

On stability measures and effects of data structure in the recognition of areas of endemism

M. DOLORES CASAGRANDA* and PABLO A. GOLOBOFF

Unidad Ejecutora Lillo/CONICET, Miguel Lillo 251, San Miguel de Tucumán (CP 4000), Tucumán, Argentina

Received 29 November 2018; revised 2 February 2019; accepted for publication 4 February 2019

Incomplete data sampling, bias, and like properties of distribution datasets that potentially introduce uncertainty in biogeographical analyses and blur biogeographical patterns; therefore, it is important to understand their influence. Despite their relevance, these problems have been largely overlooked in biogeography, where concepts such as ambiguity, stability or support have not even been defined. Here, we propose two stability measures for hypotheses of areas of endemism (AEs) and use them to explore the degree to which different structural qualities of data affect the results of analyses of endemism. Our findings suggest that different types of data incompleteness have different effects on the recovery of the species composition and the geographical or spatial structure of AEs, showing that distinct levels of sampling coverage affect the stability of results in different ways. We show that a small proportion of poorly sampled species may have a stronger impact on AEs stability than many species with medium sampling and that excluding poorly sampled species from the analyses does not guarantee more stable results. These results highlight the importance of planning data collection and indicate that, in order to obtain more stable results, focusing on completing the distribution of strongly undersampled species might be preferable to adding records of any species randomly.

ADDITIONAL KEYWORDS: data bias – distribution data – endemicity analyses – endemism – NDM/VNDM – stability index.

INTRODUCTION

In biogeography, restrictions to geographical sampling (e.g. lack of access to certain areas; for example, see Ferraro & Casagranda, 2009; Turner *et al.*, 2009) and imperfect detection (MacKenzie *et al.*, 2005; Chen *et al.*, 2013) are factors leading to incomplete and often biased datasets that poorly represent the real distribution of species (Dennis *et al.*, 1999 and references therein). Given that knowledge of species ranges is fragmentary, analysis of real distribution data implies additional methodological challenges, such as the extent to which conclusions established on the basis of possibly biased data can reflect those biases more than a biogeographical reality.

In the fields of systematics and phylogenetics there have been extensive discussions of how different aspects of data structure affect confidence in

conclusions. Among these is the problem of how taxa with many unknown characters decrease bootstrap supports (starting with Wilkinson 1994, 1995; followed by, e.g. Pol & Escapa, 2009; Goloboff & Szumik, 2015) and the problem of the appropriateness of different methods for establishing confidence in phylogenetic groups (Felsenstein, 1985; Hillis & Bull, 1993; Farris *et al.*, 1996; Rydin & Källersjö, 2002). But despite the close historical ties between biogeography and systematics, those discussions in the realm of systematics have barely permeated biogeography (with few exceptions, such as Turner *et al.*, 2009). Although some previous papers have begun to explore the influence of fragmentary data on the recognition of areas of endemism (Arias *et al.*, 2008; Casagranda *et al.*, 2009), the problem is still poorly understood, and deeper studies are needed. Likewise, the related problem of data acquisition has long been considered by ecologists (Hortal *et al.*, 2007, 2008; Dengler & Oldeland, 2010; Tessarolo & Rangel, 2014), but not been considered in historical biogeography. Thus, in

*Corresponding author. E-mail: dolores.casagranda@gmail.com.

the historical framework, concepts such as ambiguity, stability, robustness or support have so far not even been defined.

This paper presents the first attempt to evaluate stability in hypotheses of areas of endemism (AE) in the face of different aspects of data incompleteness. We also provide examples of the application of these techniques for exploring the degree to which different sampling and structural qualities affect the results of analyses of endemism. It is hoped that this contribution will provide a starting point towards the formalization of support measures in historical biogeography and a step towards understanding the influence of data sampling in biogeographical studies and the importance of planned data collection.

MATERIAL AND METHODS

DATA

We use datasets from Weirauch *et al.* (2016) and Bertelli *et al.* (2017) and data for amphibians of the Amazonian countries (Bolivia, Brazil, Colombia, Ecuador, French Guiana, Guyana, Peru, Suriname and Venezuela) compiled from Global Biodiversity Information Facility (GBIF: <https://www.gbif.org/>), to apply different types of resampling in order to simulate differences in data structure

that can result from different sampling strategies or conditions. These datasets encompass different geographical areas and differ in taxon composition, number of species and sampling density. The dataset from Weirauch *et al.* (2016) (hereafter, NAM, 'North American Mirids'; Fig. 1A) compiles 61 016 records for 1339 species of Miridae of North America; the dataset from Bertelli *et al.* (2017) (compiled by Navarro-Sigüenza and Peterson over many years; hereafter MB, 'Mexican birds') includes > 100 000 records for 780 species of birds distributed in Mexico (Fig. 1B), and the dataset compiled from GBIF (Fig. 1C) includes ~34 000 records from 2000 species of amphibians of Northern South America (hereafter, SAA, 'South American amphibians'). We selected heterogeneous datasets for the analyses in the expectation that this would allow us to interpret common results as general patterns instead of random behaviours attributable to specific traits of a particular dataset.

RESAMPLING

The incompleteness of datasets may come from different problems. These problems are interrelated only in part, showing some degree of independence. The lack of representation of some distributions may result from either a uniformly incomplete sampling (a

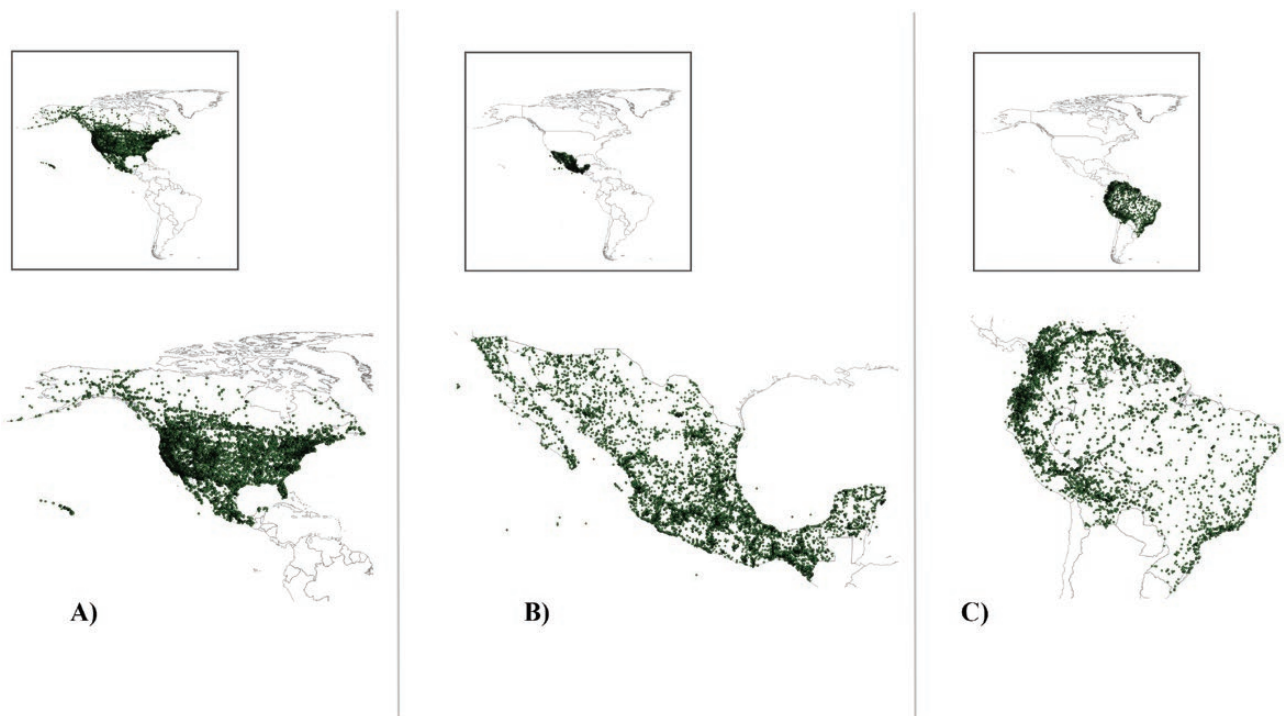


Figure 1. Distributional datasets. A, dataset from Weirauch *et al.* (2016), on North American Miridae. B, dataset from Bertelli *et al.* (2017), on Mexican birds. C, GBIF South American amphibians.

case in which we cannot speak of ‘bias’). Alternatively, some distributions may be incompletely sampled relative to others, and this bias may come from some species being sampled more strongly than others or from some regions of the study area being more thoroughly studied. Thus, we have implemented three different ways to emulate these problems in the data, to reflect these possibilities.

Case 1, poor general sampling

We consider as a case of ‘poor general sampling’ a dataset where the average number of localities recorded per species is low, meaning that the distributional range of most species is represented by only a few localities or unique records. To simulate this situation, we randomly removed records of the species in the dataset with a given probability, uniform for all species. We considered three subcases where every individual record was removed with probability 0.20, 0.45 and 0.70 (subcases 1A, 1B and 1C, respectively). The use of these three values is intended to mimic different levels of insufficient data collection, from soft to severe. Note that the removal of every individual point record is decided independently, so that different resampled datasets may have slightly different resulting numbers of records retained.

Case 2, unevenly sampled species

Distributional datasets often include unevenly sampled species, meaning that some species (by virtue of their rarity or taxonomic bias) are particularly undersampled, represented by only a small number of localities (in the SAA dataset, some species are even represented only by the type locality). We simulated three subcases of undersampled species by randomly removing records only in subsets of species; the probability of a species to be selected for record removal is 0.20, 0.45 and 0.70 (subcases 2A, 2B and 2C, respectively). For each of these subcases, we considered instances with 0.20, 0.45, and 0.70 removal probability per record in the species selected (instances I, II, and III, respectively). This produced a total of nine possible treatments.

Case 3, geographically biased sampling

Geographical biases are common in distribution datasets. Geographical sectors with good or easy accessibility (e.g. roads, rivers or paths) are more likely to be visited than those with difficult access and tend to have a higher density of records. To simulate geographical bias, we manually delimited four geographical sectors well explored and visited by amphibian collectors, using these sectors as a

constraint (Fig. 2A). We defined an inner resampling case, where records of species were removed only inside the sectors defined (Fig. 2B), and an outer resampling case, where records were removed outside the sectors defined, in order to increase the sampling imbalance (Fig. 2C). For each case, we considered subcases with 0.20, 0.45 and 0.70 removal probability per record, producing perturbations from minor to severe (subcases A–C of inner and outer resampling). It must be kept in mind that certain zones are naturally poor in species diversity or abundance, meaning that a lower concentration of records is not necessarily an unequivocal indicator of undersampling, and a higher density of records is not necessarily the result of oversampling. Therefore, some previous knowledge about the dataset and the study region is required in order to delimit well-collected areas properly. Therefore, we analysed the case of geographically biased sampling only for the SAA dataset, with which we are more familiar.

ANALYSES

We analysed the original (complete) and resampled datasets using the endemicity analyses implemented in VNDM v.3.1 (Goloboff, 2004), which identify AEs based on the criterion proposed by Szumik & Goloboff (2004). All the analyses were run on raw data, i.e. using only observed records, without extending or filling assumed species distributions. Searches were performed using 0.6 as minimal species score, two as minimal number of endemic species, and retaining overlapping subsets with a difference of 40%. Twenty replicates of resampling were run per subcase/instance, in all cases. Although most comparisons are based on the individual areas of endemism found (IAs), consensus areas (CAs) were also used to quantify some differences among original and resampled datasets (see Results section, ‘Unique areas’ and dataset instability). The CAs were built using the loose consensus method, with a 40% similarity in species composition as cut-off for merging (for details, see Aagesen *et al.*, 2013).

MEASUREMENT OF STABILITY

The degree to which the results vary in the face of the perturbation from resampling can be measured in different ways. The approach used here is based on calculating, for each of the areas found for the complete dataset, the degree to which the area is recovered in the perturbed dataset. Given that not all areas need to be recovered exactly for the resampled dataset, some measure of distance between areas is needed to assess the stability of the results in the face of resampling (in phylogenetics, there have been proposals also to measure the degree to which the groups found during

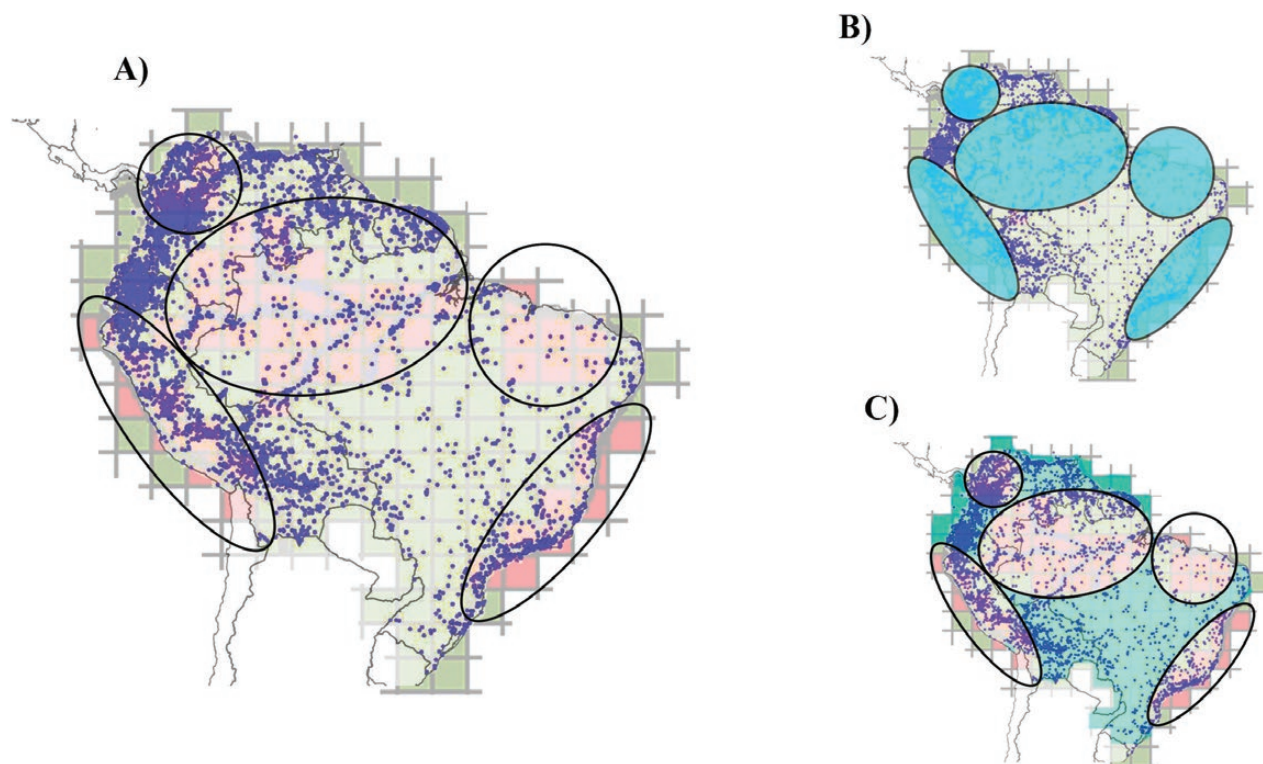


Figure 2. Geographically biased resampling. A, well-explored sectors are marked as ellipses. B, inner resampling: records are removed only from inside the well-explored areas (in light blue). C, outer resampling: records are removed from the whole area, except for the well-explored areas (light blue area).

resampling resemble the monophyletic groups; see Goloboff *et al.*, 2009; Lemoine *et al.*, 2018). Here, two measures of similarity are used to consider separately two dimensions of AEs: the geographical extension (geographical stability) and the biological composition (species stability).

The geographical stability considers the degree to which the cell composition in the two areas is similar; given that the approach of Szumik & Goloboff (2004) treats the problem as a combinatorial problem, minor differences in the inclusion or exclusion of cells (specially marginal cells) produce only minor differences in the score of endemity, thus producing variations of the ‘same’ area rather than a ‘new’ area. The similarity, C_{AB} , of cell composition in two areas A and B is given by $C_{AB} = c_{AB}/(c_{AB} + c_A + c_B)$, where c_{AB} is the number cells shared by both A and B, c_A the number of cells unique to A, and c_B the number of cells unique to B. Each of the areas for the original dataset is considered to be retrieved if the cell composition similarity is equal to or larger than a user-defined threshold (the analyses presented here used 0.65), not retrieved otherwise; the final calculation produces (for each area) the frequency of recovery. The results are output for both the IAs and for their consensus, CAs. The CAs,

in the case of resampling, groups the areas found for the original dataset that have a similar species composition (as done by Aagesen *et al.*, 2013); given that the different IAs included in a given CA may have minor differences in their cell composition, the frequency with which each of the cells is included in the underlying IAs is indicated with a colour code.

The species stability is similar to the geographical stability, but considers instead the sets of species that are (according to the criterion of Szumik & Goloboff, 2004) mapped as endemic to the area; we define s_{AB} , s_A and s_B in an analogous manner to c_{AB} , c_A and c_B , and the species similarity index is given by $S_{AB} = s_{AB}/(s_{AB} + s_A + s_B)$. Note that (by virtue of the rules to accept a species as endemic to an area), there has to be some degree of correspondence between the species distribution and the cells that compose the area, for the species to be considered as endemic; thus, this criterion takes care of the spatial similarity indirectly. As in the case of geographical similarity, one of the areas found for the original dataset is considered to have been recovered in the resampled dataset if the species similarity index is above a certain threshold (0.65, in the present examples).

Those IAs found for the resampled data not sharing a similarity > 0.65 with any of the IAs for the full

dataset were summarized by CAs and reported as ‘unique areas’ for completeness.

IMPLEMENTATION

Given the intensive computations required for this process, full automation is necessary. The required routines have been included in VNDM (Goloboff, 2004) v.3.1.

The command *autosample N* in the input file indicates that VNDM should automatically proceed to perform resampling (this requires that the results from an independent search for the full dataset are read together with the original distributions, in the form of an *.ndm file); the number *N* (1–3) indicates the type of resampling (these options are numbered as the cases 1–3 above, in the section on ‘Resampling’).

The command *autosave* instructs VNDM to save the results for resampling automatically (as image files, with summaries of results) and then quit, facilitating processing by means of batch files, determining further options with additional commands.

The command *samplespp* sets the probability of choosing a species for deletion of some records when *autosample 2* is in effect; *samplerec* sets the probability of eliminating an individual record (under *autosample 1* or *2*); *samplearea* sets the probability of eliminating a record within the last area held in memory when doing the resampling under *autosample 3* (if the subsequent number is positive), or outside of it (if the subsequent number is negative). Note that the frequency of recovery of the last area held in memory is not calculated under *autosample 3*, using the last area only as reference for enabling or disabling deactivation of records.

General options for the number of replications or comparison between areas are set with *samplerexpls*, for the number of replications, and *sampleareacut* or *samplesppcut*, for the similarity above which an area is considered as ‘recovered’ (if the subsequent number is negative, then average values are calculated, but only when above threshold; defining either of these options as zero implies that the average degree of similarity is calculated for each area). The option *autoconsense N* automatically produces consensus areas, with cut-off *N*; the loose consensus rule (see Aagesen *et al.*, 2013) is always used for autocalculating the consensus.

INTERPRETATION

Methods based on resampling can be used for evaluating several aspects of the results (see, e.g. Goloboff *et al.*, 2003; Egan, 2006; Hovenkamp, 2009). The purpose of our approach is to provide some estimation of the expected stability of the results, i.e. to give some idea of how likely the conclusions are to

change in the face of new evidence. Predicting what type of distributional evidence can be added to the dataset would require knowledge both of the true distributions and of how they are sampled in practice; for simplicity, the comparison can be made between the estimation that would have been made if the sample of records had been incomplete (i.e. as obtained by resampling) and the estimation for the complete dataset. It is thus expected that the difference between the results with some records removed (with some degree of randomness) and with the full records will provide some insight into what the difference could be between the results for the present dataset and a dataset enlarged in the future.

Another aspect of the results that needs to be established is that of support, i.e. the degree to which the evidence seems to uphold the conclusions established on the basis of the full dataset. Although related to the notion of stability, this is not exactly the same (e.g. resampling may show that conclusions contradicted by the evidence are nonetheless likely to be obtained by increased sampling; see Goloboff *et al.*, 2003 for discussion). Both the stability under resampling and the endemicity score as defined by Szumik & Goloboff (2004) seem to be related to the degree of support of the conclusions; the endemicity score increases with larger numbers of taxa showing a better conformity to an area of endemism, and the stability under resampling decreases as the distributions underlying and area of endemism are more discordant. Thus, both stability and endemicity score can be expected generally to increase as more evidence points to similar conclusion (= support). Comparison of the degree to which there is a correlation between the stability of a concluded area and the numerical value of its endemicity score provides, to some degree, validation of the extent to which these two aspects may indeed be appropriately reflecting differences in support.

RESULTS

In order to explore the effects of different sampling strategies on the identification of AEs, we calculated geographical and species stability for areas of endemism obtained under different perturbation schemes represented in the cases described in the Material and methods section (cases 1, 2 and 3).

CASE 1, POOR GENERAL SAMPLING

When analysing poor general sampling, both geographical and species stability progressively diminish with the increase of data removal, for all subcases. The higher the probability of a record to be removed, the less likely an original set of cells and the species constituting an

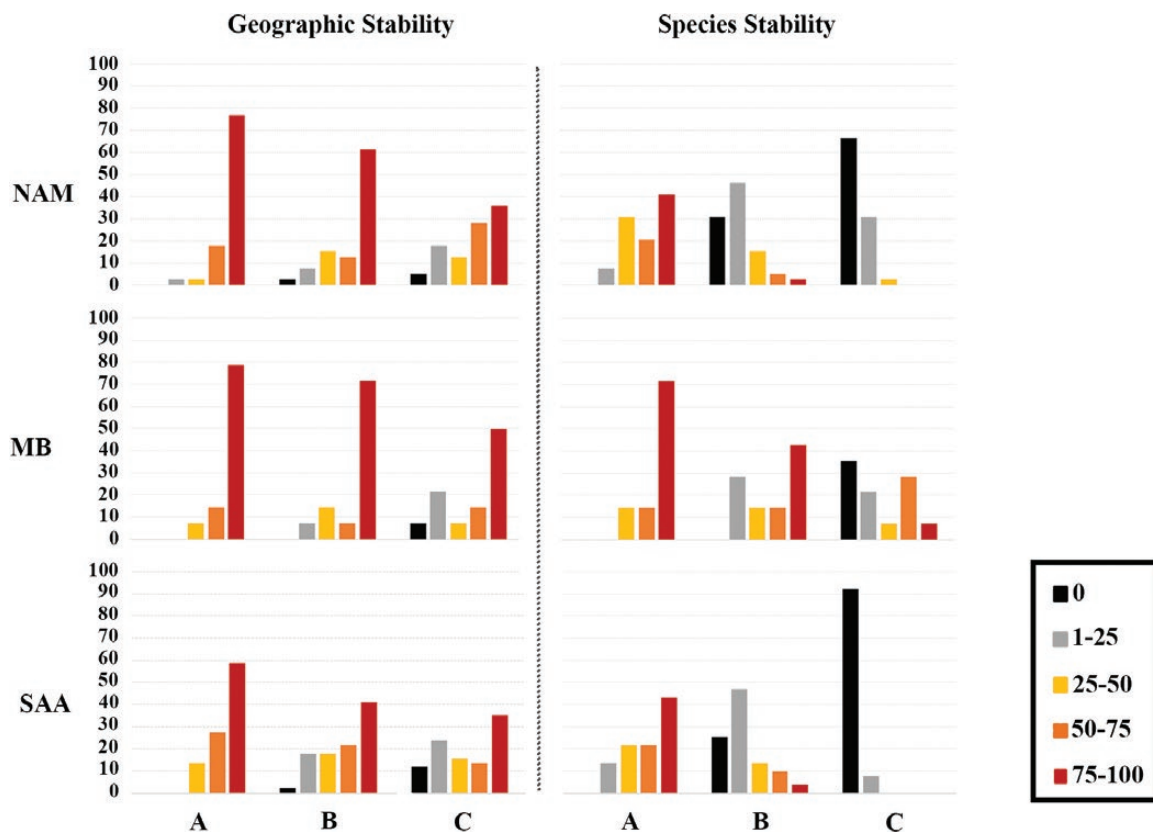


Figure 3. Poor general sampling case. Proportion of areas recovered with different values of geographical and species stability, after removal of records with a probability of 0.2, 0.45 and 0.7 (A, B and C, respectively). Bar colours represent different values of stability.

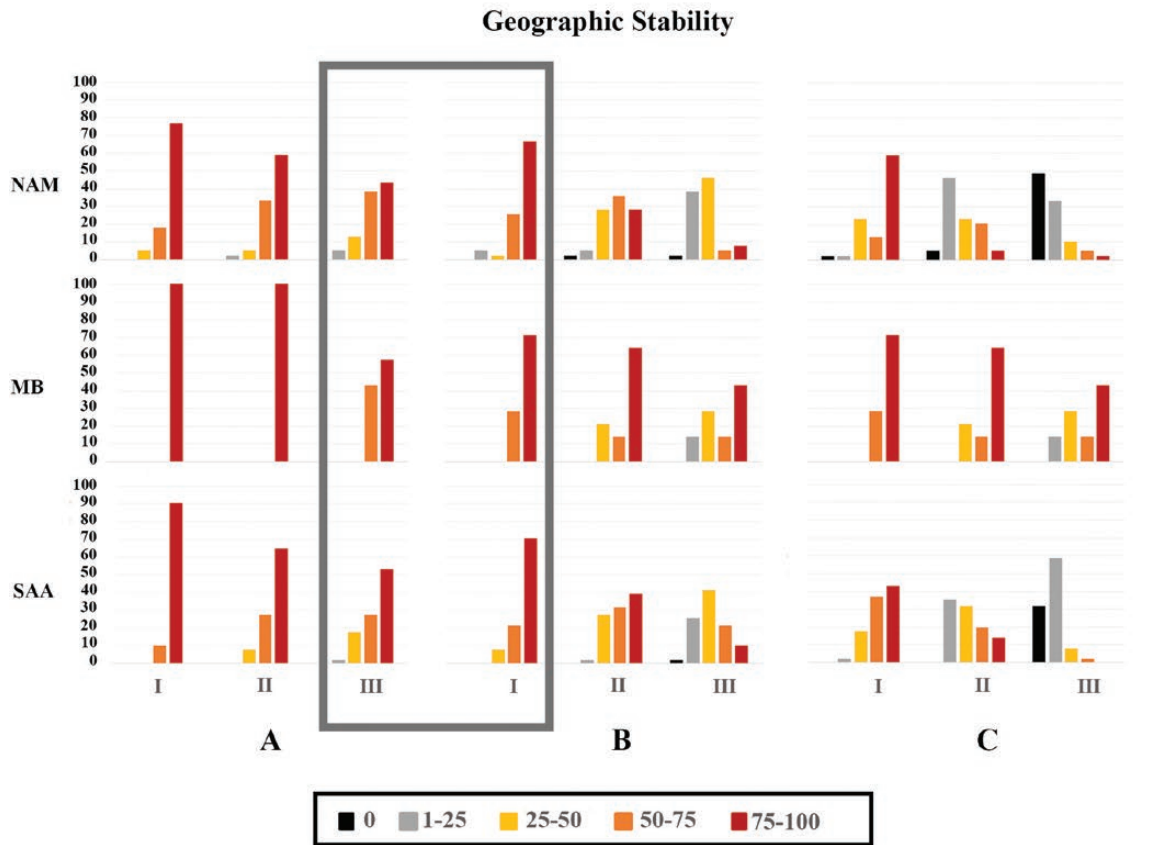
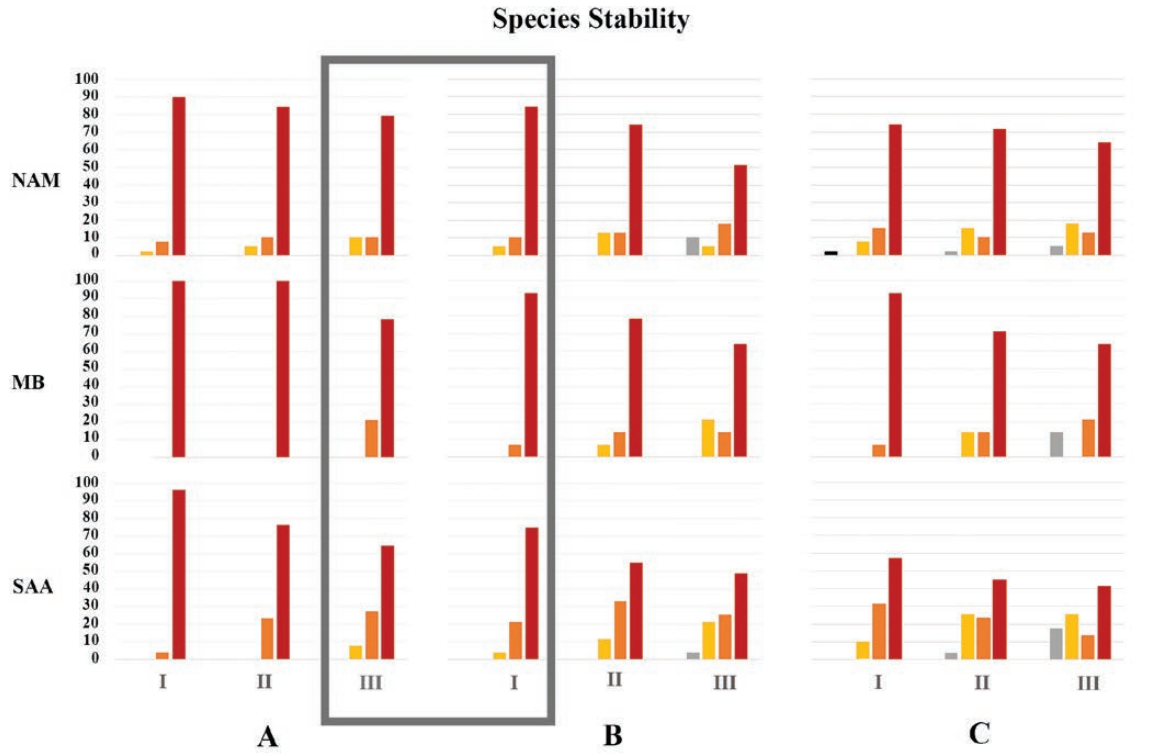
AE are to be recovered. However, geographical stability seems to be less affected than species stability. Black bars in [Figure 3](#) show the percentage of original AEs never recovered when the dataset is resampled with different probability of record removal (subcases A, B and C described above), i.e. those AEs with geographical and species stability equal to zero; but keep in mind that the 0.65 threshold is playing some role here. The difference between the number of areas and the number of endemic species never recovered is clear-cut. Although this pattern was consistently observed in all datasets, MB is slightly more stable than the other two datasets, showing the lowest number of original AEs with stability equal to zero (see [Fig. 3](#)).

CASE 2, UNEVENLY SAMPLED SPECIES

Results for undersampled species are similar to those in the previous case (for all subcases); the stability

of AEs decreases as both the percentage of species resampled and the probability of data removal are increased, as shown in [Figure 4](#). The subcases reveal some differences in the behaviour of different datasets. When analysing subcase 2A (resampling species with a probability of 0.2), NAM and SAA show similar patterns of variation for both geographical and endemic stability; however, when increasing the probability of choosing a species for record removal to 0.45 (subcase 2B), NAM performed slightly worse than SAA in terms of geographical stability. MB produces the most stable results, showing a higher stability for most AEs and with very few AEs never recovered (stability zero), even in the most aggressive subcases of resampling (see [Fig. 4A, B](#)). A small subset of very poorly sampled species (probability of 0.20 to choose a species for eliminating records, with removal probability of 0.70) affects stability more than a milder perturbation on a larger proportion of species

Figure 4. Unevenly sampled species case. Proportion of areas recovered with different values of geographical and species stability, when removing records for a subset of species (A, B and C indicate a probability of 0.20, 0.45, and 0.70, respectively, for a species to be selected for record removal) with a probability of 0.2, 0.45 and 0.7 (I, II and III, respectively). Bar colours represent different values of stability.



(case 2A-III vs. case 2B-I), as shown in [Figure 4](#) (cases framed in black rectangles). This observation could be taken as a guide when designing data-collection strategies aiming to improve dataset stability; focusing on the addition of new localities for poorly collected species would be better than randomly adding records for any species, even when the effective number of new records is the same. Following one or other strategy would apparently have different consequences on stability.

'UNIQUE AREAS' AND DATASET INSTABILITY

The number of unique areas resulting from resampled datasets increases with data perturbation, in all cases and for all datasets, being especially noticeable in cases 1 and 2 (poor general sampling and unevenly sampled species). The number of unique areas can be considered as an additional indicator of the relative stability of results; i.e. under perturbation, some datasets produce numerous unique patterns, not identified when analysing the complete dataset, whereas others produce consistent results and do not show major differences in the number and/or shape of the AEs after data removal. Considering a high number of unique areas as an indicator of instability, MB is found to produce the most stable results among the analysed datasets ([Fig. 5](#)).

CASE 3, GEOGRAPHICALLY BIASED SAMPLING (SAA ONLY)

For the SAA dataset, we identified five frequently visited/collected zones, coincident with the Guayana, Atlantic Forest, Andes and Caribbean areas, and the Amazonian portion related to the river network (see [Fig. 2](#)). These zones were used to define inner and outer resampling schemes, as described in the Material and methods section (see [Fig. 6](#)). Our results show that application of mild and strong resampling, either inside or outside the delimited areas, affects both geographical and species stability of AEs; however, geographical stability seems to be more influenced by inner resampling than by outer resampling ([Fig. 6](#)). This might be attributable to data concentration; most of the AEs are located in frequently sampled zones, which in most cases concentrate a higher density of records, thus removal of records within those zones would affect the recovery frequency. Outer resampling has slightly stronger effects on species stability ([Fig. 6](#)). The major difference between inner and outer resampling is observed when counting for unique areas. Outer resampling generates a larger number of unique areas, suggesting that accentuating sampling imbalance would generate major overall instability ([Fig. 6](#)).

STABLE AND UNSTABLE AREAS OF ENDEMISM

On closer inspection of the MB dataset, we observed that some AEs are more stable than others, always showing high recovery frequencies independently of the degree of data removal; we call these stable AEs (see AEs 4, 6, 7, 11 and 13 in [Fig. 7A](#)). This has been observed only for geographical stability, having no correlate in species stability, where the recovery frequency of areas mostly decreases with data perturbation, as intuitively expected ([Fig. 7B](#)). In contrast, unstable AEs (again, in reference only to geographical stability) are very sensitive to resampling, and their stability decrease rapidly with data perturbation (see AEs 1, 2 and 8 in [Fig. 7A](#)) or change in unexpected ways (e.g. higher probabilities of data removal produce more stable results; as examples, see AEs 0, 3 and 9 in [Fig. 7A](#)). In the examples analysed, all the stable AEs have a high endemism score (ES; [Fig. 7C](#)). This, however, is not necessary reciprocal, as demonstrated by AE5, an unstable area with a high ES; therefore, a high ES seems to be a necessary but not a sufficient condition for stability.

DELETING POORLY SAMPLED SPECIES?

In their paper, [Weirauch *et al.* \(2016\)](#) created three different matrices based on the original data, by progressively eliminating poorly sampled species and obtaining matrices including only species with more than three, five and ten records, respectively. Although the reasons for excluding information in that way are not given explicitly, it seems to be based on the idea that 'noisy' species can be misleading and should be excluded in order to improve data quality.

To test this, we analysed the reduced matrices of [Weirauch *et al.* \(2016\)](#) and compared results in terms of stability, finding that filtering poorly sampled species does not always improve results. In most cases, geographical stability changes positively when removing a priori poorly sampled species, although this trend is not straight in case 2C-II and case 2C-III, and the exception of case 2A-I (see [Fig. 8A](#)). The opposite occurs with species stability in most cases; results have decreased in stability with a priori removal of poorly sampled species, except by the resampling cases with a high probability (0.7) of species resampling (see cases 2C-II and 2C-III in [Fig. 8B](#)). This makes us think that inclusion of poorly sampled species does not necessarily diminish the quality of results.

DISCUSSION

The development of stability measures that evaluate separately the spatial extension and species

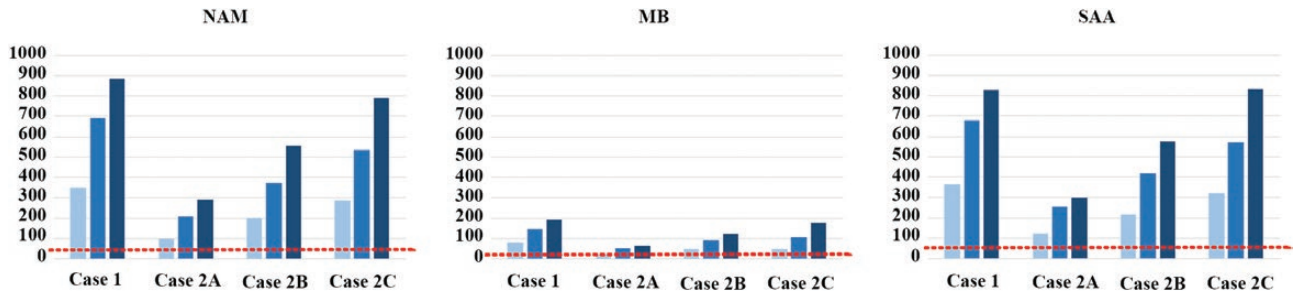


Figure 5. Unique areas. Number of unique areas found in cases 1 and 2 for each dataset. Different shades of blue indicate different probabilities of record removal: light blue 0.2, blue 0.45 and dark blue 0.70. Red dotted line indicates the total number of areas of endemism found with the original dataset.

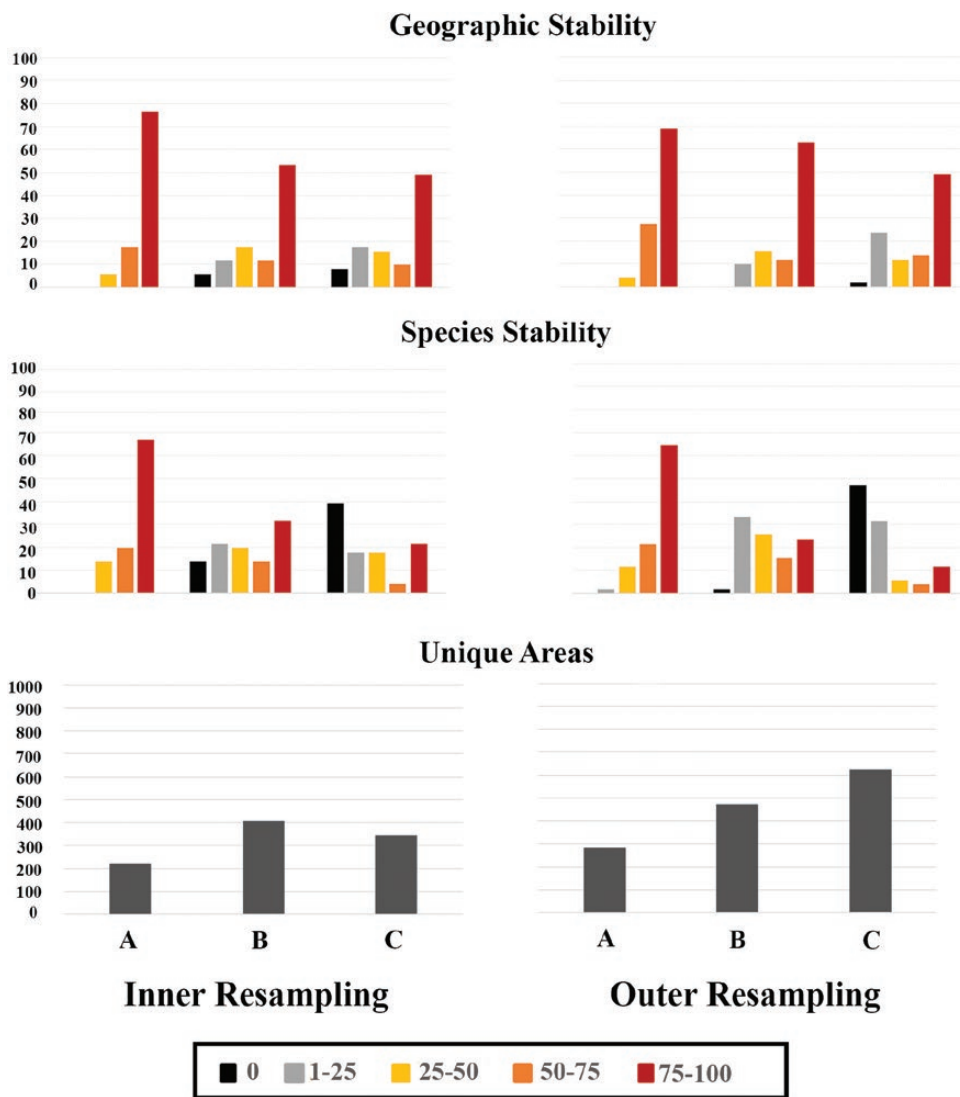


Figure 6. Geographically biased sampling case. Proportion of recovered areas in inner and outer resampling subcases, with different values of geographical and species stability under record removal probabilities of 0.2, 0.45 and 0.7 (A, B and C, respectively). Bar colours represent different values of stability. Grey bars indicate the number of unique areas found in inner and outer resampling subcases under the different record removal probabilities.

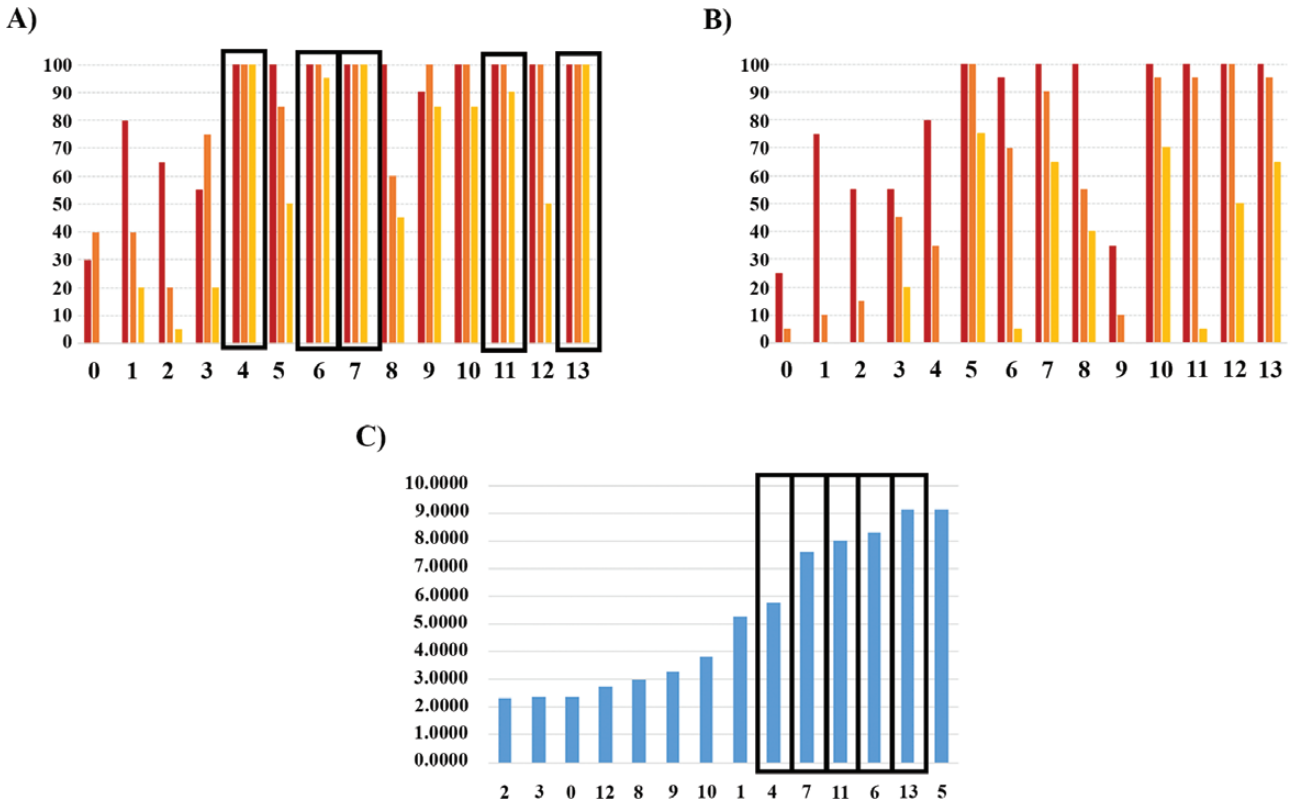


Figure 7. Stable areas. A, B, geographical (A) and species (B) stability values for each area of endemism (x -axis) are shown, for record removal probabilities of 0.2, 0.45 and 0.7 (red, orange and yellow bars, respectively). C, areas of endemism are sorted from lower to higher endemism values (y -axis). Stable areas are framed in black.

composition of AEs helps in visualization of how gaps and biases in data have a direct impact on identification and can affect recognition of the spatial and biotic dimensions of AEs to different extents. Our results suggest that data incompleteness might have a bigger impact on the recovery of the endemic species composition than on the geographical or spatial structure of AEs, with cases where the set of endemic species of an AE were liable to recovery after artificially generating loss of distributional information (see Figs 3, 4). Rather than being discouraging, this result has a positive side, suggesting that even scattered data can lead to a fair spatial description of AEs, delimiting their geographical extent and providing meaningful information to channel sampling efforts.

Our results show how the percentage of species with different degrees of sampling coverage affect stability of results (Fig. 9). The MB dataset, which produces the most stable results and includes the highest percentage (40%) of densely sampled species (we defined this as having ≥ 100 different localities). In contrast, SAA, producing rather unstable results, includes $< 3\%$ of species with a strong geographical sampling. The NAM dataset, the results for which show a stability intermediate between the other

datasets, has 10% of its species densely sampled. The correlation between sampling density and stability corroborates the importance of working with a good representation of species distribution.

The comparisons performed indicate that a small proportion of poorly sampled species causes stronger impact on AE stability than almost half of species mildly sampled (see Fig. 4). They also show that exclusion of poorly sampled species from biogeographical analyses does not necessarily provide a solution for improving the stability of the results (as shown in Fig. 8). These empirical observations suggest that focusing on completing the distribution of strongly undersampled species might be preferable to randomly adding records of any species.

According to our analyses, certain AEs seem to be geographically more stable than others, presenting a high frequency of recovery through different degrees and strategies of data removal. A better understanding of whether the stability (or lack thereof) is a natural property of AEs, or whether it is instead a consequence of data structure (gaps, distance among records, etc.), would help in depicting how data quality affects the general reliability of results. The stability measures proposed here provide a tool to quantify empirical

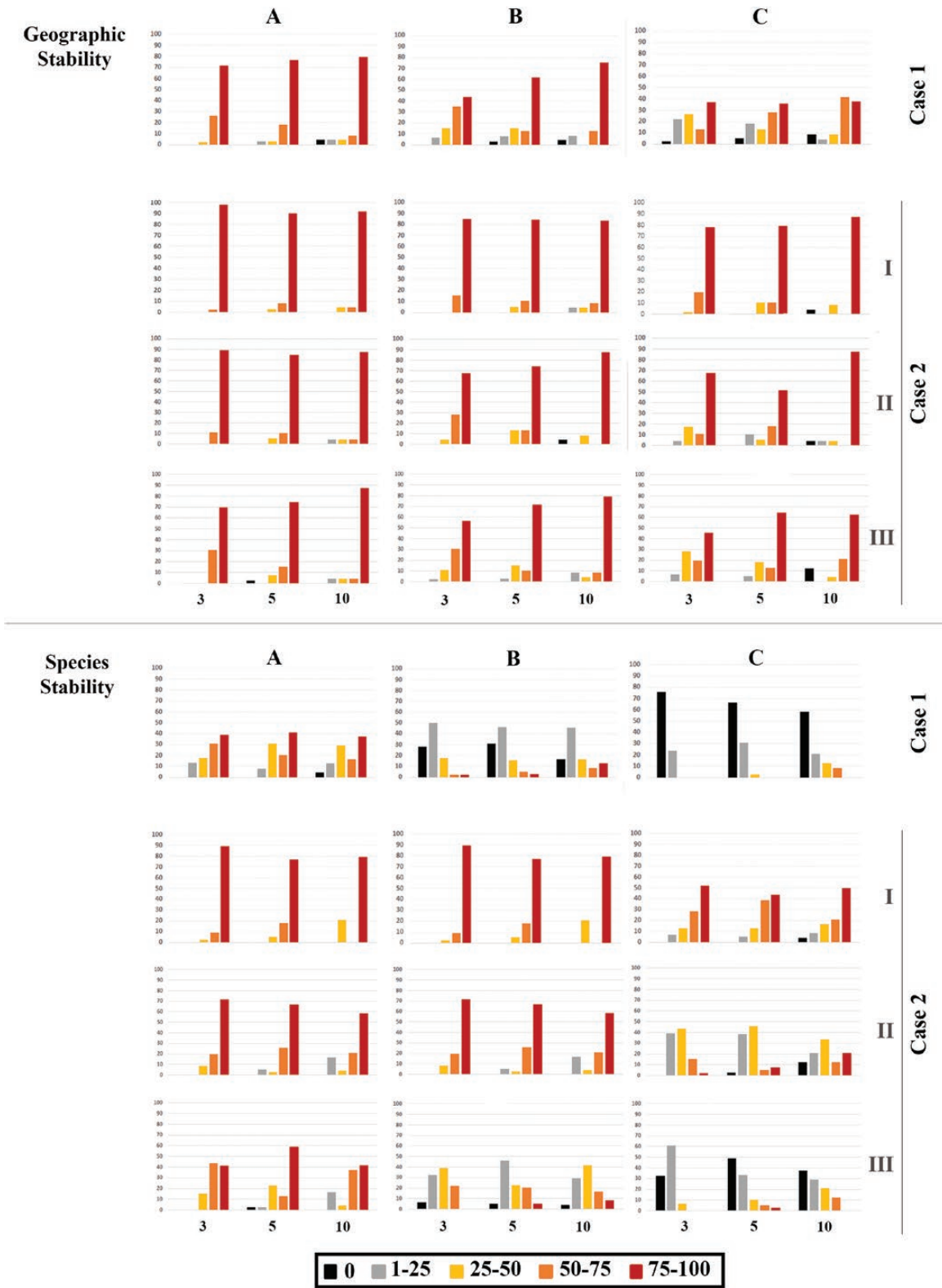


Figure 8. Reduced datasets. Proportion of areas recovered with different values of geographical (A) and species (B) stability in cases references framed 1 and 2 and their subcases, when using datasets with species with more than three, five and ten records (x-axis). The colours of bars represent different stability value ranges.

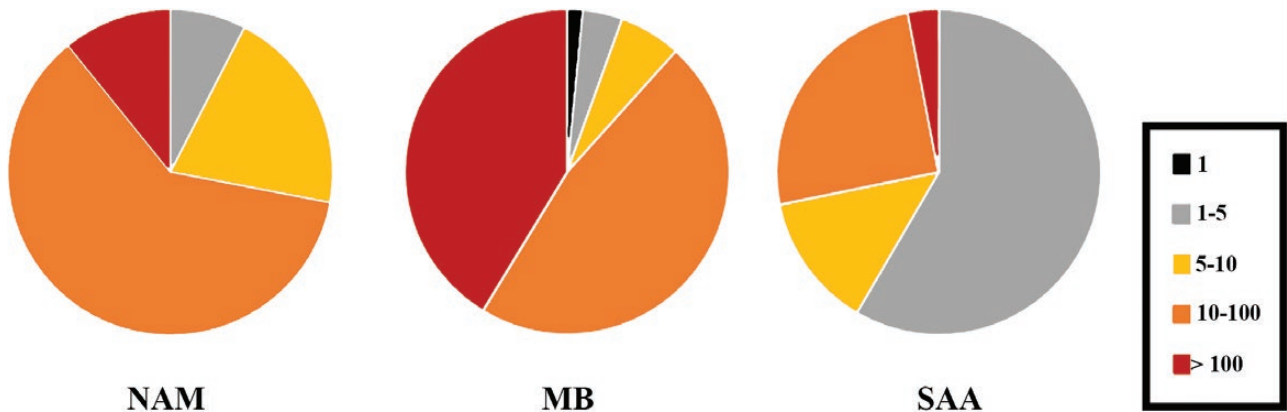


Figure 9. Species sampling per dataset. Proportion of species with different degrees of geographical sampling coverage, for NAM, MB and SAA datasets. Different colours represent different degrees of sampling coverage, expressed in number of sampled localities (see colour key).

support of biogeographical hypotheses and to understand the limits of the available evidence.

We anticipate that this paper will inspire more studies along similar lines and that the trends reported here will help to provide insights to adjust analyses in order to lead to more solid results and make more realistic interpretations of AEs. This is especially important given that the conclusions on AEs are often intended to be used for defining priorities in conservation or delimiting biogeographical schemes.

ACKNOWLEDGEMENTS

We thank Christiane Weirauch and Adolfo Navarro-Sigüenza for making their datasets available. M.D.C. thanks Claudia Szumik for discussion and comments on early ideas of this project, and Marcos Mirande, Salvador Arias and Juan Manuel Morales for discussion and comments on the first drafts of this manuscript. An initial version of this work was presented at the XXXVI International Meeting of the Willi Hennig Society (WHS). We also thank three anonymous reviewers for their helpful comments. M.D.C. thanks the WHS for financial support to assist to the Meeting, and thanks to the participants for their insights. We thank Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) for financial support (PUE 0070).

REFERENCES

- Aagesen L, Szumik C, Goloboff P. 2013. Consensus in the search for areas of endemism. *Journal of Biogeography* **40**: 2011–2016.
- Arias JS, Casagranda MD, Diaz Gómez JM. 2008. A comparison of NDM and PAE using real data. *Cladistics* **26**: 204.
- Bertelli S, Szumik C, Goloboff PA, Giannini NP, Navarro-Sigüenza AG, Peterson AT, Cracraft J. 2017. Mexican land birds reveal complexity in fine-scale patterns of endemism. *Journal of Biogeography* **44**: 1836–1846.
- Casagranda MD, Arias JS, Goloboff PA, Szumik CA, Taher LM, Escalante T, Morrone JJ. 2009. Proximity, interpenetration, and sympatry networks: a reply to Dos Santos *et al.* *Systematic Biology* **58**: 271–276.
- Chen G, Kéry M, Plattner M, Ma K, Gardner B. 2013. Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology* **101**: 183–191.
- Dengler J, Oldeland J. 2010. Effects of sampling protocol on the shapes of species richness curves. *Journal of Biogeography* **37**: 1698–1705.
- Dennis RL, Sparks TH, Hardy PB. 1999. Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation* **3**: 33–42.
- Egan MG. 2006. Support versus corroboration. *Journal of Biomedical Informatics* **39**: 72–85.
- Farris J, Albert V, Källersjö M, Lipscomb D, Kluge A. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**: 99–124.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- Ferraro DP, Casagranda MD. 2009. Geographic distribution of the genus *Pleurodema* in Argentina (Anura: Leiuperidae). *Zootaxa* **55**: 33–55.
- Goloboff P. 2004. *NDM/VNDM ver. 2.5. Programs for identification of areas of endemism*. Programs and documentation are available at: <http://www.lillo.org.ar/phylogeny/endemism/>
- Goloboff P, Catalano S, Mirande M, Szumik C, Arias JS, Källersjö M, Farris JS. 2009. Phylogenetic analysis of 73,060 taxa supports major evolutionary groups. *Cladistics* **25**: 211–230.
- Goloboff P, Farris J, Oxelmann B, Ramírez M, Szumik C. 2003. Improvements to resampling measures of group support. *Cladistics* **19**: 324–332.

- Goloboff P, Szumik C. 2015.** Identifying unstable taxa: Efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. *Molecular Phylogenetics & Evolution* **88**: 93–104.
- Hillis DM, Bull JJ. 1993.** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**: 182–192.
- Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A. 2008.** Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* **117**: 847–858.
- Hortal J, Lobo JM, Jiménez-Valverde A. 2007.** Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology* **21**: 853–863.
- Hovenkamp P. 2009.** Support and stability. *Cladistics* **25**: 107–108.
- Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. 2018.** Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**: 452–456.
- MacKenzie DI, Nichols JD, Sutton N, Kawanishi K, Bailey LL. 2005.** Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* **86**: 1101–1113.
- Pol D, Escapa IH. 2009.** Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics* **25**: 515–527.
- Rydin C, Källersjö M. 2002.** Taxon sampling and seed plant phylogeny. *Cladistics* **18**: 485–513.
- Szumik CA, Goloboff P. 2004.** Areas of endemism: an improved optimality criterion. *Systematic Biology* **53**: 968–977.
- Tessarolo G, Rangel TF, Araújo MB, Hortal J. 2014.** Uncertainty associated with survey design in species distribution models. *Diversity and Distributions* **20**: 1258–1269.
- Turner AH, Smith ND, Callery JA. 2009.** Gauging the effects of sampling failure in biogeographical analysis. *Journal of Biogeography* **36**: 612–625.
- Weirauch C, Seltmann KC, Schuh RT, Schwartz MD, Johnson C, Ann M, Soltis PS. 2016.** Areas of endemism in the Nearctic: a case study of 1339 species of Miridae (Insecta: Hemiptera) and their plant hosts. *Cladistics* **33**: 279–294.
- Wilkinson M. 1994.** Common cladistic information and its consensus representation: reduced Adams and cladistic consensus trees and profiles. *Systematic Biology* **43**: 343–368.
- Wilkinson M. 1995.** Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* **44**: 501–514.