

RESEARCH
PAPER



A comparison of methods for mapping species ranges and species richness

Catherine H. Graham^{1*} and Robert J. Hijmans²

¹Department of Ecology and Evolution, 650 Life Sciences Building, State University of New York at Stony Brook, NY 11794-5245, USA, and

²Museum of Vertebrate Zoology, University of California, 3101 Valley Life Sciences Building, Berkeley, CA 94720, USA

ABSTRACT

Aim Maps of species richness are the basis for applied research and conservation planning as well as for theoretical research investigating patterns of richness and the processes shaping these patterns. The method used to create a richness map could influence the results of such studies, but differences between these methods have been insufficiently evaluated. We investigate how different methods of mapping species ranges can influence patterns of richness, at three spatial resolutions.

Location California, USA.

Methods We created richness maps by overlaying individual species range maps for terrestrial amphibians and reptiles. The methods we used to create ranges included: point-to-grid maps, obtained by overlaying point observations of species occurrences with a grid and determining presence or absence for each cell; expert-drawn maps; and maps obtained through species distribution modelling. We also used a hybrid method that incorporated data from all three methods. We assessed the correlation and similarity of the spatial patterns of richness maps created with each of these four methods at three different resolutions.

Results Richness maps created with different methods were more correlated at lower spatial resolutions than at higher resolutions. At all resolutions, point-to-grid richness maps estimated the lowest species richness and those derived from species distribution models the highest. Expert-drawn maps and hybrid maps showed intermediate levels of richness but had different spatial patterns of species richness from those derived with the other methods.

Main conclusions Even in relatively well-studied areas such as California, different data sources can lead to rather dissimilar maps of species richness. Evaluating the strengths and weaknesses of different methods for creating a richness map can provide guidance for selecting the approach that is most appropriate for a given application and region.

Keywords

Amphibians, California, distribution model, herpetological fauna, Maxent, range maps, reptiles, species richness.

*Correspondence: Catherine Graham, Department of Ecology and Evolution, 650 Life Sciences Building, State University of New York at Stony Brook, NY 11794-5245, USA. E-mail: cgraham@life.bio.sunysb.edu

INTRODUCTION

Research in biogeography, macroecology and conservation planning is often based on the analysis of grid maps of species richness. There is considerable variation in how such maps are generated. A common approach consists of overlaying range maps of individual species. Such range maps can be expert-drawn (Rahbek & Graves, 2001; Burgess *et al.*, 2002; Brooks *et al.*, 2004;

Orme *et al.*, 2005) or predicted with species distribution models (Ferrier, 1997; Loiselle *et al.*, 2003). If sufficient point observations of species occurrences are available, species richness can also be directly generated from these points by counting the number of distinct species for cells on a grid (Hijmans & Spooner, 2001).

Given the strong interest in the analysis of species richness maps, there has been remarkably little attention given to the question of how much these maps are affected by the data and methods used

to create them, and how this variation influences inferences drawn from these maps. Guralnick & Van Cleve (2005) compared species richness patterns generated from museum and survey data and found that richness patterns varied substantially with the data source. Hurlbert & White (2005) showed that differences between richness maps created with different expert-drawn range maps and survey information influenced the correlation of environmental factors with richness patterns. In addition, the spatial resolution of richness maps affects the relative importance of different environmental correlates (Rahbek & Graves, 2001; van Rensburg *et al.*, 2002). Recently, distribution modelling has been put forth as an alternative for creating species range maps to produce richness maps (Loiselle *et al.*, 2003; Guisan & Thuiller, 2005), but such maps have not been compared to richness maps produced by other means.

Here we map species richness of the herpetological fauna of California using four types of range maps at three spatial resolutions. The types of range maps used are 'expert-drawn' and 'point-to-grid' range maps, range maps obtained with a species distribution model, and a 'hybrid' method that integrates data from the other three methods. In all these methods, the first step is the creation of grid-based range maps for individual species. These individual range maps are subsequently combined to obtain richness maps. In 'point-to-grid' mapping, range maps are obtained by overlaying point observations of species occurrences with a grid and determining presence or absence for each cell. With this method, range sizes, and ultimately species richness, are likely to be underestimated ('error of omission') in many areas (grid cells), due to insufficient and spatially biased sampling (Hijmans *et al.*, 2000). Estimated species richness will generally be highly influenced by sampling effort (i.e. related to the number of records for an area). It follows that the accuracy of the estimate of species richness should increase as the resolution of the grid decreases (i.e. larger cells). For this reason, point-to-grid mapping is often done using a grid with a relatively low resolution, which can be a serious limitation to its usefulness.

Another approach is overlaying expert-drawn range maps. Such range maps are commonly published in field guides, and often referred to as 'range maps', but we prefer the more specific term 'expert-drawn range maps' because expert-drawn range maps represent only one type of range map. The information used to generate these maps varies; some are based only on occurrence data (e.g. a polygon around known occurrences), whereas others incorporate varying degrees of knowledge, such as habitat requirements or elevational limits of the species, in essence using an informal distribution modelling approach. Compared to point-to-grid richness maps, richness maps derived from expert-drawn range maps are more likely to overestimate local richness ('error of commission'), because they are generally drawn to include all areas where a species is known to occur without excluding areas in between where the species may not exist. They tend to map the 'extent of occurrence' of species that includes the, perhaps much smaller, 'area of occupancy' (Loiselle *et al.*, 2003; Habib *et al.*, 2004; Hurlbert & White, 2005).

Range maps can also be created with species distribution models. In this approach known occurrences of a species and spatially continuous environmental data layers are used to infer

the ecological requirements of a species. The geographical distribution of a species is then predicted by mapping the area where these environmental requirements are met (Austin, 1985; Nix, 1986; Graham *et al.*, 2004a; Guisan & Thuiller, 2005). Given that modelling extrapolates species ranges from point occurrences, a successfully modelled range map should have lower omission errors than maps made from points alone. They might be less prone to error of commission because they can provide range maps of a high spatial resolution that exclude unsuitable habitat within the overall range. However, distribution records are often spatially biased (Hijmans *et al.*, 2000) and if the full environmental distribution is not captured, predicted ranges are prone to error of omission (Kadmon *et al.*, 2004). Moreover, because distribution models rely solely on environmental characteristics, and are ignorant of the effect of factors such as dispersal limitations or species interactions on ranges, models often predict the presence of species in areas where they are known to be absent. Because of this limitation, hybrid methods have been developed that capitalize on the high spatial resolution obtained from distribution modelling but limit the possible error of commission. This can be accomplished by restricting the predicted ranges using expert-drawn range maps (Loiselle *et al.*, 2003), biogeographical regions (Peterson *et al.*, 2002) or distance to the known occurrences (Rissler *et al.*, in press).

In this paper we compare richness maps created using each of the four different methods presented above: (1) point-to-grid mapping of species distribution records; (2) overlay of expert-drawn range maps; (3) overlay of modelled species ranges; and (4) a hybrid approach, which combines aspects of the three others. We also investigate how spatial resolution (the grid cell size) influences patterns of species richness generated by these methods. We conducted this study with the herpetological fauna of California because the availability of detailed range maps and the relatively high number of occurrence records allowed us to make a meaningful comparison of the four different methods.

METHODS

Point occurrence data

Point occurrence data representing specimen collection sites were obtained from the Museum of Vertebrate Zoology (MVZ) (<http://mvz.berkeley.edu>). We included records for all species that occurred in California or nearby to avoid edge effects that would have resulted if only Californian data had been used in modelling (Fig. 1a). The specimens were collected between 1861 and 2003 with 25% collected before 1940, 50% before 1959 and 75% before 1979. All data were extensively examined for errors by checking for geographical outliers with visual and overlay methods (Hijmans *et al.*, 1999) and by checking the environmental outliers in DIVA-GIS (Hijmans *et al.*, 2005).

Expert-drawn range maps

We used expert-drawn range maps from the California Wildlife Habitat Relationships System (California Department of Fish

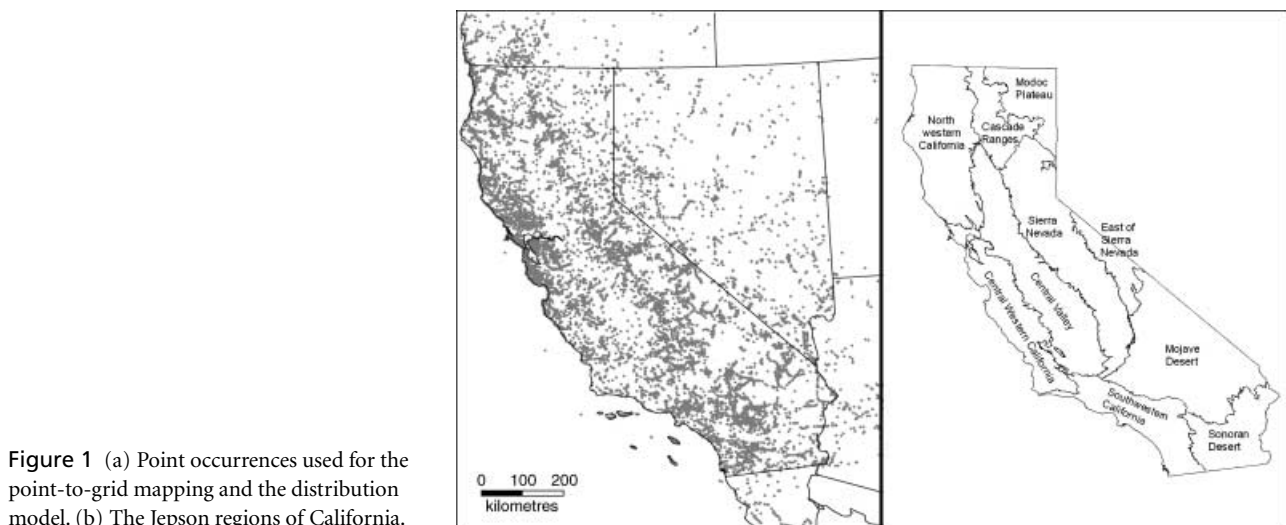


Figure 1 (a) Point occurrences used for the point-to-grid mapping and the distribution model. (b) The Jepson regions of California.

and Game & California Interagency Wildlife Task Group, 2002). These maps are based on species–habitat relationships and represent the maximum geographical extent over which the species can be reliably expected to occur given suitable habitat conditions (Airola, 1988; Davis *et al.*, 1998; California Department of Fish and Game & California Interagency Wildlife Task Group, 2000). Expert-drawn range maps focused on the currently known distribution of a species. Historical occurrences of known extinct populations were only considered for inclusion in the current distribution if the expert was moderately confident that the species could reoccupy the area within the foreseeable future without human intervention. We used Stebbins (2003) to make expert-drawn range maps for 11 species for which there were no maps available, or that needed adjustment to accommodate taxonomic revisions, notably in the genus *Batrachoseps* (Jockusch & Wake, 2002). We also edited the California Wildlife Habitat Relationship map for *Ensatina eschscholtzii* because there was a clear error in it, completely misplacing its extensive range in the Sierra Nevada.

Species distribution models

Species distribution models combine information from point occurrence data and environmental variables to predict the geographical distribution of a species. Various statistical approaches have been employed in this type of modelling. We used one modelling method, Maxent (Phillips *et al.*, 2006), which performed very well in a recently completed model comparison study (Elith *et al.*, 2006). Maxent is a machine-learning technique based on the principle of maximum entropy. It estimates the probability distribution of maximum entropy (i.e. that is most spread out, or closest to uniform) of each environmental variable across the study area. This distribution is calculated with the constraint that the expected value of each environmental variable under this estimated distribution matches the empirical average generated from environmental values associated with species occurrence data (Phillips *et al.*, 2006). The output of the Maxent model for a given species is a continuous surface of

values ranging between 0 and 100, where high values indicate a higher probability of finding a given species.

We used the point occurrence data described above and the following seven climatic variables to run models: (1) annual mean temperature; (2) mean diurnal temperature range; (3) maximum temperature of the warmest month; (4) minimum temperature of the coldest month; (5) annual precipitation; (6) precipitation seasonality (coefficient of variation across months); and (7) precipitation of the driest quarter. These variables were used because they represent variables that describe the general trends (i.e. means), variation in temperature and precipitation as well as those variables that potentially represent physiological limits for species. All variables were obtained from the WorldClim data base (Hijmans *et al.*, 2005). From this global data base we extracted data for western North America from approximately the eastern base of the Rocky Mountains between the Canadian/US border and the southern tip of Baja California. Using ArcInfo (ESRI, Redlands, CA, USA) we transformed the data to the equal-area Teale–Albers projection that is commonly used in California, using a spatial resolution of 1 km.

We divided the occurrences for each species into training (75%) and test (25%) points. Background ('pseudo-absence') points were taken for each species by random sampling from the area within a bounding box around the occurrence points. The bounding box was 1.5 times the minimum rectangular bounding box that could be drawn around the points. The test and background points were used to calculate the Cohen's κ (kappa) statistic and receiver operating characteristic (ROC) curves. Cohen's κ (Cohen, 1960) is a measure of model performance. It requires a threshold for true (species is present)/false (species is absent). We calculated κ using 100 different thresholds to determine the threshold that produced the highest κ (max- κ). We then used this threshold to transform the continuous model output to a presence/absence prediction for each species. The area under the ROC curve is commonly used to evaluate models in species distributional modelling. A ROC curve is created by plotting the true-positive fraction against the false-positive

fraction for all test points across all possible probability thresholds (Fielding & Bell, 1997). The curve goes from (0,0) to (1,1) and a model that produces a curve with a high true-positive fraction at low values of the false-positive fraction is considered good. This is commonly quantified by calculating the area under the curve (AUC). The AUC ranges from 0 to 1 where a value of 0.5 indicates that a model is no better than random and a value of 1 indicates that the model can discriminate perfectly between presence and absence records.

Hybrid approach

We developed a hybrid approach that incorporated information from the occurrence points, modelled ranges and expert-drawn range maps. First, the expert-drawn range maps were modified to include buffered occurrence points for each species; this was necessary because many occurrence points were outside the range maps. We used a buffer of 25 km. Buffer size is somewhat arbitrary, and its choice should depend on the data at hand. Based on preliminary analyses we determined that a smaller buffer would lead us to exclude many areas where a species is likely to occur, and a larger buffer size would lead us to include many areas that appear to be climatically suitable but where the species does not occur, for example because of competition or limited dispersal ability. We then cut the modelled ranges by these point-modified expert-drawn range maps in an attempt to remove error of commission (Fig. 2).

Richness maps and spatial resolution

For all four methods, we made range maps for individual species at a spatial resolution of 1 km. The expert-drawn range maps were converted from polygons to grids. To create richness maps based on expert-drawn range maps, distribution models and hybrid range maps, we combined, through summation, the presence/absence maps for individual species. A spatial resolution of 1 km is clearly too small for creating richness maps based on point occurrences alone (i.e. point-to-grid method), but we include it for comparison purposes as an example of worst

practice. Each 1-km spatial resolution range map was aggregated to 25- and 50-km resolution, such that if a species was present in any of the component cells at the finer resolution it was considered present in the aggregated (larger-resolution) cell. We then summed these lower-resolution range maps of individual species richness to create richness maps for the 25- and 50-km resolution. All these analyses were done in DIVA-GIS (Hijmans *et al.*, 2004).

RESULTS

Point occurrence data

We obtained a total of 108,854 observations (specimens) for the 159 species of terrestrial reptiles and amphibians occurring in or near California. There were 23,708 unique localities and 32,098 unique occurrences, i.e. combinations of species and localities (of which 24,397 were actually within California). The median number of observations per species was 94, the 10th percentile was 17 and the 90th percentile 505 observations per species. Within California there were point occurrences outside the expert-drawn range maps for 112 of 128 species that occur in California. Of these species, the median proportion of occurrences outside the expert-drawn range maps was 10.8%, and the maximum was 66.7%. Frequently, these occurrences were within a relatively short distance (< 25 km) of the range maps. There were only 36 species with an occurrence further than 25 km from their corresponding expert-drawn range map, and for these species the median proportion of occurrences beyond 25 km was 1.0%. The median year of collection of the records from outside the range maps was similar to those collected inside the range maps, whether calculated across all records (outside, 1971; inside, 1951) or after first calculating the medians for each species individually (outside, 1965; inside, 1961). Based on the Jepson regions of California (Hickman, 1993), there was a median of 49 observations per 1000 km² with Central Western California having the highest density of occurrence records and the Modoc Plateau the lowest (Table 1, Fig. 1b). The Jepson regions are based on a combination of the Jepson flora, describing

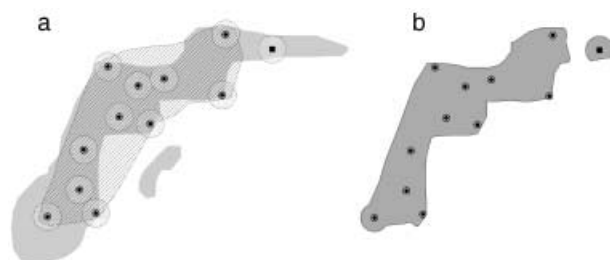


Figure 2 Schematic depiction of the hybrid richness mapping approach. (a) Occurrence data, where circles represent occurrences within and the square outside the expert-drawn range map. All occurrences were buffered. The hatched area represents the expert-drawn range map and the grey represents the modelled range map. (b) Hybrid range map which is the intersection of the expert-drawn range map, buffered points and modelled distribution.

Table 1 Availability of species occurrence data by Jepson region in California

Jepson region	Number of observations	Area (1000 km ²)	Density per 1000 km ²
Cascade Ranges	953	21	46.1
Central Western California	4973	37	133.3
East of Sierra Nevada	693	13	52.8
Great Central Valley	1902	59	32.4
Modoc Plateau	220	23	9.6
Mojave Desert	3243	74	43.8
North-western California	4661	56	83.3
Sierra Nevada	4113	63	65.1
Sonoran Desert	1339	29	45.9
South-western California	2300	34	68.0

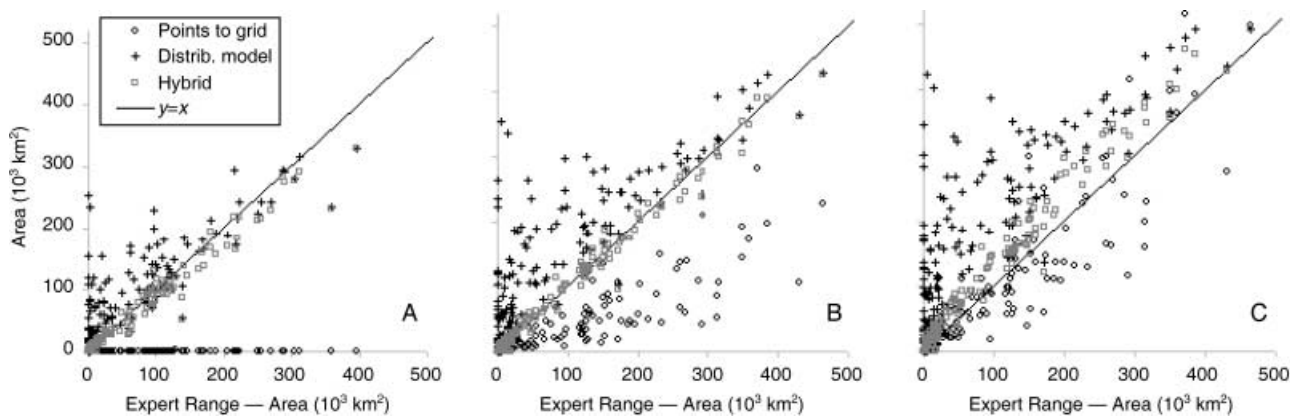


Figure 3 Comparison of range sizes of amphibian and reptile species from California for four mapping approaches and three spatial resolutions. The four different types of range maps are expert-drawn (horizontal axis), point-to-grid (circles), distribution model (crosses) or hybrid approach (squares, which used data from all other range map types), at three spatial resolutions (A, 1 km; B, 25 km; C, 50 km). The line represents $y = x$.

where plant species were known to occur, and by natural features, including broadly defined vegetation types, and geological, topographical and climatic variation (Hickman, 1993).

Species ranges

The mean value of max- κ across all models was 0.64 (SD = 0.15) and the AUC was 0.82 (SD = 0.13). Acceptable values of max- κ vary in the literature, but generally 0.6 and above are considered 'good' (Czaplewski, 1994; Fielding & Bell, 1997). The AUC result can be interpreted as the probability that the model has correctly classified presence and background points for a given species, and values above 0.75 generally indicate adequate model performance for most applications (Pearce & Ferrier, 2000).

Range sizes varied strongly by method used and spatial resolution (Table 2). The 10% quartile and median show that point-to-grid maps give the smallest ranges, the hybrid and expert-drawn range maps have similar, larger ranges, and the distribution model ranges are the largest. The 90% quartile shows a similar pattern, but modelled ranges are much closer to the hybrid and expert values. While this pattern applies at all resolutions,

differences between the approaches generally decreased with larger cell size (Table 2, Fig. 3). Only at the lowest resolution (50 km) do the range sizes of the point-to-grid range maps approach those of the expert-drawn range maps; however, the distribution model-based range maps deviate more from the expert-drawn range maps at this resolution (Fig. 3).

Species richness

Correlations among richness maps created with different methods increased with decreasing resolution (Table 3), and the differences between methods at 1-km resolution all but disappeared at 50-km resolution. Nonetheless, the spatial differences among the methods persisted, though more in terms of actual number of species predicted than in the location of regions of relatively high richness (Fig. 4). For example, all methods showed relatively high richness in the Sierra Nevada and Central Western California regions and the contact zone between the Central Western California, Mojave Desert and Sonoran Desert regions in southern California, and low richness in the north-east of the state (Figs 1b and 4).

Table 2 Comparison of range sizes (km^2) and median richness values for grid cells for species richness maps derived from four different types of range map (expert-drawn, point-to-grid, distribution model and hybrid approach using data from all others), at three spatial resolutions. The median and 10th and 90th percentiles of range sizes are reported

Range type	Spatial resolution											
	1 km				25 km				50 km			
	Expert-drawn	Point-to-grid	Distrib. model	Hybrid	Expert-drawn	Point-to-grid	Distrib. model	Hybrid	Expert-drawn	Point-to-grid	Distrib. model	Hybrid
10th percentile	1386	8	4437	1407	5625	1875	20,625	6875	10,000	5000	35,000	12,500
Median	33,602	60	70,759	31,248	55,000	16,250	128,750	61,875	80,000	40,000	180,000	92,500
90th percentile	196,665	391	212,861	171,669	268,125	106,250	298,125	267,500	347,500	237,500	367,500	332,500
Median richness	26	0	36	23	30.5	9	50	31	34	20	58	34

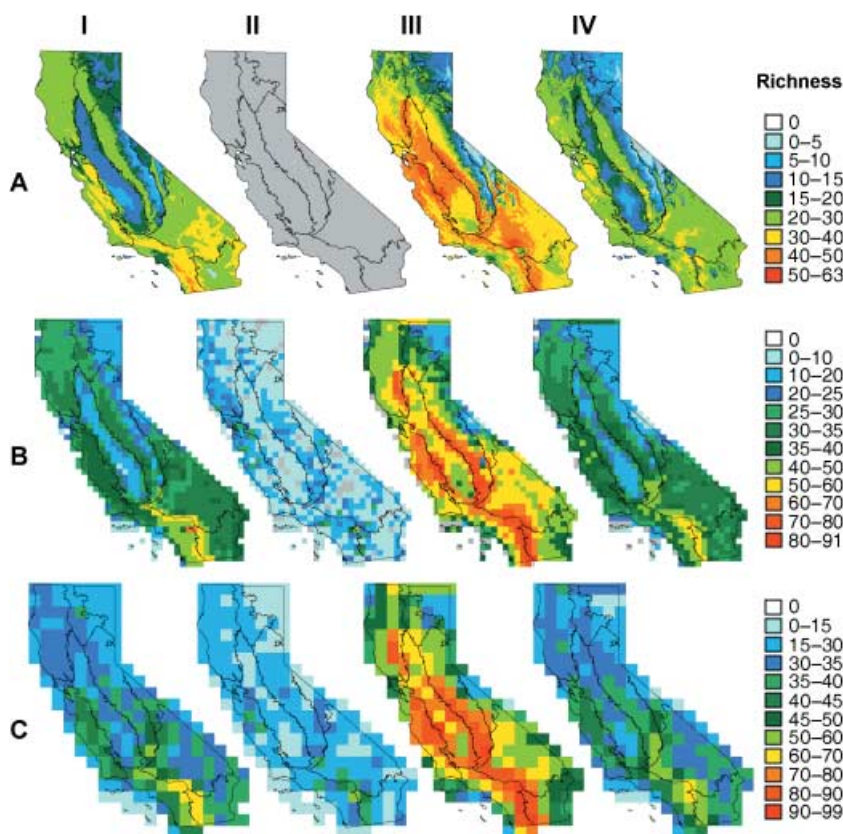


Figure 4 Species richness of amphibian and reptile species in California based on summation of four different types of range maps (I, expert-drawn; II, point-to-grid; III, distribution model; IV, hybrid approach using data from all others) at three spatial resolutions (A, 1 km; B, 25 km, C, 50 km). The polygons are the Jepson regions of California.

Table 3 Correlation coefficients among richness maps created with four different methods: expert-drawn, point-to-grid, and distribution models, and a hybrid method

	1 km			25 km			50 km		
	Point-to-grid	Distrib. model	Hybrid	Point-to-grid	Distrib. model	Hybrid	Point-to-grid	Distrib. model	Hybrid
Expert-drawn	0.038	0.445	0.794	0.487	0.635	0.839	0.727	0.729	0.868
Point-to-grid		0.035	0.056		0.393	0.492		0.622	0.717
Distrib. model			0.766			0.859			0.890

The distribution model richness map had high richness values compared to those generated by other methods (Table 2, Fig. 4), consistent with the large modelled range sizes (Table 2). Richness in the Central Western California region, especially the San Francisco Bay area, the Central Valley and, to a lesser extent, the Mojave Desert region and the contact zone between the Central Western California, Mojave Desert and Sonoran Desert regions in southern California, was markedly higher on the distribution model richness map than on that derived from the expert-drawn range maps (Fig. 5). Correlations among the expert-drawn range maps and distribution model derived richness maps increased from 0.45 at 1-km resolution to 0.73 at 50-km resolution (Table 3).

The hybrid richness map had slightly lower richness values per grid cell than the map based on expert-drawn ranges and much lower values than the richness map created with distribution

models (Table 2, Fig. 4). As the hybrid richness map draws on the expert-drawn range maps and on the distribution model range maps, it has a relatively high correlation with the richness maps derived from either of these sources, across spatial resolutions; the highest correlation, 0.89, is with the distribution model richness map at 50-km resolution (Table 3). There are substantial differences between the expert-drawn richness map and the hybrid richness map, even though species range sizes were similar (Table 2). Certain areas, such as the Central Valley and the Central Western California region, especially the San Francisco Bay area, and to a lesser extent the greater Los Angeles area in the South-western region, have higher richness on the hybrid richness map than on the expert-drawn richness map. Other areas, including the North-western region and the Modoc Plateau in the north and the Mojave and Sonoran Deserts in the south-west of the state, have higher richness on the expert-drawn

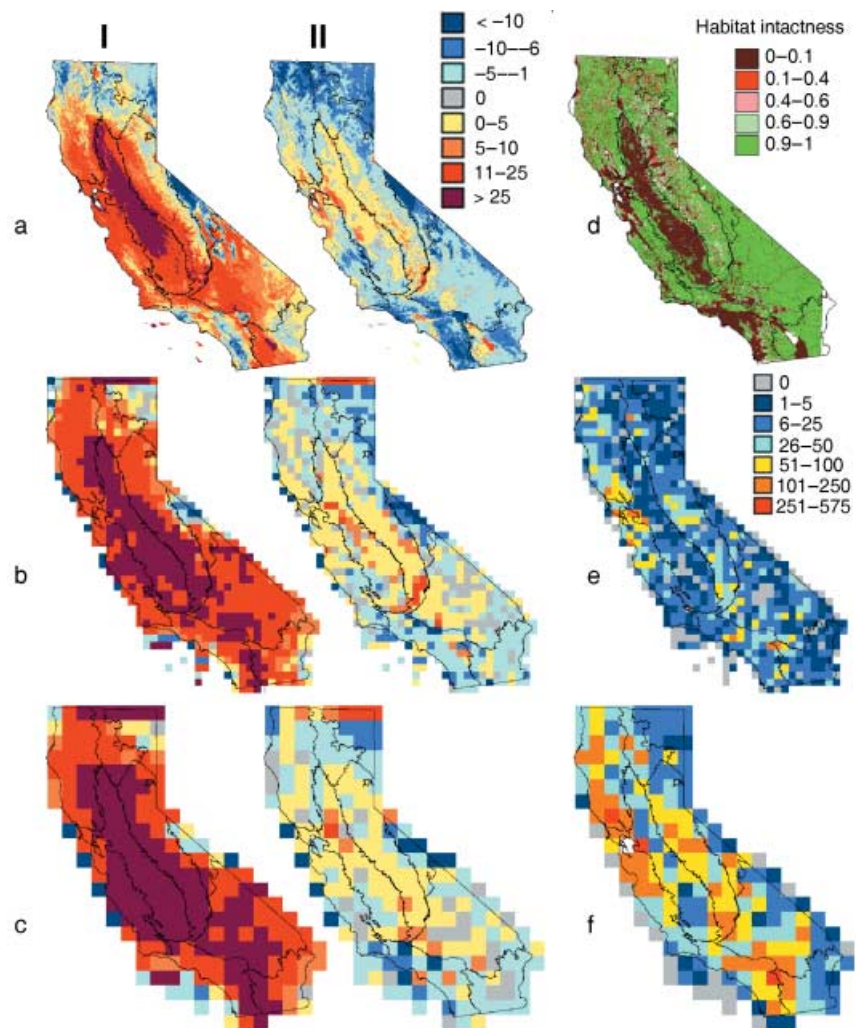


Figure 5 Difference in species richness of amphibian and reptile species in California based on the summation of four different types of range maps (I, distribution model — expert-drawn; II, hybrid approach — expert-drawn) at three spatial resolutions (a, 1 km; b, 25 km; c, 50 km); habitat intactness index (F. Davis, personal communication) (d), and the number of observations at two resolutions (e, 25 km; f, 50 km). The polygons are the Jepson regions of California.

richness map (Fig. 5a–c). These differences decrease with resolution, with very few cells having a difference of more than five species at a resolution of 50 km (Fig. 5a–c).

DISCUSSION

Each of the four methods explored in this study yielded different richness maps, highlighting the importance of considering the method used to create such maps. Richness maps had a higher correlation at lower spatial resolutions, but even at a resolution of 50 km high correlation masks pronounced regional differences that might strongly affect inference based on these maps. Progress in macroecological analysis, and other research that depends on species richness maps, would therefore be well served by continued efforts to improve knowledge of species distributions, and to account for the effect of uncertainty in the patterns analysed. However, not all analyses will be affected equally. For example, Hurlbert & White (2005) found that different environmental factors were better correlated with richness maps based on bird survey data than with maps based on expert-drawn ranges, while Mathias *et al.* (2004) found no significant differences when comparing different sets of expert-drawn range maps.

Distribution models predicted the largest range sizes of any method and corresponding higher levels of species richness, indicating that they might ‘over-predict’ the range size of a given species. Such error of commission can arise because the models are based on climate factors only and are ignorant of species interactions or limited dispersal ability (Pulliam, 2000; Anderson *et al.*, 2002; Guisan & Thuiller, 2005; Soberón & Peterson, 2005). California is topographically complex and has major mountain chains such as the Sierra Nevada and Coastal mountains, which are separated by the large Central Valley. Such topography is likely to generate barriers to dispersal for many of the amphibians in California, which often have low dispersal ability and may therefore not exist in all suitable habitats (Wake, 1997; Jockusch *et al.*, 2001; Jockusch & Wake, 2002). While over-prediction is undesirable in many applications where precise estimates of richness are required it has been illuminating in several ecological and evolutionary applications (e.g. Peterson *et al.*, 1999; Raxworthy *et al.*, 2003; Graham *et al.*, 2004b). In the context of richness mapping it may be interesting to compare the predicted and observed species richness across space to identify regions of discrepancy between the two measures and to test abiotic and biotic hypotheses leading to this discrepancy.

The expert-drawn range maps for California attempt to capture *current* species distributions. The large number of distribution records outside (but close to) the ranges indicates that the expert maps do not exactly delineate ranges. Most of the occurrence data base, also thoroughly checked by experts, was not yet available when the expert-drawn maps were created because retrospective georeferencing of the MVZ collection had not been completed. It might be that in some cases the species was considered extinct in locations where it was previously collected, but our analysis suggests that this is not generally the case, because the records outside the range maps tended to be more recent than the ones inside the range maps. Nevertheless, the attempt to capture current distributions may explain some of the differences between richness maps derived from expert-drawn and modelled species ranges. Richness based on modelled ranges is higher in the Central Valley where land use is dominated by agriculture and in the urban Los Angeles area (Fig. 5d). While experts are likely to have excluded such areas as suitable habitat for many species, distribution models indicate that parts of the Central Valley have suitable conditions for many species based on climate. In this case, richness maps based on distribution models may provide a more accurate estimate of historical species richness in the Central Valley, while expert-drawn range maps may better portray the current distribution.

A second difference between richness maps based on modelled ranges and those based on expert-drawn range maps is likely to stem from biases in the point occurrence data used in the distribution models. In general, models tend to predict lower richness as compared to range maps in the north-west (North-western and Madoc Plateau regions) and south-west (Mojave and Sonoran Desert regions). For these regions we had few point occurrences relative to other parts of California, indicating that they are relatively under-sampled (Fig. 5e,f), perhaps a reflection of relying solely on a collection housed in central west California. Such under-sampling can influence species distribution models; if there are no records of occurrence in an area with an environment that is distinct from areas from where there are records, then a model will generally predict this environment to be unsuitable (Kadmon *et al.*, 2004). On the other hand, in cases where species only use environmentally restricted parts of such areas, the modelled ranges are likely to be more accurate than the expert-drawn range maps.

It is likely that the point-to-grid method does not capture actual species richness unless a rather large grid cell size is used. While they may serve as a conservative benchmark at large grid sizes, the other methods for creating richness maps seem more promising unless observational data are very dense. Nonetheless, point occurrence data can provide valuable additional information to expert-drawn range maps. First, expert-drawn range maps tend to be somewhat static; once created they are difficult to revise or evaluate by others than the experts who created them. In contrast, through on-going georeferencing (Wieczorek *et al.*, 2004) and species occurrence data-sharing initiatives (Stein & Wieczorek, 2004) additional occurrence records are continuously becoming available. These new records may represent areas outside the expert-drawn range maps and provide valuable information

about species ranges. Further, because much of this information comes from museum collections each point can — to some extent — be evaluated for veracity (Graham *et al.*, 2004a). Second, expert-drawn range maps, in the case of California and perhaps more generally, reflect current species distributions. For areas with a long sampling history, records of occurrence that fall outside these maps may provide information on the historical range (i.e. pre-anthropogenic disturbance) of a given species.

The hybrid approach uses all sources of information: point occurrences, expert-drawn range maps and modelled distributions. By limiting models with point-modified expert maps we can capitalize on the strengths of each method. The high spatial resolution of ranges predicted with a species distribution model can be used, but errors of commission outside of the known range of a species are removed. In the case of the California data, the modification of range maps may have resulted in the inclusion of areas that currently are not suitable for a given species because of alteration of the habitat by humans. Whether species distribution and richness maps should aim at representing the current situation, a potential future situation or the situation at some point in the past depends on the application for which the maps are made.

There are several sources of variability and error associated with point occurrence, expert-drawn and modelled ranges, and the magnitude of these errors is generally unknown and likely to vary on a case-by-case basis. Point occurrence data will almost always underestimate the true area of occupancy for a species because of sampling bias and stochasticity of occupancy of locations over time. Further, the availability of point occurrences varies and is likely to be higher in California than in many other regions, especially in the species-rich tropical countries. The nature of expert-drawn range maps depends upon their intended use and on the expertise available, and they are usually much larger in extent than the true area of occupancy. While the expert-drawn range maps used here may provide relatively accurate distributions, such maps are likely to be much less accurate for many other regions. Finally, the benefit of the hybrid mapping technique used here is dependent upon both the magnitude of under- or over-estimation of expert-drawn and modelled range maps and the relative spatial disjunction of these ranges.

We have not explored the possibility of varying the method used in different areas within a given region or assessing the uncertainty associated with a given method. For example, where an expert-drawn range map predicts that a species is present but a distribution model predicts its absence, an expert-drawn range map might be accepted as correct if there is reason to believe that the sampling was biased in geographical or environmental space, whereas the model result might be accepted in other cases. Further, overall sampling effort in terms of environmental coverage, based on point occurrences in a geographical region across all species, might be used to indicate confidence in richness patterns since both expert maps and distribution models are likely to be affected by a lack of data. Information about historical barriers and/or competitors could be used in lieu of expert-drawn range maps to limit species distribution models.

Even in relatively well studied and sampled areas such as California, different methods and data sets lead to rather distinct

maps of species richness. We cannot readily evaluate which method was most accurate because we do not have perfect knowledge about how species are distributed. Nonetheless, by exploring different methods of creating richness maps we can evaluate the weakness of different approaches and this may help us determine how to create a richness map for a given application and region. In most cases, this is likely to be a hybrid approach that combines the high spatial resolution of distribution modelling with expert opinion.

ACKNOWLEDGEMENTS

We would like to thank the Museum of Vertebrate Zoology for occurrence data.

REFERENCES

- Airola, D.A. (1988) *Guide to the California wildlife habitat relationships system*. California Department of Fish and Game, Sacramento.
- Anderson, R.P., Peterson, A.T. & Gomez-Laverde, M. (2002) Using niche-based GIS modelling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3–16.
- Austin, M.P. (1985) Continuum concept, ordination methods, and niche theory. *Annual Review of Ecology and Systematics*, **16**, 39–61.
- Brooks, T.M., Bakarr, M.I., Boucher, T., Da Fonseca, G.A.B., Hilton-Taylor, C., Hoekstra, J.M., Moritz, T., Olivieri, S., Parrish, J., Pressey, R.L., Rodrigues, A.S.L., Sechrest, W., Stattersfield, A., Strahm, W. & Stuart, S.N. (2004) Coverage provided by the Global Protected-Area System: is it enough? *Bioscience*, **54**, 1081–1091.
- Burgess, N.D., Rahbek, C., Larsen, F.W., Williams, P. & Balmford, A. (2002) How much of the vertebrate diversity of sub-Saharan Africa is catered for by recent conservation proposals? *Biological Conservation*, **107**, 327–339.
- California Department of Fish and Game & California Inter-agency Wildlife Task Group (2000) *Standards and guidelines for CWHR species models*. Technical Report No. 31. California Wildlife Habitat Relationships System. California Department of Fish and Game, Sacramento.
- California Department of Fish and Game & California Inter-agency Wildlife Task Group (2002) *California wildlife habitat relationships system*. Wildlife and Habitat Data Analysis Branch, Sacramento.
- Calsbeek, R., Thompson, J.N. & Richardson, J.E. (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California floristic province. *Molecular Ecology*, **12**, 1021–1029.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Czaplewski, R.L. (1994) *Variance approximations for assessments of classification accuracy*. USDA Forest Service Rocky Mountain Forest and Range Experiment Station Research Paper (RM-316). Fort Collins, Colorado.
- Davis, F.W., Storms, D.M., Hollander, A.D., Thomas, K.A., Stine, P.A., Odion, D., Borchert, M.I., Thorne, J.H., Gray, M.V., Walker, R.E., Warner, K. & Graae, J. (1998) *The California gap analysis project: final report*. University of California Santa Barbara.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettman, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Ferrier, S. (1997) Biodiversity data for reserve selection: making best use of incomplete information. *National parks and protected areas: selection, delimitation and management* (ed. by P.J. Pigram and R.C. Sundell), pp. 315–329. Centre for Water Policy Research, University of New England, New England.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004a) New developments in museum-based informatics and application in biodiversity analysis. *Trends in Ecology & Evolution*, **19**, 497–503.
- Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J. & Moritz, C. (2004b) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in Dendrobatid frogs. *Evolution*, **58**, 1781–1793.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models? *Ecology Letters*, **8**, 993–1009.
- Guralnick, R. & Van Cleve, J. (2005) Strengths and weaknesses of museum and national survey data sets for predicting regional species richness: comparative and combined approaches. *Diversity and Distributions*, **11**, 349–359.
- Habib, L.D., Wiersma, Y.F. & Nudds, T.D. (2004) Effects of errors in range maps on estimates of historical species richness of mammals in Canadian national parks. *Journal of Biogeography*, **30**, 375–380.
- Hickman, J.C. (1993) *The Jepson manual: higher plants of California*. University of California Press, Berkeley.
- Hijmans, R.J. & Spooner, D.M. (2001) Geographic distribution of wild potato species. *American Journal of Botany*, **88**, 2101–2112.
- 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 Evolution*, **46**, 291–296.
- Hijmans, R.J., Garrett, K.A., Huamán, Z., Zhang, D.P., Schreuder, M. & Bonierbale, M. (2000) Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology*, **14**, 1755–1765.
- Hijmans, R.J., Guarino, L., Bussink, C., Mathur, P., Cruz, M., Barrantes, I. & Rojas, E. (2004) *DIVA-GIS, version 4. A geographic information system for the analysis of biodiversity data. Manual*. (Available at <http://www.diva-gis.org>)

- 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.
- Hurlbert, A.H. & White, E.P. (2005) Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8**, 319–327.
- Jockusch, E.L. & Wake, D.B. (2002) Falling apart and merging: diversification of slender salamanders (Plethodontidae: *Batrachoseps*) in the American West. *Biological Journal of the Linnean Society*, **76**, 361–391.
- Jockusch, E.L., Yanev, K.P. & Wake, D.B. (2001) Molecular phylogenetic analysis of slender salamanders, genus *Batrachoseps* (Amphibia: Plethodontidae), from central coastal California with descriptions of four new species. *Herpetological Monographs*, **15**, 54–99.
- 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.
- 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.
- Mathias, P.V.C., Mendonca, C.V., Rangel, T.F.L.V.B. & Diniz, J.A.F. (2004) Sensitivity of macroecological patterns of South American parrots to differences in data sources. *Global Ecology and Biogeography*, **13**, 193–198.
- Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of elapid snakes of Australia* (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A., Webster, A.J., Ding, T., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Blackburn, T.M., Gaston, K.J. & Owens, I.P.F. (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature*, **436**, 1016–1019.
- Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.
- Peterson, A.T., Soberon, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberon, J., Buddemeier, R.H. & Stockwell, D.R.B. (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences USA*, **98**, 4534–4539.
- Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. & Peterson, A.T. (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837–841.
- van Rensburg, B.J., Chown, S.L. & Gaston, K.J. (2002) Species richness, environmental correlates and, and spatial scale: a test using South African birds. *The American Naturalist*, **159**, 566–577.
- Rissler, L.J., Hijmans, R.J., Graham, C.H., Moritz, C. & Wake, D.B. (in press) Phylogeographic lineages and species comparisons in conservation analyses: A case study of California herpetofauna. *The American Naturalist*.
- Soberon, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- Stebbins, R.C. (2003) *Western reptiles and amphibians*. Peterson Field Guides. Houghton Mifflin Company, Boston.
- Stein, B.R. & Wiecezorek, J.R. (2004) Mammals of the World: MaNIS as an example of data integration in a distributed network environment. *Biodiversity Informatics*, **1**, 14–22.
- Wake, D.B. (1997) Incipient species formation in salamanders of the *Ensatina* complex. *Proceedings of the National Academy of Sciences USA*, **94**, 7761–7767.
- Wiecezorek, J.R., Guo, Q. & Hijmans, R.J. (2004) The point-radius method for georeferencing point localities and calculating associated uncertainty. *International Journal of Geographic Information Science*, **18**, 745–767.

BIOSKETCHES

Catherine Graham is a macroecologist/landscape ecologist whose work focuses on investigating the effect of spatial and temporal arrangement of habitats on ecological and evolutionary processes and the use of this information to inform conservation policy.

Robert Hijmans is a geographer with interests in biogeography and land use in developing countries. He currently works on species distribution modelling methods, and on temporal changes in the distribution of vertebrates in California. He also develops data bases and computer tools for georeferencing and analysing species distribution data.

Editor: Tim Blackburn