Global Change Biology

Global Change Biology (2010), doi: 10.1111/j.1365-2486.2010.02239.x

The data void in modeling current and future distributions of tropical species

KENNETH J. FEELEY*†‡ and MILES R. SILMAN*

*Department of Biology, Wake Forest University, Winston-Salem, NC 27106, USA, †Department of Biological Sciences, Florida International University, Miami, FL 33199, USA, ‡Center for Tropical Plant Conservation, Fairchild Tropical Botanic Garden, Coral Gables, FL 33156, USA

Abstract

Conserving biodiversity in the face of climate change requires a predictive ecology of species distributions. Nowhere is this need more acute than in the tropics, which harbor the majority of Earth's species and face rapid and large climate and land-use changes. However, the study of species distributions and their responses to climate change in high diversity tropical regions is potentially crippled by a lack of basic data. We analyzed a database representing more than 800 000 unique geo-referenced natural history collections to determine what fraction of tropical plant species has sufficient numbers of available collections for use in the habitat or niche models commonly used to predict species responses to climate change. We found that more than nine out of 10 species from the three principle tropical realms are so poorly collected ($n<20$ records) that they are essentially invisible to modern modeling and conservation tools. In order to predict the impact of climate change on tropical species, efforts must be made to increase the amount of data available from tropical countries through a combination of collecting new specimens and digitizing existing records.

Keywords: bioclimatic niche envelope, climate change, conservation biogeography, habitat modeling, natural history collections, species migrations, tropical plants

Received 24 November 2009; revised version received 23 March 2010 and accepted 26 March 2010

Introduction

The tropics support the vast majority of the world's terrestrial biodiversity. However, many tropical species are potentially threatened by global climate change (Bradshaw *et al*., 2008; Malhi *et al*., 2008). One of the principle mechanisms through which climate change affects biodiversity is by shifting the distribution of areas with suitable climatic conditions, thereby forcing species to migrate in order to remain within their bioclimatic niches or 'envelopes' (e.g., Walther *et al*., 2002, 2005; Thomas *et al*., 2004; Beckage *et al*., 2008; Lenoir *et al*., 2008). Even if species are capable of migrating, these range shifts often result in reduced habitat areas, increased risks of extinction, and loss of local biodiversity (Thomas *et al*., 2004; Malcolm *et al*., 2006; Colwell *et al*., 2008; Feeley & Silman, 2009a). One widely cited study, estimated that a $>2\,^{\circ}$C increase in temperature will cause over one-third of Amazonian tree species to become 'committed to extinction' based on reductions in the amount of suitable habitat available under future climate (Thomas *et al*., 2004). It is important to note that this study was based on analyses of just nine Amazonian tree species and thus its generality can be questioned. It is clear that we must expand studies examining relationships between species distributions and climate to include a much broader and representative sample of species if we hope to understand, predict, and eventually mitigate the effects of global change on tropical biodiversity.

Correspondence: Miles R. Silman, e-mail: silmanmr@wfu.edu, tel. +1 336 758 5596, fax +1 336 758 6008; Kenneth J. Feeley, e-mail: kjfeeley@gmail.com

One popular technique for investigating species' responses to climate change is species distribution modeling or SDM (Guisan & Zimmermann, 2000; Peterson & Vieglais, 2001; Elith *et al*., 2006; Elith & Leathwick, 2009). In SDMs, species' climatic niches or envelopes are defined by assessing the relationship between known occurrences and climate. The species' distributions can then be mapped in relation to current climatic variables and also projected into the future under various global change scenarios. Differences between the current and projected distributions are then used to predict rates of habitat loss (or gain) and the associated extinction risks. This information is a valuable tool in designing management strategies and/or protected areas aimed at minimizing future species loss (Hannah *et al*., 2007; Pressey *et al*., 2007; Thuiller, 2007; Lee & Jetz, 2008).

Unfortunately, in the diverse tropics the study of species distributions and their responses to climate change is potentially crippled by a lack of basic data (Tobler *et al*., 2007; Collen *et al*., 2008). In more developed countries, abundant data on species occurrences are available from national inventories (e.g., the USAs Forest Inventory and Analysis; http://fia.fs.fed.us/). In contrast, most inventories from the tropics are sparse, forcing ecologists to rely primarily on locations extracted from museum and collection records (Graham *et al*., 2004). Using these 'presence-only' data, a minimum of 20–50 collections per taxa are generally required to produce accurate species distribution models (e.g., Hernandez *et al*., 2006; Loiselle *et al*., 2008; Wisz *et al*., 2008). Due to the paucity of digitized collections, very few tropical species meet this criterion. Here, we illustrate this lack of basic data and discuss the implications for habitat modeling and conservation.

Methods

To quantify the availability of tropical collections data for habitat modeling, we collated all vascular plant records available through the Global Biodiversity Information Facility (GBIF: http://www.gbif.org) for the three principal tropical realms: tropical South America, equatorial Africa (including Madagascar), and Southeast Asia (including New Guinea; Fig. 1). GBIF is the world's largest online depository of collection records (Yesson *et al*., 2007) and allows access to data from various herbaria and natural history collections from throughout the world. For South America, we augmented the dataset with additional collection records acquired through SpeciesLink (http://splink.cria.org.br) which distributes natural history data from Brazilian collections and herbaria (see Supporting Information for a list of herbaria contributing collections data to this study).

For downloading collection records, we used automatic filters to eliminate records with obvious geo-referencing errors (in the case of GBIF, we filtered for records 'without known coordinate issues' and in SpeciesLink, we filtered for records whose coordinates were 'not suspect'). We further screened the data to remove records that had not been identified to species. Finally, we identified and removed duplicate records by screening for unique combinations of species name, latitude, and longitude (rounded to nearest 0.01° = approximately 1 km at the equator; rounding performed to catch duplicate records with geo-referencing data presented at different resolutions).

From the resulting dataset, we tabulated the total number of collection records available per species within each of the three tropical realms and the number of species meeting minimum collection size criteria of $\geq$1, $\geq$5, $\geq$20, $\geq$50, and $\geq$100 unique geo-referenced records.

Results

We analyzed nearly 800 000 unique geo-referenced records from more than 100 000 vascular plant species. A total of 546 203 collections were available through GBIF and SpeciesLink for 70 161 species from tropical South America, 138 556 collections were available from GBIF for 20 965 species from equatorial Africa, and 87 035 collections were available from GBIF for 17 460 species from Southeast Asia.

Across the three tropical realms, 38% of the included species are known from only a single record, the median number of records per species is 2, and just 32.2%, 8.6%, 2.3%, and 0.6% of species are represented by $\geq$5, 20, 50, and 100 samples, respectively. From South America, the percentage of species represented by $\geq$5, 20, 50, and 100 samples is 33.0%, 9.4%, 2.7%, and 0.7%, respectively. From Africa, the percentage of species meeting these criteria are 35.6%, 7.3%, 1.3%, and 0.2% and from Southeast Asia, the percentages are 24.1%, 5.3%, 1.0%, and 0.1% (Fig. 2; for the numbers and percentages of species meeting minimum sample size criteria on a continuous scale, see Fig. S1).

Discussion

Mitigating the loss of species due to climate change requires a predictive ecology of species distributions. Nowhere is this more acute than in the tropics, which harbor the majority of Earth's species and face rapid and large climate and land-use changes (Bradshaw *et al*., 2008; Malhi *et al*., 2008). Unfortunately, we know very little about most tropical species (Collen *et al*., 2008). Analyzing a database of ca. 1 million unique

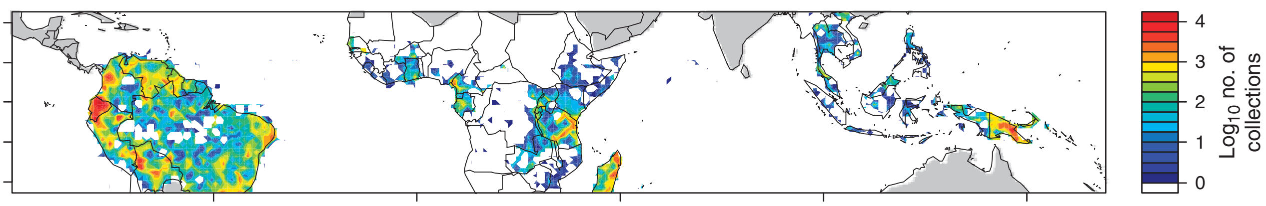

Fig. 1 Filled contour plot (resolution = 1°) indicating the number of unique geo-referenced vascular plant collection records identified to species available from GBIF and SpeciesLink within tropical South America, Equatorial Africa, and Southeast Asia. Gray countries were not included. GBIF, Global Biodiversity Information Facility.

  doi: 10.1111/j.1365-2486.2010.02239.x

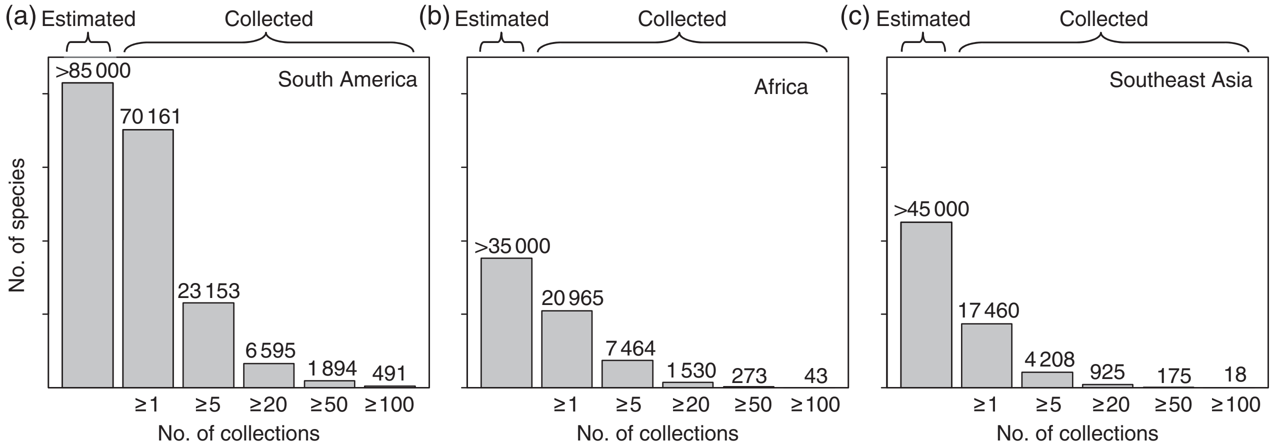

Fig. 2 The number of vascular plant species from tropical South America (a), Equatorial Africa including Madagascar (b), and Southeast Asia including New Guinea (c) estimated for each region (Raven, 1988) and represented by ≥ 1, 5, 20, 50, or 100 unique geo-referenced collections available through GBIF or SpeciesLink. GBIF, Global Biodiversity Information Facility.

geo-referenced records, we find that most plant species collected from the tropics are represented by only one or a few collections and thus cannot be included in even the most advanced species distribution models (Wisz *et al*., 2008). And the situation is even worse than these numbers indicate. Estimated plant diversity in the Neotropics, equatorial Africa, and Southeast Asia exceeds 85 000, 35 000, and 40 000 species, respectively (Raven, 1988), in which case more than 1/3 of the species from these realms are not represented by a single record accessible through the major online data portals (Fig. 2), creating a nearly insuperable barrier to ecologists, distribution/niche modelers, and conservation practitioners seeking to use the data. More specimens clearly exist in physical herbaria but are of only limited value to conservation practitioners until digitized and made publicly available through major online clearing houses.

Can we make generalizations about the effects of global change on tropical plant communities based on just the small proportion of species for which sufficient data exist? If the included species are effectively random samples from across the species pool then the answer is arguably ‘Yes.’ However, the number of records collected is likely associated with various species traits potentially making the most collected species unrepresentative. For example, widespread or highly abundant species are likely to be disproportionately collected over rarer or more localized species (Schwartz *et al*., 2006). From our South American records, no species represented by $\geq$100 collections has a latitudinal range of $<5^{\circ}$, while the minimum latitudinal range of species using our minimum cut-off, $\geq$20 collections, is reduced to $<1^{\circ}$. Rare habitat specialists are at the greatest risk of extinction due to climate change and land-use and thus extinction risks predicted from mostly widespread generalists may be understated. Furthermore, even for relatively well-collected species, the locations of available collections may not accurately represent actual distributions since collection intensity can vary greatly by region (Yesson *et al*., 2007), and even within regions due to collecting biases (Kadmon *et al*., 2004; Tobler *et al*., 2007). Also, there may be geographic biases in which samples are made publicly-available online (Yesson *et al*., 2007). For example, large expanses of all three realms, including entire countries, do not have a single available record (Figs 1 and 3) reflecting both biases in collection efforts and differences in programs to make existing collections publicly available online. These underrepresented areas contrast with other regions, such as Ecuador, where relatively large numbers of collections are available online (Figs 1 and 3) due to the sustained efforts of local herbaria, including the Museo de Historia Natural (QCNE), the Pontificia Universidad Católica del Ecuador (QCA), and the Universidad Central del Ecuador (QAP) and several international institutions including the Missouri Botanical Garden (http://www.mobot.org/), the University of Aarhus (Denmark; http://herb42.bio.au.dk/aau_herb/default.php), and the Smithsonian Tropical Research Institute (http://www.stri.org/).

There are several factors which will increase or decrease the final number of species with adequate numbers of collections to be included in studies using species distribution models. Distribution models rely fundamentally on the association between species occurrences and habitat variables and thus are strongly influenced by geo-referencing errors. In creating our records database, we did not apply stringent filters to identify and eliminate records with possible geo-referencing errors. As stronger filters are applied to help improve the accuracy of models, the number of records available per species will invariably decrease. In a recent study modeling the distributions of Amazonian and Andean plant species, the number of species meeting our minimum collection size criteria ($\geq$20 records) dropped by more than 50% after applying filters to minimize geo-referencing errors (Feeley & Silman, 2009b). Importantly, filters may exclude species nonrandomly. In the study cited above, species from the tropical Andes were disproportionately excluded compared with Amazonian species because the impacts of geo-referencing errors are magnified in areas with steep topography (Feeley & Silman, 2009b).

The number of available records can also be modified based on the resolution of the environmental and/or taxonomic data. Most niche models are based on presence of species per climate/habitat grid cell and thus the number of occurrences depends on grid cell size or resolution (Graham *et al.*, 2004; Guisan *et al.*, 2007). If resolution of current or future variables is coarse, many records will be classified as functional duplicates and eliminated. In contrast, if resolution is improved, more records will be classified as 'unique,' increasing the number of records available per species. Likewise, as taxonomic nomenclature is standardized between collections and heterotypic (taxonomic) synonyms resolved, multiple 'species' will be combined, increasing the number of records per species.

Finally, the number of records will be increased if researchers take the time to visit physical herbaria or access other individual online databases that are not included in major online clearing houses. The importance of using multiple data sources to obtain collection records is illustrated in our study by the thousands of collections for South America that would have been overlooked if we had only obtained records through GBIF and had not included SpeciesLink (Fig. 3). Factors such as these will all affect the final number of species that can be included in ecological or modeling studies; however, it is likely that any improvement will be minor in relation to diversity and the overall lack of data.

What is required to fill the data void and answer questions about tropical biodiversity and climate change? Modeling techniques have improved greatly and are already reaching an asymptote in the lower limit of collections required for accurate projections (Wisz *et al.*, 2008). We argue that the answer lies principally in increasing the amount of raw data available on species occurrences. This involves the unglamorous but crucial work of increasing the number of herbarium collections from the tropics through efforts aimed at establishing and continuing large-scale inventory programs, and also by supporting the efforts of individual researchers and herbaria. The availability and quality of existing data must also be improved; large numbers of collections sit unused or are not databased and are thus unavailable for general ecological research while other records are available but lack geo-referencing and are thus 'unusable' from a modeling perspective. Important efforts are underway to digitize all type specimens, but even these will only increase the number of species represented by one or a few (due to synonymy) electronic records. Finally, ongoing efforts to create a free and open exchange of biodiversity collections and data must be fostered. Increasing basic data on species' distributions makes biodiversity management in the tropics a predictive exercise, one clearly preferable to the alternative of having future ecologists examine extinctions retrospectively.

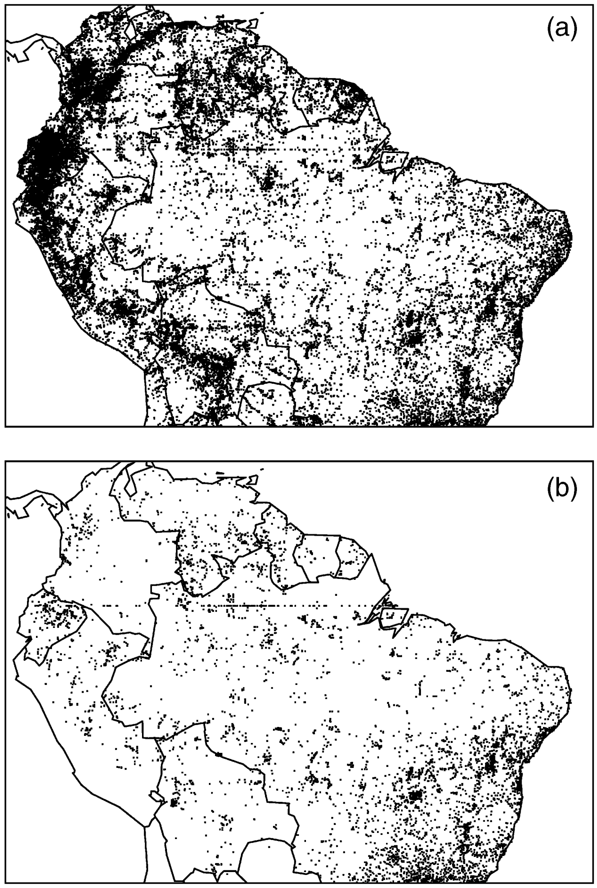

Fig. 3 The collection locations of online herbarium records available for South America from (a) GBIF and (b) SpeciesLink. GBIF, Global Biodiversity Information Facility.

Acknowledgements

We thank the Global Biodiversity Information Facility, Species Link, and all contributing herbaria for making their data publicly available and facilitating studies of ecology and biogeography. We thank the members of the Silman lab and four reviewers for their insightful comments. This research was supported by the Gordon and Betty Moore Foundation Andes to Amazon Program, NSF DEB-0237684 to M. R. S., and from the Amazon Conservation Association. This publication is a contribution of the Andes Biodiversity and Ecosystem Research Group (http://www.andesconservation.org).

References

Beckage B, Osborne B, Gavin DG, Pucko C, Siccama T, Perkins T (2008) A rapid upward shift of a forest ecotone during 40 years of warming in the green mountains of Vermont. *Proceedings of the National Academy of Sciences*, **105**, 4197–4202.

Bradshaw CJA, Sodhi NS, Brook BW (2008) Tropical turmoil: a biodiversity tragedy in progress. *Frontiers in Ecology and the Environment*, **7**, 79–87.

Collen B, Ram M, Zamin T, McRae L (2008) The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science*, **1**, 75–88.

Colwell RK, Brehm G, Cardelus CL, Gilman AC, Longino JT (2008) Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science*, **322**, 258–261.

Elith J, Graham CH, Anderson RP *et al*. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Feeley KJ, Silman MR (2009a) Biotic attrition from tropical forests correcting for truncated temperature niches. *Global Change Biology*, available at http://dx.doi.org/10.1111/j.1365-2486.2009.02085.x.

Feeley KJ, Silman MR (2009b) Modelling Andean and Amazonian plant species responses to climate change: the effects of geo-referencing errors and the importance of data filtering. *Journal of Biogeography*, doi: 10.1111/j.1365-2486.2010.02197.x.

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.

Guisan A, Graham CH, Elith J, Huettmann F (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332–340.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hannah L, Midgley G, Andelman S *et al*. (2007) Protected area needs in a changing climate. *Frontiers in Ecology and the Environment*, **5**, 131–138.

Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401–413.

Lee TM, Jetz W (2008) Future battlegrounds for conservation under global change. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1261–1270.

Lenoir J, Gegout JC, Marquet PA, de Ruffray P, Brisse H (2008) A significant upward shift in plant species optimum elevation during the 20th century. *Science*, **320**, 1768–1771.

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105–116.

Malcolm JR, Liu CR, Neilson RP, Hansen L, Hannah L (2006) Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology*, **20**, 538–548.

Malhi Y, Roberts JT, Betts RA, Killeen TJ, Li W, Nobre CA (2008) Climate change, deforestation, and the fate of the Amazon. *Science*, **319**, 169–172.

Peterson AT, Vieglais DA (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience*, **51**, 363–371.

Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA (2007) Conservation planning in a changing world. *Trends in Ecology and Evolution*, **22**, 583–592.

Raven P (1988) Our diminishing tropical forests. In: *Biodiversity* (ed. Wilson EO), pp. 119–122. National Academies Press, Washington, DC.

Schwartz MW, Iverson LR, Prasad AM, Matthews SN, O'Connor RJ (2006) Predicting extinctions as a result of climate change. *Ecology*, **87**, 1611–1615.

Thomas CD, Cameron A, Green RE *et al*. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Thuiller W (2007) Biodiversity: climate change and the ecologist. *Nature*, **448**, 550–552.

Tobler M, Honorio E, Janovec J, Reynel C (2007) Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodiversity and Conservation*, **16**, 659–677.

Walther G-R, Beiner S, Burga CA (2005) Trends in the upward shift of alpine plants. *Journal of Vegetation Science*, **16**, 541–548.

Walther G-R, Post E, Convey P *et al*. (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

Yesson C, Brewer PW, Sutton T *et al*. (2007) How global is the global biodiversity information facility? *PLoS One*, **2**, e1124.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The number of tropical plant species from South America, Africa, and Southeast Asia represented by a given number of collections and for a given minimum sample size criteria.
Table S1. List of herbaria contributing collections data to this research accessed through GBIF and SpeciesLink.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.