Geospatial Health 2022; volume 17:1033

Ecological characterization of a cutaneous leishmaniasis outbreak through remotely sensed land cover changes

Verónica Andreo,[1,2] Juan Rosa,[3] Karina Ramos,[4] O. Daniel Salomón[2,5]

[1]Instituto de Altos Estudios Espaciales Mario Gulich, UNC-CONAE, Falda del Cañete, Córdoba; [2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires; [3]Instituto de Medicina Regional, Universidad Nacional del Nordeste, Resistencia, Chaco; [4]Programa de Leishmaniasis, Ministerio de Salud de la Provincia de Corrientes, Corrientes; [5]Instituto Nacional de Medicina Tropical (INMeT), ANLIS CG Malbrán, Puerto Iguazú, Misiones, Argentina

Abstract

In this work we assessed the environmental factors associated with the spatial distribution of a cutaneous leishmaniasis (CL) outbreak during 2015-2016 in north-eastern Argentina to understand its typical or atypical eco-epidemiological pattern. We combined locations of human CL cases with relevant predictors derived from analysis of remote sensing imagery in the framework of ecological niche modelling and trained MaxEnt models with cross-validation for predictors estimated at different buffer areas relevant to CL vectors (50 and 250 m radii). To account for the timing of biological phenomena, we considered environmental changes occurring in two periods, 2014-2015 and 2015-2016. The remote sensing analysis identified land cover changes in the surroundings of CL cases, mostly related to new urbanization and flooding. The distance to such changes was the most important variable in most models. The weighted average map denoted higher suitability for CL in the outskirts of the city of Corrientes and in areas close to environmental changes. Our results point to a scenario consistent with a typical CL outbreak, *i.e.* changes in land use or land cover are the main triggering factor and most affected people live or work in border habitats.

Correspondence: Oscar Daniel Salomón, Instituto Nacional de Medicina Tropical - ANLIS 'Dr. Carlos G Malbrán', Almafuerte y Ámbar s/n. 3370. Puerto Iguazú, Misiones, Argentina.
Tel.: +54.3757.425001 / +54.3757.425002.
E-mail: odanielsalomon@gmail.com

Key words: Cutaneous leishmaniasis; remote sensing; land use and land cover change; species distribution models; disease ecology; Argentina.

Acknowledgements: the authors would like to thank Analía Araujo who carried out the sand flies trappings and geo-located the 25 cases reported as CL to the Corrientes health authorities. We appreciate the information and advice provided by Horacio Lucero and Javier Liotta regarding molecular diagnosis procedures. We also thank NASA for making Landsat data freely available and all the developers and contributors to the free and open source software used in this study. This work was funded by the National Agency for the Promotion of Research, Technological Development and Innovation through the grant PICT 2017-2663.

See online Appendix for additional materials.

Received for publication: 30 June 2021.
Revision received: 15 November 2021.
Accepted for publication: 16 November 2021.

Geospatial Health 2022; 17:1033
doi:10.4081/gh.2022.1033

Introduction

Leishmaniases are globally distributed parasitoses caused by Trypanosomatidae of the genus *Leishmania* that are transmitted by Phlebotominae insects. This group of diseases is reported as the highest increase in prevalence among the neglected tropical diseases (NTDs) between 1990 and 2016 (Hotez, 2018). In the Americas the main clinical forms include visceral leishmaniasis (VL) with confirmed transmission in 12 countries at an incidence of ~3500 cases/year over the last 20 years, and cutaneous (CL) - mucocutaneous leishmaniasis (MCL), which is endemic in 18 countries with ~55,000 cases/year in the same period (PAHO, 2019). In Honduras, Nicaragua and El Salvador, the latter figure includes around 810 cases during 2018 of an atypical form (ACL), a disease caused by the VL parasite transmitted by its phlebotomine vector, and usually presenting as a papular clinical expression that differs from the ulcerated one of the typical CL (Sandoval *et al*., 2018).

In the South Cone of South America, CL has been endemic since pre-hispanic times, with episodic outbreaks, generally associated with environmental changes and border effects in rural and peri-urban landscapes (Quintana *et al*., 2012; Salomón *et al*., 2016; Salomón, 2019). VL, on the other hand, appeared and spread in the region only during the last decades, and as an urban disease. After its initial introduction in a city, VL produces a pattern of human cases scattered in time and space or a persistent

plateau of infection (Salomón *et al*., 2015; Bruhn *et al*., 2018). Consistent with the transmission scenarios described above, sporadic CL outbreaks have been reported in north-eastern Argentina, as well as the recent emergence of urban VL. Peaks in human incidence of CL due to *L. braziliensis* have shown to be associated with changes in land use and people inhabiting forest edges, with *Nyssomia neivai* (previously *Lutzomyia intermedia*) and *Ny. whitmani* as main vectors (Salomón *et al*., 2016). Human VL due to *L. infantum* is associated with urban environments, with *Lu. longipalpis* as vector and dogs as reservoir (Salomón *et al*., 2008; Fernandez *et al*., 2013).

Corrientes, a city in north-eastern Argentina, shows the same CL and VL epidemiological patterns with *Lu. longipalpis* as vector in the urban area since 2008 and *Ny. neivai* in the peri-urban outskirts and rural landscapes (Salomón *et al*., 2009). During 2015-2016 the health authorities of the province reported a CL outbreak to the National Surveillance System (Table 1). However, recent research suggests that this was an outbreak due to the parasite that produces VL, *i.e*. an atypical urban CL outbreak despite the typical CL presentation, and that it could be the case of other CL outbreaks as well (Acosta-Soto *et al*., 2020). This new scenario would imply a significant increase in costs for leishmaniasis control programmes in terms of differential diagnosis and follow up of potential lethal visceralization of the human cases. Clarifying this issue is thus of great interest for public health strategies and individual case management. To elucidate the typical or atypical epidemiological pattern of the CL outbreak in the city of Corrientes, we used satellite imagery to detect and map environmental changes, as well as to derive different environmental variables that we then combined with geo-coded CL cases in the framework of ecological niche modelling (ENM) to retrospectively model and map the risk of CL and discuss the results from an eco-epidemiological point of view.

Materials and methods

Study area

The study was carried out in the city of Corrientes (27°28'08' S, 58°49'50' W) and its surroundings in Corrientes Province, Argentina (Figure 1). The city is located on the eastern shore of the Paraná River and belongs to the Chaco ecoregion (Arana *et al*., 2017). The urban area is approximately 60 km^2 and has 328,868 inhabitants (INDEC, 2010) and has a complex and heterogeneous urban landscape, which includes a downtown area with high population density and a well-developed infrastructure that includes peri-urban areas with patches of houses mixed with green areas (agricultural lands, forest remnants and secondary vegetation). The study area included Riachuelo, a small rural town with almost

Table 1. Leishmaniasis cases in Corrientes Province, Argentina in the period 1955-2019.

Disease	Period									
	1955-74	1975-94	1995-04	2013	2014	2015	2016	2017	2018	2019
CL	2.2	6.8	11.7	8	4	70	49	12	9	8
MCL	*	*	*	0	0	0	0	0	3	2
VL	0	0	0	8	5	4	2	0	0	0

CL, cutaneous leishmaniasis; MCL, mucocutaneous leishmaniasis; VL, visceral leishmaniasis. *No discrimination between CL and MCL. Data obtained from the National Surveillance System (https://www.argentina.gob.ar/salud/epidemiologia/boletines) either by period (average cases/year) or by year.

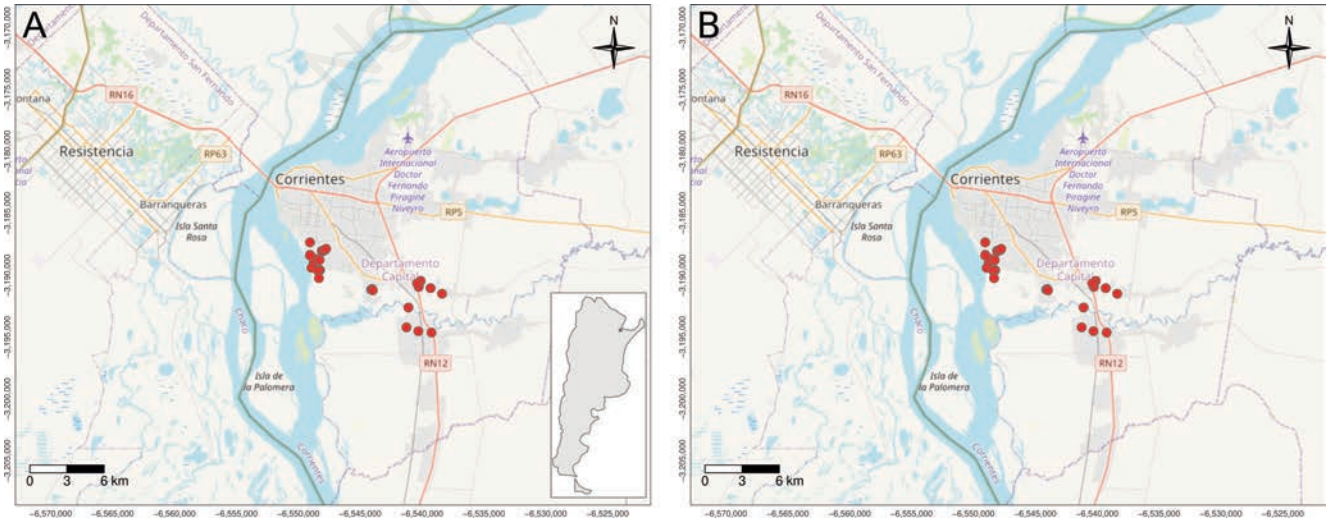

Figure 1. City of Corrientes and surrounding area. A) Red dots are the cutaneous leishmaniasis cases from the period 2015-2016 reported to Corrientes Health authorities (n=25). B) Red dots are the cases digitized from the report by Acosta-Soto *et al.* (2020) (n=74 of 81 reported by the authors).

4000 inhabitants (INDEC, 2010), situated south of Corrientes City. In general, the area has a subtropical climate without dry seasons. The temperature varies from 8 to 21°C in the winter and from 20 to 32°C in the summer. Precipitation is abundant throughout the year with annual values in the range 1400-1900 mm (SMN, 2020).

Human cases and vector data

We used two sources of CL human cases: one obtained jointly with agents of the Leishmaniasis Program in Corrientes Province during 2015-2016, and another obtained through digitization from a recently published map (Acosta-Soto *et al*., 2020). Both datasets correspond to the same outbreak, but they were treated separately since we did not have access to more detailed information as to perform a proper anamnesis and check for repeated cases. Our original dataset consisted of 25 human cases reported as CL to the health authorities in Corrientes Province that were confirmed through visits to the patients' residences and characterized by the standardized clinical description and diagnosis procedures (Ministerio de Salud, 2007). All patients had typical CL ulcers and molecular analyses performed on samples from three patients confirmed *L. braziliensis* as the causative agent (H. Lucero, pers. com.). Furthermore, phlebotomine sampling was carried out in February 2017 for three consecutive nights using one CDC-like light traps active for 12 hours around the patient residences.

Acosta-Soto *et al*. (2020) reported 80 CL cases from the same outbreak but did not provide a list of coordinates, so to be able to compare the models with both datasets, we geo-referenced their map with satellite images and Open Street map data (https://planet.osm.org and https://www.openstreetmap.org) and digitized the CL cases reported. Since there seemed to be more than one CL case at some of the locations, we recovered 74 out of the 80 cases they analysed. Importantly, our original 25 CL cases roughly coincided in space and time with the core of the outbreak as reported by Acosta-Soto *et al*. (2020).

Remote sensing data

We downloaded three Landsat 8 scenes (path 226, row 79) corresponding to the dates 2014/07/19, 2015/05/19 and 2016/06/22. We selected cloud-free images roughly belonging to the same period of the year to minimize the effect of seasonal fluctuations. Images were drawn from the United States Geological Service (USGS) web site (https://earthexplorer.usgs.gov/) as level 1 TP products belonging to the tier 1 collection (Figure 1 in the Appendix).

The Landsat 8 images were atmospherically corrected to obtain surface reflectance. Since images showed no evident displacement across dates, co-registration was not needed. For each date we estimated several spectral indices: the normalized difference vegetation index (NDVI), the enhanced vegetation index (EVI), the normalized difference water index (NDWI), the land surface water index (LSWI) and the normalized difference built-up index (NDBI). These indices are commonly used to characterize vegetation cover, water content in plants and/or over surface and built-up features. Furthermore, we obtained different texture measures, such as contrast, correlation, variance and entropy based on EVI, LSWI and NDBI. Texture refers to the relation of grey level values among neighbouring pixels.

Data processing

All derived environmental information was used as input to perform a supervised classification using the Random Forest approach (Breiman, 2001), one of the most robust and commonly used machine learning classifiers. To perform supervised classifications for each date, we selected ground truth samples for 5 different classes: urban, water, bare soil, forest and low vegetation. Since some classes are subject to yearly changes, the ground truth data was obtained separately for each year (Table 1 in the Appendix). Sample selection was aided by very high resolution (VHR) imagery from Google® and Bing® within QGIS 3.10 (QGIS Development Team, 2019). The Random Forest classifier was run with 600 trees and default maximum number of features. The data were randomly split in five-folds to perform cross validation. All variables were standardized to a 0-1 scale before running the classification. In all cases, we obtained overall accuracy (OA) values in the range of 0.90-0.95. The F1 score for land cover classes varied between 0.81 and 0.98; the lowest values corresponding to urban and low vegetation in 2015 (Figure 2 in the Appendix).

To identify land cover changes we applied the Change Vector Analysis (CVA) algorithm (Malila, 1980), for which we used the Brightness and Greenness features of the tasselled cap transform performed over Landsat 8 bands (Liu *et al*., 2014). This analysis results in a map of angles and a map of magnitude of change, which are then combined into four classes from which we estimated distance maps. Paired t-tests were performed to confirm that the environment as seen by NDVI, LSWI and NDBI was indeed different among years where environmental changes were detected.

To extract environmental variables for sites with reported CL human cases, we created buffers with 50 and 250 m radii following Quintana *et al*. (2020). We estimated average and standard deviation (SD) for all variables described above as well as the most common type of change, the number of different land cover classes and the most common land cover class. Moreover we estimated the Simpson diversity index (Simpson, 1949) and the interspersion, a measure that reflects the percentage of cells that differ from the central cell or pixel. All remote sensing and processing using geographical information systems (GIS) were done and automated in GRASS GIS 7.8 (GRASS Development Team, 2020). All the variables derived from satellite images processing are described in more detail in Table 2 under Appendix.

Modelling and analysis

MaxEnt is a machine learning method that estimates the potential geographic distribution of biological phenomena by finding the probability distribution of maximum entropy (closest to uniform), subject to the constraint of the expected values of the environmental predictors (Phillips *et al*., 2006; 2017). MaxEnt was developed for presence-only data by contrasting presences against background locations (Phillips *et al*., 2006; Merow *et al*., 2013), and has been shown to outperform other algorithms, even when used with few positive records (Hernandez *et al*., 2006; van Proosdij *et al*., 2016). This algorithm has been widely used for species distribution modelling, but also for risk prediction of infectious diseases (Joshi and Miller, 2021), including CL (Chavy *et al*., 2019).

We used MaxEnt 3.4.1 (https://biodiversityinformatics.amnh.org/open_source/maxent/) to model the distribution of CL cases as a function of environmental variables and generate maps of favourable conditions for CL occurrence and potential risk. Since cases spanned over a period of approximately two years and we assumed that they are triggered by previous environmental changes, we grouped the environmental variables in four sets according to buffer size and year, *i.e*. 2014-50 m, 2014-250 m, 2015-50 m and 2015-250 m. Sets for 2014 included environmental

changes detected between 2014 and 2015, before the CL outbreak, and sets for 2015 included environmental changes during the outbreak between 2015 and 2016. Given that the number of variables in each set was larger than the number of reported cases in our dataset, we performed an *a priori* removal of highly correlated variables (r>0.85) using the method proposed by Naimi *et al*. (2014), *i.e*. identifying pairs of highly correlated variables and removing the one with the highest variance inflation factor (VIF). We ran the MaxEnt algorithm with the remaining variables and with default settings (500 iterations, regularization =1 and allowing linear *l*, quadratic *q*, product *p* and hinge *h* feature classes). We used 10,000 random background points that we obtained after masking a 300-m radius area from each CL case in both datasets. We created 4 random folds to perform training with cross-validation. Since the datasets differed in number and location of cases, the random background points were obtained separately. Once the default models were fitted, we proceeded to perform feature selection. We first removed remaining correlated variables (r>0.7) via Jackknife tests. Second, also via Jackknife tests, we removed variables with importance values lower than 10%. In both cases, the model performance was checked with regard to removing or keeping a variable and this was based on the area under the receiver operating characteristic (ROC) curve, the AUC.

After obtaining models with the best set of important and uncorrelated environmental variables, we used a genetic algorithm to perform hyper-parameter tuning with optimization. This process allows to adjust the complexity of models built with MaxEnt through the inclusion of additional feature classes (*i.e*. transformations of the original predictor variables), as well as with a regularization multiplier that contributes to select those features and to reduce overfitting (Merow *et al*., 2013). We allowed regularization and feature classes to vary and we assessed the overall discrimination ability of each model on the basis of the train and test AUC. Regularization was allowed to vary from 0.2 to 5 in steps of 0.2 and the feature classes combinations tested were: l, lq, lh, lp, lqp and lqph. From this step, we obtained one optimized model per variable set, *i.e*. 2014-50 m, 2014-250 m, 2015-50 m and 2015-250 m. These models were finally run using the best set of hyper-parameters obtained and projected over the study area to obtain suitability maps for the occurrence of CL. The procedure described was followed separately for the two CL datasets. All the modelling was done in R with the *SDMtune* package (Vignali *et al*., 2020).

Model evaluation and ensemble

We performed a cross model evaluation, *i.e*. we used CL cases described by Acosta-Soto *et al*. (2020) to evaluate performance of models trained with our original 25 points and vice-versa. For each data set, we generated 250 extra points to be used as background in the evaluation. We extracted the predicted probability for the pixel beneath the presence or background coordinates plus the four neighbouring pixels to minimize potential geo-location and projection errors. We used both threshold dependent and independent measures. The AUC was used as a threshold independent measure. For the threshold dependent evaluation, we first determined the

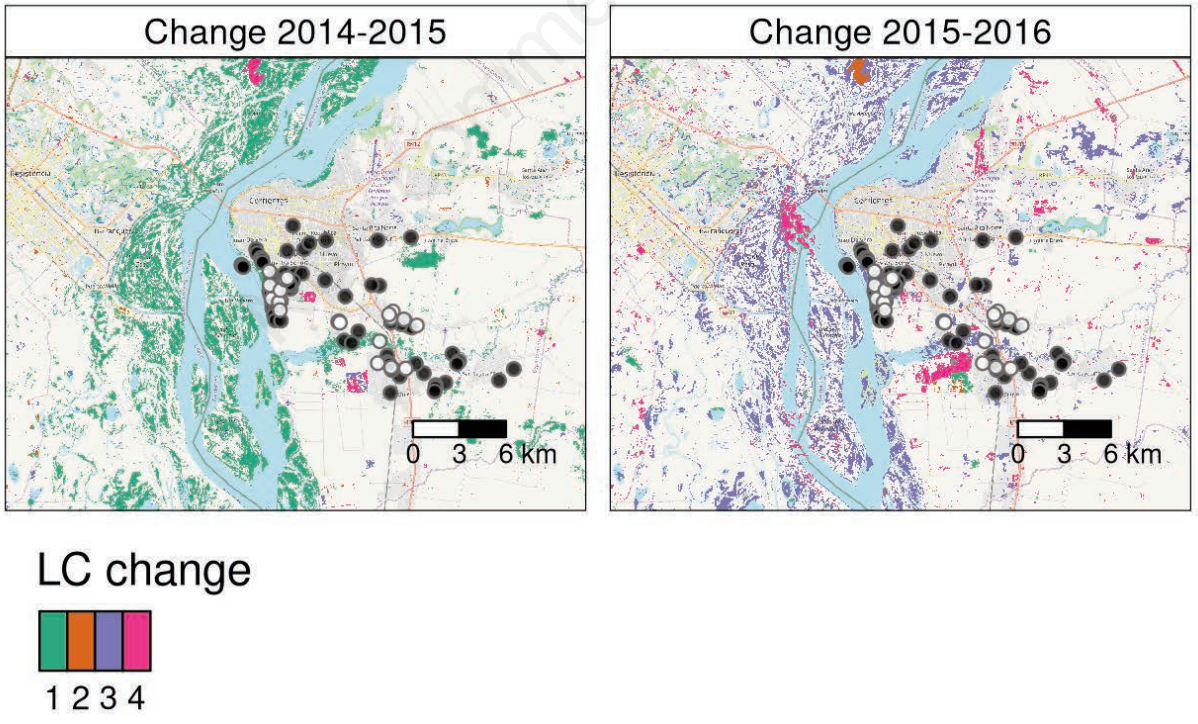

Figure 2. Land cover changes for the city of Corrientes and its surroundings during 2014-2016. Environmental changes detected by Change Vector Analysis between 2014 and 2015 (left) and 2015 and 2016 (right). White dots correspond to our dataset (n=25) with black dots representing cutaneous leishmaniasis cases, (n=74) digitized from the report by Acosta-Soto *et al.* (2020). Land cover (LC) change classes are as follows: 1=moisture reduction, 2=chlorophyll increase, 3=moisture increase, 4=bare soil increase.

cut-off values using common criteria: minimum occurrence prediction, mean occurrence prediction, 10% omission, sensitivity=specificity and maximum sensitivity plus specificity (Liu *et al*., 2005, 2013). For each threshold we estimated different model accuracy metrics (Table 3 in the Appendix). The threshold-dependent evaluation was performed with the R *SDMTools* package (VanDerWal *et al*., 2019). We used the maximum sensitivity plus specificity cut-off values to create binary maps and compare predicted areas for CL occurrence in models trained with different datasets. To obtain a final probability map, we performed model ensemble by averaging the two best models' predictions weighted by their AUC values (Araújo and New, 2007).

Results

The CL cases reported by Acosta-Soto *et al*. (2020) cover a larger area than those surveyed with Corrientes health authorities during 2015-2016. In general, the types of land cover represented in each CL dataset are similar but Acosta-Soto *et al*. (2020) report some more over the urban and peri-urban fabric (Figure 1). In any case, both datasets show a similar spatial disposition surrounding environmental changes detected in the south of Corrientes City and Riachuelo (Figure 2). In total, approximately 17% of the study area suffered some type of environmental change (12% in the period 2014-2015 and 11% in the period 2015-2016). Clearly, the largest change detected was related to riparian areas, which from 2014 to 2015 became drier and, from 2015 to 2016, flooded. Another relevant change detected was the installation of a new urbanization in the south of Corrientes City (a green area that becomes highly reflective). All classes of changes identified showed significant differences between the years with respect to NDVI, NDBI and LSWI when compared through paired t-tests (Figures 3 and 4 in the Appendix). Figure 3 shows the counts of the three most captured phlebotomine species at the 25 locations with CL cases reported by the health authorities. The most common sand fly species was *Ny. neivai*, a known vector of *L. braziliensis*. Indeed it was found at 18 out of 25 sites, while *Lu. longipalpis* was only found at 5 sites.

The final MaxEnt models obtained after variable selection and calibration are shown in Figures 4 and 5. Models fitted with our original dataset (n=25) showed very low probabilities of CL occurrence in the core urban area of Corrientes City. In fact, the highest probabilities seemed to be related to the outskirts of the city and areas close to detected environmental changes (Figure 2). Models using environmental variables derived from 2014 and 2014-2015 predicted much smaller areas of high probability for presence of CL compared to the reported by Acosta-Soto *et al*. (2020). Indeed, predictive maps obtained from models fitted with the dataset digitized from this report (n=74), showed large areas with high probabilities, and these were sometimes observed in the core urban area (Figure 5).

Table 2 summarizes the cross-validation performance measures and the best hyper-parameter settings resulting from calibration after variable selection for the two CL datasets considered. The predictive performance of final models (based on the training data set) was very good in terms of AUC. In general, AUC values were higher for models fitted with our original dataset and models were also simpler in terms of the number of feature classes included. The difference in training and testing AUC within the cross-

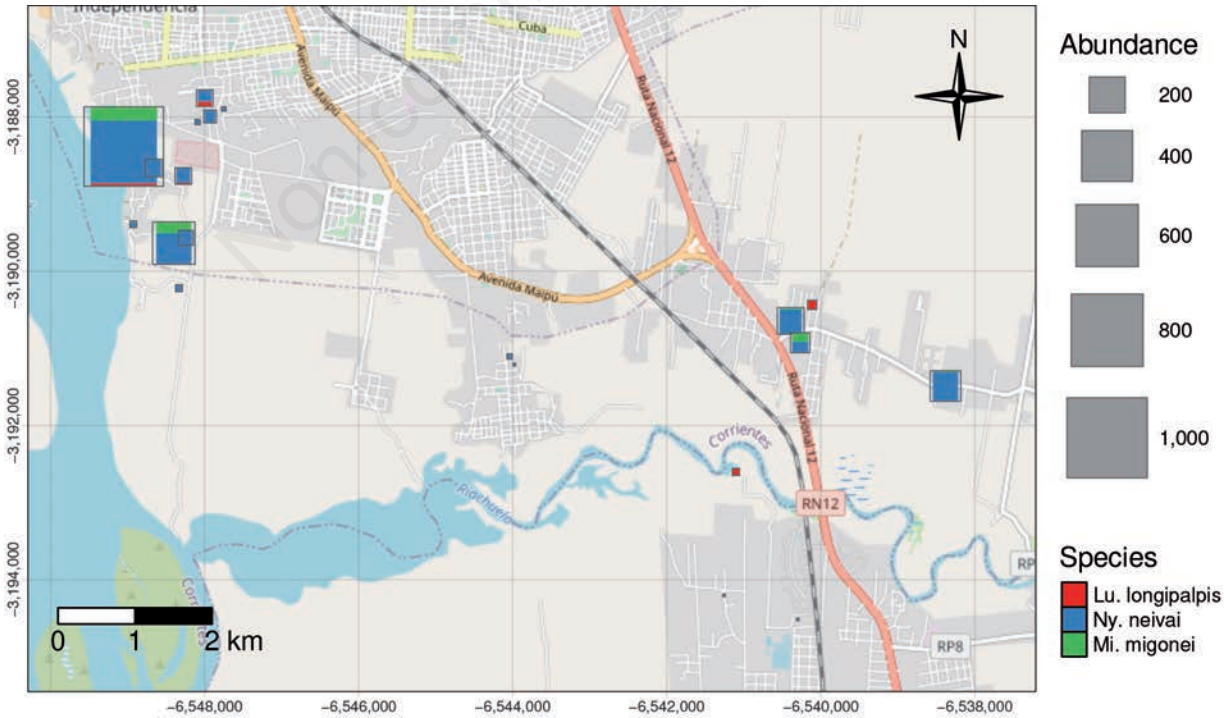

Figure 3. Abundance and composition of the three most common vector species in the 25 locations from where cases of cutaneous leishmaniasis have been reported.

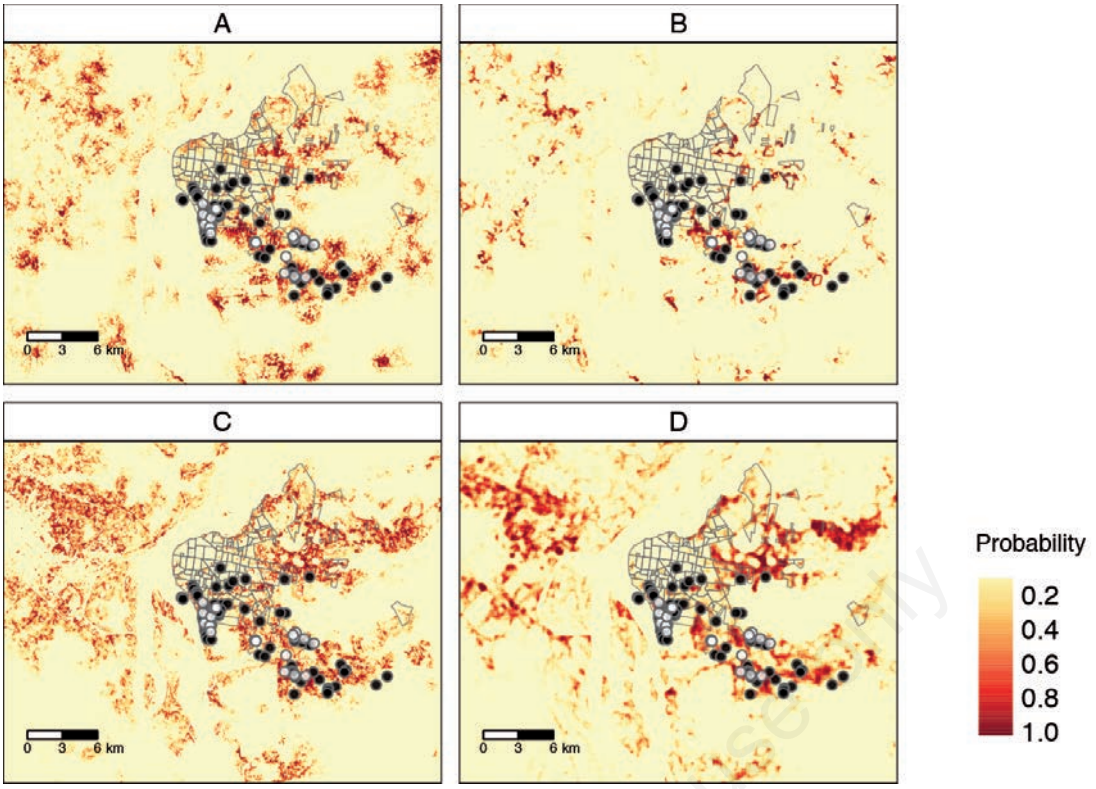

Figure 4. Predictive maps obtained from the best models built with 25 cutaneous leishmaniasis cases for 2014-2015 and 2015-2016 for areas of various sizes. Time and radii of areas investigated: A) 2014-2015-50 m; B) 2014-2015-250 m; C) 2015-2016-50 m; D) 2015-2016-250 m.

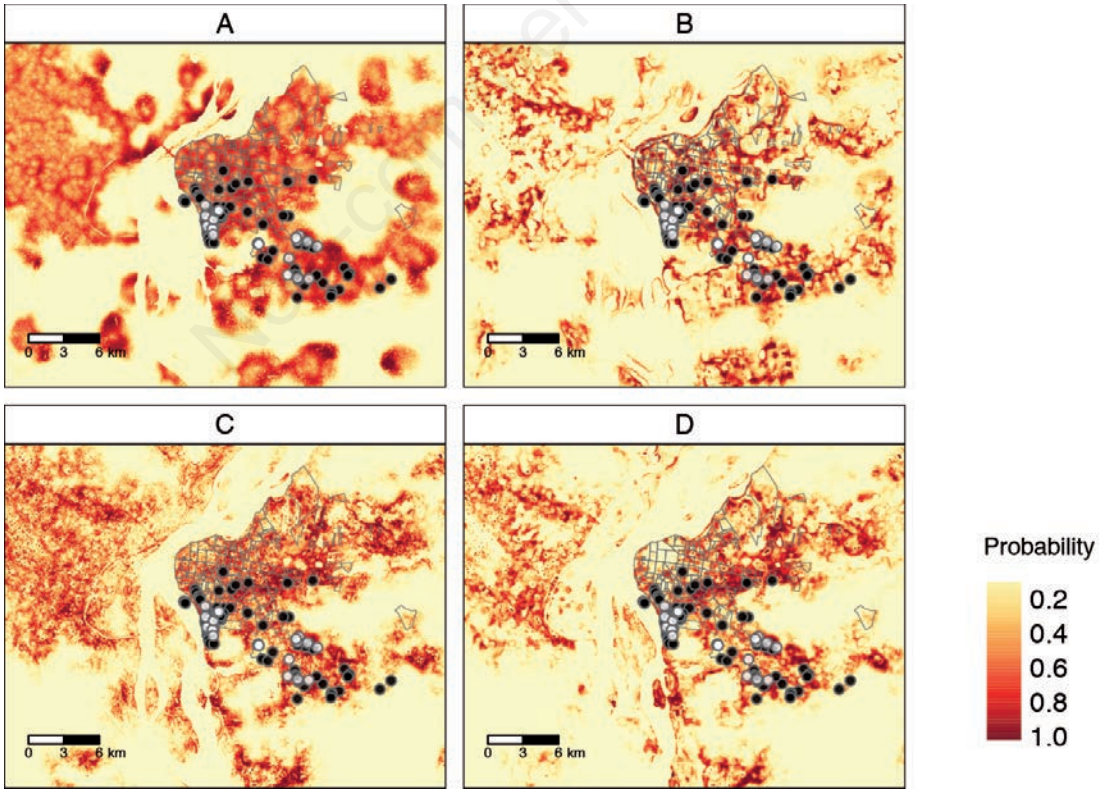

Figure 5. Predictive maps obtained from the best models built with 74 cutaneous leishmaniasis cases for 2014-2015 and 2015-2016 for areas of various sizes. Time and radii of areas investigated: A) 2014-2015-50 m; B) 2014-2015-250 m; C) 2015-2016-50 m; D) 2015-2016-250 m.

validation repetitions was also lower for models built with only 25 data points.

The distance to environmental changes appeared in all models and was the most important variable in 7 of the 8 best models (importance values ranging from 25% to more than 40%). The different environmental changes detected were related to changes in vegetation cover (increase or decrease), changes in the brightness of land covers (*i.e.* from vegetated areas to bare soil or built-up) and changes related to water surface or water content (especially flooding in riparian areas). Other relevant variables were interspersion of mode and texture measures, such as contrast or entropy obtained from different spectral indices (see Figures 5 and 6 in the Appendix).

Figures 6 and 7 show the type of relationship between the most important variable in each model and the predicted probability of occurrences. For models trained with n=25 and environmental data

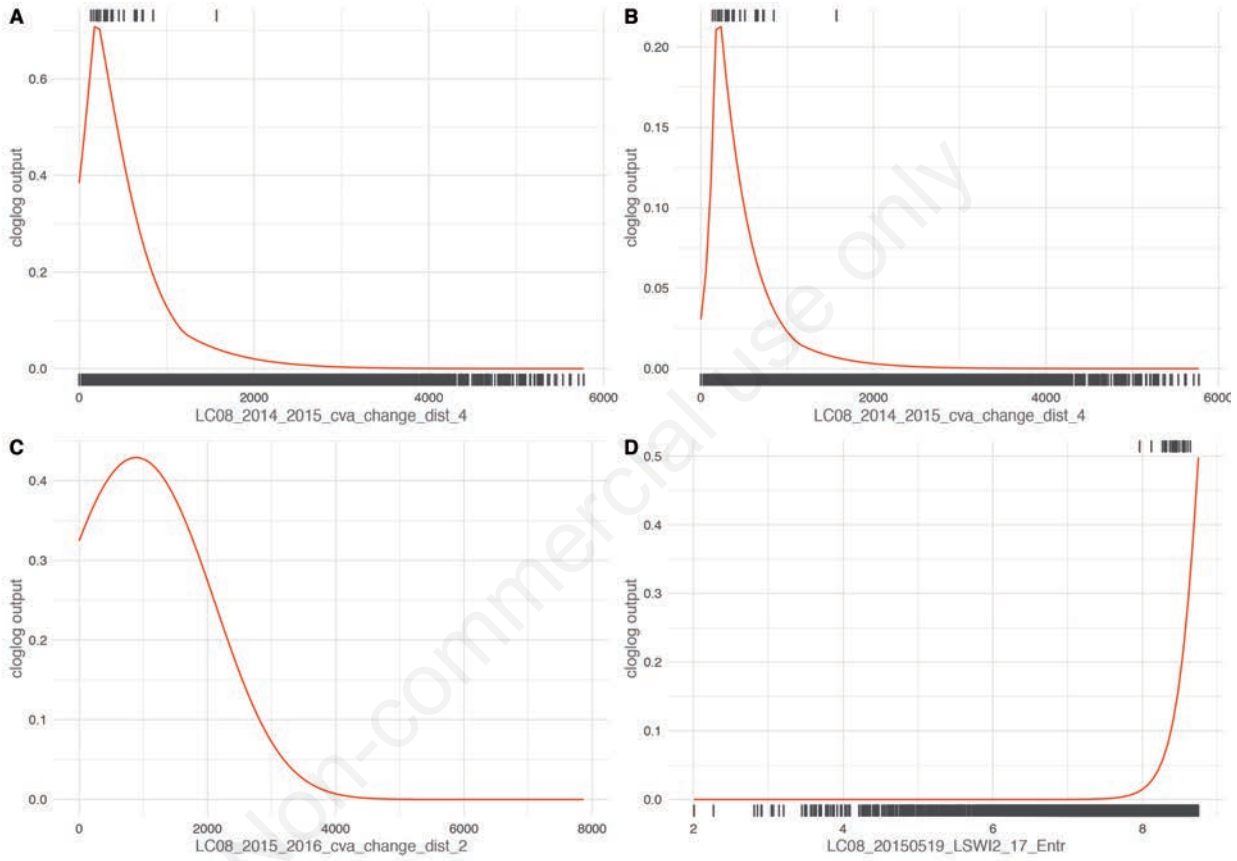

Figure 6. Response curves for the most important variable in models trained with 25 cutaneous leishmaniasis cases for 2014-2015 and 2015-2016 with variables obtained from areas of various sizes. Time and radii of areas from which variables were obtained: A) 2014-2015-50 m; B) 2014-2015-250 m; C) 2015-2016-50 m; D) 2015-2016-250 m.

Table 2. Performance and hyper-parameters of the best models built with n=25 and n=74.

Model	Feature class	Regularization	Train AUC	Test AUC	Diff. AUC
N=25					
2014_50 m	lqph	1.80	0.967	0.937	0.031
2014_250 m	lh	1.00	0.979	0.959	0.020
2015_50 m	lq	0.40	0.950	0.922	0.029
2015_250 m	lqp	0.80	0.944	0.933	0.011
N=74					
2014_50 m	lqp	0.40	0.842	0.807	0.034
2014_250 m	lqph	0.40	0.907	0.873	0.034
2015_50 m	lqph	0.80	0.891	0.837	0.054
2015_250 m	lqph	0.40	0.920	0.870	0.050

Models calibrated with k-fold cross-validation (k=4). AUC, area under the receiver-operator curve; l, linear; q, quadratic; p, product; h, hinge.

from 2014-2015, the most important variable displayed the same type of response regardless of using 50 or 250 m buffer sizes, *i.e.* the probabilities were highest at 250-300 m from an environmental disturbance. Something similar happens in panel C (Figure 6), where the probability of CL occurrence increases up to a distance close to 1000 m and then decreases again.

In the case of models trained with n=74, the response curve displayed high values close to the change, which decreased to rapidly increase towards a distance around 400-700 m and then decreased again in three out of the four models (Figure 7). Response curves for all variables retained as important in the eight best models are included in the Appendix (Figures 7-14).

The models that were trained with the dataset derived from the work by Acosta-Soto *et al*. (2020) showed better AUC values when evaluated with n=25 than those trained with n=25 and evaluated with n=74 (Table 3). This is clearly explained by the spatial distribution of cases in one set compared with the other, *i.e.* the distribution of CL cases reported by Acosta-Soto *et al*. (2020) contained that of the 25 cases reported by the Corrientes health authorities and included some more urban and peri-urban occurrences that the model trained with n=25 did not 'learn' and hence, could not predict. Receiver-operator curves (ROC) are shown in Figure 15 in the Appendix.

The performance of models as evaluated with threshold dependent metrics was again better for models trained with 74 CL cases (Table 4). We chose the maximum sensitivity + specificity criteria to build presence/absence maps since it provided, in general, low omission errors and acceptable overall accuracy values, while maximizing both sensitivity and specificity metrics. The table with all the criteria evaluated and performance metrics is included in the Appendix (Table 3).

The cut-off thresholds were much lower for models trained with n=25 and evaluated with n=74 than for those trained with 74 CL cases and evaluated with 25. The former were in the 0.02-0.04 range while the latter spanned 0.33-0.54. The effect of applying these thresholds can be compared in Figure 16 in the Appendix. In general, because of the low thresholds applied to predictions from

Table 3. Evaluation measures of the best models fitted for two sample sizes (n=25 and n=74).

Model/Period	Test AUC (n=74)	Test AUC (n=25)
2014_50 m	0.743	0.851
2014_250 m	0.724	0.912
2015_50 m	0.828	0.905
2015_250 m	0.800	0.915

Models fitted with n=25 were evaluated with the 74 CL cases digitized from the report by Acosta-Soto *et al.* (2020) (middle column). Models fitted with n=74 were evaluated with the 25 cases reported as CL by the Corrientes Health Authorities (right column).

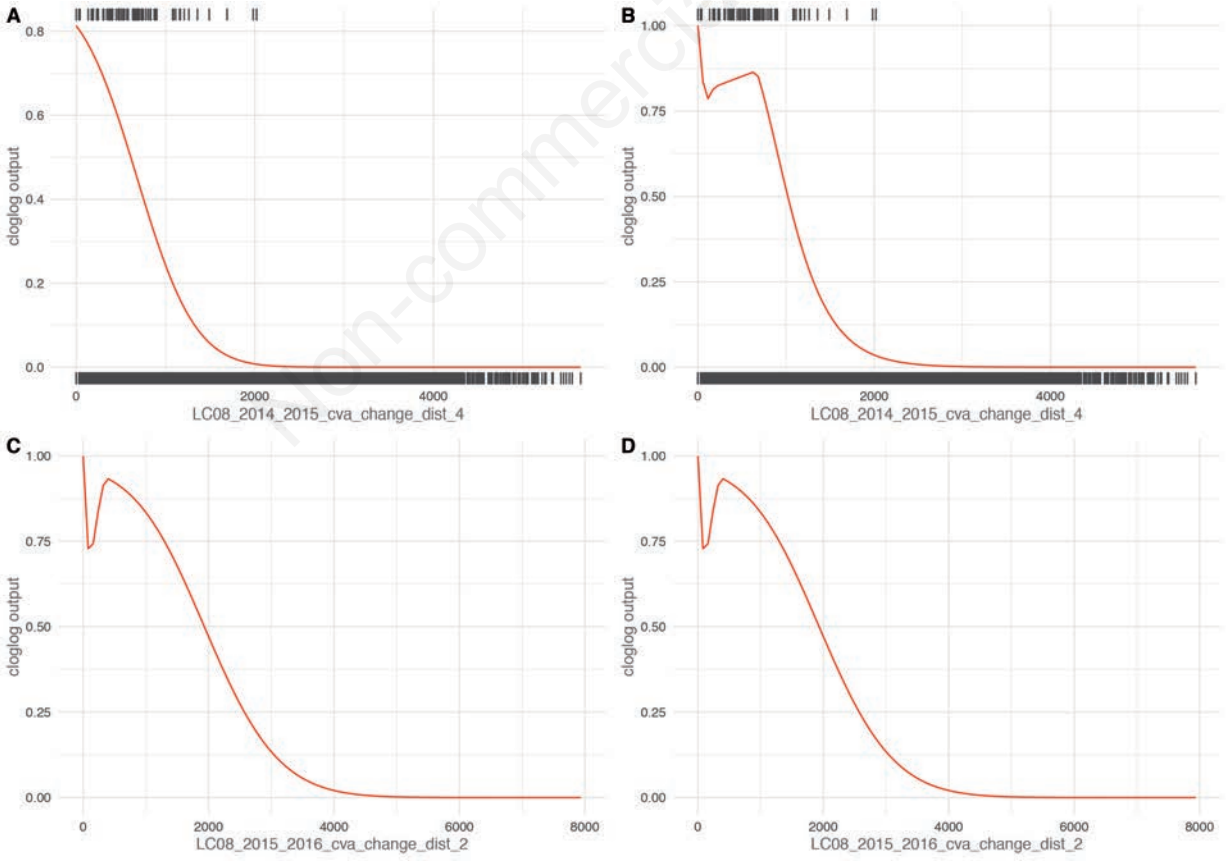

Figure 7. Response curves for the most important variables in models trained with 74 cutaneous leishmaniasis cases for 2014-2015 and 2015-2016 with variables obtained from areas of various sizes. Time and radii of areas from which variables were obtained: A) 2014-2015-50 m; B) 2014-2015-250 m; C) 2015-2016-50 m; D) 2015-2016-250 m.

models trained with n=25, the areas of CL occurrence were larger than those resulting after applying thresholds to models trained with n=74. In any case, the pattern/shape of the predicted CL occurrence area was similar.

The model trained with n=74 using environmental variables from 2015-2016 in buffers of 250 m can be considered the best according to threshold dependent and independent evaluation metrics (Tables 3 and 4). The best model for the dataset consisting of 25 CL cases, also included variables from 2015 - 2016, but with buffer areas of 50 m instead. This later model had lower AUC and lower threshold-related metrics than the former. However, to consider uncertainty in input data and predictions, we built a final ensemble map by averaging the predictions of both models weighted by their respective testing AUC values (Figure 8). The weighted average map maintained high resemblance with the best model prediction, denoting higher suitability for CL occurrence in the outskirts of Corrientes city and areas close to environmental changes. The models, though trained with different datasets, displayed a good degree of agreement according to the standard deviation (Figure 8). The areas with greater level of uncertainty included the south-eastern part of Corrientes City and the Northeast of Riachuelo.

Discussion

In this study we looked for the environmental determinants associated with the 2015-2016 CL outbreak in Corrientes City and its surroundings and explored the capacity of remote sensing image analysis and ENM to predict the potential distribution of CL outbreaks. Acosta-Soto *et al*. (2020) have recently proposed that the outbreak under study was an atypical CL urban outbreak due to *L. infantum*; hence displaying a spatial pattern consistent with VL instead of typical CL. Here, we have attempted to elucidate the typical or atypical eco-epidemiological pattern of this CL outbreak given its implications for public health strategies.

In general, the models that we trained with the two datasets differed mainly in the variables that we identified as important and regarding the extent of highly suitable areas predicted (Figures 4 and 5). In any case, the distance to different environmental changes appeared either as the most important variable or in the top five most important variables in all models. This is highly consistent with typical CL outbreaks in peri-urban and rural areas after disturbances such as deforestation, urbanization or flooding (Salomón *et al*., 2006a,b). Indeed, these are the main types of environmental changes that we detected through remote sensing image analysis (Figure 2). Even small scale changes within peri-urban areas such as ground movements for a new building or park might create favourable conditions for increased vector abundance in endemic areas and hence a potentially higher risk of transmission (Quintana *et al*., 2010; Gouveia *et al*., 2012). Change of class 2 in models from 2015-2016 seems to be reflecting these types of small modifications and appeared to be a highly important variable in three of the best models that we obtained. Other variables retained in the best models were related to vegetation cover, water or humidity and habitat heterogeneity in the areas surrounding the CL cases. These associations are consistent with previous works that addressed the relationship of CL vector species with NDVI (Chanampa *et al*., 2018) and landscape changes (Quintana *et al*., 2010) in north-western Argentina and elsewhere (Wasserberg *et al*., 2003) at equivalent spatial scales.

The timing of variables in the best models (according to the evaluation AUC, Table 3) is consistent with the fact that the core of the outbreak occurred mainly in the second half of 2015 and, to a lesser extent, in the first half of 2016 (Figure 1 in Acosta-Soto *et al*., 2020). Hence, variables recorded before the epidemic period as well as environmental changes detected between May 2015 and June 2016, were better predictors than those from one year before the outbreak, according to the lags between environmental variables-vector population dynamics and incubation periods (Ferreira de Souza *et al*., 2015; Talmoudi *et al*., 2017; Gutiérrez-Torres, 2020).

The best model according to the evaluation AUC values and threshold dependent metrics was found to be one including variables measured at 250 m radii areas (Table 3). A previous study found that both *Lu. longipalpis* and *Ny. whitmani*'s abundance was related to environmental variables recorded at this scale (Quintana *et al*., 2020). Hence, this model and the predictive map generated from it (Figure 5D), might indirectly explain the presence of the vectors of both *L. infantum* and *L. braziliensis*. The phlebotomine trapping performed close to the 25 cases, however, showed that the most abundant species in such sites was *Ny. neivai* (Figure 3). Moreover, previous studies carried out in the city of Corrientes during the CL outbreak (Berrozpe *et al*., 2017, 2019), show that *Lu. longipalpis* is dominant only in the most urbanized parts of the city, while *Ny. neivai*, the vector associated with typical CL, is the dominant species in the peri-urban and rural areas where most of the reported CL cases occur. Even though the vectors of CL have been found to be infected with *L. infantum* elsewhere (Moya *et al*., 2017; Thomaz-Soccol *et al*., 2018), their role as vectors of this par-

Table 4. Performance metrics obtained for the threshold criteria maximum sensitivity plus specificity for the best models fitted for two sample sizes (n=25 and n=74).

Model	Threshold value	AUC	Omission rate	Sensitivity	Specificity	Overall accuracy
N=25						
2014_50 m	0.040	0.702	0.257	0.743	0.660	0.679
2014_250 m	0.020	0.697	0.419	0.581	0.812	0.759
2015_50 m	0.030	0.766	0.135	0.865	0.668	0.713
2015_250 m	0.040	0.745	0.203	0.797	0.692	0.716
N=74						
2014_50 m	0.540	0.806	0.160	0.840	0.772	0.778
2014_250 m	0.330	0.862	0.040	0.960	0.764	0.782
2015_50 m	0.450	0.866	0.040	0.960	0.772	0.789
2015_250 m	0.340	0.880	0.040	0.960	0.800	0.815

AUC, area under the receiver-operator curve.

asite is not yet clear and the main vector of *L. infantum* in the region is still believed to be *Lu. longipalpis*, which has not been associated focally with deforestation or flood events. Instead, VL-*Lu. longipalpis* have spread regionally along with human massive migrations and unplanned urbanizations associated with developmental projects producing large-scale environmental changes, such as constructions involving gas pipelines or highways (Correa Antonialli *et al*., 2007; Pasquali *et al*., 2019).

Acosta-Soto *et al*. (2020) described the CL outbreak they reported on as urban, however, only one of the cases reported by them might fully fit this description when observing its position regarding the distribution and density of shops and offices in the city (see Figure 17 in Appendix and https://ide.corrientes.gob.ar/). All the other CL cases were mainly located in peri-urban or mostly rural areas (Berrozpe *et al*., 2017), more coincident with typical CL transmission scenarios, while isolated cases of CL could have urban residences but having been infected elsewhere. Furthermore, if only one sample was sequenced, as seems to be concluded from the main limitation described in the discussion of the above-mentioned article, this single case could be the urban one. That it would possibly be the only one from a suspected VL case with marrow aspirate, strengthens this supposition.

The use of human cases for predictive modelling has certain limitations since the exact location where transmission occurred is hardly ever known and usually the patient residential address is recorded instead. In an endemic area this bias might be reduced, but as shown here, it might determine the difference between an urban atypical VL outbreak and a typical peri-urban/rural CL outbreak. Furthermore, as with most machine learning modelling approaches, MaxEnt is constrained by the input occurrence data. Hence, when projected over space, it will predict higher suitability in areas with environmental characteristics similar to those of input presences. This might be the reason why models trained with n=25 yielded very high AUC values in the training with cross-validation but a poor performance when evaluated with the data from the report by Acosta-Soto *et al*. (2020). In any case, models trained with n=74 do not show much higher AUC values when evaluated with our CL dataset. In order to address these caveats, we performed a weighted average with the best model from each dataset to obtain a final ensemble prediction of suitability for CL occurrence. Moreover, to account for uncertainty among these two models' predictions, we estimated the map of standard deviation, which showed that the models were mostly congruent. Importantly, none of the models predicted high CL probabilities in the core urban area.

Although several previous studies have associated CL vectors and cases to climatic and environmental variables at different scales (Gomez-Bravo *et al*., 2017; Chanampa *et al*., 2018; Berrozpe *et al*., 2019; Chavy *et al*., 2019; Valero and Uriarte, 2020), this is, to the best of our knowledge, the first time that environmental changes detected by means of automatic satellite image analysis have been used as predictors of CL cases. Since fine scale local changes might affect the focal distribution of disease vectors, such as sand flies, further refinements in terms of spatial resolution could be carried out with Sentinel 2 data and time series.

The conclusions of the study by Acosta-Soto *et al*. (2020) deserve confirmation because the eco-epidemiological and clinical pattern of the outbreak is consistent with CL (open ulcers, cases clustered in space and time, in recently modified environments, close to forests and more prevalent in males (as they are involved in activities within primary forests). Importantly, it is not consistent with what we know about ACL in the region (few small non-ulcerated papules, histologically different from typical CL), scattered cases more frequent in 5- to15-year olds and in females (or almost equally distributed between sexes). Moreover, in addition to the information on the identification of *L. braziliensis* in patients from the same outbreak, the literature on ACL cases in the Americas molecular diagnosis uses reference strains of *L. infantum*, *L. braziliensis*, *L. mexicana* or *L. amazonensis*. The inclusion of *L. amazonesis* as a control is essential, as this species belongs to the subgenus *Leishmania* like *L. infantum*, produces CL typical skin lesions in the southern states of Brazil (Silveira *et al*., 1990) and was isolated from free-ranging monkeys in the Corrientes-Riachuelo outbreak area (Martinez *et al*., 2020). However, only *L.*

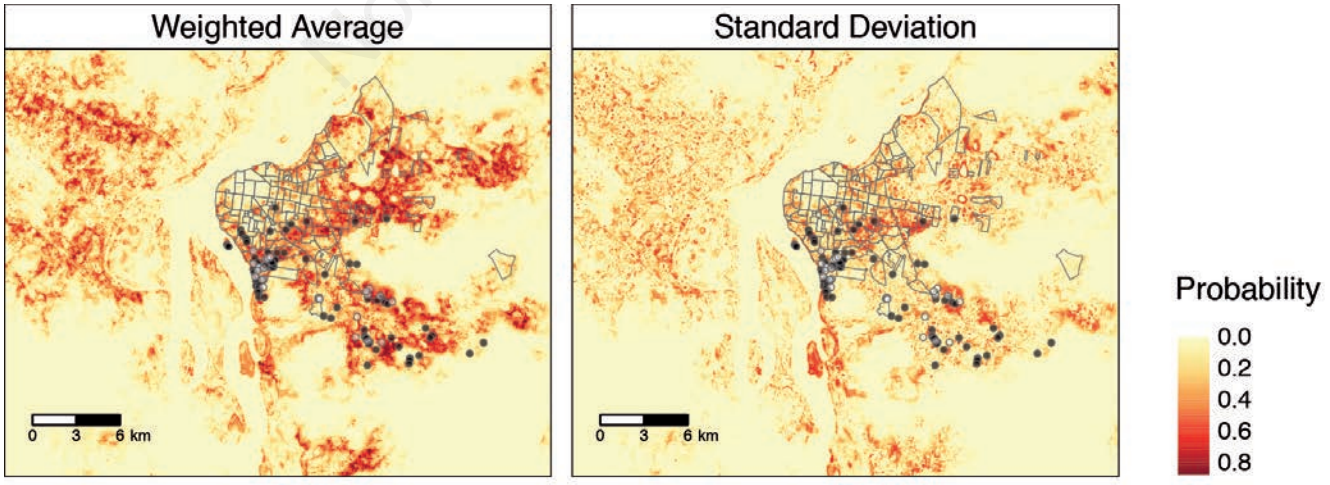

Figure 8 Weighted average predicted suitability for the occurrence of cutaneous leishmaniasis and standard deviation based on the selected best models.

infantum controls were used in Acosta-Soto *et al*. (2020), so mixed infections cannot be ruled out. Furthermore, the procedures performed during sample collection, slide preservation and processing to avoid potential cross-contamination, as explained in detail in similar publications, were not described, nor was the Genbank accession number of the sequences obtained, which would allow traceability and evaluation of the percentage of homology of the sequences with *L. braziliensis* and *L. amazonensis*. Indeed, due to the selection force exerted on the internal transcribed spacer (ITS)-1 sequence, where what is relevant is the length of the DNA fragment and some specific DNA bases linked to RNA splicing, and not the total sequence *per se*, there is a reasonable doubt that mutations occurring within this fragment would not affect biological functionality but could generate different percentages of sequence homology.

In relation to the CL outbreak scenario or description of atypical cases, as already indicated, Acosta-Soto *et al*. (2020) clarify that it was possible 'to conduct a search for the parasite in only one of the samples out of all of the cases diagnosed, because of the unavailability of slices' and that 'of the 81 patients infected during the outbreak, 80 were diagnosed based on direct swabs of tissue taken from the lesions, whereas bone marrow aspirate and serological studies via immunochromatography were used in the remaining patient'. However, the latter procedures are not recommended for patients without suspected VL.

Conclusions

Given the environmental changes detected and their relevance to explain the outbreak, the peri-urban and rural distribution of CL cases (either surrounding or being surrounded by environmental changes) and the higher abundance of CL vectors in the area where cases occurred, the scenario is mostly consistent with a typical CL outbreak (Salomón, 2019). To identify an apparently typical CL outbreak as cases of atypical CL due to *L. infantum*, as it has been suggested and potentially extrapolated to also include other CL outbreaks, would imply a huge increase in costs for local control programmes and health systems all over the continent. Hence, given that the CL outbreak under study was identified as typical by means of clinical and molecular diagnose and also appears to be typical regarding its spatial and environmental association, the suggestion that Corrientes 2015-2016 outbreak might have been caused by *L. infantum*, requires clinical, epidemiological and parasitological re-evaluation and confirmation.

References

Acosta-Soto L, Encinas ES, Deschutter EJ, Pasetto RAL, Petri de Odriozola EMA, Bornay-Llinares FJ, Ramos-Rincón JM, 2020. Autochthonous Outbreak of Cutaneous Leishmaniasis due to Leishmania infantum in Corrientes Province, Argentina. Am J Trop Med Hyg 102:593-7.

Arana MD, Martínez GA, Oggero AJ, Natale ES, Morrone JJ, 2017. Map and Shapefile of the Biogeographic Provinces of Argentina. Zootaxa 4341:420-2.

Araújo MB, New M, 2007. Ensemble forecasting of species distributions. Trends Ecol Evol 22:42-7.

Berrozpe P, Lamattina D, Santini MS, Araujo AV, Utgés ME, Salomón OD, 2017. Environmental suitability for Lutzomyia longipalpis in a subtropical city with a recently established visceral leishmaniasis transmission cycle, Argentina. Mem I Oswaldo Cruz 112:674-80.

Berrozpe PE, Lamattina D, Santini MS, Araujo AV, Torrusio SE, Salomón OD, 2019. Spatiotemporal dynamics of Lutzomyia longipalpis and macro-habitat characterization using satellite images in a leishmaniasis-endemic city in Argentina. Med Vet Entomol 33:89-98.

Breiman L, 2001. Random Forests. Machine Learning 45:5-32.

Bruhn FRP, Morais MHF, Cardoso DL, Bruhn NCP, Ferreira F, Rocha CMBM, 2018. Spatial and temporal relationships between human and canine visceral leishmaniases in Belo Horizonte, Minas Gerais, 2006-2013. Parasite Vector 11:372.

Chanampa M del M, Gleiser RM, Hoyos CL, Copa GN, Mangudo C, Nasser JR, Gil JF, 2018. Vegetation Cover and Microspatial Distribution of Sand Flies (Diptera: Psychodidae) in an Endemic Locality for Cutaneous Leishmaniasis in Northern Argentina. J Med Entomol 55:1431-9.

Chavy A, Nava AFD, Luz SLB, Ramírez JD, Herrera G, Santos TV dos, Ginouves M, Demar M, Prévot G, Guégan J-F, Thoisy B de, 2019. Ecological niche modeling for predicting the risk of Cutaneous Leishmaniasis in the Neotropical moist forest biome. PLOS Neglect Trop D 13:e0007629.

Correa Antonialli SA, Torres TG, Paranhos Filho AC, Tolezano JE (2007) Spatial analysis of American Visceral Leishmaniasis in Mato Grosso do Sul State, Central Brazil. J Infection 54:509-14.

Fernández MS, Santini MS, Cavia R, Sandoval AE, Pérez AA, Acardi S, Salomón OD, 2013. Spatial and temporal changes in Lutzomyia longipalpis abundance, a Leishmania infantum vector in an urban area in northeastern Argentina. Mem I Oswaldo Cruz 108:817-24.

Ferreyra de Souza RA, Andreoli RV, Kayano MT, Carvalho AL, 2015. American Cutaneous Leishmaniasis cases in the metropolitan region of Manaus, Brazil: association with climate variables over time. Geospatial Health 10:40-7.

Gómez-Bravo A, German A, Abril M, Scavuzzo M, Salomón OD, 2017. Spatial population dynamics and temporal analysis of the distribution of Lutzomyia longipalpis (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in the city of Clorinda, Formosa, Argentina. Parasite Vector 10:352.

Gouveia C, de Oliveira RM, Zwetsch A, Motta-Silva D, Carvalho BM, de Santana AF, Rangel EF, 2012. Integrated Tools for American Cutaneous Leishmaniasis Surveillance and Control: Intervention in an Endemic Area in Rio de Janeiro, RJ, Brazil. Interdiscipl Perspect Infect Dis 2012:568312.

GRASS Development Team, 2020. Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.8. Open Source Geospatial Foundation. Available from: http://grass.osgeo.org

Gutiérrez-Torres JD, 2020. Temporal lagged relationship between a vegetation index and cutaneous leishmaniasis cases in Colombia: an analysis implementing a distributed lag nonlinear model. Parasitol Res 119:1075-82.

Hernandez PA, Graham CH, Master LL, Albert DL, The ADL, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 5:773-85.

Hotez PJ, 2018. Human Parasitology and Parasitic Diseases: Heading Towards 2050. Adv Parasit 100:29-38.

INDEC, 2010. Censo Nacional de Población, Hogares y Viviendas. Instituto Nacional de Estadística y Censo. Available from: https://www.indec.gob.ar/indec/web/Nivel4-Censo Provincia-3-999-18-021-2010

Joshi A, Miller C, 2021. Review of machine learning techniques for mosquito control in urban environments. Ecol Inform 61:101241.

Liu C, Berry PM, Dawson TP, Pearson RG, 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28:385-93.

Liu C, White M, Newell G, 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40:778-89.

Liu Q, Liu G, Huang C, Liu S, Zhao J, 2014. A tasseled cap transformation for Landsat 8 OLI TOA reflectance images. 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, pp 541-544.

Malila WA, 1980. Change vector analysis: an approach for detecting forest changes with Landsat. LARS Symposia 326-335.

Martínez MF, Kowalewski MM, Giuliani MG, Acardi SA, Salomón OD, 2020. Molecular characterization of Leishmania species in free-ranging howler monkeys in northeastern Argentina. Acta Trop 210:105534.

Merow C, Smith MJ, Silander JA, 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36:1058-69.

Ministerio de Salud, 2007. Manual de normas y procedimientos de Vigilancia y Control de Enfermedades de Notificación Obligatoria. Ministerio de Salud de la República Argentina, Buenos Aires, Argentina.

Moya SL, Giuliani MG, Santini MS, Quintana MG, Salomón OD, Liotta DJ, 2017. Leishmania infantum DNA detected in phlebotomine species from Puerto Iguazú City, Misiones province, Argentina. Acta Trop 172:122-4.

Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG, 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37:191-203.

PAHO, 2019. Leishmaniases. Epidemiological Report of the Americas, December 2019. PAHO, Washington, D.C.

Pasquali AKS, Baggio RA, Boeger WA, González-Britez N, Guedes DC, Chaves EC, Thomaz-Soccol V, 2019. Dispersion of Leishmania (Leishmania) infantum in central-southern Brazil: Evidence from an integrative approach. PLOS Neglect Trop D 13:e0007639.

Phillips S, Anderson R, Schapire R, 2006. Maximum Entropy Modeling of Species Geographic Distributions. Ecoll Model 190:231-59.

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME, 2017. Opening the black box: an open-source release of Maxent. Ecography 40:887-93.

QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation. Available from: http://qgis.osgeo.org

Quintana MG, Fernández MS, Salomón OD, 2012. Distribution and abundance of phlebotominae, vectors of leishmaniasis, in Argentina: spatial and temporal analysis at different scales. J Trop Med 2012:1-16.

Quintana MG, Salomón OD, De Grosso MSL, 2010. Distribution of phlebotomine sand flies (Diptera: Psychodidae) in a primary forest-crop interface, Salta, Argentina. J Med Entomol 47:1003-10.

Quintana MG, Santini MS, Cavia R, Martinez MF, Liotta DJ, Fernandez MS, Perez AA, Direni Mancini JM, Moya SL, Giuliani MG, Salomón OD (2020) Multiscale environmental determinants of Leishmania vectors in urban and rural context. Parasite Vector 13:502.

R Core Team (2020) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/

Salomón O, Sinagra A, Nevot M, Barberian G, Paulin P, Estevez J, Riarte A, Estevez J, 2008. First visceral leishmaniasis focus in Argentina. Mem I Oswaldo Cruz 103:109-11.

Salomón OD, 2019. Instructions on how to make an Outbreak of American Cutaneous Leishmaniasis. J Trop Med Health 03:18.

Salomón OD, Feliciangeli MD, Quintana MG, Afonso MM dos S, Rangel EF, 2015. Lutzomyia longipalpis urbanisation and control. Mem I Oswaldo Cruz 110:831-46.

Salomón OD, Mastrángelo AV, Santini MS, Liotta DJ, Yadon ZE, 2016. La eco-epidemiología retrospectiva como herramienta aplicada a la vigilancia de la leishmaniasis en Misiones, Argentina, 1920-2014. Pan Am J Public Health 40:29-39.

Salomón OD, Orellano PW, Quintana MG, Perez S, Estani SS, Acardi S, Lamfri M, 2006a. Transmisión de la Leishmaniasis Tegumentaria en la Argentina. Medicina (Buenos Aires) 66:211-9.

Salomón OD, Ramos LK, Quintana MG, Acardi SA, Santini MS, Schneider A, 2009. Distribución de vectores de Leishmaniasis Visceral en la provincia de Corrientes, 2008. Medicina (Buenos Aires) 69:625-30.

Salomón OD, Sosa-Estani S, Ramos K, Orellano PW, Sanguesa G, Fernández G, Sinagra A, Rapasciolli G, 2006b. Tegumentary leishmaniasis outbreak in Bella Vista City, Corrientes, Argentina during 2003. Mem I Oswaldo Cruz 101:767-74.

Sandoval Pacheco CM, Araujo Flores GV, Favero Ferreira A, Sosa Ochoa W, Ribeiro da Matta VL, Zúniga Valeriano C, Pereira Corbett CE, Dalastra Laurenti M, 2018. Histopathological features of skin lesions in patients affected by non-ulcerated or atypical cutaneous leishmaniasis in Honduras, Central America. Int Jf Exp Pathol 99:249-57.

Servicio Meteorológico Nacional, 2020. Estadísticas Climáticas. Ciudad de Corrientes. Available from: https://www.smn.gob.ar/estadisticas

Silveira TG, Teodoro U, Arraes SM, Lonardoni MV, Dias ML, Shaw JJ, Ishikawa EA, Lainson R, 1990. An autochthonous case of cutaneous leishmaniasis caused by Leishmania (Leishmania) amazonensis Lainson & Shaw, 1972 from the north of Paraná State, Brazil. Mem Inst Oswaldo Cruz 85:475-6.

Simpson, E, 1949. Measurement of diversity. Nature 163:688.

Talmoudi K, Bellali H, Ben-Alaya N, Saez M, Malouche D, Chahed MK, 2017. Modeling zoonotic cutaneous leishmaniasis incidence in central Tunisia from 2009-2015: Forecasting models using climate variables as predictors. PLoS Neglect Trop D 11:e0005844.

Thomaz-Soccol V, Gonçalves AL, Piechnik CA, Baggio RA, Boeger WA, Buchman TL, Michaliszyn MS, Rodrigues dos Santos D, Celestino A, Aquino J, Leandro A de S, Paz OL de S da, Limont M, Bisetto A, Shaw JJ, Yadon ZE, Salomón OD, 2018. Hidden danger: Unexpected scenario in the vector-parasite dynamics of leishmaniases in the Brazil side of triple border (Argentina, Brazil and Paraguay). PLoS Neglect Trop D 12:e0006336.

Valero NNH, Uriarte M, 2020. Environmental and socioeconomic risk factors associated with visceral and cutaneous leishmaniasis: a systematic review. Parasitol Res 119:365-84.

VanDerWal J, Falconi L, Januchowski S, Shoo L, Storlie C, 2019. SDMTools - Species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. Available from: https://CRAN.R-project.org/package=SDMTools

van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N, 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39:542-52.

Vignali S, Barras A, Braunisch V, 2020. SDMtune: species distribution model selection. Available from: https://github.com/ConsBiol-unibern/SDMtune

Wasserberg G, Abramsky Z, Kotler BP, Ostfeld RS, Yarom I, Warburg A, 2003. Anthropogenic disturbances enhance occurrence of cutaneous leishmaniasis in Israel deserts: patterns and mechanisms. Ecol Appl 13:868-81.