



Comment

Cite this article: Gutiérrez EE. 2016 Ecological niche modelling requires real presence data and appropriate study regions: a comment on Medone *et al.* (2015). *Phil. Trans. R. Soc. B* **371**: 20160027.
<http://dx.doi.org/10.1098/rstb.2016.0027>

Accepted: 19 February 2016

Subject Areas:

health and disease and epidemiology

Author for correspondence:

Eliécer E. Gutiérrez

e-mail: gutierrez@si.edu

The accompanying reply can be viewed at
<http://dx.doi.org/10.1098/rstb.2016.0188>.

Ecological niche modelling requires real presence data and appropriate study regions: a comment on Medone *et al.* (2015)

Eliécer E. Gutiérrez^{1,2}

¹PNPD-Ecologia, Departamento de Zoologia, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília DF, 70910-900, Brazil

²Division of Mammals, National Museum of Natural History, MRC 108, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA

Medone *et al.* [1] used ecological niche modelling (ENM) analyses to assess the climatic suitability expected by the year 2050 in Venezuela and Argentina for two vectors of Chagas disease, the kissing bugs *Rhodnius prolixus* and *Triatoma infestans*. Based on these analyses, and on epidemiological data, Medone *et al.* [1] projected changes in the rate at which susceptible humans acquire Chagas disease and in the incidence of the disease in both countries. Medone *et al.* [1] concluded that by 2050 the climatic suitability for these vectors would decrease in areas in which a high-to-moderate transmission risk is currently observed. Herein, I argue that the nature of both the presence data and the study regions used by Medone *et al.* [1] limit, and perhaps even invalidate, the biological interpretations derived from their ENM analyses.

At the core of ENM is a process termed ‘model calibration’, mathematically characterizing the environmental conditions suitable for a species. Model calibration is achieved by means of contrasting the environmental conditions at sites in which the focal species is known to be present (presence data) with those at sites in which such species is absent (absence data) or at sites in which such species has not been observed (pseudo-absence data). Regardless of which of these data types are used, and of the algorithm chosen to model species’ niches, a fundamental requirement of ENM is the use of real presence data [2].

Low-quality presence data preclude reliable modelling of species’ niches [3–5]. Two errors are widely recognized, namely incorrect taxonomic identifications leading to the use of presence data not belonging to the focal species [3,5] and assignment of incorrect latitude or longitude during the georeferencing process [4]. These errors flaw the model calibration process by assuming environmental conditions not matching those of sites in which the focal species is or can be present. Common steps to assure high-quality data for ENM analyses include: (1) examine museum specimens to both confirm the taxonomic identification of the focal species and obtain associated locality data [3,6], (2) georeference with as many sources as necessary (e.g. maps, GIS software, gazetteers, field notes, interviews with specimen collectors; [7]), (3) exclude data with high georeferencing uncertainties [4]; and (4) account analytically for uncertainties generated in the georeferencing process [8–10]. Medone *et al.* [1] did not carry out any of these procedures. Instead, they digitized the expert-drawn range maps of Carcavallo *et al.* [11]. Then, they randomly selected sites within these maps to obtain a large number of data points for each species (*R. prolixus*, $n = 1240$; *T. infestans*, $n = 2350$). Medone *et al.* [1] calibrated the models for each species based on the resulting datasets as if they represented real presence data, even though they recognized that ‘... not all these points are actual presence points ...’. The data so generated are not adequate for ENM analyses because an unknown fraction of them does not represent real presence data.

The expert-drawn range maps [11] employed by Medone *et al.* [1] are not reliable. Such maps were made on the basis of unspecified, subjective criteria

(e.g. ‘... according to current experience and information...’ [11]). The literature cited as the source of information for these maps is insufficient to allow accurate depiction of the wide ecogeographic distribution of *R. prolixus* and *T. infestans*. Moreover, some of the cited literature lack geographical information of individual specimen localities (e.g. [12,13]). The inclusion of small areas covered by seawater for some species (e.g. *Alberprosenia malheiroi* and *R. paramensis*) suggests that the drawing technique used to make these maps, including those of the focal species, was rudimentary and error-prone. This kind of error cannot easily be detected and corrected when the erroneously included areas match smaller bodies of water or terrestrial habitats unsuitable for the species. Even if the range maps of Carcavallo *et al.* [11] were made according to adequate standards, still the information extracted from them do not represent real presence data. Although range maps can be adequate for some general purposes (e.g. illustrating species distributions in field guides), the scale and resolution to which they are typically drawn preclude their use as source of presence data for ENM analyses. That is, because range maps reflect the extent of occurrence and not necessarily the area of occupancy of species, they are prone to false positives, i.e. to include areas with unsuitable environmental conditions for the focal species [14]. A search for false positives in the datasets of Medone *et al.* [1] by extracting elevation values from a GIS file (30" resolution; [15]) allows flagging sites at elevations of 3000–3600 m for *R. prolixus* and of 4300–4650 m for *T. infestans* that are at odds with their own assertion that the elevation range of these species is 0–2600 m and 0–4100 m, respectively. Another example of possible false positives involve ‘data’ for *R. prolixus* for sites in Panama and southern Costa Rica, which are within areas where the species has not been found to occur [16].

Medone *et al.* [1] claimed to have considered ‘...the possible prediction errors that may result [from their data-

gathering procedure], and how they compare with prediction errors resulting from the use of confirmed presence occurrences...’ First and foremost, no comparison of the prediction errors caused by use of unconfirmed occurrences was presented throughout the article or its electronic supplementary material. By contrast, the percentage of confirmed occurrences correctly predicted by the models of each species were indeed reported (*R. prolixus*, 84.17%; *T. infestans*, 93.50%). Unfortunately, the *p*-values (at the selected threshold) associated with these predictions were not reported. Moreover, neither the geographical projection of the models nor the percentage of the study region predicted as suitable by each model was reported. Hence, the low omission rates obtained might have resulted simply from extremely broad predictions caused by pervasive positional errors in the ‘presence’ data.

The choice of study regions determines the performance of ENM analyses [17–19]. The basic principle to appropriately choose study regions is that they should not include areas that cannot be accessed by the species owing to its limited dispersal abilities or to the presence of barriers to dispersal [17–19]. Operational strategies have been proposed to minimize possible violations of this principle [7,18], but it is unclear whether Medone *et al.* [1] implemented them. More importantly, the geographical coordinates or polygons that define their chosen regions were not reported, making it impossible for other authors to both assess the appropriateness of their study regions and replicate their ENM analyses.

The use of inadequate presence data and unknown study regions for model calibrations render the results obtained by Medone *et al.* [1] unreliable for interpretations of biological phenomena. Unless these issues are further and satisfactorily clarified, their results could potentially misguide the design of future research, health policies and epidemiological programmes to control the vectors of Chagas disease.

References

- Medone P, Ceccarelli S, Parham PE, Figuera A, Rabinovitch JE. 2015 The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. *Phil. Trans. R. Soc. B* **370**, 20130560. (doi:10.1098/rstb.2013.0560)
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011 *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.
- Lozier JD, Aniello P, Hickerson MJ. 2009 Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *J. Biogeogr.* **36**, 1623–1627. (doi:10.1111/j.1365-2699.2009.02152.x)
- Feeley KJ, Silman MR. 2010 Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. *J. Biogeogr.* **37**, 733–740. (doi:10.1111/j.1365-2699.2009.02240.x)
- Romero D, Olivero J, Márquez AL, Báez JC, Real R. 2014 Uncertainty in distribution forecasts caused by taxonomic ambiguity under climate change scenarios: a case study with two newt species in mainland Spain. *J. Biogeogr.* **41**, 111–121. (doi:10.1111/jbi.12189)
- Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT. 2004 New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* **19**, 497–503. (doi:10.1016/j.tree.2004.07.006)
- Gutiérrez EE, Borja RA, Anderson RP. 2014 Can biotic interactions cause allopatry? Niche models, competition, and distributions of South American mouse opossums. *Ecography* **37**, 741–753. (doi:10.1111/ecog.00620)
- Wieczorek J, Guo Q, Hijmans RJ. 2004 The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inform. Sci.* **18**, 745–767. (doi:10.1080/13658810412331280211)
- Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG. 2013 Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37**, 191–203. (doi:10.1111/j.1600-0587.2013.00205.x)
- Velásquez-Tibatá J, Graham CH, Munch SB. 2015 Using measurement error models to account for georeferencing error in species distribution models. *Ecography* **39**, 305–316. (doi:10.1111/ecog.01205)
- Carcavallo RU, Galíndez Girón I, Jurberg J, Lent H. 1999 Geographical distribution and alti-latitudinal dispersion of Triatominae. In *Atlas of Chagas’ disease vectors in the Americas*, vol. 3 (eds RU Carcavallo, I Galíndez Girón, J Jurberg, H Lent), pp. 747–792. Rio de Janeiro, Brasil: FIOCRUZ.
- Lent H, Wygodzinsky P. 1979 Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of chagas’ disease. *Bull. Am. Mus. Nat. Hist.* **163**, 123–520. (<http://hdl.handle.net/2246/1282>)
- Carcavallo RU, Rabinovitch JE, Tonn RJ (eds). 1985 *Factores biológicos y ecológicos en la enfermedad de*

- Chagas*. Buenos Aires, Argentine Republic: Organización Panamericana de la Salud and Ministerio de Salud y Acción Social.
14. Habib LD, Wiersma YF, Nudds TD. 2003 Effects of errors in range maps on estimates of historical species richness of mammals in Canadian national parks. *J. Biogeogr.* **30**, 375–380. (<http://www.jstor.org/stable/3554565>)
 15. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978. (doi:10.1002/joc.1276)
 16. Hashimoto K, Schofield CJ. 2012 Elimination of *Rhodnius prolixus* in Central America. *Parasites & Vectors* **5**, 45. (doi:10.1186/1756-3305-5-45)
 17. Anderson RP, Raza A. 2010 The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *J. Biogeogr.* **37**, 1378–1393. (doi:10.1111/j.1365-2699.2010.02290.x)
 18. Barve N, Barve V, Jiménez-Valverde A, Maher SP, Peterson AT, Soberón J, Villalobos F. 2011 The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Modell.* **222**, 1810–1819. (doi:10.1016/j.ecolmodel.2011.02.011)
 19. Anderson RP. 2015 Modeling niches and distributions: it's not just 'click, click, click'. *Biogeografía* **8**, 4–27.