes data and environmental factors [26, 36–38]. However, RF models has never been applied to routinely acquired potential breeding sites data to investigate their complex relationships with urban landscapes.

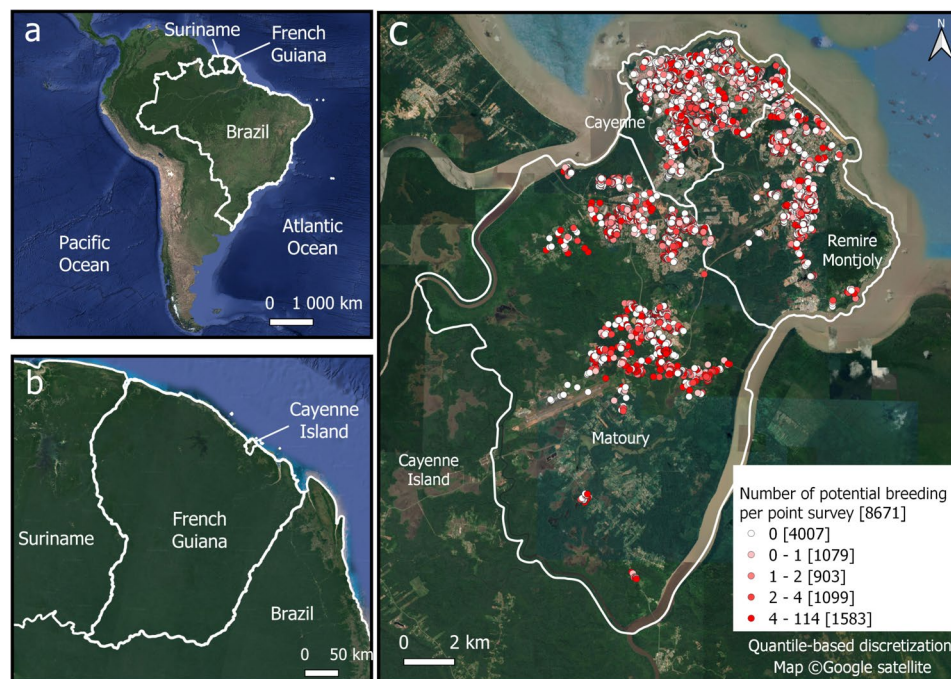
Hence, using descriptive landscape variables such as texture, landscape metrics, buildings and vegetation heights, this paper proposes an original approach that aims to enhance the identification of relationships between urban landscapes and potential *Ae. aegypti* breeding sites at a fine spatial resolution. Our study analyzes the relationships between urban landscape

variables, especially those derived from very high resolution (VHR) satellite imagery, and *in-situ* data on potential breeding sites in cities of French Guiana (South America), while evaluating the capability of these variables to be used in predictive models.

## Materials and methods

### Geographical context

The study area is the Cayenne Island, the largest urban region of French Guiana, a French overseas territory of 83,800 km<sup>2</sup> and 285,000 inhabitants (2020) located in the northeast of South America, sharing borders with Brazil and Suriname (Fig. 1a, b). ‘Cayenne Island’ is the name given to the peninsula surrounded by the estuaries of the Cayenne River to its west, the Mahury River to its east, and the Atlantic Ocean on its northeast. The peninsula is composed of the cities of Cayenne, Matoury, and Remire-Montjoly covering only 0.25% (i.e. 206.91 km<sup>2</sup>) of the French Guiana territory but hosting nearly half of the French Guiana population (126,223 inhabitants in 2020). French Guiana is covered by 95% of forest, and has an equatorial climate, characterized by year-round high temperatures and high humidity with an annual mean temperature of 26 °C, an annual mean humidity of 80–90% and between 2000 and 4000 mm of annual precipitation. The year is divided into four seasons: a long rainy season from early April to mid-July, a lengthy dry season from mid-July to mid-November, a short rainy season from mid-November to mid-February, and a brief



**Fig. 1** Study Site: (a) French Guiana located in South America; (b) French Guiana, neighboring countries and Cayenne Island; (c) Number of potential breeding sites by prospected location during 2022. The three municipalities of the Cayenne Island (Cayenne, Rémire-Montjoly and Matoury) are delimited with solid white lines

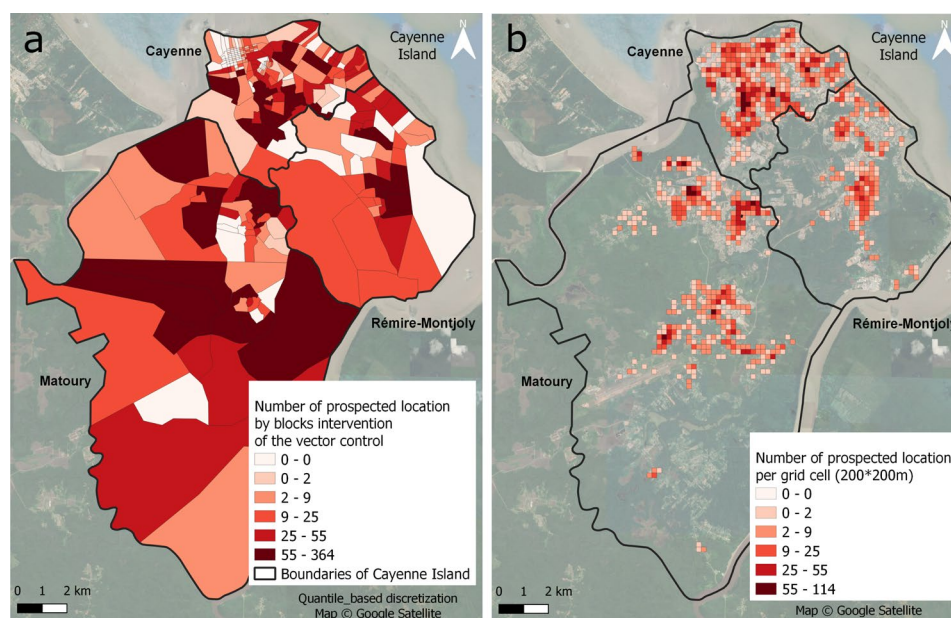
dry season from mid-February to early April called “little summer of March.” French Guiana regularly faces dengue outbreaks [8], with the latest one that lasted almost a year and a half, from January 2020 to June 2021. During this last epidemic event, nearly 10,900 clinically evocative cases seen in consultations were estimated, and 6195 probable or confirmed cases were identified.

### Entomological data

Various efforts are coordinated by government agencies to prevent, monitor and address dengue outbreaks, including regular field surveys aiming to identify and characterize mosquito breeding sites. A dataset of *Ae. aegypti* breeding sites in French Guiana for 2022 was obtained from the ‘*Direction de la Démoustication et des Actions Sanitaires*’ (Direction of mosquito control and health actions) of the ‘*Collectivité Territoriale de Guyane*’ (regional administration - CTG). Each month, randomly selected addresses are inspected for the presence or absence of *Ae. aegypti* larvae. Containers that store water with larvae are recorded as being positive breeding sites, while negative breeding sites are containers with water but no larvae, or without water during the time of prospection but that could have water in the future. This survey is carried out throughout French Guiana, with the territory divided into geographical sectors, subdivided into ‘blocks’ used to plan the monthly field surveys. Depending on the size of the blocks, the number of locations to be visited each month is determined. Only locations visited by agents with geographical coordinates located in Cayenne Island were included in the final dataset. If a container is positive for larvae during a survey,

mechanical control or biological treatment are applied and the control action is registered. To characterize the routine visits carried out by vector-control agents, a representation of the number of locations prospected per block is provided in Fig. 3a.

The dataset included 56,985 records acquired from January 4, 2022 to December 30, 2022, 18,576 being located on the island of Cayenne, with 8775 locations actually visited by the agents. Indeed, 52.76% of the sites located on the Cayenne Island could not be inspected due to the absence of inhabitants or a refusal from the owners to conduct the survey. Then, 104 records, corresponding to tire sales zones, garages, and cemeteries, were excluded from the analyses as these places: (i) are known to structurally produce a very high quantity of potential breeding sites and undergo higher sampling efforts, introducing biases in the models; (ii) are not associated with specific urban landscape types, introducing random noise in the modeling step. The effects of such biases and random noise have been revealed by preliminary models that were tested on the entire dataset. As a result, 8,671 records were considered in the study. The number of potential breeding sites per location was obtained by adding positive and negative breeding sites (Fig. 1c). Breeding sites are classified by the CTG into one of 19 types: containers over 200 L, containers under 200 L, small waste, large waste, tires, barrels, building materials, large volumes, manholes, gutters, troughs, green plants, other plants, wells, pits, watercraft, protection, canals, and others. These types were then grouped into larger categories based on their similarity and type of management based on experts’ knowledge (Table 1).



**Fig. 3** Entomological data sampling effort over (a) blocks, and (b) grid cells for 2022 on Cayenne Island

**Table 1** Types and categories of breeding sites

Categories	Types	Comments	Management
Containers	containers over 200 L	containers with a volume of more than 200 L	emptying / covering / insecticide treatment
	containers under 200 L	containers with a volume of less than 200 L	
	barrels	containers with a capacity smaller than barrels barrels of 200 L	
Plants	pots plants		regular maintenance of pots and garden
	other plants	ground plants	
Waste	building materials		storage/ removal maintenance
	tires		
	watercraft		
	small waste	can be removed manually	
	big waste	can not be removed manually	
Water Management Infrastructure	protection	tarpaulin, plastic, fabric... used to cover an object	maintenance of water management infrastructures
	gutters		
	canals		
	manholes		
Large Volume	troughs		emptying / covering
	pit		
	large volumes	pool, large water basin	
Others	others	unclassified types	

Pre-processing of entomological data was carried out with the QGIS software (v. 3.16) [39].

### Satellite-derived and geospatial data

#### Satellite imagery

Pléiades satellite images of the study area were acquired on July 20, 2022. Pléiades images consist of a panchromatic band (0.5 m resampled spatial resolution) and 4 spectral bands (blue, green, red, and near-infrared bands at 2 m spatial resolution). The very high spatial resolution of Pléiades images allows for the observation of fine details within the defined area, including urban vegetation and building arrangement.

#### Textural indices

In this paper, image texture was used to help characterize the arrangement of the urban elements and therefore different urban types (e.g., dense housing, commercial areas, urban parks) [40]. The unsupervised FOTOTEX algorithm [41] has been applied to the Pléiades imagery

panchromatic band with a 201-pixel (100.5 m) window size using a block method (each window is processed sequentially block by block in the image, see [42] for more details), generating textural indices indicative of different urban patterns, depending on the distribution and configuration of urban elements. The window size was considered large enough to include several repetitions of patterns of interest and therefore, characterize different types of urban areas. The FOTOTEX algorithm [41] consists in applying a Fourier Transform (FT) to each window, then applying a principal component analysis (PCA), with the windows as statistical individuals, and the spatial frequencies from FT as variables. The first three components of the PCA were used as textural indices. This algorithm has been used in vegetation contexts [42–46], and has recently been applied to identify urban areas [41]. However, this algorithm has never been used in combination with breeding site data and our hypothesis is that such an approach could enhance our knowledge of the links between urban landscape and breeding site distribution, and therefore facilitate the remote sensing-based prediction of breeding site distribution. In this study, the FOTOTEX algorithm was performed using the python package “*fototex 1.5.9*” [41].

#### Spectral indices

Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) were derived from the Pléiades images. The NDVI is calculated using the equation  $NDVI = (NIR - RED) / (NIR + RED)$  where RED and NIR correspond to the reflectance measure in Bands 3 and 4, respectively [47]. The NDVI is widely used for identifying vegetation in remote sensing [48], including green spaces in urban environments such as parks and gardens. These vegetated areas can create favorable conditions for *Ae. aegypti* breeding by retaining rainwater or offering shaded sites for rest [49, 50]. The NDWI is calculated using the equation  $NDWI = (GREEN - NIR) / (GREEN + NIR)$  where GREEN corresponds to the reflectance measure in Band 2 [51, 52]. The NDWI is effective in detecting water bodies in urban environments [53], including stagnant water in basins, pools, and canals, providing insights into potential breeding sites for *Ae. aegypti* [30].

#### Vegetation and buildings

Urban vegetation presence was extracted using the NDVI calculated on the Pléiades image, applying a threshold of 0.20 to keep only the vegetation [54]. The layer related to individual buildings was provided by the French national geographic agency IGN (BD TOPO® 2022), based on cadastre and satellite image photo-interpretation. A masking technique was applied to determine the heights of vegetation and buildings and thus provide information

on the vertical structure of urban landscape elements. This involved overlaying the vegetation and building layers with a 2015 LIDAR-derived digital elevation model (1-meter spatial resolution) obtained from CTG. The process filtered out only the elevation information for areas where vegetation or buildings are present.

### Extraction of analysis variables

#### Spatial units of analysis

The following analyses were carried out over a 200 m regular grid cell to ensure sufficient spatial resolution for capturing the diverse characteristics of the urban landscape. QGIS was used to create a regular grid and the R Stats software v. 4.2.1 [55] was used to generate breeding sites variables and urban landscape variables for each cell of the grid. This choice will be discussed in the Discussion section.

#### Breeding site variables

For each grid cell, the absolute number of potential breeding sites (regardless of breeding site types and categories) was calculated by summing positive and negative sites. This was also done by considering separately the different types and categories of breeding sites. Using an analysis grid helps reduce the variations in sampling effort associated with the dataset [56]. A normalized number of potential breeding sites was also calculated by dividing the absolute number of potential breeding sites by the number of prospected locations per cell (Additional file 1). Only modeling performance using the absolute number of potential breeding sites is presented in the Results section due to the similarity of the results.

#### Urban landscape variables

Summary statistics of urban landscape variables were computed within each grid cell: Mean NDVI, Mean

NDWI, Mean Vegetation Heights, Mean Buildings Heights, Mean Textural component 1, 2, and 3. The building density (number of buildings per grid cell), and the average building size per cell (total building surface divided by the number of buildings in each cell) were computed using the building layer.

Landscape metrics have been calculated over vegetation and building layers within each grid cell to capture information about urban landscape composition and configuration. This helped quantify the spatial organization of patterns within a defined geographic area. Specific landscape metrics (see Table 2) have been selected based on their potential to relate to the urban landscape and breeding sites of *Aedes* mosquitoes [7, 57]. Thus, by analyzing these several landscape metrics describing the configuration of urban landscape, we could identify links between description of urban landscapes and the spatial repartition of potential breeding sites. For example, more fragmented vegetation could enhance the number of potential breeding sites. In this study, landscape metrics were calculated using the “*landscapemetrics*” R package [58]. Table 3 lists the satellite-derived and geospatial data, their sources, the methods used to create them and the urban landscapes variables derived from them in this study.

#### Multiple factor analysis

Multiple Factor Analysis (MFA), a multivariate statistical technique, was performed to explore and analyze relationships among multiple quantitative and/or qualitative datasets [59]. The analysis consists of grouping variables, depending on study objectives, and in balancing the contribution of the groups in the analysis – i.e. in compensating the contribution differences due to variable contribution themselves and/or to the number of variables per group. This approach aims to favor

**Table 2** Landscape metrics used in this study [58]

Landscape metrics	Information type	Description	Units	Interpretation
Total class area (CA)	Area and Edge metric	Dominance of one class over the defined area	Hectare	Absolute measure of composition
Largest patch index (LPI)	Area and Edge metric	Percentage of the landscape covered by the corresponding largest patch of each class	Percentage	Measure of dominance
Total class area percentage (CPLAND)	Core Area metric	Percentage of the core area of class to the total landscape	Percentage	Relative measure of composition
Landscape shape index (LSI)	Aggregation metric	Compactness of the class based on the length of the class within the spatial unit	No units	Patch distribution (compact or dispersed)
Patch density (PD)	Aggregation metric	Number of patches of a class within the spatial unit	Number for 100 hectares	Fragmentation
Mean shape (SHAPE_MN)	Shape metric	Ratio between the actual perimeter and the hypothetical minimum perimeter of the patch	No units	Complexity of patch shapes
Mean of Contiguity (CONTIG_MN)	Shape metric	Assigning an adjacency value to each pixel in a patch according to a nine-cell focal filter matrix. Larger and more connections between patch result in larger contiguity index values	No units	Connectivity

**Table 3** Geospatial data, source information, and variables extracted for urban landscape characterization

Satellite-Derived and Geospatial Data	Information source	Method for layer creations	Remarks	Variables extracted by grid cell	Category of explanatory variable
Pléiades Multispectral	Multispectral	Equations		Mean NDVI Mean NDWI	Spectral Indices
Pléiades Panchromatic	Panchromatic	Algorithm FOTOTEX		Texture PC1 Texture PC2 Texture PC3	Textural Indices
Vegetation	NDVI threshold	Landscape metrics		Class area Percentage of class Landscape Shape Index Landscape Patch Index Patch Density Mean Shape Index	Landscape metrics over vegetation
Buildings	BDTOPO	Masking	With MNH from CTG	Mean Height Vegetation Mean Height Buildings	Heights
		Landscape metrics		Class area Percentage of class Landscape Shape Index Landscape Patch Index Patch Density Mean Shape Index	Landscape metrics over buildings
		Counting buildings over each grid cell Total area of buildings divided by the number of buildings for each grid cell		Number of buildings Mean size of buildings	Buildings informations

the identification of the links between variables, within groups (like a classical principal component analysis (PCA) – if variables are quantitative, as it is the case here) but also between the different groups. As for PCA, the factorial axes represent the main orthogonal directions along which the data vary the most and the variable contributions indicate to which extent each variable contributes to the factorial axes (the sum of the variable contributions for a given axis is equal to 1), while the squared cosines ( $\cos^2$ ) measure to what extent the variables are well represented on each axis (the sum of the  $\cos^2$  of a given variable over all factorial axes is equal to 1). This analysis aims to identify variables characterizing urban landscapes that are the most discriminant regarding variables characterizing *Ae. aegypti* breeding sites, and to select some of them for modeling purposes. This approach was applied to two groups of variables: one grouping the 24 breeding site variables (i.e., number of potential breeding sites, number of potential breeding sites of each type and regrouped category; Table 1) and another one grouping the 24 urban landscape variables (i.e., spectral indices, textural indices, landscape metrics over vegetation and landscape metrics over buildings, heights, and buildings information) (Table 3). In that way, we can identify if some types of breeding sites were

associated with specific urban landscape variables. The selection of urban landscape variables for modeling relies on identifying variables with contribution values above the average contribution and squared cosine ( $\cos^2$ ) values above 0.1, considering the first three factorial axes. Then a selection of one variable per group of explanatory variables was performed to do the prediction. In this study, MFA was conducted using R “*FactorMineR*” package [60].

### Modeling and analysis

Non-linear models were used to establish the relationships between different types of potential breeding sites (dependent variable) and urban landscape variables (independent variables). We applied Random Forest (RF), which combine multiple decision trees for regression [61]. RF can model complex relationships between explanatory variables and the target variable with great robustness, allowing to deal with non-linearly separable data (more precisely, it is a “piecewise linear model”). RF is a non-parametric model that imposes no limitations and makes no assumption about the distribution of the data, allowing greater flexibility [62]. It showed its efficiency for regression and prediction applied to spatio-temporal distribution or abundance of *Ae. aegypti* (see *Introduction* and [37, 38]). However, its application to

potential breeding sites has not been explored yet. Different RF models were built, considering different explanatory variables grouping: variable categories indicated in Table 3 (texture indices, spectral indices, heights, landscape metrics related to vegetation, landscape metrics related to buildings, building information); variables selected according to MFA (referred to as MFA variables hereafter); and the complete set of variables. A range of response variables were tested, which included all potential breeding sites, as well as all types and categories of potential breeding sites (cf. Table 1). Models were applied to grid cells that include a minimum of five home visits, resulting in a total of 417 grid cells (51% of the grid cells with at least one record) analyzed. This threshold of five home visits was chosen to ensure that the analyzed grid cells had a sufficient sample size for reliable and robust statistical analysis. In this study, RF models were computed using R “*randomForest*” package [63].

To assess the RF models’ performance, a cross-validation procedure was applied, by performing a random selection of 70% of the 417 grid cells to build the training set and keeping 30% of the cells to test the model. Such a procedure was repeated 100 times, with replacement of the cells for the random selection (bootstrapping). For relatively rare breeding site types that are only present in a few cells of the study area, such a procedure can lead to building models with too few cells. Consequently, we made sure that each individual model was based on enough cells with a presence of the specific type. This was done by removing cross-validation iterations that used, for training, less than 70% of the cells where the considered breeding site type or category was observed.

The Normalized Root Mean Square Error (NRMSE) was computed to evaluate the prediction quality while reducing the impact of extreme values and considering the overall distribution of values in the data. As the response variables did not follow a normal distribution, the NRMSE based on a min-max range was used to incorporate a measure of spread of the data. The NRMSE provides a relative measure of the model’s effectiveness, where a low NRMSE value indicates a more accurate fit of the model. The coefficient of determination ( $R^2$ ), was used to assess the performance of regression models over observations and results of predictions made on each validation sample [64]. Figure 2 summarizes the methodological framework of this study.

## Results

### Descriptive analysis of entomological data

#### Sampling effort

The number of locations visited by vector-control agents, when aggregated by survey block, ranged from 0 to 364, with an average of 30 homes (Fig. 3a). When aggregated by grid cells (Fig. 3b) excluding cells with no visited

houses, sampling effort ranges from 1 to 114 houses, with an average of 10 homes.

#### Potential breeding sites

Analyses revealed a strongly positively skewed distribution of potential breeding site numbers, with a mean and a median of 2.68 and 1 sites per house, respectively. On Cayenne Island, 5 of the 19 breeding site types, account for 78.4% of all breeding sites: green plants (36.4%), containers under 200 L (18.0%), small waste (10.0%), tires (7.4%), and barrels (6.6%) (Fig. 4a). Both the median and the third quartile of the logarithmic number of potential breeding sites are very low, ranging from 0 to 1.4 and 0 to 1.95, respectively, regardless of the type. Building materials or green plants present the highest maximum values of potential breeding sites (Fig. 4b). The number of potential breeding sites per grid cell also display a strongly positively skewed distribution (Fig. 5a), with a mean absolute value of 28, a maximum of 372, and a median of 13. The spatial representation of the number of potential breeding sites per grid cell (Fig. 5b) shows higher values over residential areas in Cayenne, Matoury, and Remire Montjoly in contrast with lower values in peri-urban areas of these towns.

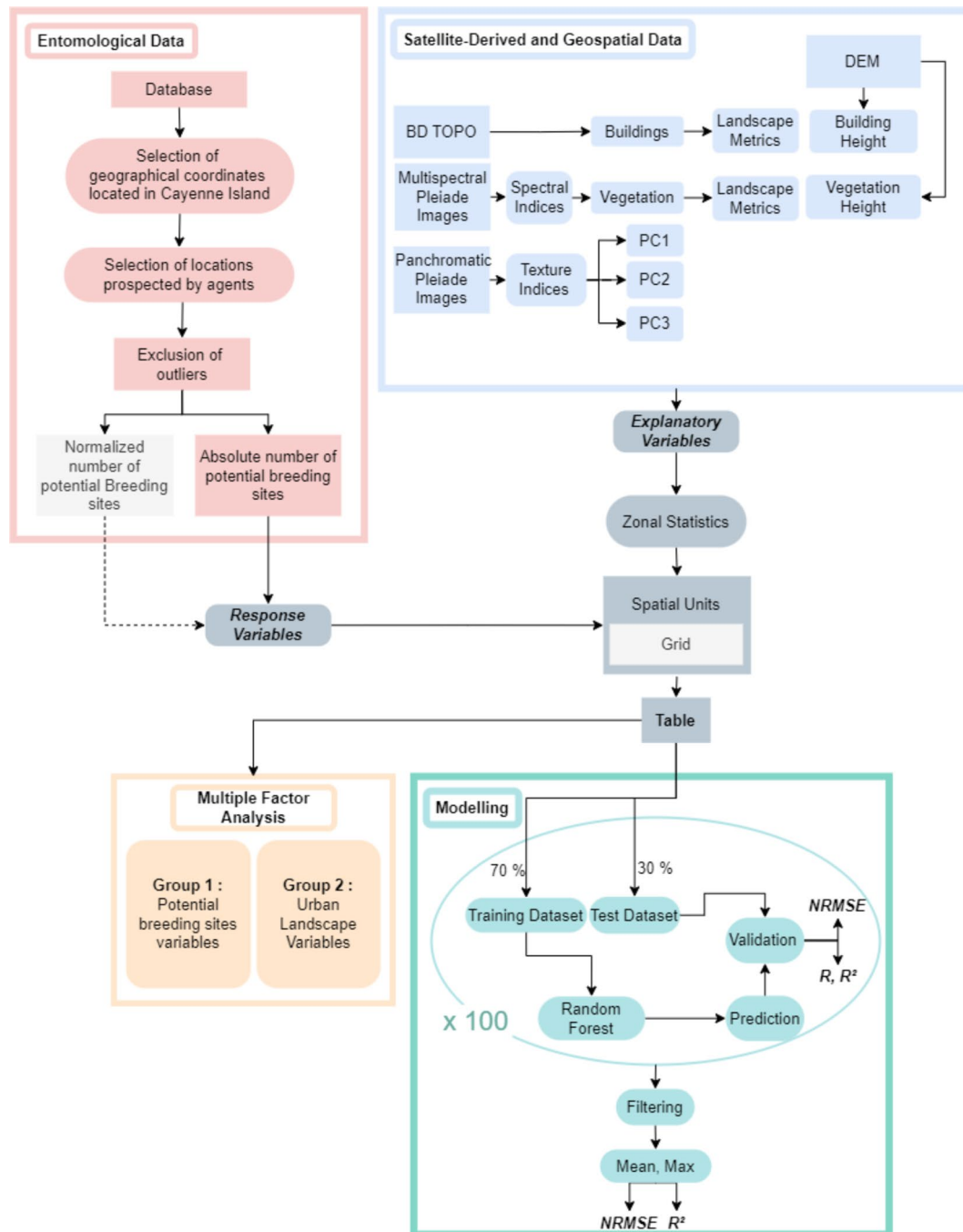
#### Extraction of analysis variables

All extracted variables are represented geographically in Additional files 2, 3, 5 and their histogram is provided in Additional file 7.

#### Textural indices

FOTOTEX analyses resulted in three principal component analysis (PCA) axes, accounting for 87.03% of the total variance of the image (i.e., 69%, 18% and 0.03% respectively). Frequencies are concentrated along the value 1 of the first axis (Fig. 6a). The second axis displays frequencies ranging from intermediate to high values, even reaching very high frequencies (996 cycles.km<sup>-1</sup>). The relationship between specific urban landscapes and their position along the two PCA axes is illustrated in Fig. 6b. By analyzing the windows with the most extreme values along the bisector in each angular sector (corresponding to a section of a given angle of the factorial plan), we observe that windows with urban patterns are present close to values of 1 on the Axis 1 and along the Axis 2. FOTOTEX RGB composite (Fig. 7a, b) allows for a rapid assessment and overview of the spatial organization of urban landscapes based on Pléiades imagery. Different types of urban areas are well-discriminated, especially with dense or very dense urban areas in pink/red, areas gathering houses with gardens in purple/blue, and peri-urban areas with isolated houses dominated by the presence of vegetation in green (Fig. 7b).





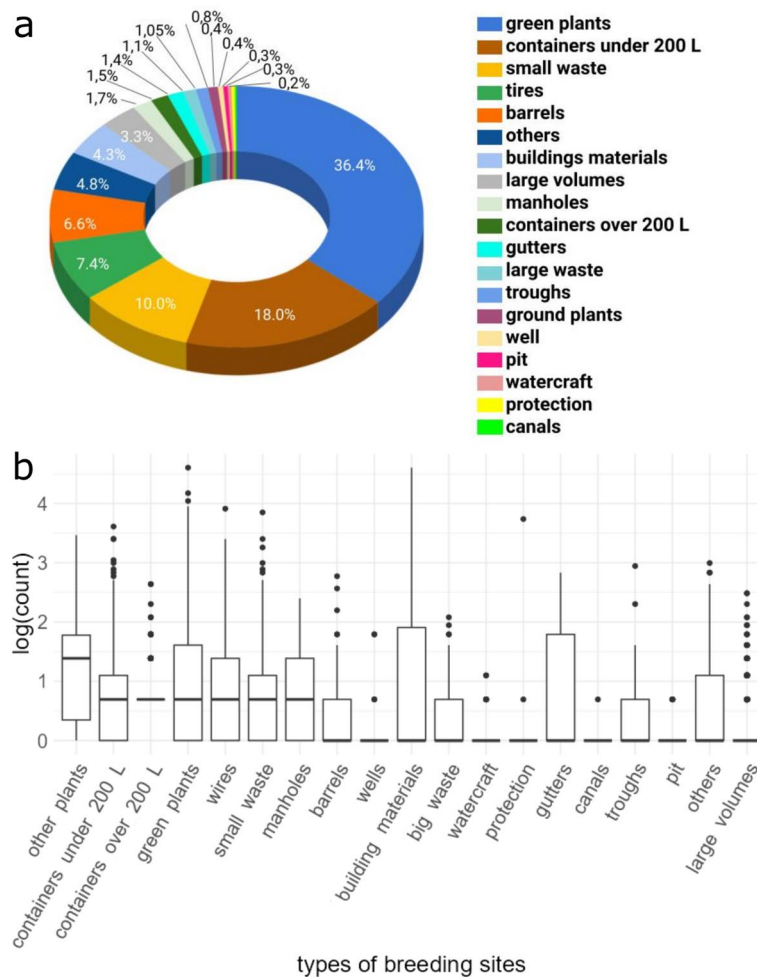
**Fig. 2** Methodological framework of this study

**Landscape metrics**

**Vegetation**

When applied to vegetation, the total class area, percentage of the class area, and the largest patch index metrics all collectively show an increasing surface of vegetation from the center of Cayenne towards the neighboring towns of Remire-Montjoly and Matoury, closer to forested areas (Fig. 8a). Landscape shape index suggests that Cayenne’s vegetation has a higher connectivity between

patches than the south of Matoury (Fig. 8b). Patch density indicates higher fragmentation, especially in Cayenne’s center, north of Remire-Monjoly, and in four clusters in Matoury (Fig. 8c). Mean shape shows that Remire Montjoly and Matoury have on average higher values per grid, indicating a higher diversity of patch shapes than in Cayenne.



**Fig. 4** (a) Proportion of breeding sites per type in Cayenne Island; (b) Logarithmic count of potential breeding sites, categorized by types and ordered by median value

**Buildings**

Class area, and percentage of class show a greater presence of the building class in Cayenne and in the residential areas of Matoury than in Remire Montjoly and in the peri-urban areas bordering the forests in Matoury (Fig. 8d). Largest patch index indicates a pronounced presence of largest patches of buildings in Cayenne, north of Matoury and in peri-urban areas in South Matoury. Landscape shape index shows more compactness of buildings, in Cayenne and the residential areas of Matoury. Then patch density shows a greater fragmentation in Matoury and Remire Montjoly than in Cayenne. Mean shape indicates higher diversity of patch shapes, especially in Cayenne and residential areas of Matoury (Fig. 8e).

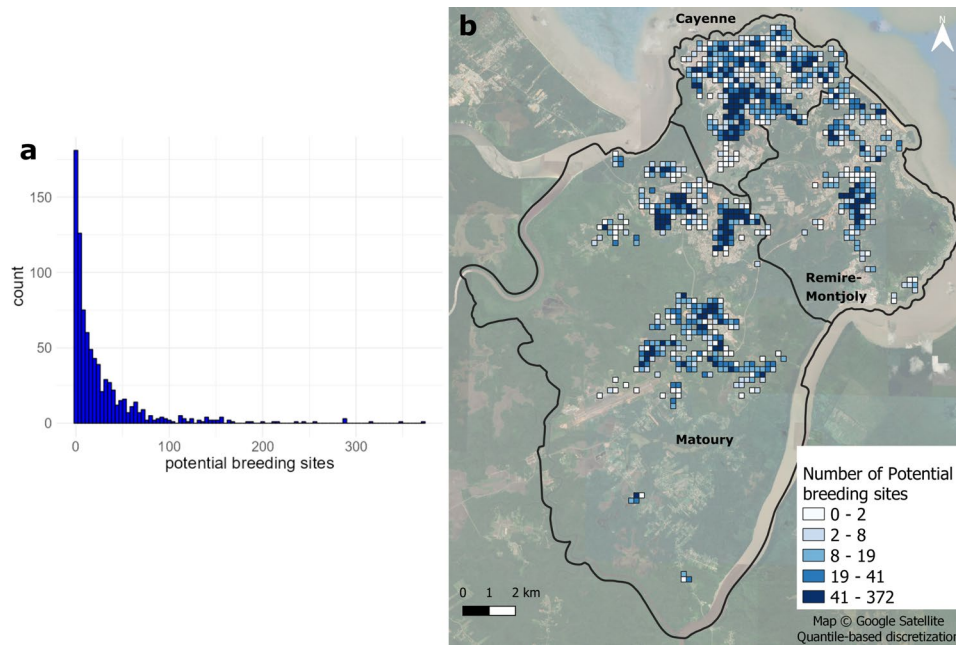
Finally, mean of contiguity shows strong heterogeneity of values across the landscape, indicating a challenge in identifying clear connectivity patterns among grid cells for vegetation and buildings (Fig. 8f).

**Multiple factor analysis**

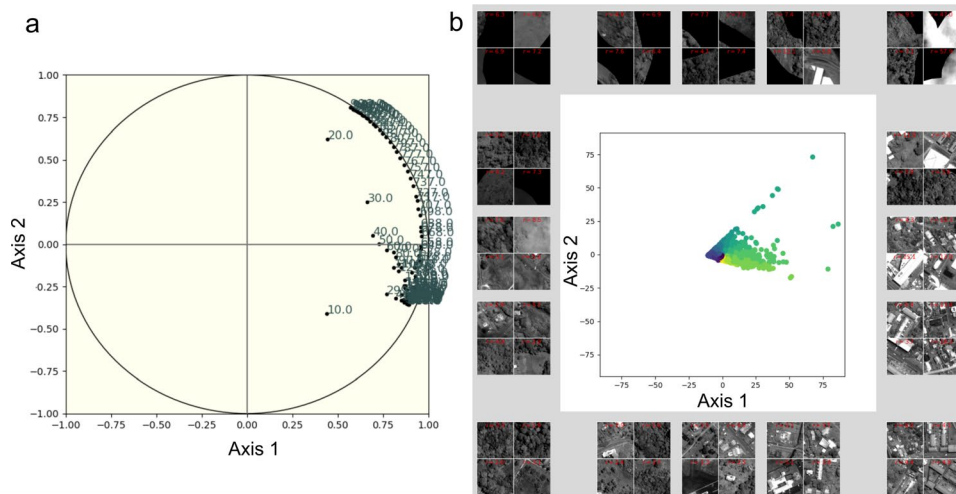
MFA three principal axes account for 40.8% of the total variance (21.9%, 12.4%, and 6.5% respectively for Axes 1, 2, and 3). Variables considered as significantly contributory to the first three axes (i.e. with contributions above the mean contribution of all variables for each axis) are presented in Fig. 9.

**Axis 1**

Landscape variables with the most influence on variance are total building class area and landscape shape index (vegetation and buildings), along with the number of buildings and the percentage of vegetation (Fig. 9a). NDVI, largest patch index, NDWI and the first component of texture indices (texture PC1) also significantly contribute to the first axis. In contrast, for breeding site variables, the potential breeding sites is the most contributory variable, followed by potential breeding sites of the categories containers and plants. These are followed



**Fig. 5** Number of potential breeding sites per grid cell as (a) a histogram and (b) geographically



**Fig. 6** (a) Representation of the spatial frequencies (expressed in cycles.km<sup>-1</sup>) on the 1st factorial plane defined by the first two factorial axes; (b) Analysis window projections on the first factorial plane defined by the first two factorial axes. Pléiades © CNES 2022 Distribution Airbus DS

by three specific types of breeding sites: containers under 200 L, green plants and small waste.

**Axis 2**

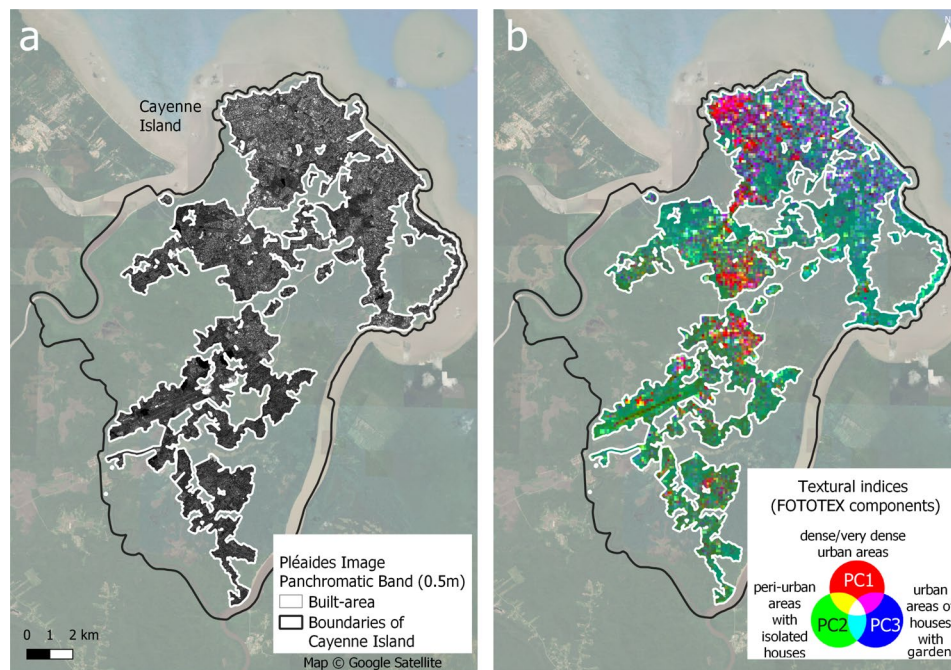
The potential number of breeding sites is also the most contributory variable, followed directly by NDVI (Fig. 9b). The next four contributory variables are breeding site-related variables (i.e. all potential breeding sites, containers, barrels, containers under 200 L). Variables associated to vegetation, buildings and the breeding sites “waste”, “small waste” and “plants” are less contributory. The first component of texture is in the 9th position for

urban landscape variables. Notably, vegetation height is part of the most contributory variables on Axis 2.

**Axis 3**

Mean building size and number of buildings are the most contributory urban variables (Fig. 9c).

The representation of the variables on the factorial plans allows identifying positive correlations when variables are clustered within the same quadrant and negative correlations when variables are positioned in opposite quadrants relative to the axes’ origins (Fig. 10a). Consequently, three groups can be distinguished: one made of urban landscape variables positively correlated with Axis



**Fig. 7** (a) Panchromatic band of the Pléiades image where built-up areas have been delineated and FOTOTEX has been applied. (b) Textural indices produced by FOTOTEX: RGB composite images with the first three Principal Components (PC) - Component 1 in the red band (69% of variance), C2 in the green band (18%), and C3 in the blue band (0.03%). Pléiades © CNES 2022 Distribution Airbus DS

1, another made of urban landscape variables positively correlated with Axis 2, and a third group containing breeding site variables, perpendicular to the urban variable groups. On the correlation circle with Axes 2 and 3, potential breeding sites appear to be correlated with the percentage of vegetation, but with low representational quality (Fig. 10a).

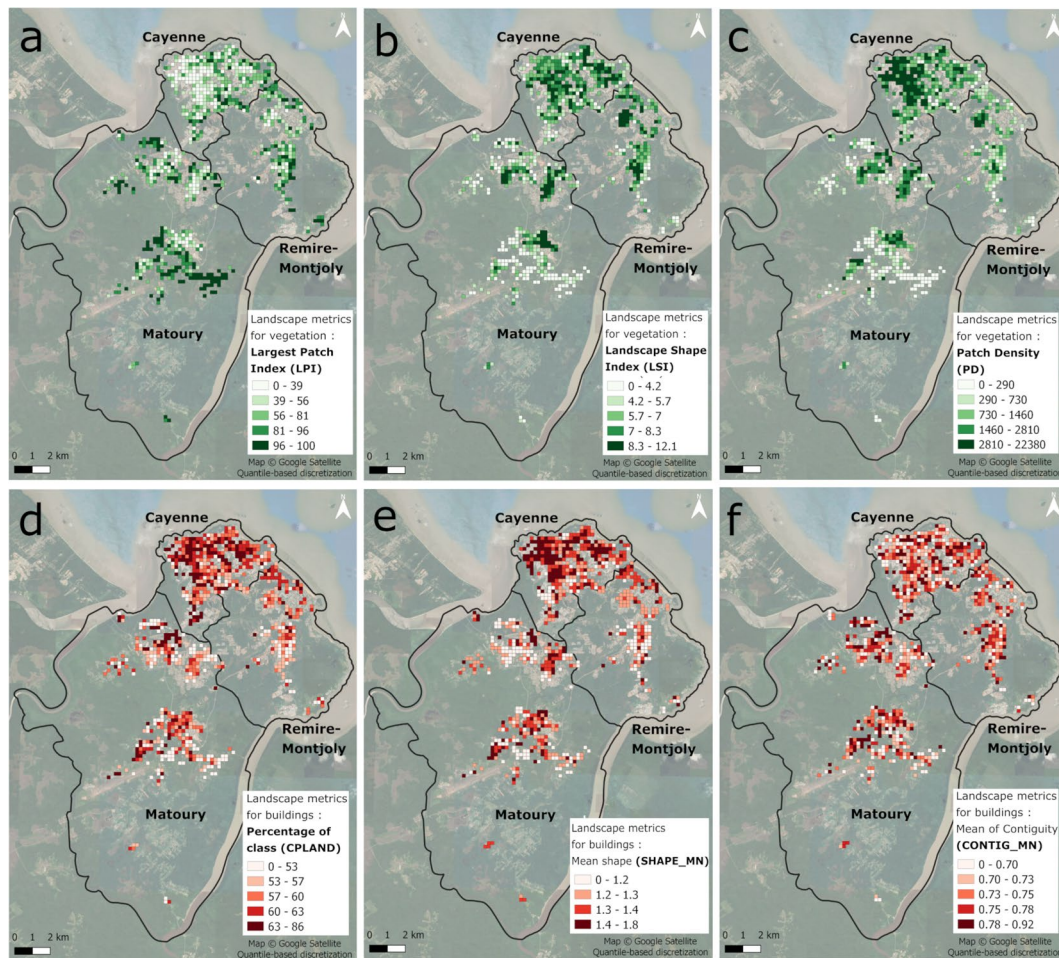
Consequently, for RF models, the selected variables were building class area, number of buildings, vegetation landscape patch index, NDVI, NDWI, the first component of texture, and the height of vegetation.

### Modeling results

Random Forest (RF) models provided mean  $R^2$  values ranging from 0 to 0.20 for the categories of breeding sites (Fig. 11a) and from 0 to 0.33 for the types of breeding sites (Fig. 11b). The average NRMSE between observations and predictions is 0.14 (range: 0.05–0.65). Detailed analysis of the RF models for potential breeding sites shows that the “large volume” breeding sites type has the highest average prediction result with “MFA variables”, with an  $R^2$  of 0.33 and an NRMSE of 0.13 (Fig. 11a). When using the variables selected from the MFA, the  $R^2$  for “all potential breeding sites” is 0.15, while it is 0.18 when using only “buildings information”, with the NRMSE both at 0.14. “Containers of less than 200 liters” have a  $R^2$  of 0.16 and an NRMSE of 0.11 with “MFA variables” and with only “buildings information” a  $R^2$  of 0.14 and NRMSE of 0.12. Next, the category

of “Containers” grouping several types of breeding sites provide an  $R^2$  of 0.20 and an NRMSE of 0.11 (Fig. 11b). The categories of explanatory variables with the highest  $R^2$  are “building information” and “MFA variables”, except for category “Water Management Infrastructure” and for types “gutters”, “canals”, “tires”, “troughs”, “buildings materials”, “watercraft”, “pit”, and “other plants”. Texture, heights and landscape metrics, when taken individually, provide low predictive powers ( $R^2$  below 0.10). Some types of breeding sites like “gutters”, “canals”, and “tires” have higher  $R^2$  values with “landscape metrics for vegetation” variables but these values remain low (between 0.05 and 0.10). Many types of breeding sites have very low  $R^2$  values (below 0.02) and rather higher NRMSE values: “others plants”, “protection”, “pit”, “watercraft”, and “buildings materials”. Overall, there is considerable variability in model performance across breeding sites types and categories, with both low  $R^2$  values and NRMSE values. When analyzing the maximum performance values of the RF models (Additional file 8), for “large volumes” the  $R^2$  is 0.94 with “buildings information”, closely followed by MFA variables, although NRMSE are higher (around 3). For “all potential breeding sites” or “barrels”, the maximum  $R^2$  is 0.35 with “buildings information” and NRMSE values are under 1, regardless of the predictor variables considered.

When a model was trained on the complete dataset with the set of variables selected based on MFA, the number of potential breeding sites can be predicted for



**Fig. 8** Landscape metrics applied to vegetation and buildings (a) Largest Patch Index (LPI); (b) Landscape Shape Index (LSI); (c) Patch Density (PD); (d) Percentage of class (CPLAND); (e) Mean Shape (SHAPE\_MN); and (f) Mean of Contiguity (CONTIG\_MN)

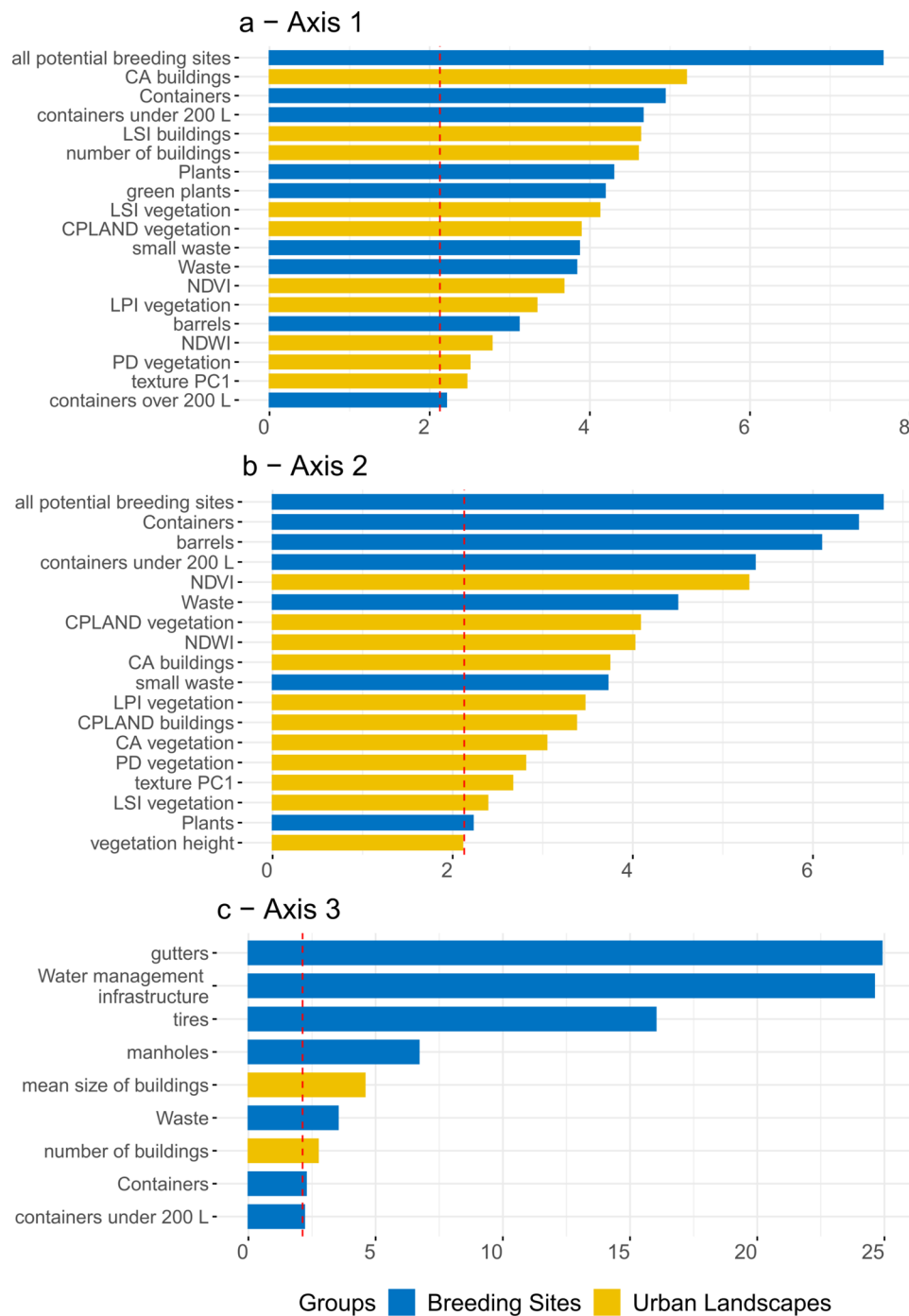
the entire urbanized area of the Cayenne Island (Fig. 12a). In this case, the  $R^2$  between the observed values used for the model and the predicted values is 0.90 (Fig. 12d). Residential areas with high densities of buildings are the areas where predicted potential breeding sites are higher (over 43 per grid cell) while commercial areas concentrate less predicted potential breeding sites (under 22 per grid cell). Peri-urban areas with isolated houses show values between 7 et 22, with some areas exhibiting numbers of potential breeding sites above 43. The model did not predict any cells with zero potential breeding sites.

## Discussion

This study aimed to better understand how the distribution of *Aedes aegypti* mosquitoes' potential breeding sites relates to urban landscapes, with the goal to assess to what extent urban environmental variables can spatially predict the number of breeding sites. Targeting specifically potential breeding sites, our study tries to assess the number of available breeding sites in a given territory. In fact, as the presence of larvae depends on many factors

and is not always easy to establish in the field, predicting positive breeding sites only can result in underestimating the actual capacity of charge of the environment, i.e. the capacity of the territory to “produce” mosquitoes. We consider that to improve vector control strategies, it is interesting to focus on potential breeding sites rather than positive ones to better estimate breeding sites availability and target priority areas during inter-epidemic periods, or optimize efforts and resources during epidemic periods. Moreover, mechanistic models like the one developed by Tran et al. [65] require the capacity of charge of the environment as an input, corresponding, in practice, to the number of potential breeding sites per surface unit (e.g., per ha).

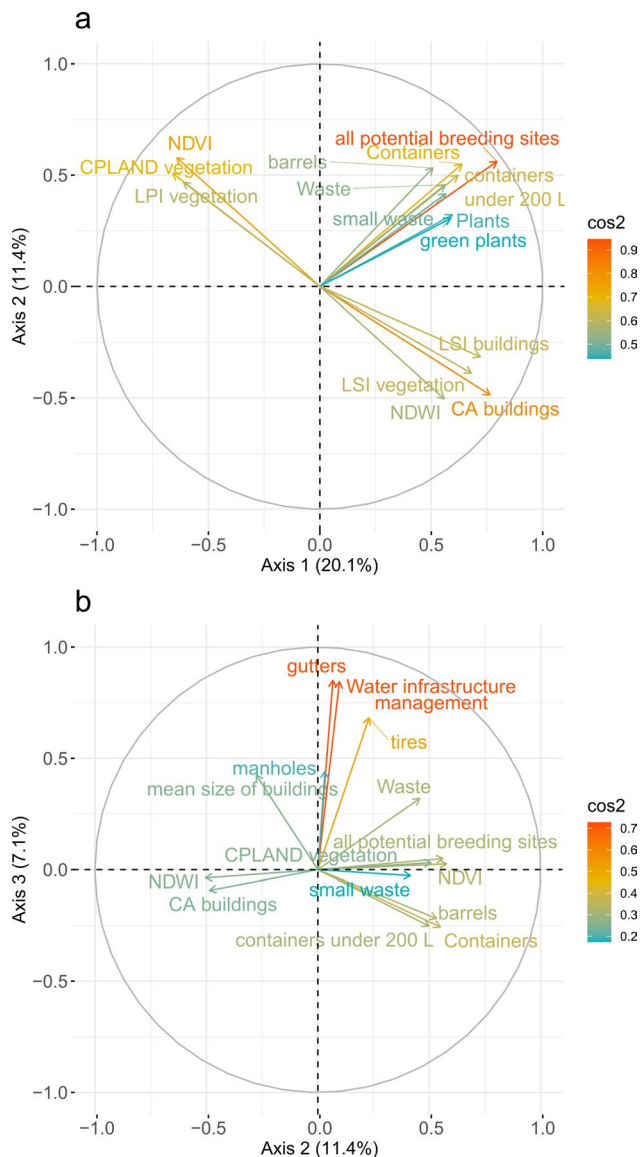
The French Guiana entomological data proved to be heterogeneous, with a high variety of breeding site types and categories over the study area (Figs. 4 and 5, Additional file 6). MFA performed using quantitative variables did not reveal strong linear relationships between urban and potential breeding site variables. However, the factorial plane defined by the axes 2 and 3 of the MFA shows



**Fig. 9** Contributions of quantitative variables above the mean contribution of each axis (red dashed line) to (a) Axis 1 (b) Axis 2 and (c) Axis 3. CA: Total Class Area, CPLAND: Total class area percentage, LPI: Largest patch index, LSI: Landscape shape index, PD: Patch density, SHAPE: Mean shape, CONTIG: Mean of Contiguity

a more contrasted distribution of variables than the one defined by the axes 1 and 2 (Fig. 10), which led us to use the MFA to select the most contributory variables to build the prediction model. Moreover, to take advantage of the non-linear relationships between variables related to breeding sites and environmental variables, a RF algorithm was implemented to see if potential breeding sites

could be predicted. We showed that the selected variables from MFA and the set of variables related to buildings resulted in the best performances of the RF models for predicting the number of all potential breeding sites (Fig. 11a). This can be explained by the fact that these models integrate the number of buildings and the mean building size, variables which have a strong link with the



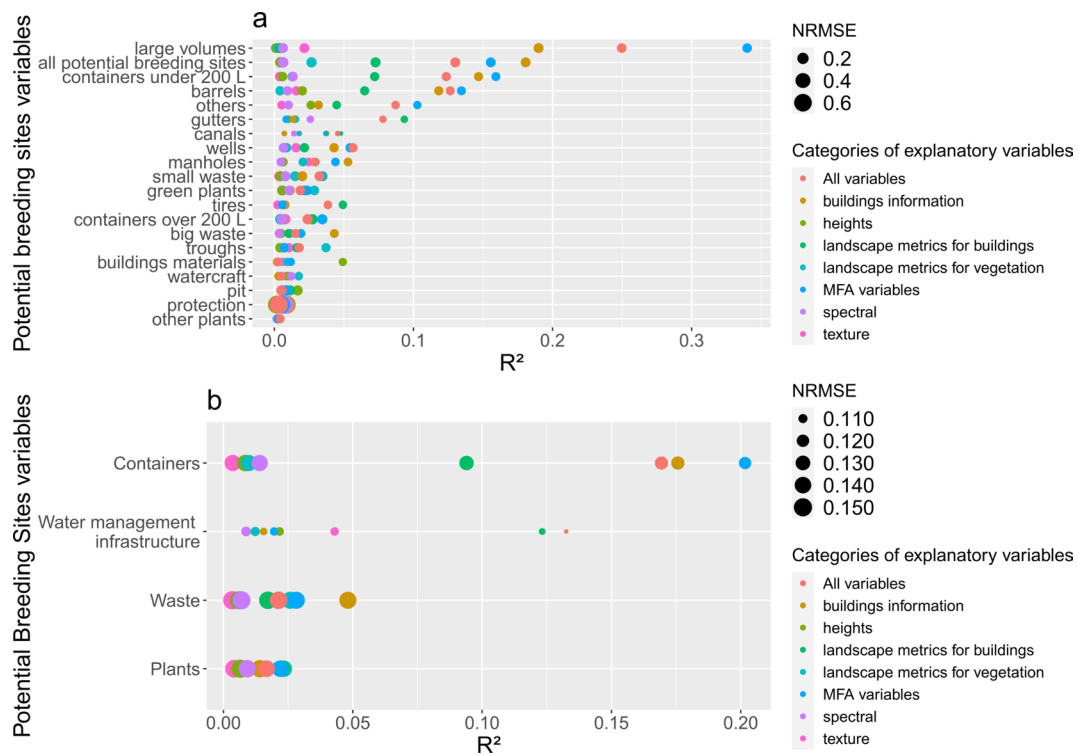
**Fig. 10** Variables, with quality of the representation ( $\cos^2$ ) represented by the arrow color, and correlation circle, for factorial plans defined by (a) Axes 1 and 2 and (b) Axes 2 and 3

number of potential breeding sites [66]. In the present study, the other environmental variables did not appear significantly linked to potential breeding sites (as a whole or considered by type or category). The low performance values of RF models over several types of potential breeding sites could be explained by the strong variability in the data. Even when applying a 70/30 split with bootstrapping for cross-validation, and by taking into account a minimum of non-null cells for training, the quantity of data used by the models is very variable depending on the types of breeding sites. Among the 417 cells used in the modeling process, only 20 have canals, 23 with protection devices (like tarpaulin), 39 with watercraft, 50 with pits, and 21 with “other plants” (not in pots). Even with a

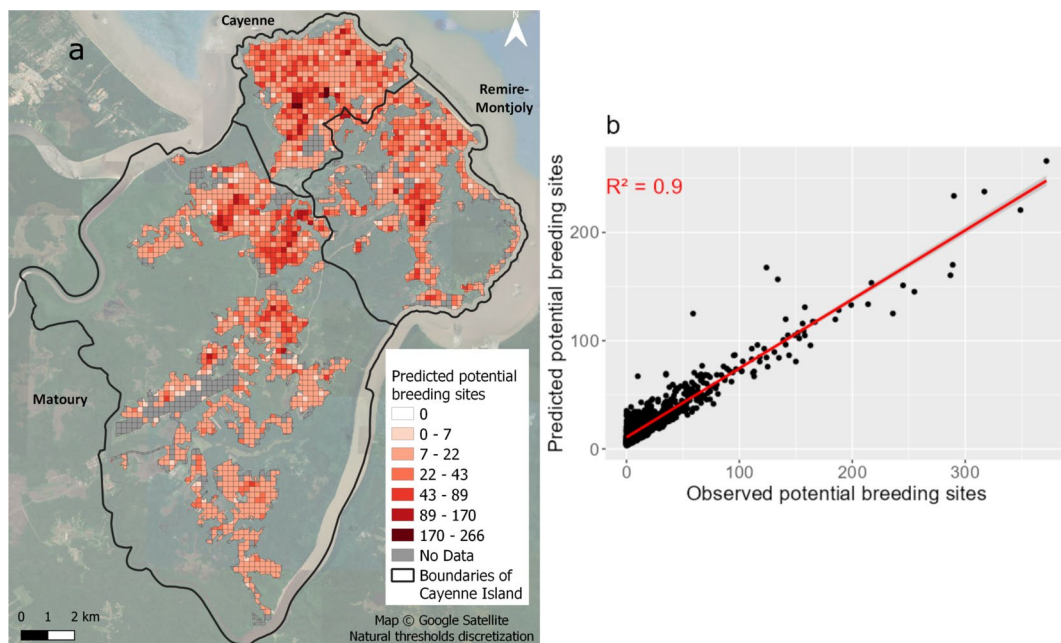
greater availability in the study area, some types of breeding sites still have model performances that remain very low (small waste, green plants, tires) thus demonstrating the difficulty of identifying predictive variables relevant for all types of breeding sites. We compared the modeling performance using absolute versus normalized number of potential breeding sites (absolute number of potential breeding sites divided by the number of houses visited in each cell) to take into account the sampling effort. We found that no matter the type of breeding sites considered, both  $R^2$  and NRMSE consistently yielded lower values for normalized (mean range from 0 to 0.17) (Additional file 4) than for absolute number of potential breeding sites.

Our study suggests that although satellite image texture analysis can discriminate between neighborhood types [41], among the texture variables, the first FOTOTEX principal component was the only one significantly contributory, according to the MFA. The first component of FOTOTEX concentrates the majority of the information related to urban texture, accounting for 69% of the variance. It corresponds to a high building densities (red band in Fig. 7), and consequently can be considered as equivalent to the number or buildings derived from BD TOPO database. As such, a variable derived from remote sensing can be produced in any city, making the prediction method presented in this paper reproducible, whereas topographic databases such as BD TOPO are not available in every country. Even though the first component of texture is used in RF models as one of the MFA selection variables and shows one of the best average prediction performances on test sets, performance of texture variables, when considered as one group of explanatory variables, was found to be insufficient.

In our study, unlike in Espinosa et al. [67] and Arduino et al. [33], we focused on the direct relationship between potential breeding sites and landscape descriptors through the use of a regular grid, avoiding the need for a previous segmentation of the urban space based on its characteristics, which is a complex task. We consider that a more detailed and fine-resolution description of urban areas better captures the complexity and heterogeneity of the landscape hence allowing a better estimation of potential breeding site distribution. Albrieu-Llinàs et al. [34] seem to be the only paper that tried to integrate the urban landscape diversity in relation to positive breeding sites, using it to segment the urban areas. In our study, the landscape patch index (LPI) used on vegetation data and the percentage of building class per grid cell, were amongst the most contributory variables, being well represented in the MFA to use in modeling potential breeding site distribution (Fig. 9a). This shows that in urban areas, not only buildings but also urban vegetation contributes to explain the number and distribution



**Fig. 11** RF models mean R<sup>2</sup> and NRMSE values when considering different response variables and different groups of explanatory variables for (a) types of potential breeding sites, and (b) categories of potential breeding sites



**Fig. 12** (a) Predicted values of potential breeding sites per grid cell for Cayenne Island; (b) Observed vs. predicted values of potential breeding sites from the RF model using MFA variables for the full dataset

of potential breeding sites. This can be explained by the fact that vegetation can support the creation of potential breeding sites, like tree hollows or plants creating water retention [68, 69], and more generally by the fact

that vegetation is a key element of the urban landscape characterization. Landscape metrics associated with fragmentation (PD), complexity (SHAPE) and contingency (CONTIG) did not appear to be linked to potential



breeding sites in the MFA analysis and also as explanatory predictors in the RF models. Landscape metrics related to composition (CA, CPLAND), dominance (LPI) and compactness (LSI) show more significant relationships with potential breeding site distribution. This study brings the first responses regarding the influence of composition of urban landscapes on the distribution of *Aedes aegypti* breeding sites.

While studies found that *Aedes* mosquitoes can be found on all floors of buildings [70, 71], mosquito density tends to decrease with increasing building levels [72]. We found no study that examined the relationship between the number of potential breeding sites and building or vegetation height. In our study, while vegetation height was contributory and represented in MFA analysis, the height of buildings did not prove significant in explaining or predicting potential breeding sites. This may be attributed to either the absence of a meaningful relationship between them, or too small variations in building heights in the area to have an impact on potential breeding site distribution (value of buildings between 0 and 18.4 m) or to the mismatch between the 2015 digital terrain model data and the 2022 breeding site data.

Climatic variables such as temperature, precipitation and relative humidity were not included in this study for three reasons. First, there is no compelling evidence in the literature of a clear link between those variables and the distribution of potential breeding sites, when compared to variables that characterize urban landscapes [73, 74]. Second, small spatial heterogeneity in temperature and rainfall across the study area would not impact population practices (e.g., water storage, plant watering) and the subsequent potential breeding site presence and abundance. Finally, the spatial resolution of existing climatic data would have been much lower than the resolution of the urban variables that were used in our analyses.

The present study shows that relationships between urban landscapes and potential breeding sites are difficult to identify and characterize, in particular due to the difficulty of extracting variables that effectively capture the heterogeneity of the urban landscapes. The determination of the suitable scale of analysis and of the spatial units for urban landscape variable calculation is another reason why this question is difficult [75]. For potential breeding site data, the 200 m grid appeared to be effective at reducing the influence of outliers and sampling effects. It also helps reduce spatial autocorrelation that occurs when buffers are used around each sampled house. Such a grid is relevant as it permits to work with spatial units (cells) presenting uniform size and form, at a spatial resolution that both permits to work at a local scale, while identifying repetitions of urban patterns. The blocks (i.e. the division of the territory used by vector control agents) (Fig. 3) were considered as spatial units

of analysis, but appeared less relevant than grid cells due to strong heterogeneity in block size and shape. Despite the importance of the spatial nature of response and predictor variables (geospatial data), models used in the study did not explicitly incorporate the spatial locations of the grid cells or any other explicit spatial structures (which can be considered through spatial random effect). The explicit integration of variables on spatial structure or potential neighborhood effects could ameliorate the results [76, 77].

Finally, this entomological dataset made of data routinely acquired by vector control services was not designed specifically for such a study. Using this dataset may have introduced biases as they were collected for immediate and efficient operational needs, leading to oversampling or undersampling over specific geographic areas and/or in specific environmental contexts. Despite the relatively low predictive power ( $R^2$ ) of the RF model, the model demonstrated a high goodness of fit ( $R^2=0.9$ ) when using the entire dataset for Cayenne Island. Such a high  $R^2$  value, compared to the relatively low  $R^2$  obtained with cross-validation, may however confirm the model over-fitting and its poor transferability to new (“unknown”) contexts. Such an issue is common in regression models [78]. A complementary cross-validation, exploiting additional field data on the cells that were predicted and not used for the model training, could be carried out. Although RF is applicable when the number of variables is high, due to its potential for prediction, some studies highlighted the importance of selecting a limited number of predictors before making a prediction [79]. This was the purpose of selecting variables using MFA in this study. For future studies, other methods of variable selection could be tested to improve our results. However, the predictions appear coherent and are promising for further analysis. This kind of prediction could be directly used as an input (environmental charge capacity) of the mechanistic model based on the bio-ecology of *Aedes* mosquitoes to model population dynamics [65]. This would significantly objectivize the environmental charge capacity estimation which is, in practice, mainly estimated by expertise in order to implement such a mechanistic model. Unlike previous studies using positive breeding sites, the originality of this study lies in the direct use of potential breeding sites. In French Guiana, potential breeding sites data were used in research work for the first time, and the protocol must be adapted to the specific needs of this study in terms of predictive modeling. We expect that vector control will benefit from this research, helping update the design of entomological sampling to improve vector control strategies based on house prospections, by integrating knowledge about urban landscapes. The recent study by Rodriguez Conzalez et al. [57] confirms that combining Very

High-Resolution (VHR) imagery and landscape metrics has the potential to improve vector control surveillance strategies, determining an optimal spatial repartition of ovitraps in complex urban environments. A similar approach could be carried out to improve sampling strategies in surveillance based on house prospection.

## Conclusion

We investigated the relationships between urban landscapes descriptors and potential breeding site distribution with a focus on the composition and three-dimensional structure of urban environments at a fine spatial scale, avoiding prior application of classification methods to better capture the complexity and heterogeneity of the landscape. The originality of this study is also to focus on the relevance of considering potential breeding sites for a more comprehensive understanding of the capacity of the territory to “produce” *Aedes* mosquitoes. We used a multifactorial analysis to explore relationships between potential breeding sites and urban landscape variables and select the most discriminatory variables. Such variable selection yielded improved performances in Random Forest models for predicting potential breeding sites. However, the challenge of predicting breeding site number as a function of their type persists, particularly due to the wide disparity in the numbers of breeding sites observed from one type of site to another. The study also highlighted the difficulty of identifying predictive variables relevant to all breeding site types. While texture variables alone did not exhibit sufficient performance, they proved valuable when combined with other variables extracted from remote sensing, showing the importance of both buildings and urban vegetation in explaining the distribution of potential breeding sites. The RF model achieved an excellent goodness of fit, but also overfitting, suggesting a difficulty in its generalization to new situations. This could be addressed by testing other cross-validation methods, finding better ways of filtering data (to minimize noise such as the errors in breeding site categorization and geolocalisation) and selecting predictors, and by enriching the learning set with data provided by different contexts. Although this research has limitations, it offers a first approach in the use of routinely collected data on breeding sites by vector control agencies in a research framework highlighting the need to include an objective remotely sensed characterization of the urban environment to improve vector control strategies.

## Abbreviations

LULC	Land use and land cover
NDVI	Normalized difference vegetation index
NDWI	Normalized difference water index
VHR	Very High Resolution
CTG	Collectivité Territoriale de Guyane
FT	Fourier Transform

PCA	Principal Component Analysis
CA	Total Class Area
CPLAND	Total class area percentage
LPI	Largest patch index
LSI	Landscape shape index
PD	Patch density
SHAPE_MN	Mean shape
CONTIG_MN	Mean of Contiguity

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12942-024-00378-3>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8

## Acknowledgements

Pléiades images were provided by DINAMIS (“Dispositif Institutionnel National d’Accès Mutualisé en Imagerie Satellitaire”). Pléiades © CNES 2022 Distribution AIRBUS DS. We thank the “Collectivité Territoriale de Guyane” (CTG) for providing breeding site dataset. We thank Loreinzia Clarke and Patrice Azor for constructive discussions, data sharing and precious help in understanding the dataset. We are deeply grateful to the CTG agents for their daily work on the field, providing data for this study.

## Author contributions

CT carried out the conception and design of the study, image processing, analysis and interpretation of data, and led the writing of the manuscript. ER and RD contributed to the conception of the study, participated in result interpretation and manuscript revision. AT, TC, RM, contributed to the conception of the study, and to manuscript revision. ND, JR contributed to manuscript revision. JR, JR provided data, participated in data analysis and results interpretation. All authors read and approved the final manuscript.

## Funding

This work has been funded by the French space agency CNES and the French Occitanie Region. It also received the support of the European PROGYSAT Interreg project.

## Data availability

The urban datasets generated and analyzed, as well as the R codes developed during the current study, are available in the Forge IRD repository, [https://forge.ird.fr/espace-dev/personnels/teillet/aedes\\_breeding\\_sites\\_modelling.git](https://forge.ird.fr/espace-dev/personnels/teillet/aedes_breeding_sites_modelling.git).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>ESPACE-DEV, Univ Montpellier, IRD, Univ Guyane, Univ Reunion, Univ Antilles, Univ Avignon, Maison de la Télédétection, 500 rue Jean-François Breton, Montpellier Cedex F-34093, France  
<sup>2</sup>CIRAD, UMR TETIS, Maison de la Télédétection, 500 rue Jean-François Breton, Montpellier, Cedex F-34093, France  
<sup>3</sup>TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Maison de la Télédétection, 500 rue Jean-François Breton, Montpellier, Cedex F-34093, France  
<sup>4</sup>Direction de la Démoustication, Collectivité Territoriale de Guyane (CTG), 4179 Route de Montabo, Cayenne, Guyane française 97300, France  
<sup>5</sup>International Joint laboratory Sentinela, FIOCRUZ, UnB, IRD, Maison de la Télédétection, 500 rue Jean-François Breton, Montpellier, Cedex F-34093, France

Received: 1 February 2024 / Accepted: 22 June 2024

Published online: 07 July 2024

### References

- Kolimenakis A, Heinz S, Wilson ML, Winkler V, Yakob L, Michaelakis A, et al. The role of urbanisation in the spread of *Aedes* mosquitoes and the diseases they transmit—A systematic review. *PLoS Negl Trop Dis*. 2021;15(9):e0009631.
- Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Jit M*, editor. *eLife*. 2015;4:e08347.
- Duvallet G, Fontenille D, Robert V, editors. *Entomologie médicale et vétérinaire* [Internet]. Marseille: IRD; 2017. <http://www.documentation.ird.fr/hor/fdi:010069823>.
- Lwande OW, Obanda V, Lindström A, Ahlm C, Evander M, Näslund J et al. Globe-Trotting *Aedes aegypti* and *Aedes albopictus*: Risk Factors for Arbovirus Pandemics. *Vector-Borne Zoonotic Dis* [Internet]. 2019; <https://www.liebert-pub.com/doi/https://doi.org/10.1089/vbz.2019.2486>.
- LaDeau SL, Allan BF, Leisham PT, Levy MZ. The ecological foundations of transmission potential and vector-borne disease in urban landscapes. *Funct Ecol*. 2015;29(7):889–901.
- Abílio AP, Abudasse G, Kampango A, Candrinho B, Sítio S, Luciano J, et al. Distribution and breeding sites of *Aedes aegypti* and *Aedes albopictus* in 32 urban/peri-urban districts of Mozambique: implication for assessing the risk of arbovirus outbreaks. *PLoS Negl Trop Dis*. 2018;12(9):e0006692.
- Marti R, Li Z, Catry T, Roux E, Mangeas M, Handschumacher P, et al. A Mapping Review on Urban Landscape factors of Dengue retrieved from Earth Observation Data, GIS techniques, and Survey questionnaires. *Remote Sens*. 2020;12(6):932.
- Bonifay T, Le Turnier P, Epelboin Y, Carvalho L, De Thoisy B, Djossou F, et al. Review on Main arboviruses circulating on French Guiana, an Ultra-peripheral European Region in South America. *Viruses*. 2023;15(6):1268.
- Talaga S. Écologie, diversité et évolution des moustiques (Diptera Culicidae) de Guyane française: implications dans l'invasion biologique du moustique *Aedes aegypti* (L.) [Internet] [phdthesis]. Université de Guyane; 2016. p. 217 p. <https://hal.inrae.fr/tel-02795002>.
- Epelboin Y, Chaney SC, Guidez A, Habchi-Hanriot N, Talaga S, Wang L, et al. Successes and failures of sixty years of vector control in French Guiana: what is the next step? *Mem Inst Oswaldo Cruz*. 2018;113:e170398.
- Guidez A, Pocquet N, Restrepo J, Mathieu L, Gaborit P, Issaly J, et al. Spatio-temporal multiple insecticide resistance in *Aedes aegypti* populations in French Guiana: need for alternative vector control. *Mem Inst Oswaldo Cruz*. 2021;115:e200313.
- SPF. Bulletin de santé publique dengue en Guyane. Septembre 2023. [Internet]. 2023. <https://www.santepubliquefrance.fr/regions/guyane/documents/bulletin-regional/2023/bulletin-de-sante-publique-dengue-en-guyane-septembre-2023>.
- Paraná VC, Feitosa CA, da Silva GCS, Gois LL, Santos LA. Risk factors associated with severe dengue in Latin America: a systematic review and meta-analysis. *Trop Med Int Health*. 2024;29(3):173–91.
- Zhu Z, Zhou Y, Seto KC, Stokes EC, Deng C, Pickett STA, et al. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sens Environ*. 2019;228:164–82.
- Lorenz C, Chiaravalloti-Neto F, de Oliveira Lage M, Quintanilha JA, Parra MC, Dibo MR, et al. Remote sensing for risk mapping of *Aedes aegypti* infestations: is this a practical task? *Acta Trop*. 2020;205:105398.
- Yu D, Fang C. Urban Remote sensing with spatial Big data: a review and renewed perspective of Urban studies in recent decades. *Remote Sens*. 2023;15(5):1307.
- Kopecká M, Szatmári D, Rosina K. Analysis of Urban Green spaces based on Sentinel-2A: Case studies from Slovakia. *Land*. 2017;6(2):25.
- Yin J, Dong J, Hamm NAS, Li Z, Wang J, Xing H, et al. Integrating remote sensing and geospatial big data for urban land use mapping: a review. *Int J Appl Earth Obs Geoinf*. 2021;103:102514.
- Liu T, Yang X. Monitoring land changes in an urban area using satellite imagery, GIS and landscape metrics. *Appl Geogr*. 2015;56:42–54.
- Xing H, Meng Y. Measuring urban landscapes for urban function classification using spatial metrics. *Ecol Indic*. 2020;108:105722.
- Mathan M, Krishnaveni M. Monitoring spatio-temporal dynamics of urban and peri-urban land transitions using ensemble of remote sensing spectral indices—a case study of Chennai Metropolitan Area, India. *Environ Monit Assess*. 2019;192(1):15.
- de Almeida CR, Teodoro AC, Gonçalves A. Study of the Urban Heat Island (UHI) using Remote Sensing Data/Techniques: a systematic review. *Environments*. 2021;8(10):105.
- Liang B, Weng Q. Characterizing Urban Landscape by using Fractal-based texture information. *Photogramm Eng Remote Sens*. 2018;84(11):695–710.
- Hoang ND. Automatic impervious surface area detection using image texture analysis and neural Computing models with Advanced optimizers. *Comput Intell Neurosci*. 2021;2021:e8820116.
- Liao HY, Wen TH. Extracting urban water bodies from high-resolution radar images: measuring the urban surface morphology to control for radar's double-bounce effect. *Int J Appl Earth Obs Geoinf*. 2020;85:102003.
- Scavuzzo JM, Trucco F, Espinosa M, Tauro CB, Abril M, Scavuzzo CM, et al. Modeling dengue vector population using remotely sensed data and machine learning. *Acta Trop*. 2018;185:167–75.
- Lorenz C, Castro MC, Trindade PMP, Nogueira ML, de Oliveira Lage M, Quintanilha JA, et al. Predicting *Aedes aegypti* infestation using landscape and thermal features. *Sci Rep*. 2020;10(1):21688.
- Andreo V, Cuervo PF, Porcasi X, Lopez L, Guzman C, Scavuzzo CM. Towards a workflow for operational mapping of *Aedes aegypti* at urban scale based on remote sensing. *Remote Sens Appl Soc Environ*. 2021;23:100554.
- Estallo EL, Lamfri MA, Scavuzzo CM, Almeida FFL, Introini MV, Zaidenberg M, et al. Models for Predicting *Aedes aegypti* Larval Indices based on Satellite images and climatic variables. *J Am Mosq Control Assoc*. 2008;24(3):368–76.
- Estallo EL, Sangermano F, Grech M, Ludueña-Almeida F, Frías-Céspedes M, Ainete M, et al. Modelling the distribution of the vector *aedes aegypti* in a central Argentine city. *Med Vet Entomol*. 2018;32(4):451–61.
- Bailly S, Machault V, Beneteau S, Palany P, Girod R, Lacaue JP et al. Modeling spatiotemporal *Aedes aegypti* risk in French guiana using meteorological and remote sensing data. 2021;2021.08.02.21261373.
- Machault V, Yébakima A, Etienne M, Vignolles C, Palany P, Tourre Y, et al. Mapping entomological dengue risk levels in Martinique using high-resolution remote-sensing Environmental Data. *ISPRS Int J Geo-Inf*. 2014;3(4):1352.
- Arduino M, de Mucci B, Santos LF, de Soares LM. S. Importance of microenvironment to arbovirus vector distribution in an urban area, São Paulo, Brazil. *Rev Soc Bras Med Trop* [Internet]. 2020;53. <http://www.scielo.br/j/rsbmt/a/bLLXZtWfNdY3RrpJ7cWJzx/?lang=en&format=html>.
- Albrieu-Ljinás G, Espinosa MO, Quaglia A, Abril M, Scavuzzo CM. Urban environmental clustering to assess the spatial dynamics of *Aedes aegypti* breeding sites. *Geospatial Health* [Internet]. 2018; <https://www.geospatialhealth.net/index.php/gh/article/view/654>.
- Kache PA, Santos-Vega M, Stewart-Ibarra AM, Cook EM, Seto KC, Diuk-Wasser MA. Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases. *Nat Ecol Evol*. 2022;6(11):1601–16.
- Chen S, Whiteman A, Li A, Rapp T, Delmelle E, Chen G, et al. An operational machine learning approach to predict mosquito abundance based on socio-economic and landscape patterns. *Landscape Ecol*. 2019;34(6):1295–311.
- Mudele O, Bayer FM, Zanandrez L, Eiras AE, Gamba P. Modeling the Temporal Population Distribution of *Ae. aegypti* Mosquito using Big Earth Observation Data [Internet]. arXiv; 2019. <http://arxiv.org/abs/1911.08979>.
- Rahman MS, Ekalaksananan T, Zafar S, Poolphol P, Shipin O, Haque U, et al. Ecological, Social, and other Environmental determinants of Dengue Vector Abundance in Urban and Rural areas of Northeastern Thailand. *Int J Environ Res Public Health*. 2021;18(11):5971.
- QGIS.org, 2024. QGIS Geographic Information System. QGIS Association. [Internet]. <http://www.qgis.org>.

40. Herold M, Liu X, Clarke KC. Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogramm Eng Remote Sens*. 2003;69(9):991–1001.
41. Teillet C, Pillot B, Catry T, Demagistri L, Lyszczarz D, Lang M, et al. Fast unsupervised Multi-scale characterization of Urban landscapes based on Earth Observation Data. *Remote Sens*. 2021;13(12):2398.
42. Couteron P, Pellissier R, Nicolini EA, Paget D. Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images. *J Appl Ecol*. 2005;42(6):1121–8.
43. Couteron P, Barbier N, Gautier D. Textural ordination based on Fourier spectral decomposition: a method to analyze and compare landscape patterns. *Landscape Ecol*. 2006;21:555–67.
44. Barbier N, Couteron P, Gastelly-Etchegorry JP, Proisy C. Linking canopy images to forest structural parameters: potential of a modeling framework. *Ann Sci*. 2012;69(2):305–11.
45. Lang M, Alleaume S, Luque S, Baghdadi N, Feret JB. Landscape Structure Estimation using Fourier-Based Textural Ordination of High Resolution Airborne Optical Image. In: *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium* [Internet]. Valencia: IEEE; 2018. pp. 6600–3. <https://ieeexplore.ieee.org/document/8518640/>.
46. Lang M. Caractérisation de l'hétérogénéité spatiale de milieux naturels à partir d'imagerie optique très haute résolution spatiale: cas d'application aux milieux méditerranéens de garrigue. 2019.
47. Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. In. 1974. <https://ntrs.nasa.gov/citations/19740022614>.
48. Huang S, Tang L, Hupy JP, Wang Y, Shao G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J Res*. 2021;32(1):1–6.
49. Honório NA, Castro MG, de Barros FSM, Magalhães M, de Sabroza AFM. The spatial distribution of *Aedes aegypti* and *Aedes albopictus* in a transition zone, Rio De Janeiro, Brazil. *Cad Saúde Pública*. 2009;25(6):1203–14.
50. Huynh TTT, Minakawa N. A comparative study of dengue virus vectors in major parks and adjacent residential areas in Ho Chi Minh City, Vietnam. *PLoS Negl Trop Dis*. 2022;16(1):e0010119.
51. Gao Bcai. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ*. 1996;58(3):257–66.
52. Xu H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int J Remote Sens*. 2006;27(14):3025–33.
53. Chen F, Chen X, Van de Voorde T, Roberts D, Jiang H, Xu W. Open water detection in urban environments using high spatial resolution remote sensing imagery. *Remote Sens Environ*. 2020;242:111706.
54. Hashim H, Abd Latif Z, Adnan NA. Urban vegetation classification with NDVI threshold value method with very high resolution (VHR) Pleiades Imagery. *Int Arch Photogramm Remote Sens Spat Inf Sci*. 2019;XLII–4/W16:237–40.
55. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. 2023. <https://www.R-project.org/>.
56. Varela S, Anderson RP, García-Valdés R, Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* [Internet]. 2014;37(11). <https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1111/j.1600-0587.2013.00441.x>.
57. Rodríguez González C, Guzman C, Andreo V. Using VHR satellite imagery, OBIA and landscape metrics to improve mosquito surveillance in urban areas. *Ecol Inf*. 2023;77:102221.
58. Hesselbarth MHK, Sciaini M, With KA, Wiegand K, Nowosad J. Landscape-metrics: an open-source R tool to calculate landscape metrics. *Ecography*. 2019;42(10):1648–57.
59. Abdi H, Williams LJ, Valentín D. Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Comput Stat*. 2013;5(2):149–79.
60. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25:1–18.
61. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
62. Segal MR. Machine Learning Benchmarks and Random Forest Regression. 2004; <https://escholarship.org/uc/item/35x3v9t4>.
63. Liaw A, Wiener M. Classification and Regression by randomForest. 2002.
64. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021;7:e623.
65. Tran A, Mangeas M, Demarchi M, Roux E, Degenne P, Haramboure M, et al. Complementarity of empirical and process-based approaches to modelling mosquito population dynamics with *Aedes albopictus* as an example—application to the development of an operational mapping tool of vector populations. *PLoS ONE*. 2020;15(1):e0227407.
66. Madewell ZJ, Sosa S, Brouwer KC, Juárez JG, Romero C, Lenhart A, et al. Associations between household environmental factors and immature mosquito abundance in Quetzaltenango, Guatemala. *BMC Public Health*. 2019;19(1):1729.
67. Espinosa MO, Polop F, Rotela CH, Abril M, Scavuzzo CM. Spatial pattern evolution of *Aedes aegypti* breeding sites in an Argentinean city without a dengue vector control programme. *Geospatial Health* [Internet]. 2016;11(3). <http://geospatialhealth.net/index.php/gh/article/view/471>.
68. Darriet F. Des moustiques et des hommes: chronique d'une pullulation annoncée [Internet]. Marseille: IRD; 2014. (Didactiques). <http://www.documentation.ird.fr/hor/fdi:010063026>.
69. Meena A, Choudhary N. Container breeding preference of *Aedes albopictus* in urban environment. 2019.
70. Lau KW, Chen CD, Lee HL, Izzul AA, Asri-Isa M, Zulfadli M, et al. Vertical distribution of *Aedes* mosquitoes in multiple storey buildings in Selangor and Kuala Lumpur, Malaysia. *Trop Biomed*. 2013;30(1):36–45.
71. Jayathilake TAHDG, Wickramasinghe MB, de Silva BGDNK. Oviposition and vertical dispersal of *Aedes* mosquitoes in multiple storey buildings in Colombo district, Sri Lanka. 2015.
72. Ab Hamid N, Mohd Noor SN, Isa NR, Md Rodzay R, Bachtiar Effendi AM, Hafisool AA, et al. Vertical Infestation Profile of *Aedes* in selected urban high-rise residences in Malaysia. *Trop Med Infect Dis*. 2020;5(3):114.
73. Sarraf MS, Tripathi NK, Faruque FS, Bajwa UI, Kitamoto A, Souris M. Mapping urban and peri-urban breeding habitats of *Aedes* mosquitoes using a fuzzy analytical hierarchical process based on climatic and physical parameters. *Geospat Health*. 2014;8(3):685.
74. Espinosa M, Weinberg D, Rotela CH, Polop F, Abril M, Scavuzzo CM. Temporal dynamics and spatial patterns of *Aedes aegypti* breeding sites, in the context of a Dengue Control Program in Tartagal (Salta Province, Argentina). *PLoS Negl Trop Dis*. 2016;10(5):e0004621.
75. Wu J, Jelinski D, Luck M, Tueller P. Multiscale Analysis of Landscape Heterogeneity: Scale Variance and Pattern Metrics. *Geogr Inf Sci*. 2000;6:6–19.
76. Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*. 2018;6:e5518.
77. Benito B, BlasBenito/spatialRF. spatialRF: Easy Spatial Regression with Random Forest [Internet]. Zenodo; 2021. <https://zenodo.org/records/4745208>.
78. Kuhn M, Johnson K. Measuring Performance in Regression Models. In: Kuhn M, Johnson K, editors. *Applied Predictive Modeling* [Internet]. New York, NY: Springer; 2013. pp. 95–100. [https://doi.org/10.1007/978-1-4614-6849-3\\_5](https://doi.org/10.1007/978-1-4614-6849-3_5).
79. Borup D, Christensen BJ, Mühlbach NS, Nielsen MS. Targeting predictors in random forest regression. *Int J Forecast*. 2023;39(2):841–68.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.