Journal of Biogeography (J. Biogeogr.) (2010) 37, 733–740

ORIGINAL ARTICLE

Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering

Kenneth J. Feeley[1,2]* and Miles R. Silman[1]*

[1]Department of Biology, Wake Forest University, Winston Salem, NC 27106, USA, [2]Department of Biological Sciences, Florida International University, Miami, FL 33199; and Center for Tropical Plant Conservation, Fairchild Tropical Botanic Garden, Coral Gables, FL 33156, USA

ABSTRACT

Aim Species distribution models are a potentially powerful tool for predicting the effects of global change on species distributions and the resulting extinction risks. Distribution models rely on relationships between species occurrences and climate and may thus be highly sensitive to georeferencing errors in collection records. Most errors will not be caught using standard data filters. Here we assess the impacts of georeferencing errors and the importance of improved data filtering for estimates of the elevational distributions, habitat areas and predicted relative extinction risks due to climate change of nearly 1000 Neotropical plant species.

Location The Amazon basin and tropical Andes, South America.

Methods We model the elevational distributions, or 'envelopes', of 932 Amazonian and Andean plant species from 35 families after performing standard data filtering, and again using only data that have passed through an additional layer of data filtering. We test for agreement in the elevations recorded with the collection and the elevation inferred from a digital elevation model (DEM) at the collection coordinates. From each dataset we estimate species range areas and extinction risks due to the changes in habitat area caused by a 4.5 °C increase in temperature.

Results Amazonian and Andean plant species have a median elevational range of 717 m. Using only standard data filters inflates range limits by a median of 433 m (55%). This is equivalent to overestimating the temperature tolerances of species by over 3 °C – only slightly less than the entire regional temperature change predicted over the next 50–100 years. Georeferencing errors tend to cause overestimates in the amount of climatically suitable habitat available to species and underestimates in species extinction risks due to global warming. Georeferencing error artefacts are sometimes so great that accurately predicting whether species habitat areas will decrease or increase under global warming is impossible. The drawback of additional data filtering is large decreases in the number of species modelled, with Andean species being disproportionately eliminated.

Main conclusions Even with rigorous data filters, distribution models will mischaracterize the climatic conditions under which species occur due to errors in the collection data. These errors affect predictions of the effects of climate change on species ranges and biodiversity, and are particularly problematic in mountainous areas. Additional data filtering reduces georeferencing errors but eliminates many species due to a lack of sufficient 'clean' data, thereby limiting our ability to predict the effects of climate change in many ecologically important and sensitive regions such as the Andes Biodiversity Hotspot.

*Correspondence: Kenneth J. Feeley and Miles R. Silman, Department of Biology, Wake Forest University, 1834 Wake Forest Rd, Winston-Salem, NC 27106, USA.
E-mail: kjfeeley@gmail.com, silmanmr@wfu.edu

Keywords
Bioclimatic niches, climate change, collection records, conservation biogeography, data filters, extinction risk, global warming, habitat distribution models, herbarium data, range maps.

doi:10.1111/j.1365-2699.2009.02240.x

Journal of Biogeography

INTRODUCTION

Given the imminent dangers to biodiversity from global climate change, there is increasing interest in mapping the potential distribution of species under various climate change scenarios. One popular method is to use the locations of observations or collection/herbarium records to model species distributions as functions of the current climate (as interpolated from distributed networks of meteorological stations, e.g. WorldClim or CRU; Hijmans *et al.*, 2005) and then to project the distributions into the future under predicted climatic conditions. Results from such studies have played a key role in highlighting the potential impacts of climate change, including the high risk of extinction for many species due to the rapid reduction of climatically suitable habitat area (e.g. Thomas *et al.*, 2004; Malcolm *et al.*, 2006).

This 'climatic envelope' approach can suffer from many known challenges, including small sample sizes and spatially biased sampling, as well as the inability to incorporate factors other than climate (e.g. edaphic conditions, inter-specific interactions, evolution or adaptation to environmental changes, etc.). These limitations are widely recognized, have been relatively well studied (Stockwell & Peterson, 2002; Loiselle *et al.*, 2003, 2008; Pearson & Dawson, 2003, 2004; Hampe, 2004; Araújo & Guisan, 2006; Guisan *et al.*, 2006, 2007; Hernandez *et al.*, 2006; Araújo & Luoto, 2007; Tobler *et al.*, 2007; Feeley & Silman, in press), and are accounted for in some of the newer species distribution models (Engler *et al.*, 2004; Elith *et al.*, 2006). However, another set of limitations that has received surprisingly little attention is inaccuracies in the collections data themselves. In distribution modelling, spatial coordinates are used to extract climatic variables from interpolated raster surfaces such that errors in georeferencing will translate into inaccurate characterizations of species bioclimatic niches. This error will propagate to erroneous distribution maps and hence erroneous estimates of extinction risks due to climate change (Graham *et al.*, 2008).

In order to minimize the impacts of georeferencing errors, most studies perform some degree of data filtering, or cleaning, prior to analysis. Most of these filters eliminate only extreme or obvious errors such as those placing samples in bodies of water or in areas that are otherwise deemed 'impossible'. Many georeferencing errors are more subtle and will not be caught using these coarse filters. For example, probably the most common source of georeferencing error is that collectors may not record the exact geographic coordinates of where specimens were encountered and instead only note the general locality and/or approximate coordinates (Wieczorek *et al.*, 2004), often relying on existing maps and geographic landmarks that would allow others to return to the general area of the collection. This is especially true for specimens collected prior to the popularization of GPS technology, which first became available in 1993 and only became available without the addition of 'selective availability', or intentionally introduced error, in 2000 (Parkinson & Spilker, 1996). When collection records are digitized, coordinates are therefore inferred, commonly from the 'locale' or closest populated area. As a consequence, the specimen will have geographic coordinates that are displaced by up to several kilometres. Errors of this nature will not be caught by most data filters but may still have important effects on characterizations of climatic niches given that even relatively small horizontal offsets can result in large changes in elevation and climate (Wieczorek *et al.*, 2004; Rowe, 2005).

In addition to locality, many collection records also list the elevation at which the specimen was encountered. It may therefore be possible to conduct an additional screen for georeferencing errors by comparing the recorded elevation and the elevation of the listed collection coordinates. Disagreement between these two independent measures of elevation would indicate a possible georeferencing error and that the record should not be included in the analyses (Graham *et al.*, 2004; Parra *et al.*, 2004). Use of this filter should allow for more accurate characterizations of species' elevational distributions which are tightly linked to temperature ranges and tolerances. However, other climatic parameters such as precipitation and seasonality, as well as non-climatic environmental parameters such as soil type, are not as tightly correlated with elevation and as such the listed and actual collection elevations may still show strong agreement even if the georeferencing errors place the sample in an area of vastly different climate. This problem will be especially pronounced in areas of low topographic relief where a very large horizontal displacement is needed to cause a noticeable elevational displacement. For example, in the relatively flat Amazon basin georeferencing errors may reach several hundreds of kilometres with little difference between the true point of collection and the listed erroneous collection coordinates. In more mountainous regions the amount of horizontal displacement in filtered samples will probably be decreased but the non-temperature climate data inferred from collection coordinates may still be unreliable given the rapid rate of change in most climatic parameters. For example, a specimen collected along the wet eastern flanks of the Andes but listed as having been collected from the dry western flanks may still exhibit little or no difference in the inferred and actual elevations, and thus will not be caught by an elevational filter. Despite this limitation, the proposed elevational filter is still a potentially valuable tool, and improving characterizations of species temperature distributions will improve the ability to predict species responses to increasing temperatures.

Here we assess the importance of georeferencing errors and the application of the elevational data filter for mapping the elevational distributions of nearly 1000 well-collected Amazonian and Andean plant species. After applying standard data filters to herbarium collection records we estimate the elevational distribution of each species based on the elevations extracted from a digital elevation model (DEM) using the recorded geographic coordinates. We then apply the elevational filter to eliminate records with large disagreements between recorded and DEM-inferred elevations and recalculate the distribution of each species. In order to illustrate some of the potential hazards of georeferencing errors and the impor-

tance of data filtering, we project both sets of species distributions into the future based on a simple scenario of global warming and compare the predicted changes in habitat area and associated extinction risks calculated using the two datasets.

MATERIALS AND METHODS

Herbarium records for 35 of the most common Neotropical plant families were downloaded through the Global Biodiversity Information Facility (GBIF) data portal (http://www.gbif.org/; specific databases accessed are listed in Table S1 in the Supporting Information). The elevation at which each specimen was collected was inferred by extracting the elevation at the recorded coordinates from a Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) (http://www2.jpl.nasa.gov/srtm/; ground scale of 30 arcsec and vertical accuracy of *c.* ± 16 m). All records were then screened using standard data filters (e.g. Hijmans *et al.*, 1999; Graham *et al.*, 2004; Graham & Hijmans, 2006). First, only records with coordinates and only records 'without known coordinate issues' (as defined by GBIF) were included. We also excluded any records with obvious geographic or elevational errors (e.g. those occurring over bodies of water or at elevations > 6000 m). Furthermore, we only included specimens collected from the 'Tropical and Subtropical Moist Broadleaf Forests' and 'Montane Grasslands and Shrublands' biomes of South America as defined by the World Wildlife Fund (Olson *et al.*, 2001). This final filter limited our focus to just Amazonian and Andean plant species and excluded any records from other geographic areas or biomes (e.g. Central American forests or the Atacama or Sechura deserts). The resultant set of records constitutes what we will refer to as our 'full' dataset (it should be noted, however, that this is not truly a full dataset in that it was screened to the same rigour as applied in most studies).

We next created a 'clean' dataset by applying an additional layer of data filtering and eliminating any records that did not have recorded elevations or for which the recorded elevations differed from the DEM-inferred elevation by more than ± 100 m.

For each species with ≥ 30 specimens in both the full and clean datasets, we then estimated the elevational distribution, or 'envelope'. The lower and upper limits of each species' distribution was set as the 2.5% and 97.5% quantiles, respectively, of the DEM-inferred elevations in each dataset (the use of 95% quantiles rather than absolute maxima and minima is in essence another data filter because it minimizes the influence of outliers which are potentially caused by georeferencing errors). We then measured the extent of land area (a) occurring within the elevational ranges of each species as modelled from the full and clean datasets (areas tabulated only within the study biomes).

To illustrate the effects of georeferencing errors and the importance of date filtering on models of species responses to climate change, we predicted each species' future distribution under a simple global warming scenario by increasing the upper and lower limits of the elevational envelopes by 800 m. This increase in elevational range limits is the predicted response of species to a 4.5 °C increase in temperature given the established adiabatic lapse rate of 5.6 °C per 1000 m (Terborgh & Weske, 1969; Bush *et al.*, 2004) and assuming no limits to migration. According to IPCC climate models (Christensen *et al.*, 2007), 4.5 °C is the approximate mean increase in temperature predicted for tropical South America over the next 100–150 years. We also predicted each species' future distribution assuming zero migration by increasing only the lower elevational limits of the distributions while maintaining the upper elevational limits constant. Finally, we estimated the percentage change in habitat area [$\Delta a = 100 \times (a_{\text{future}} - a_{\text{current}})/a_{\text{current}}$] as a measure of relative extinction risk (i.e. the greater the percentage of habitat area lost, the greater the risk of extinction; MacArthur & Wilson, 1967) and calculated the percent error (e) in the extinction risks as estimated from the full versus clean datasets under both migration scenarios as $e = 100 \times \log_{10}[(100 + \Delta a_{\text{clean}})/(100 + \Delta a_{\text{full}})]$. Negative e values indicate that the use of the full dataset predicts a greater loss of habitat than predicted using the clean dataset. e could not be calculated for any species predicted to lose 100% of habitat based on one but not both datasets (i.e. $e = \pm\infty$). In our analyses, a complete loss of habitat is predicted in the no-migration scenario for species with an elevational range of < 800 m (i.e. less than the projected upward shift).

RESULTS

A total of 151,247 collection records representing 1882 species from the 35 focal plant families passed the initial data filtering and had a sufficient number of collections (i.e. $n \geq 30$) to be included in our full dataset. Many of these records had either no recorded elevation (118) or had a disagreement between the recorded and DEM-inferred elevation exceeding ± 100 m (71,107) and were thus not included in the clean dataset. As a consequence, the final datasets each included 932 species, with a total of 107,349 specimens in the full dataset and 63,572 specimens in the clean dataset (Table S2).

Georeferencing errors had significant effects on our estimates of species distributions. For the majority of species, the use of the full dataset misrepresented species as occurring at different elevations and across a wider range of elevations than estimated using the clean data (Table S2, Fig. 1). Elevational midpoints were moved upslope by georeferencing errors for almost all Amazonian species and downslope for most Andean species (Fig. 1a,b). Based on the clean data, species have a median elevational range of 717 m; georeferencing errors in the full dataset inflated range estimates by a median of 433 m or 55% (Fig. 2). The overprediction of species' habitat ranges is due primarily to georeferencing errors and is not simply an artefact of greater sample sizes in the full dataset (see Fig. S1).

As a consequence of differences in the location and extent of elevational distributions, georeferencing errors significantly

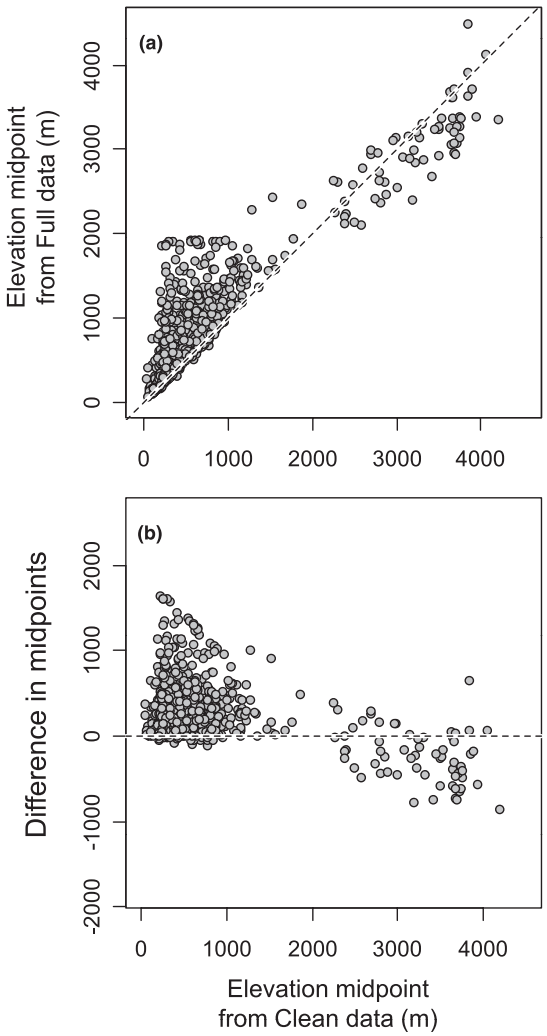

Figure 1 Comparison of species range midpoints. (a) Estimates of midpoints of species elevational ranges as modelled using the full versus the clean datasets (n = 932 for each dataset). (b) Difference in elevational midpoints (full – clean) versus the midpoint of elevational range as estimated from the clean dataset.

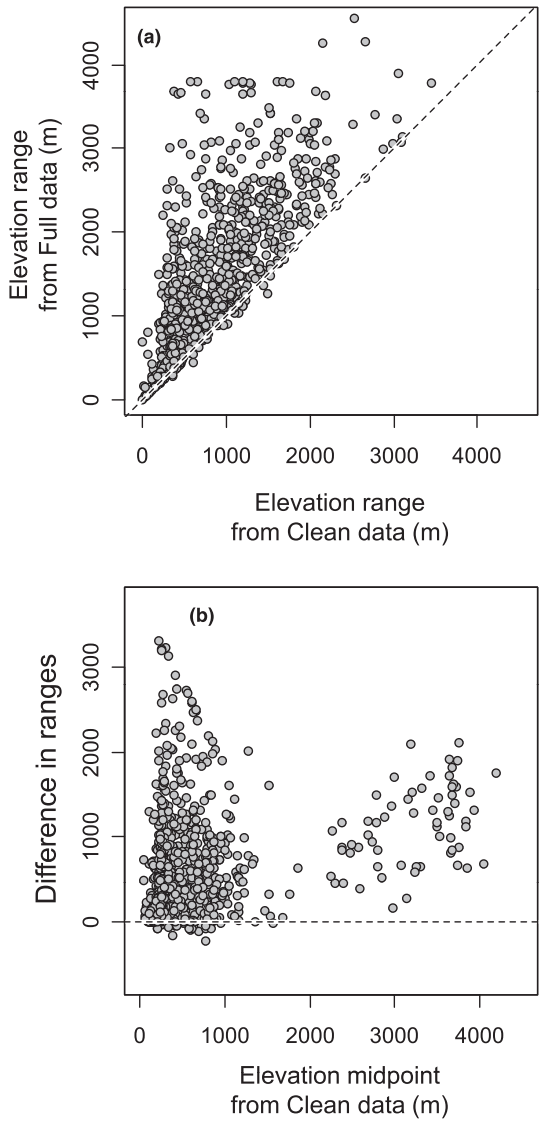

Figure 2 Comparison of species elevational ranges. (a) Estimates of species elevational ranges as modelled using the full versus the clean datasets (n = 932 for each dataset). (b) Difference in elevational ranges (full – clean) versus the midpoint of elevational range as estimated from the clean dataset.

changed estimates of species habitat areas (Fig. 3). The full data overestimated habitat areas in 83.1% of the species (775) by an average of +51.7% and underestimated habitat areas in 14.7% of the species (137) by an average of just −3.7% (Fig. 3b).

As a consequence of overpredicting habitat areas, the use of the full dataset also tended to underpredict the losses of habitat area due to future climate change (Fig. 4). Using the clean dataset, 96% of species are predicted to lose habitat area due to a 4.5 °C increase in temperature (i.e. $\Delta a_{\text{clean, mig}} < 0$) even when assuming perfect migration (i.e. +800 m in species lower and upper elevational limits; some Andean species are predicted to gain habitat area as they migrate up from the slopes onto the larger Altiplano). The median predicted change in habitat area is −91.8%. If the elevational filter is not applied, habitat change, and thus the risk of extinction due to climate change, is underpredicted (i.e. $e < 0$) in 87% of the species losing habitat. The average underprediction is 9%, and in some species the degree of underprediction is as high as 50% (Fig. 4a,c).

If species do not migrate (i.e. +800 m in just the lower limits of species elevational envelopes), all species are predicted to lose habitat area and 55% of species are predicted to lose their entire habitat area (and thus face certain extinction). As above, if the elevational data filter is not applied, habitat loss is underpredicted for the majority of species (70%). Using the full dataset, only 27% of species are predicted to experience complete habitat loss. On average, habitat loss based on the full dataset is 16% less than is predicted using the cleaned data (Fig. 4b,d).

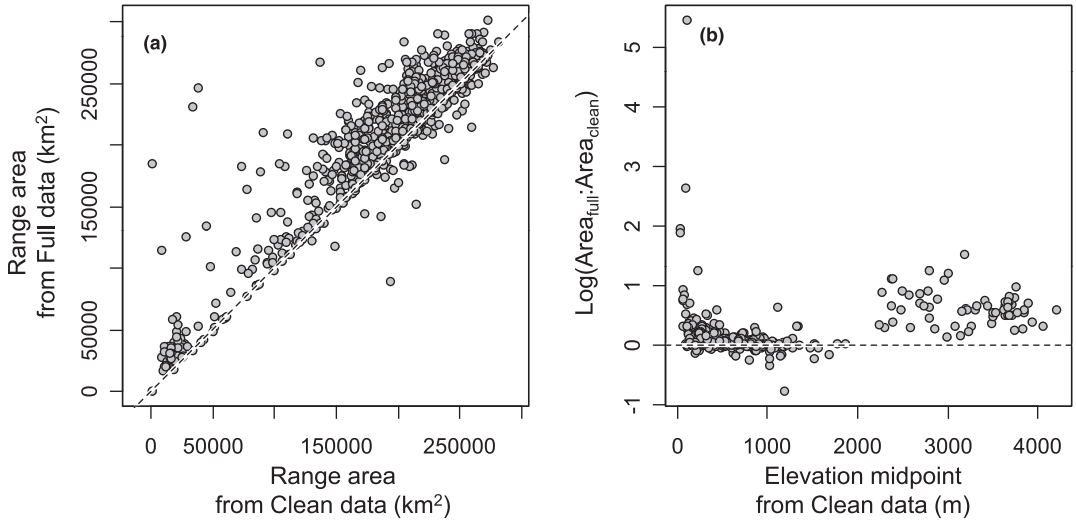

Figure 3 Change in modelled range areas estimated from full and clean datasets. (a) Estimates of species range areas (km^2) as modelled using the full versus the clean dataset (n = 932 for each dataset). (b) Log ratio of range area estimates versus the midpoint of elevational range as estimated from the clean dataset. Log ratios > 0 indicate that the use of unfiltered data inflates range area estimates.

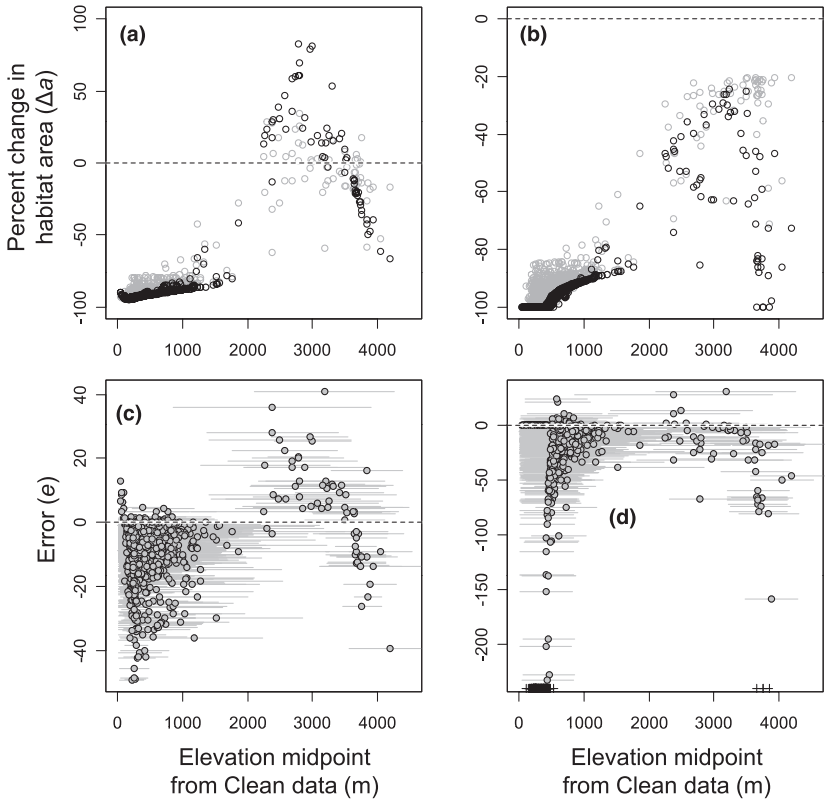

Figure 4 Changes in habitat areas and errors under scenarios of warming. The top panels (a,b) indicate the percentage change in habitat area (Δa) predicted based on the full (grey points) and clean (black points) datasets (n = 932 for each dataset) versus the midpoint of the species current elevational range (as based on the clean dataset) for Amazonian and Andean plant species with a 4.5 °C increase in temperatures under scenarios of perfect (a) and no (b) migration. Negative values indicate a loss of habitat area and positive values indicate a predicted increase in habitat area. The bottom panels (c,d) show the corresponding error (e, see text) in the rate of habitat change as predicted using the clean versus full datasets under scenarios of perfect (c) and no (d) migration. Grey horizontal lines show the extent of each species' elevational range and points indicate the midpoints (midpoints and ranges based on the clean data). Positive values indicate that the use of the full dataset overpredicts habitat loss when compared with predictions based on cleaned data. The points along the bottom of panel (d) indicate species for which a complete loss of habitat ($\Delta a = -100$) was predicted based on the clean but not full dataset ($e = -\infty$); no species was predicted to lose all of its habitat area with perfect migration.

DISCUSSION

We mapped the distributions of more than 900 Neotropical plant species using data that had been filtered to the rigour employed in most comparable studies and again using an additional data filter testing for agreement in two independent measures of collection elevation (recorded and DEM-inferred). We found that the species ranges as mapped from these two datasets differed markedly. These differences in range in turn influenced our predictions of how species will respond to climate change, highlighting the importance of rigorous data filtering and increased efforts to minimize georeferencing errors.

In our study, georeferencing errors that were not caught by the standard data filters inflated species ranges by a median of > 50% (Fig. 1, Table S2). This result becomes even more striking when translated into terms of temperature. The measured lapse rate in the tropical Andes is -5.6 °C km^{-1} and is stable in time and space (Bush & Silman, 2004). Based on the clean data, the median temperature tolerance for Amazonian and Andean plant species is 4.0 °C. Based on the full dataset, the median temperature tolerance for species is 7.1 °C. To put this in perspective, the 3.1 °C increase in temperature tolerances due to georeferencing errors in our full dataset is only slightly less than the predicted change in Amazonian temperatures over the next 100 years, and is greater than half the difference between full glacial and interglacial temperatures in the tropics (Ballantyne *et al.*, 2005).

The inflation of species temperature tolerances and range areas translates into large and important errors, generally underestimates, in predictions of extinction risk due to global warming (Fig. 4). For example, using the cleaned data we estimated that with a 4.5 °C increase in mean annual temperatures over 75% of species (700 of 932) are predicted to lose over 90% of their habitat areas even if they are capable of perfect migration (using a species–area curve with $z = 0.25$, a 90% reduction in habitat area equates to an approximate 50% risk of extinction). If, however, only standard data filters are applied, the number of species predicted to experience this degree of habitat loss is reduced to < 50% (452). Furthermore, in 33% (21 of 64) of the strictly Andean species included in our analyses (defined here as having a lower elevational limit above 500 m), the effects of georeferencing errors were so great that our predictions of changes in habitat area due to global warming (Δa_{mig}) switched from a net loss to a net gain of habitat area (five species), or vice versa (16 species; Fig. 4a). In other words, for many species georeferencing errors can prevent us from accurately predicting even if species will 'benefit' or be harmed by a warming climate (Fig. 4c). The fact that georeferencing errors can have such dramatic effects on our estimates of species distributions and responses to climate change may call into question the results of some previous studies, and clearly highlights the need for improved data quality and/or screening. Because of the effects of even small errors in geolocation, plant collection records from mountainous areas not based on exact GPS points, or extensive and detailed field notes, give little or misleading information for niche models and plant extinction risks under climate change. Indeed, a more informative approach for conservation may be to look at a series of hypothetical species distributions ranging from narrow to broad across the gradients of interest and then infer actual species extinctions from the modelled behaviour under scenarios of climate change.

While the use of data filters is clearly important if we hope to accurately map species ranges and predict their responses to future temperature changes, there is a significant drawback. Namely, the number of records that can be included in the model decreases with each successive criterion, or filter,

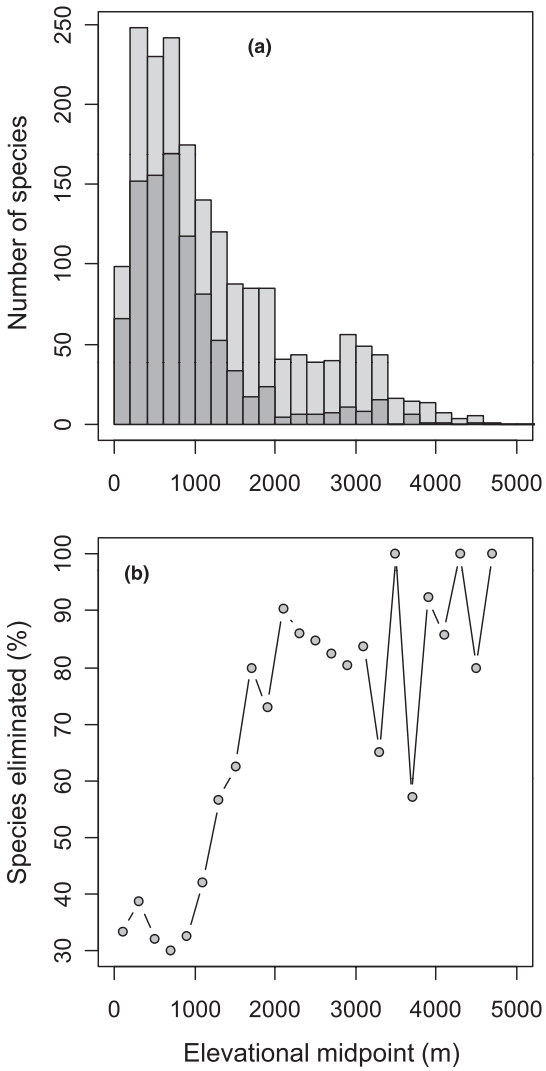

Figure 5 Changes in data available for species under differing filtering scenarios. (a) Number of species that meet the collection criteria of ≥ 30 records after applying the standard filters (light grey) and the additional elevational data filter (dark grey) versus the midpoint of the elevational range. (b) The percentage of species excluded by the elevational data filter versus the midpoint of the elevational range.

that must be satisfied. This poses a serious problem, especially in areas such as the tropics where the majority of collected species are represented by just a handful of georeferenced records even before the application of any filters. For example, from the 35 families included in this study, we originally downloaded collections data for 16,744 species. The records for these species had already passed an initial level of data filtering in that they were all georeferenced and had 'no known coordinate issues'. After applying additional data filters to eliminate records occurring in 'impossible' locales or outside our study biomes, only 1882 species (11.2%) satisfied our collection criteria of ≥ 30 records. The application of our additional elevation filter further reduced the number of available species to 932 (5.6%). Making matters worse, the elimination of species was highly non-random. The final data filter eliminated $< 40\%$ of lowland Amazonian species but $> 75\%$ of Andean species left after applying the standard filters (Fig. 5). As a consequence, Andean species are severely underrepresented in this study, as is probably also the case in virtually all previous studies (e.g. Thomas *et al.*, 2004). The Andes Biodiversity Hotspot is one of the most diverse regions on Earth and supports many endemic species of high conservation priority (Myers, 1988; Myers *et al.*, 2000), yet the lack of usable data prevents us from being able to predict the effects of climate change and will hinder conservation efforts. Immediate efforts are needed to increase the quality and number of data available from this and other underrepresented systems.

ACKNOWLEDGEMENTS

We thank the Global Biodiversity Information Facility and all contributing herbaria for making their data publicly available and for facilitating studies of biogeography. We received financial support from the Gordon and Betty Moore Foundation Andes to Amazon program, NSF DEB-0237684 to M.R.S., and the Amazon Conservation Association. We thank W. Jetz, J. Rapp, J. Wyatt and the Andes Biodiversity and Ecosystem Research Group (ABERG) for helpful comments on the manuscript.

REFERENCES

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33,** 1677–1688.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16,** 743–753.

Ballantyne, A.P., Lavine, M., Crowley, T.J., Liu, J. & Baker, P.B. (2005) Meta-analysis of tropical surface temperatures during the Last Glacial Maximum. *Geophysical Research Letters*, **32,** L05712.

Bush, M.B. & Silman, M.R. (2004) Observations on Late Pleistocene cooling and precipitation in the Neotropics. *Journal of Quaternary Science*, **18,** 677–684.

Bush, M.B., Silman, M.R. & Urrego, D.H. (2004) 48,000 years of climate and forest change in a biodiversity hot spot. *Science*, **303,** 827–829.

Christensen, J.H., Hewitson, B., Busuioc, A. *et al.* (2007) Regional climate projections. *Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change* (ed. by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller), pp. 847–940. Cambridge University Press, Cambridge.

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

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41,** 263–274.

Feeley, K.J. & Silman, M.R. (in press) Biotic attrition from tropical forests correcting for truncated temperature niches. *Global Change Biology*.

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15,** 578–587.

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

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & the NCEAS Predicting Species Distributions Working Group (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45,** 239–247.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43,** 386–392.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F. & NCEAS Species Distribution Modelling Group (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13,** 332–340.

Hampe, A. (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13,** 469–471.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography*, **29,** 773–785.

Hijmans, R.J., Schreuder, M., De la Cruz, J. & Guarino, L. (1999) Using GIS to check co-ordinates of genebank accessions. *Genetic Resources and Crop Evaluation*, **46,** 291–296.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25,** 1965–1978.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. & Williams, P.H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, **17,** 1591–1600.

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35,** 105–116.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ.

Malcolm, J.R., Liu, C.R., Neilson, R.P., Hansen, L. & Hannah, L. (2006) Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology*, **20,** 538–548.

Myers, N. (1988) Threatened biotas: 'hot spots' in tropical forests. *The Environmentalist*, **8,** 187–208.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403,** 843–858.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, K.R. (2001) Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience*, **51,** 933–938.

Parkinson, B.W. & Spilker, J.J., Jr (1996) *The global positioning system: theory and applications*. American Institute of Aeronautics and Astronautics, Washington, DC.

Parra, J.L., Graham, C.C. & Freile, J.F. (2004) Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, **27,** 350–360.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12,** 361–371.

Pearson, R.G. & Dawson, T.P. (2004) Bioclimate envelope models: what they detect and what they hide – response to Hampe (2004). *Global Ecology and Biogeography*, **13,** 471–473.

Rowe, R.J. (2005) Elevational gradient analyses and the use of historical museum specimens: a cautionary tale. *Journal of Biogeography*, **32,** 1883–1897.

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148,** 1–13.

Terborgh, J. & Weske, J.S. (1969) Colonization of secondary habitats by Peruvian birds. *Ecology*, **50,** 765–782.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427,** 145–148.

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.

Wieczorek, J., Guo, Q.G. & Hijmans, R.J. (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, **18,** 745–767.

SUPPORTING INFORMATION

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

Figure S1 Estimates of species elevational ranges as modelled using the full versus the clean datasets controlling for differences in sample sizes.

Table S1 List of all herbaria and collections contributing plant collection records used in this study as accessed through the Global Biodiversity Information Facility (12/2007–1/2008).

Table S2 The number of collections and current elevational ranges for each of the 932 Amazonian and Andean plant species as derived from the full and clean datasets.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Kenneth J. Feeley is an Assistant Professor of Plant Conservation Biology at Florida International University and the Fairchild Tropical Botanic Garden. This study was conducted as part of his post-doctoral research fellowship with the Andes Biodiversity and Ecosystem Research Group based in the Wake Forest University, Department of Biology.

Miles R. Silman is a community ecologist interested in understanding the factors that influence species distributions and the forces that promote and maintain diversity in tropical ecosystems. He is an Associate Professor in the Department of Biology at Wake Forest University and a principal investigator in the Andes Biodiversity and Ecosystem Research Group.

Editor: Robert Whittaker