RESEARCH ARTICLE

Estimating snakebite incidence from mathematical models: A test in Costa Rica

Carlos A. Bravo-Vega^{1*}, Juan M. Cordovez¹, Camila Renjifo-Ibáñez², Mauricio Santos-Vega¹, Mahmood Sasa³

1 Research Group in Mathematical and Computational Biology (BIOMAC), Department of biomedical engineering, University of los Andes, Bogotá, Colombia, 2 Corporación colombiana de investigación agropecuaria, AGROSAVIA, Bogotá, Colombia, 3 Instituto Clodomiro Picado and Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica

* ca.bravo955@uniandes.edu.co

Abstract

Background

Snakebite envenoming is a neglected public health challenge that affects mostly economically deprived communities who inhabit tropical regions. In these regions, snakebite incidence data is not always reliable, and access to health care is scare and heterogeneous. Thus, addressing the problem of snakebite effectively requires an understanding of how spatial heterogeneity in snakebite is associated with human demographics and snakes' distribution. Here, we use a mathematical model to address the determinants of spatial heterogeneity in snakebite and we estimate snakebite incidence in a tropical country such as Costa Rica.

Methods and findings

We combined a mathematical model that follows the law of mass action, where the incidence is proportional to the exposed human population and the venomous snake population, with a spatiotemporal dataset of snakebite incidence (Data from year 1990 to 2007 for 193 districts) in Costa Rica. This country harbors one of the most dangerous venomous snakes, which is the Terciopelo (*Bothrops asper*, Garman, 1884). We estimated *B. asper* distribution using a maximum entropy algorithm, and its abundance was estimated based on field data. Then, the model was adjusted to the data using a lineal regression with the reported incidence. We found a significant positive correlation ($R^2 = 0.66$, p-value < 0.01) between our estimation and the reported incidence, suggesting the model has a good performance in estimating snakebite incidence.

Conclusions

Our model underscores the importance of the synergistic effect of exposed population size and snake abundance on snakebite incidence. By combining information from venomous snakes' natural history with census data from rural populations, we were able to estimate snakebite incidence in Costa Rica. The model was able to fit the incidence data at fine



Citation: Bravo-Vega CA, Cordovez JM, Renjifo-Ibáñez C, Santos-Vega M, Sasa M (2019) Estimating snakebite incidence from mathematical models: A test in Costa Rica. PLoS Negl Trop Dis 13(12): e0007914. https://doi.org/10.1371/journal. pntd.0007914

Editor: Maria Elena Bottazzi, Baylor College of Medicine, UNITED STATES

Received: June 29, 2019

Accepted: November 9, 2019

Published: December 2, 2019

Copyright: © 2019 Bravo-Vega et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are held in figshare repository: https://doi.org/10.6084/m9. figshare.8311994.

Funding: This study was funded by Mahmood Sasa: Vicerrectoría de Investigación, University of Costa Rica (http://www.vinv.ucr.ac.cr/), Project ID: B8A53. This study was also funded by Carlos Bravo-Vega: Colciencias, Colombia (https://www. colciencias.gov.co/), Application 727 for doctoral students. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

administrative scale (district level), which is fundamental for the implementation and planning of management strategies oriented to reduce snakebite burden.

Author summary

Snakebite envenoming is an important health issue that affects mostly rural populations in tropical countries. Antivenom is the only effective treatment, but it is not always available in every medical center located in rural areas. Additionally, the number of snakebite cases is usually not known, or severely underestimated for several tropical countries. As a consequence, antivenom distribution and disease management are hard to optimize, then the estimation of snakebite incidence has been proposed as a key factor to improve snakebite management. To the date, studies have used statistics and community-based surveys to estimate incidence, but these approaches are still challenging. Here our goal is to develop a mathematical model that can estimate the geographic patterns of snakebite incidence. By applying a simple model, which is based on the law of mass action, we successfully captured the spatial variation of snakebite incidence in Costa Rica. The inputs for this model are venomous snakes' census and presence records, and human population data. Our work gives the basis for the development of more complex models based on the law of mass action to estimate several epidemiological variables associated with snakebite in rural populations.

Introduction

Snakebite envenoming is mainly a disease of rural populations that affects up to 2.7 million people and kills as many as 95,000 worldwide per year [1-4]. The highest snakebite mortality rates occur in tropical countries with the lowest per capita government expenditure on health and the lowest per capita GDP [5,6]. The public health challenge of this Neglected Tropical Disease (NTD) is that the most afflicted communities are the least able to access or afford effective antivenom treatment. The development of an appropriate response to snakebite requires a reasonable estimation of its burden at local and regional scales, and the understanding of the drivers of its spatiotemporal patterns. Understanding these elements could contribute to: 1) public health planning by government authorities, 2) optimization of production and distribution of antivenom, 3) the design of control strategies to minimize snakebite incidence, and 4) correct training of the medical staff who will treat envenomed patients [7]. However, as it happens with other NTDs, epidemiological data of snakebite incidence is not readily available for several countries, especially in Tropical regions [8,9].

The most cost-effective way to obtain reliable NTDs epidemiological data is via medical records, especially in countries where these reports are mandatory [10]. However, many governments still do not require health centers to inform of snakebite cases, making epidemiological data scarce or incomplete [2]. In addition, countries with large rural and indigenous populations tend to exhibit an underestimation in their records due to population disbelief in modern medicine and a limited access to health care [2,10,11]. For example, in West Bengal, a community-based study using door-to-door surveys estimated that only 22% of snakebite victims sought medical attention, and that medical centers only reported 7% of the total cases [12]. Likewise, another community-based study performed in eastern Nepal estimated a total of 4078 snakebites, but the whole country only reported 480 cases for the same year (89% of the cases were not reported) [13]. However, such studies require numerous interviewers and resources that are not readily available in most instances. Therefore the improvement of the surveillance systems and the estimation of snakebite burden is central for the control and management of this NTD [8].

Statistical inference has emerged as an effective tool for estimating snakebite burden without detailed epidemiological surveys. For example, several meta-analyses have built linear statistical models to infer epidemiological parameters in countries where data is not available [2,6,14]. Although estimations of snakebite incidence and mortality can also be made with this approach, they only provide broad approximations based on inferences, and do not take into account the processes underlining human-snake interactions [15]. Compartmental epidemiological models, another tool for estimating epidemiological variables, are based on the interactions between populations involved in disease' epidemiology [15,16]. These models have been used widely for several NTDs such as Chagas disease, Dengue and Malaria, but not for snakebite [17,18]. One of the main assumptions in these models is the law of mass action, namely: the encounters between two populations that are moving randomly are proportional to the multiplication of the abundance of both populations [19]. Using the law of mass action, the incidence for several diseases can be defined as proportional to the total encounters between an infectious population and a susceptible population [15].

As an effort to evaluate the possible use of compartmental epidemiological models in the prediction of snakebite incidence, we formulated a model based on the law of mass action to estimate the snakebite burden in Costa Rica. We selected this country because of its highly reliable snakebite epidemiological data, that allows final evaluation of our model estimations (Fig 1) [14]. To do so, we predicted the distribution of *B. asper* in the country using niche modeling based on collection records. This species is considered the most medically important species of venomous snakes in the country and it is responsible for over 70% of the snakebites [20,21] (see below). Then, we conducted fieldwork to assess *B. asper* relative abundance and to calibrate the spatial model. Finally, we evaluated the performance of the model capturing the spatial variation of reported snakebite incidence.

Materials and methods

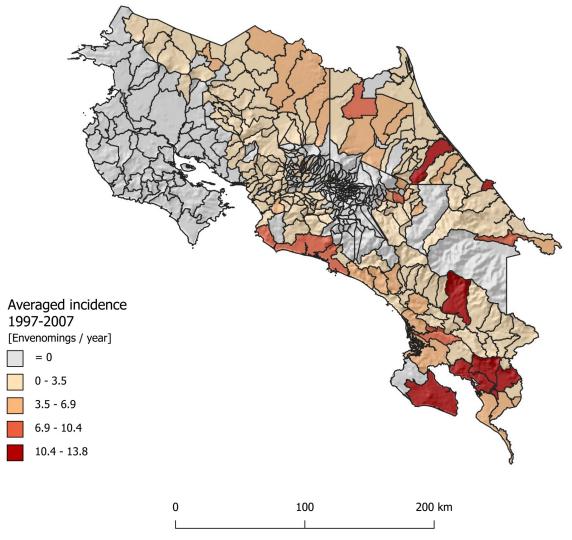
Study area

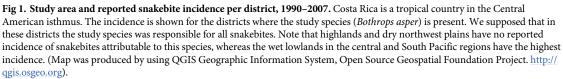
Costa Rica (Fig 1) is a country with a reliable snakebite dataset, where the public health system incorporates medical centers throughout its territory and a well-established social security system [22]. In this country, snakebite reporting is mandatory, and statistics on snake envenomation are consistent [23]. In addition, Costa Rica is home to the Instituto Clodomiro Picado at the University of Costa Rica, a research center and antivenom production facility that ensures that antivenom is accessible and available throughout the national territory. All these characteristics make Costa Rica a great country to asses, parameterize, and evaluate a mathematical model to predict snakebite incidence.

Applying the law of mass action to predict snakebite

The law of mass action states that the number of encounters between two populations are proportional to the multiplication of their sizes [15]. We can apply this law to snakebite, where the total encounters between venomous snakes and exposed humans will be proportional to the size of each population. Then, if we multiply these encounters by the conditional probability that an encounter ends in a bite, we can estimate snakebite incidence:

$$I_i = \theta \times \beta \times S_i \times V_i \tag{1}$$





https://doi.org/10.1371/journal.pntd.0007914.g001

Where the subindex *i* denotes the administrative unit in which the model will be applied, θ is the conditional probability that snakebite occurs during an encounter, β is the contact rate between susceptible humans (S_i) and venomous snakes (V_i), and I_i is the predicted incidence.

One limitation of the model proposed in Eq 1 is quantifying the abundance of venomous snakes in the field (V_i), which requires considerable sampling effort [24]. To overcome this limitation, we can multiply the contact rate between susceptible humans and venomous snakes (β) with the abundance of venomous snakes (V_i) to get a new variable (F_i) which represents the encounter frequency with venomous snakes per human per unit of time. This new variable is easier to estimate in the field because it only requires census of venomous snakes normalized by man-hour of sampling effort instead of absolute snake abundance. The proposed model incorporating the new variable is:

$$I_i = \alpha + \theta \times F_i \times S_i \tag{2}$$

Where the sub-index *i* denotes the administrative unit sampled–in this case districts, which are the finer administrative unit in Costa Rica, α is the intercept, F_i is the encounter frequency with venomous snakes, and S_i is the susceptible human population calculated from the reported rural population from 2000 census data (http://www.inec.go.cr/). The new parameter α allows the average of the regression residuals to be equal to zero [25]. Therefore, the *sum* of the predicted values will be equal to the *sum* of the observed values (i.e. The predicted *national* incidence will be equal to the reported *national* incidence). We used yearly snakebite incidence data from 1990 to 2007 reported by the Social Security of Costa Rica (Caja Costarricense del Seguro Social, CCSS), and we estimated the encounter frequency (F_i) based on fieldwork.

Estimation of encounter frequency with venomous snakes (F_i)

Out of 146 species of snakes in Costa Rica, 24 are clinically relevant, including seven coral snakes (family Elapidae) and 17 pit vipers (family Viperidae) [26]. Among the pit vipers, the terciopelo (*Bothrops asper*, Fig 2) is responsible for more than 70% of snakebite cases in the region, due to its high relative abundance and its capacity to adapt to perturbed environments [20,21]. This species is distributed along the humid lowlands from Mexico and Central America to northern South America, where is commonly found in agricultural fields or in close proximity to human settlements [21,27]. We chose this species to test the proposed model because it causes most snakebites.

To estimate F_i (Eq 2), we established the distribution of *B. asper* in Costa Rica and then we determined the encounter frequency by searching for *B. asper* within its distribution (see description of fieldwork below).



Fig 2. Study species: Terciopelo pit viper, Bothrops asper. This species is responsible for most snakebite envenomation throughout its distribution. It is distributed in humid lowlands from southern Mexico to northeastern South America. Photography was taken by Bravo-Vega CA.

https://doi.org/10.1371/journal.pntd.0007914.g002

Distribution of B. asper in Costa Rica

To estimate the potential distribution of this species, we performed environmental niche modeling with Maxent (Maximum entropy algorithm [28]) because of its high performance when only occurrence data is available, and its ability to discriminate between suitable and unsuitable areas. We obtained *B. asper* presence data from locality records at the Instituto Clodomiro Picado, and we used environmental data from the Bioclim database on the WorldClim Server (http://www.worldclim.org/bioclim) [29]. The species distribution model was run using the dismo package in R [30,31]. To evaluate model precision, we computed a ROC (Receiver Operating Characteristic) curve based on 1000 random "pseudo-absence" background points and calculated true and false positive rates. With these rates, the area under the ROC curve (AUC) was computed, and finally the model performance was compared against a randomized prediction using the same environmental data [28].

To generate a presence/absence map based on Maxent results, we selected the threshold that has zero omission rate with the maximum predicted presence area [32], and we removed areas above 1200 m.a.s.l., the highest altitude limit reported for *B. asper* in Costa Rica [33]. The distribution maps were created using Quantum GIS (QGIS) software [34].

Sampling encounter frequency (field work)

A series of active search surveys for snakes were carried out during the dry season (January to April) and wet season (August to November) in 2017 at 12 sampling sites in Costa Rica, divided in six in the Pacific versant and six in the Caribbean versant (Fig 3). Sites were evenly distributed among three altitude ranges: 0–400 m, 400–800 m, and over 800 m. This selection was done because populations of *B. asper* varies among the two versants [35], and their abundance could decrease with elevation [21,36]. Additionally, with selected sampling sites we covered most of the ecoregions where *B. asper* is present in Costa Rica: The Caribbean zone, the north zone, the central pacific zone and the south pacific zone [37]. We only missed the Guatuso plain area. Given that these ecoregions are not severely heterogeneous, we assumed that the measured encounter frequency in selected sampling sites could be extrapolated throughout predicted distribution.

At each site, we sought for *B. asper* in the night during new moon periods across three land cover types: open perturbed areas (*i.e.* pastures or crops), closed canopy perturbed areas (*i.e.* plantain, oil palm or sugar cane crops), and unperturbed forest. The sampling effort was at least ten person-hours per land cover type per site. The number of observed individuals was recorded and normalized by man-hours of active searching, and then we extrapolated this frequency of encounters based on predicted distribution, altitude thresholds and versant. Finally, we computed the spatial-averaged encounter frequency per each district (F_i).

Model calibration

We assumed that *B. asper* was responsible for all snakebite cases because it has a very high incidence of envenomation. To calibrate our model, we performed linear regression between $F_i \times S_i$ (obtained as described above) and the reported incidence by the CCSS using a 99% confidence and prediction intervals. Our sample size consisted of 193 districts where snakebites were reported. The slope of the regression is the parameter θ , and the intercept is the parameter α (Eq 2). To evaluate the model, we computed r-squared using a 99% confidence level and we did a Pearson's product moment correlation test between model residuals and predicted values.

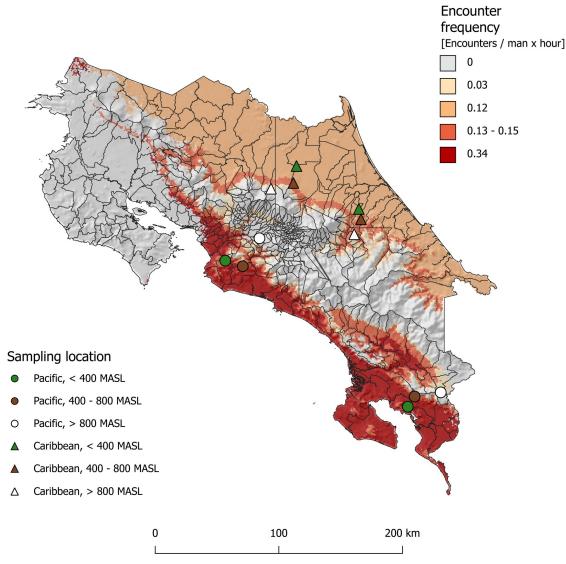


Fig 3. Sampling locations and predicted encounter frequency of *Bothrops asper* **over its distribution.** Each symbol corresponds with a sampling location. Circles are sampling locations in the Pacific versant, while triangles represent locations in the Caribbean versant. Green symbols are lowlands locations (Altitude below 400 m.a.s.l.), brown symbols are midlands locations (Altitude between 400 m.a.s.l.) and 800 m.a.s.l.), and white symbols are highlands locations (Altitude above 800 m.a.s.l.). Gray areas are places where the environmental niche model predicted absence for the study species. The other map colors show the value of the extrapolated encounter frequency. Encounter frequency was higher in the Pacific versant than in the Caribbean, and it tends to decrease with altitude. Map was produced by using QGIS Geographic Information System, Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

https://doi.org/10.1371/journal.pntd.0007914.g003

To test model performance against different hypothesis, we compared it to four different models: *i*) A model with the intercept set equal to zero, *ii*) A linear model between the incidence and rural population, *iii*) A linear model between the incidence and the district-averaged encounter frequency, and *iv*) A multiple linear model of incidence as a function of rural population and the district-averaged encounter frequency. To compare the five models, we used r-squared and the Akaike information criterion (AIC). This criterion accounts for the trade-off between model prediction performance and the number of parameters, where the model with the lowest AIC will be the preferred one [38].

Results

Expected distribution of B. asper in Costa Rica

We estimated the presence of *B. asper* in 55% of the Costa Rican continental territory, encompassing mostly low and wet areas. The most important bioclimatic variables in the species distribution model (variables that account for the 80% of the variation of the prediction) were precipitation of driest quarter, precipitation of coldest quarter, minimum temperature of the coldest month, maximum temperature of the warmest month, precipitation of the vettest quarter, annual precipitation, annual mean temperature and mean temperature of the coldest quarter. The AUC value of the model is 0.79, which suggests that it is better at predicting presence points than a random model (AUC = 0.5), but is less than perfect (AUC = 1).

At field sites, we employed a total sampling effort of 562 person-hours, finding 186 snakes, of which 85 were *B. asper* (relative abundance of 0.45). Interestingly, the high relative abundance of 0.7 with respect to venomous snakes (Families *Elapidae* and *Viperidae*) supports the assumption that most snakebites are caused by *B. asper*. National averaged encounter frequency of *B. asper* is 0.15 vipers per person-hour, and it tends to decrease with elevation, with higher values in the southern Pacific lowlands (see map of extrapolated values in Fig 3).

Modeling snakebite incidence in Costa Rica

An average of 438.5 snakebite cases were reported per year (minimum: 360 in 1991, maximum: 517 in 2006). Most cases occurred in the South Pacific and Caribbean lowlands and were distributed in elevations below 1400 m.a.s.l (Fig 1): These records could be attributed to *B. asper*. All models showed a significant p-value for the regression, but the proposed model (Eq 2, Model 1 in Table 1) was the best model (Highest r-squared and lowest AIC).

The overall snakebite incidence estimated by model 1 was 438.5 (ranging from 361.29 to 515.82) cases per year. We found a significant association (r-squared = 0.66, p<0.01) between the reported snakebite incidence and predicted venomous snake encounters ($F_i \times S_i$, where F_i is the district averaged encounter frequency with *B. asper* and S_i is the rural population per district) (Fig 4). Since the residuals of this model did not vary across fitted values (p>0.01 for the Pearson's product moment correlation test), the linear regression met the assumptions of normality and heteroscedasticity.

The prediction interval for the model was \pm 4.41 cases. Seven districts (4% of the total used districts) (Fig 4) exhibited deviations between predicted and reported incidence that were greater than this interval, six underestimated by the model (Puerto Jiménez, Bratsi, Santa Teresita, Buenos Aires, Limón, and Tayutic), and one overestimated by it (Ciudad Quesada).

Table 1. Comparison between different models to estimate snakebite incidence.

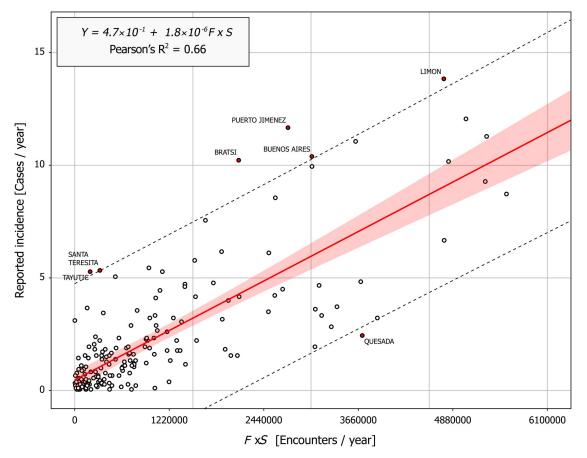
Model	Structure	R -squared	AIC	Predicted national incidence (Cases/year)	Range for predicted national incidence (Cases/year)
Model 1*	$I_i = \alpha + \theta \times F_i \times S_i$	0.665	741.44	438.5	361.29-515.82
Model 2	$I_i = \theta \times F_i \times S_i$	0.648	749.21	382.3**	344.99-419.68
Model 3	$I_i = \alpha + \hat{\theta} \times S_i$	0.344	871.3	438.5	313.38–563.73
Model 4	$I_i = \alpha + \hat{\theta} \times F_i$	0.172	916.26	438.5	331.42-545.69
Model 5	$I_i = \alpha + \theta_1 \times F_i + \theta_2 \times S_i$	0.523	811.49	438.5	324.86-552.24

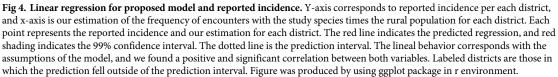
All models showed a significant p-value < 0.01.

* Proposed model, which has the highest r-squared and the lowest AIC.

** Model 2, which is the model without intercept, is the only model that cannot estimate the national incidence.

https://doi.org/10.1371/journal.pntd.0007914.t001





https://doi.org/10.1371/journal.pntd.0007914.g004

Discussion

Our model successfully captures and estimates snakebite incidence. This is the simplest model that can be performed by using the law of mass action, which is the basis for the development of more complex models [15]. These complex models can estimate relevant variables in public health such as rate of patients recovery, mortality, and disability [15,16,19]. Such approaches have been successfully used for other Neglected Tropical Diseases, including diseases with more complicated transmission dynamics than snakebite [39,40]. Thus, our model constitutes a base model to understand snakebite epidemiology, and it could be expanded to estimate mortality, morbidity, and economic burden to ultimately optimize medical responses to snakebite [41,42]. Our model accurately estimated the national incidence of snakebite in Costa Rica and its range (Model prediction: 438.5 ± 77.2 yearly cases; Reported incidence: 438.5 yearly cases, minimum: 359 in 1991, maximum: 518 in 2006), captured 66% of the variation of district data, and had a significant r-squared value similar to that reported for more complex spatial regressions of other health problems (range 0.54-0.7) [43,44]. Therefore, our model suggests that the spatial distribution of snakebite is mostly driven by a combination of environmental

factors affecting venomous snakes' distribution and their association with exposed human population size.

The distribution of venomous snakes has been used as an input to determine the effect of climate change on snakebite burden [45], and to determine vulnerable human populations [46]. Our algorithm uses those distributions to understand the geographical variation of snakebite, but more importantly we can estimate incidence in places where data is scarce. The fact that the law of mass action helps to account for snakebite spatial heterogeneity in our model, confirms that incidence is directly affected by abundance of venomous snakes. Our results underscore the importance of incorporating the abundance—and not the number—of medically important snake species as a risk estimator. Recently, studies oriented to identify vulnerable populations using diversity of dangerous venomous snakes as risk index do not highlight Latin America as vulnerable [46]. We suggest that Latin America has several vulnerable populations related to fewer but more abundant venomous snakes' species.

Mathematical model performance

Our species distribution model adequately predicts the distribution of *B. asper* in Costa Rica, as high AUC values (>0.75) indicate a reliable distribution estimation [47]. Our estimated distribution agrees with previous studies in which the distribution range of *B. asper* was estimated based on presence records at natural history museums [26,48]. Additionally, our encounter frequency estimations of *B. asper* across an altitudinal gradient corroborate previous studies that suggest that *B. asper* prefers the lowlands [21,48]. The highest encounter frequency was in the humid Pacific lowlands during the dry season, close to water bodies. Because we did not measure population sizes, we cannot evaluate whether differences in encounter frequency between the Caribbean and Pacific lowlands are attributed to demographic disparities. We hypothesize that the marked dry season in the Pacific restricts available water bodies, which may motivate *B. asper* to aggregate near water, and in turn raise the likelihood of human-snake encounters in the Pacific lowlands [21,49].

The models based on the law of mass action showed the strongest performance predicting snakebite incidence in Costa Rica. In fact, model 1 and model 2 share the greatest r-squared values with the lowest AIC, but only model 1 estimated well the national incidence (View Table 1). Then models based on this law estimates better snakebite incidence than other models based on *F* and *S*. In addition, given that the maximum reported incidence for a district was 15 cases per year, the value for the intercept in the proposed model (Model 1) was clearly close to zero (0.47 cases per year). Then, our proposed model outperformed the other models. The slope for our model, the parameter θ , represents the conditional probability that an encounter with a venomous snake ends in a bite. Although this parameter is significantly different from zero, its value is very low (1.8 x 10–6); possibly because our estimation is based on an active search to estimate the venomous snakes encounter frequency. The encounter frequency between an exposed human population and venomous snakes is much smaller, which increases the conditional probability. Despite the low value of θ , our model continues to estimate the incidence of snakebite adequately.

In a few districts, estimated and reported incidences did not match. This could be explained by two conditions which are not mutually exclusive. First, some districts may have included reports of snakebite that occurred in neighboring districts. Data collected by the CCSS is based on the attending medical center, rather than on the site of envenomation. For example, Puerto Jiménez, Limón and Buenos Aires are large towns with at least one CCSS health clinic, so patients from surrounding rural districts could be referred to those centers [50]. Second, the model assumes spatial homogeneity in encounter frequency and susceptible human population distribution within each district, which is obviously an approximation. Districts could include areas with elevations where *B. asper* is not distributed, but human population can be living in elevations where *B. asper* is found: Our model will underestimate incidence in these places. For example, a large portion of Tayutic, Bratsi and Santa Teresita districts is located in highlands where *B. asper* is not found, whereas human population is concentrated at lower elevations [50]. This could explain why reported incidence in these districts is higher than estimated incidence. Conversely, overestimations might reflect a bias due to spatial heterogeneity of the human population. For example, the 36% of the population of Quesada are catalogued as rural, but Quesada contains the biggest city in northern Costa Rica [51]. *B. asper* is distributed across this district, but most rural people work in urban areas and are not prone to encounters, leading to a higher estimated incidence than the reported by medical centers [52].

Our model has the potential to estimate snakebite incidence in places where data is not available despite the few mismatches discussed previously. This model estimated incidence adequately by only using presence records for venomous snakes, measurements of encounter frequency with venomous snakes, and census data for rural human populations. This methodology is simpler than performing community-based studies, so by using this model a rapid and affordable estimation for snakebite incidence can be done. This model could be used for other species of venomous snakes once encounter frequency surveys are performed. Additionally, this model can be improved: Incorporating spatial data on population density will resolve issues with rural/urban heterogeneity, as well as including map layers of socioeconomic factors will highlight particularly at-risk populations. Given that the estimation of snakebite burden has been flagged as a critical goal to improve public health policy, the proposed model can play a crucial role in the planning and management done to minimize snakebite burden in areas where epidemiological data is scarce [7,8].

Limitations and challenges

Although our model performed well at estimating snakebite incidence, our methodology exhibits different challenges: 1) It requires reliable occurrence data, which is not always available for tropical venomous snake species [53,54]. 2) A baseline incidence dataset is needed to calibrate and train the model, so its application in countries where snakebite incidence is poorly reported may be challenging. 3) Our methodology relies on the estimation of the encounter frequency (F). This estimation requires a high field work effort and qualified personnel in searching venomous snakes.

In the presence of the conditions described above, our model has the potential to be used in other regions with snakebite envenomation. Distribution of venomous snakes can be estimated by using data for other countries, and the model could be calibrated by using a subset of districts with reliable data, or using snakebite data for neighbor countries. Such extrapolations must be done carefully, because snakes' distribution could change drastically between regions; for example in Costa Rica *B. asper* reaches a maximum of 1200 m.a.s.l. [26], but in Ecuador it is found up to 1900 m.a.s.l. [55].

Finally, to use the model in places where snakebite is caused by multiple species, such as Colombia, where *Bothrops asper* and *Bothrops atrox* cause the majority of envenomations, we recommend to measure encounter frequency for each species separately. Then, to calibrate the model it would be necessary to perform a multiple lineal regression since the parameter θ may vary between species. This parameter represents the probability that an encounter with a snake species ends in a bite. The model accounting for multiple species would be:

 $I_i = \alpha + \sum_{j=1}^{N} (\theta_j \times F_{j,i} \times S_i)$, where *j* is the sub-index for each snake species, $F_{j,i} \times S_i$ are the independent variables, and I_i is the response variable.

In summary, our study highlights the effectiveness of mathematical models estimating snakebite incidence and the importance of including venomous snakes' biology to understand snakebite epidemiology. The lack of this biological information currently becomes a challenge to fit this kind of models. Therefore, it is important to move forward in the knowledge of the biology and natural history of venomous snakes (distribution, population dynamics, habitat use, etc.) [21,54]. Even though this model can estimate incidence, it will never substitute a proper epidemiological surveillance program. Such programs must be adopted by public health authorities in countries affected by snakebite to minimize impact in at-risk populations, and to eliminate the neglect of snakebite as a tropical disease [8].

Data availability

The data that support the findings of this study are openly available in figshare at <u>https://doi.org/10.6084/m9.figshare.8311994</u> [56].

Supporting information

S1 Fig. Presence locations for *Bothrops asper*. Map was produced by using QGIS Geographic Information System, Open Source Geospatial Foundation Project. <u>http://qgis.osgeo.org</u>. (TIF)

S2 Fig. (a) Rural population by district. (b) Estimated average encounter frequency by district. Maps were produced by using QGIS Geographic Information System, Open Source Geospatial Foundation Project. <u>http://qgis.osgeo.org</u>. (TIF)

S3 Fig. (a) Reported incidence by district. (b) Estimated incidence by district. Maps were produced by using QGIS Geographic Information System, Open Source Geospatial Foundation Project. <u>http://qgis.osgeo.org</u>. (TIF)

(TIF)

S4 Fig. Analysis of omission/commission for the maximum entropy model. Note that the omission on training and test samples is close to the predicted omission, so the model fulfills the assumptions of maximum entropy. Figure was produced by using r environment. (TIF)

S5 Fig. Receiver operating characteristic (ROC) curve for the maximum entropy model. Figure was produced by using r environment. (TIF)

Acknowledgments

We thank to Fabian Bonilla and Miguel Solano for their unconditional help during field work, Jenny Stynoski for her help during the writing of this manuscript, Instituto Clodomiro Picado for their advising during the study, and Organization for Tropical Studies (OTS) for their support during the field work at their biological stations.

Author Contributions

Conceptualization: Carlos A. Bravo-Vega, Juan M. Cordovez, Mahmood Sasa.

Data curation: Carlos A. Bravo-Vega, Mahmood Sasa.

- Formal analysis: Carlos A. Bravo-Vega, Juan M. Cordovez, Mauricio Santos-Vega, Mahmood Sasa.
- Funding acquisition: Carlos A. Bravo-Vega, Juan M. Cordovez, Camila Renjifo-Ibáñez, Mahmood Sasa.
- Investigation: Carlos A. Bravo-Vega, Juan M. Cordovez, Camila Renjifo-Ibáñez, Mauricio Santos-Vega, Mahmood Sasa.

Methodology: Carlos A. Bravo-Vega, Mahmood Sasa.

Project administration: Carlos A. Bravo-Vega, Mahmood Sasa.

Resources: Juan M. Cordovez, Mahmood Sasa.

Software: Carlos A. Bravo-Vega.

Supervision: Juan M. Cordovez, Camila Renjifo-Ibáñez, Mahmood Sasa.

Validation: Juan M. Cordovez, Camila Renjifo-Ibáñez, Mauricio Santos-Vega, Mahmood Sasa.

Visualization: Juan M. Cordovez, Mauricio Santos-Vega.

Writing – original draft: Carlos A. Bravo-Vega, Mahmood Sasa.

Writing – review & editing: Juan M. Cordovez, Camila Renjifo-Ibáñez, Mauricio Santos-Vega, Mahmood Sasa.

References

- 1. WHO. WHO | Prevalence of snakebite envenoming. WHO [Internet]. 2017 [cited 2018 Oct 18]; http://www.who.int/snakebites/epidemiology/en/
- Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, et al. The Global Burden of Snakebite: A Literature Analysis and Modelling Based on Regional Estimates of Envenoming and Deaths. PLOS Med [Internet]. 2008 Nov 4; 5(11):e218. https://doi.org/10. 1371/journal.pmed.0050218 PMID: 18986210
- 3. Chippaux JP. Snakebite envenomation turns again into a neglected tropical disease! J Venom Anim Toxins Incl Trop Dis. 2017; 23(1):1–2.
- WHO. Rabies and envenomings : a neglected public health issue : report of a Consultative Meeting. Who [Internet]. 2007;(January):32. http://www.who.int/bloodproducts/animal_sera/Rabies.pdf
- Harrison RA, Hargreaves A, Wagstaff SC, Faragher B, Lalloo DG. Snake envenoming: a disease of poverty. PLoS Negl Trop Dis. 2009 Dec; 3(12):e569. https://doi.org/10.1371/journal.pntd.0000569 PMID: 20027216
- Chippaux J-P. Incidence and mortality due to snakebite in the Americas. Gutiérrez JM, editor. PLoS Negl Trop Dis. 2017 Jun; 11(6):e0005662. https://doi.org/10.1371/journal.pntd.0005662 PMID: 28636631
- Chippaux J-P. Estimating the Global Burden of Snakebite Can Help To Improve Management. PLoS Med [Internet]. 2008 Nov 4 [cited 2018 Oct 18]; 5(11):e221. http://dx.plos.org/10.1371/journal.pmed. 0050221 PMID: 18986211
- Gutiérrez JM, Warrell DA, Williams DJ, Jensen S, Brown N, Calvete JJ, et al. The Need for Full Integration of Snakebite Envenoming within a Global Strategy to Combat the Neglected Tropical Diseases: The Way Forward. PLoS Negl Trop Dis [Internet]. 2013 Jun 13; 7(6):e2162. https://doi.org/10.1371/ journal.pntd.0002162 PMID: 23785526
- Hansson E, Cuadra S, Oudin A, de Jong K, Stroh E, Torén K, et al. Mapping Snakebite Epidemiology in Nicaragua–Pitfalls and Possible Solutions. PLoS Negl Trop Dis [Internet]. 2010 Nov 23; 4(11):e896. https://doi.org/10.1371/journal.pntd.0000896 PMID: 21124884
- 10. Gutiérrez JM, Burnouf T, Harrison RA, Calvete JJ, Brown N, Jensen SD, et al. A Call for Incorporating Social Research in the Global Struggle against Snakebite. de Silva HJ, editor. PLoS Negl Trop Dis

[Internet]. 2015 Sep 17 [cited 2018 Oct 18]; 9(9):e0003960. https://dx.doi.org/10.1371/journal.pntd. 0003960 PMID: 26379235

- Vasquez J, Jimenez SL, Gomez IC, Rey JP, Henao AM, Marin DM, et al. Snakebites and ethnobotany in the Eastern region of Antioquia, Colombia—the traditional use of plants. J Ethnopharmacol. 2013 Mar; 146(2):449–55. https://doi.org/10.1016/j.jep.2012.12.043 PMID: 23396114
- Sinha A, Bhattacharya S, Ram R, Dasgupta U, Ram A, Majumder D. Epidemiological profile of snake bite in South 24 Parganas district of West Bengal with focus on underreporting of snake bite deaths. Indian J Public Health [Internet]. 2014 [cited 2018 Oct 18]; 58(1):17. http://www.ncbi.nlm.nih.gov/ pubmed/24748352 https://doi.org/10.4103/0019-557X.128158
- Sharma SK, Khanal B, Pokhrel P, Khan A, Koirala S. Snakebite-reappraisal of the situation in Eastern Nepal. Toxicon [Internet]. 2003 Mar 1 [cited 2019 May 15]; 41(3):285–9. http://www.ncbi.nlm.nih.gov/ pubmed/12565750 https://doi.org/10.1016/s0041-0101(02)00289-1
- Hansson E, Sasa M, Mattisson K, Robles A, Gutiérrez JM. Using Geographical Information Systems to Identify Populations in Need of Improved Accessibility to Antivenom Treatment for Snakebite Envenoming in Costa Rica. PLoS Negl Trop Dis [Internet]. 2013 Jan 31; 7(1):e2009. https://doi.org/10.1371/ journal.pntd.0002009 PMID: 23383352
- 15. Brauer F, Castillo-Chávez C. Mathematical models in population biology and epidemiology. Springer; 2012. 508 p.
- Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Virulence [Internet]. 2013 May 15 [cited 2018 Oct 18]; 4(4):295–306. http://www.ncbi.nlm.nih.gov/pubmed/23552814 https://doi. org/10.4161/viru.24041
- Velasco-Hernández JX. A model for Chagas disease involving transmission by vectors and blood transfusion. Theor Popul Biol [Internet]. 1994 Aug [cited 2018 Oct 18]; 46(1):1–31. http://www.ncbi.nlm.nih. gov/pubmed/8079195 https://doi.org/10.1006/tpbi.1994.1017
- Luz PM, Struchiner CJ, Galvani AP. Modeling Transmission Dynamics and Control of Vector-Borne Neglected Tropical Diseases. Brooker S, editor. PLoS Negl Trop Dis [Internet]. 2010 Oct 26 [cited 2018 Oct 18]; 4(10):e761. http://dx.plos.org/10.1371/journal.pntd.0000761 PMID: 21049062
- **19.** Edelstein-Keshet L. Mathematical Models in Biology. 1st ed. New York: Society for Industrial and Applied Mathematics SIAM; 2005.
- Sasa M, Vazquez S. Snakebite envenomation in Costa Rica: a revision of incidence in the decade 1990–2000. Toxicon. 2003 Jan; 41(1):19–22. <u>https://doi.org/10.1016/s0041-0101(02)00172-1</u> PMID: 12467657
- Sasa M, Wasko DK, Lamar WW. Natural history of the terciopelo Bothrops asper (Serpentes: Viperidae) in Costa Rica. Toxicon. 2009 Dec; 54(7):904–22. https://doi.org/10.1016/j.toxicon.2009.06.024 PMID: 19563822
- 22. Warf B. Do You Know the Way to San José? Medical Tourism in Costa Rica [Internet]. Vol. 9, Journal of Latin American Geography. University of Texas Press; 2010 [cited 2018 Oct 18]. p. 51–66. https://www.jstor.org/stable/25765284
- 23. Sasa M, Vazquez S. Snakebite envenomation in Costa Rica: a revision of incidence in the decade 1990–2000. Toxicon [Internet]. 2003 Jan [cited 2018 Oct 18]; 41(1):19–22. <u>http://www.ncbi.nlm.nih.gov/pubmed/12467657 https://doi.org/10.1016/s0041-0101(02)00172-1</u>
- 24. Willson JD, Winne CT, Todd BD. Ecological and methodological factors affecting detectability and population estimation in elusive species. J Wildl Manage [Internet]. 2011 Jan 28 [cited 2018 Oct 18]; 75 (1):36–45. http://doi.wiley.com/10.1002/jwmg.15
- Eisenhauer JG. Regression through the Origin. Teach Stat [Internet]. 2003 [cited 2019 Mar 4]; 25(3). http://web.ist.utl.pt/~mcasquilho/compute/errtheory/,regression/regrthroughorigin.pdf
- 26. Solórzano L. A, Instituto Nacional de Biodiversidad (Costa Rica). Serpientes de Costa Rica : distribución, taxonomía e historia natural [Internet]. INBio; 2004 [cited 2018 Oct 18]. 791 p. https://books. google.com.co/books/about/Serpientes_de_Costa_Rica.html?id=w4BgAAAAMAAJ&redir_esc=y
- Wasko DK, Sasa M. Food resources influence spatial ecology, habitat selection, and foraging behavior in an ambush-hunting snake (Viperidae: Bothrops asper): an experimental study. Zoology [Internet]. 2012 Jun [cited 2018 Oct 18]; 115(3):179–87. http://www.ncbi.nlm.nih.gov/pubmed/22440190 https:// doi.org/10.1016/j.zool.2011.10.001
- Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol Modell [Internet]. 2006; 190(3):231–59. http://www.sciencedirect.com/science/article/pii/ S030438000500267X
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol [Internet]. 2005 Dec 1 [cited 2018 Oct 18]; 25(15):1965–78. http://doi.wiley.com/10.1002/joc.1276

- R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2008. http://www.r-project.org
- Hijmans RJ, Phillips S, Leathwick J, Maintainer JE. Package "dismo" [Internet]. 2017 [cited 2018 Oct 18]. https://cran.r-project.org/web/packages/dismo/dismo.pdf
- Norris D. Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil [Internet]. Vol. 7, Mongabay.com Open Access Journal-Tropical Conservation Science. 2014 [cited 2018 Oct 18]. www.tropicalconservationscience.org
- Wilson LD, Townsend JH, Johnson JD, Murphy JB (James B. Conservation of Mesoamerican amphibians and reptiles. Eagle Mountain Press; 2010. 812 p.
- 34. QGIS Development Team. QGIS Geographic Information System [Internet]. Open Source Geospatial Foundation Project. [cited 2018 Oct 18]. https://qgis.org/es/site/
- Saldarriaga-Córdoba M, Parkinson CL, Daza JM, Wüster W, Sasa M. Phylogeography of the Central American lancehead Bothrops asper (SERPENTES: VIPERIDAE). Chiang T-Y, editor. PLoS One [Internet]. 2017 Nov 27 [cited 2018 Oct 18]; 12(11):e0187969. https://dx.doi.org/10.1371/journal.pone. 0187969 PMID: 29176806
- 36. Saldarriaga-Córdoba MM, Sasa M, Pardo R, Méndez MA. Phenotypic differences in a cryptic predator: Factors influencing morphological variation in the terciopelo Bothrops asper (Garman, 1884; Serpentes: Viperidae). Toxicon [Internet]. 2009 Dec 1 [cited 2018 Oct 18]; 54(7):923–37. http://www.ncbi.nlm.nih. gov/pubmed/19505490 https://doi.org/10.1016/j.toxicon.2009.05.031
- Solano G, Aguilar LA, Lizano MV. Zonificación forestal de Costa Rica y estado Poblacional de especies forestales, basado en el Inventario Nacional Forestal e instrumentos de monitoreo y manejo de bosques naturales. [Internet]. 2017 [cited 2019 Apr 30]. http://www.sirefor.go.cr/?wpfb_dl=19
- Bozdogan H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika [Internet]. 1987 Sep [cited 2019 Jan 22]; 52(3):345–70. http://link.springer.com/10.1007/BF02294361
- Alison Kealey A, Robert Smith R. Neglected Tropical Diseases: Infection, Modeling, and Control. J Health Care Poor Underserved [Internet]. 2010 Feb [cited 2018 Oct 18]; 21(1):53–69. <u>http://www.ncbi.</u> nlm.nih.gov/pubmed/20173255 https://doi.org/10.1353/hpu.0.0270
- Racloz V, Ramsey R, Tong S, Hu W. Surveillance of Dengue Fever Virus: A Review of Epidemiological Models and Early Warning Systems. Anyamba A, editor. PLoS Negl Trop Dis [Internet]. 2012 May 22 [cited 2018 Oct 18]; 6(5):e1648. <u>http://dx.plos.org/10.1371/journal.pntd.0001648</u> PMID: 22629476
- Basu S, Andrews JR, Poolman EM, Gandhi NR, Shah NS, Moll A, et al. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemio-logical modelling study. Lancet [Internet]. 2007 Oct 27 [cited 2018 Oct 18]; 370(9597):1500–7. http://www.ncbi.nlm.nih.gov/pubmed/17964351 https://doi.org/10.1016/S0140-6736(07)61636-5
- 42. Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. Stat Med [Internet]. 1999 Dec 15 [cited 2018 Oct 18]; 18(23):3263–82. http://www.ncbi.nlm.nih.gov/pubmed/10602150 https://doi.org/10.1002/(sici)1097-0258(19991215) 18:23<3263::aid-sim315>3.0.co;2-3
- 43. McCann CM, Baylis M, Williams DJL. The development of linear regression models using environmental variables to explain the spatial distribution of Fasciola hepatica infection in dairy herds in England and Wales. Int J Parasitol [Internet]. 2010 Aug 1 [cited 2018 Oct 18]; 40(9):1021–8. http://www.ncbi. nlm.nih.gov/pubmed/20227416 https://doi.org/10.1016/j.ijpara.2010.02.009
- Kleinschmidt I, Bagayoko M, Clarke GP, Craig M, Le Sueur D. A spatial statistical approach to malaria mapping. Int J Epidemiol [Internet]. 2000 Apr [cited 2018 Oct 18]; 29(2):355–61. <u>http://www.ncbi.nlm.</u> nih.gov/pubmed/10817136 https://doi.org/10.1093/ije/29.2.355
- Yañez-Arenas C, Townsend Peterson A, Rodríguez-Medina K, Barve N. Mapping current and future potential snakebite risk in the new world. Clim Change [Internet]. 2016; 134(4):697–711. <u>https://doi.org/ 10.1007/s10584-015-1544-6</u>
- 46. Longbottom J, Shearer FM, Devine M, Alcoba G, Chappuis F, Weiss DJ, et al. Vulnerability to snakebite envenoming: a global mapping of hotspots. Lancet (London, England) [Internet]. 2018 Aug 25 [cited 2019 Sep 17]; 392(10148):673–84. http://www.ncbi.nlm.nih.gov/pubmed/30017551
- **47.** Phillips SJ, Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography (Cop) [Internet]. 2008 Apr 1 [cited 2018 Sep 20]; 31(2):161–75. http://doi.wiley.com/10.1111/j.0906-7590.2008.5203.x
- Campbell JA, Lamar WW. The venomous reptiles of the Western Hemisphere. Comstock Pub. Associates; 2004. 870 p.

- Waylen PR, Caviedes CN, Quesada ME, Waylen PR, Caviedes CN, Quesada ME. Interannual Variability of Monthly Precipitation in Costa Rica. J Clim [Internet]. 1996 Oct 1 [cited 2019 May 15]; 9(10):2606– 13. http://journals.ametsoc.org/doi/abs/10.1175/1520-0442%281996%29009%3C2606%3AIVOMPI% 3E2.0.CO%3B2
- Ortiz-Malavasi E. Atlas Digital de Costa Rica está a disposición del público. Investig Núm 23(1) [Internet]. 2015 Aug 1 [cited 2018 Oct 18]; https://repositoriotec.tec.ac.cr/handle/2238/7477
- 51. INEC IN de E y censos. IX Censo Nacional de Población CARACTERÍSTICAS SOCIALES Y DEMO-GRÁFICAS [Internet]. San José; [cited 2018 Oct 18]. http://www.inec.go.cr/sites/default/files/ documentos/inec_institucional/publicaciones/02_ix_censo_nacional_de_poblacion._caracteristicas_ sociales_y_demograficas.pdf
- 52. Rodríguez Sánchez G, Alvarado Briceño Julio Benavides Espinoza Jeimy Calvo rojas Andrea Carrizales Alvarado S, Hernández Fallas Roberto Carlos Ortega Meléndez Daniel Sancho Murillo S. Diagnóstico situacional del Cantón de San Carlos [Internet]. [cited 2018 Oct 18]. https://www.munisc.go.cr// documentos/Secciones/20/DiagnósticosituacionaldelCantóndeSanCarlos.pdf
- 53. Loiselle BA, Jørgensen PM, Consiglio T, Jiménez I, Blake JG, Lohmann LG, et al. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J Biogeogr [Internet]. 2007 Sep 6 [cited 2018 Oct 18]; 0(0):070908043732003–??? http://doi. wiley.com/10.1111/j.1365-2699.2007.01779.x
- Lynch JD. El contexto de las serpientes en Colombia con un análisis de las amenazas en contra de su conservación. Rev Colomb Cienc [Internet]. 2012; 36(140):435–49. http://www.scielo.org.co/pdf/ racefn/v36n140/v36n140a09.pdf
- 55. Valencia J, Garzón-Tello K, Barragán-Paladínes M. Serpientes venenosas del Ecuador: sistemática, taxonomía, historia natural, conservación. Envenenamiento y aspectos antropológicos [Internet]. 1st ed. Quito: Fundación Herpetológica Gustavo Orcés; 2016 [cited 2018 Oct 18]. http://biblioteca.udla.edu.ec/client/en_US/default/search/detailnonmodal?qu=Barragán-Paladines%2C+María+Elena&d=ent%3A%2F%2FSD_ILS%2F0%2FSD_ILS%3A29363~0&ic=true&te=ILS&ps=300
- Bravo-Vega C, Cordovez J, Renjifo-Ibañez C, Santos-Vega M, Sasa M. Data from: Estimating snakebite incidence from mathematical models: A test in Costa Rica. figshare. Dataset. [Internet]. figshare; 2019 [cited 2019 Jun 22]. https://figshare.com/articles/Data_and_code_used_for_this_study/ 8311994/2