Estimating patterns of reptile biodiversity in remote regions

Gentile Francesco Ficetola¹, Anna Bonardi¹, Roberto Sindaco² and Emilio Padoa-Schioppa¹

¹Dipartimento di Scienze dell’Ambiente e del Territorio, Università degli Studi di Milano-Bicocca, Piazza della Scienza 1, 20126, Milano, Italy; ²Museo Civico di Storia Naturale, I-10022, Carmagnola, Italy

ABSTRACT

Aim The incompleteness of information on biodiversity distribution is a major issue for ecology and conservation. Researchers have made many attempts to quantify the amount of biodiversity that still remains unknown. We evaluated whether models that integrate ecogeographical variables with measures of the effectiveness of sampling can be used to estimate biodiversity patterns (species richness) of reptiles in remote areas that have received limited surveys.

Location The Western Palaearctic (Europe, Northern Africa, the Middle East and Central Asia).

Methods We gathered data on the distribution of turtles, amphisbaenians and lizards. We used regression models integrating spatial autocorrelation (spatial eigenvector mapping and Bayesian autoregressive models) to analyse species richness, and identified relationships between species richness, ecogeographical features and large-scale measures of accessibility.

Results The two regression techniques were in agreement. Known species richness was dependent on ecogeographical factors, peaking in areas with high temperature and annual actual evapotranspiration, and intermediate cover of natural vegetation. However, richness declined sharply in the least accessible areas. Our models revealed regions where reptile richness is likely to be higher than currently known, particularly in the biodiversity hotspots in the south of the Arabian Peninsula, the Irano-Anatolian region, and the Central Asian mountains. An independent validation data set, with distribution data collected recently throughout the study region, confirmed that combining accessibility measures with ecogeographical variables allows a good estimate of reptile richness, even in remote areas that have received limited monitoring so far. Some remote regions that support very rich communities are covered very little by protected areas.

Main conclusions Integrating accessibility measures into species distribution models allows biologists to identify areas where current knowledge underestimates the actual richness of reptiles. Our study identifies regions requiring future biodiversity research, proposes a novel approach to biodiversity prediction in poorly studied areas, and identifies potential regions for conservation.

Keywords Amphisbaenians, biodiversity hotspots, conservation biogeography, ecogeography, lizards, predictive models, spatial autocorrelation, species richness, survey bias, turtles.
Information is a major challenge within conservation biogeography, limiting our ability to analyse biodiversity patterns and set conservation priorities (Lomolino, 2004; Cardoso et al., 2011; Mora et al., 2011). Unfortunately, quantifying the amount of biodiversity that remains unknown in a given area is challenging (Raxworthy et al., 2003; Mora et al., 2011).

Species distribution models based on relationships between biodiversity distribution and environmental features help to elucidate the factors determining biodiversity and to identify conservation targets even when information is incomplete (Raxworthy et al., 2003; Qian & Rickles, 2008; Thullier et al., 2011). These models, being based on the available species distribution data, may be affected by poor sampling. Integrating data on the effectiveness of surveys may greatly improve the performance of models of species richness and species distribution (Kadmon et al., 2004; Bini et al., 2006; Phillips et al., 2009; Barbosa et al., 2010). Unfortunately, measures of the effectiveness of surveys are often lacking, particularly for large-scale data sets, where data are obtained from a variety of heterogeneous sources. Accessibility can be a major source of sampling bias. Monitoring in remote areas is often limited, and so biodiversity can be significantly underestimated (Nelson et al., 1990; Margules & Pressey, 2000; Bini et al., 2006; Sánchez-Fernández et al., 2008; Barbosa et al., 2010; Kent & Carmel, 2011). The accessibility of a region can thus be a useful proxy of the effectiveness of sampling in biogeographical studies. Recently, Nelson developed a global-scale measure of accessibility, quantified as the travel time from the nearest city using land-based or water-based transport (Nelson, 2008; Uchida & Nelson, 2010). However, we are not aware of studies integrating large-scale measures of accessibility into species distribution models.

The aim of this study was to analyse the richness of reptile species (considering turtles, amphisbaenians and lizards) in the Western Palaearctic, and to evaluate the importance of accessibility in determining known species richness at the biogeographical scale. The Western Palaearctic is a large biogeographical region, encompassing Europe, North Africa, the Middle East and Central Asia (Fig. 1); it contains hotspots of reptile biodiversity and endemism (Mittermeier et al., 2004; Cox et al., 2006; Sindaco & Jeremčenko, 2008), but is heavily threatened by growing human pressure (Brooks et al., 2006; Cox et al., 2006; Ficetola & Padoa-Schioppa, 2009). The area includes easily accessible, human-dominated regions (e.g. Western Europe), in which extensive inventories of reptile biodiversity have been performed, but also remote regions (Fig. 1a) where biodiversity data remain sparse (Sindaco & Jeremčenko, 2008). Furthermore, we evaluated whether integrating accessibility into species distribution models may help biologists to improve estimates of species richness, even in remote areas that have received limited surveys so far. In addition, reptiles are among the terrestrial vertebrates for which distribution data are more limited, and the only ones for which no comprehensive assessment of conservation status has been completed (Hoffmann et al., 2010). The results of our study may improve knowledge of reptile biodiversity in the study area, and provide large-scale information that can be useful for conservation planning.

MATERIALS AND METHODS

Study area and data sets

We considered the ‘Western Palaearctic’ according to Sindaco & Jeremčenko (2008). This region includes the western portion of the Palaearctic realm [Europe, the Arabian Peninsula, western Asia (west of the Indus Valley), the former Soviet Republics of Central Asia and Russia (west of the Ural Mountains)] and several countries of northern Africa (including the Sahara Desert north of approximately 22° N) (Fig. 1). Our data set included more than 67,000 distributional records collected before 2008, showing the known distribution of the 480 native species of terrestrial turtles, amphisbaenians and lizards occurring in the area (Sindaco & Jeremčenko, 2008). The number of records per cell cannot be considered an index of the sampling effort, because in the best-surveyed areas (e.g. Europe), data were obtained from comprehensive sources such as atlases, summarizing a much larger number of observations. Presence records of each species were mapped on a grid of 3530 cells with a resolution of 1° × 1° (the resolution of data with less accurate coordinates). Taxonomy followed the checklist of Sindaco & Jeremčenko (2008), which was based on the critical review of about 880 papers. The list includes species recognized on the basis of biological, evolutionary and morphological species concepts (Uetz, 2010). Distribution maps of each of the 480 species used for analyses are available in Sindaco & Jeremčenko (2008). Reptile richness in each cell was then calculated by overlapping the distribution maps of all the 480 species. We also considered coastal cells and islands, as they harbour multiple endemic species and can constitute important biodiversity areas, when taking into account their limited surface area (Cox et al., 2006; Ficetola & Padoa-Schioppa, 2009).

As environmental variables, we considered two geographical variables – cell surface occupied by non-aquatic environments (calculated on the basis of the GlobCover land cover; Bicheron et al., 2008) and elevation range (calculated on the basis of a 30-arc-second digital elevation model; Hijmans et al., 2005), and three climatic variables – annual mean temperature, annual actual evapotranspiration (annual AET hereafter), and annual summed precipitation (New et al., 2002), which are considered to be major drivers of reptile biodiversity (Qian & Rickles, 2008; Powney et al., 2010). Potential evapotranspiration can also have an important role (Rodríguez et al., 2005), but was not considered because it is highly collinear to the other climatic variables. Furthermore, we considered the average accessibility of each cell (Nelson, 2008) and the percentage of each cell occupied by natural vegetation, calculated on the basis of GlobCover (Bicheron et al., 2008). We also calculated the percentage of each cell covered by protected areas, on the basis of the World Database on Protected Areas (http://protectedplanet.net/) (Fig. 1b). All variables were upscaled.
Accessibility represents the travel time to the nearest city (population > 50,000 people) using land-based or water-based transport on a regular raster grid. It is based on a cost–distance algorithm computing the 'cost' of travelling between two locations, assuming that travelling time is affected by parameters such as road/railway network, presence of paths, navigable water bodies, slope and political boundaries. For each 1° × 1° cell, we obtained the average accessibility (in hours), from the 30′ resolution accessibility grid layer (Nelson, 2008; Fig. 1a). A few cells (3%) had average accessibility ≤ 1 h. To avoid these cells acting as outliers, we assumed that all areas within 1 h of the nearest city may be surveyed with similarly high sampling effort, and were assigned an accessibility of 1 h. The picture shows the average predictions of spatial eigenvector mapping and Bayesian conditional autoregressive model. Predictions of the two models are shown in Fig. S2 in Appendix S2. The locations of major cities are also shown: Alg, Algiers; Alm, Almaty; Ba, Baghdad; Cai, Cairo; Kar, Karachi; Mar, Marrakech; Mos, Moscow; Par, Paris; Riy, Riyadh; San, Sana’a; Te, Tehran; Tel, Tel Aviv; Tri, Tripoli.
were linked using a king’s connection scheme and we gener-
tated eigenvectors using Moran’s eigenvector maps (Dray et al.,
2006; Borcard et al., 2011). We then selected the eigenvectors
that best reduced the spatial autocorrelation of residuals, and
included them as spatial predictors in generalized linear mod-
els (GLM) (Griffith & Peres-Neto, 2006; Dormann et al.,
2007). Comparisons among statistical methods showed that
this implementation of SEVM is flexible and efficient, even
when analysing non-normal data (Dormann et al., 2007). To
take overdispersion into account, we used a quasi-Poisson
family and tested significance using an F test (Crawley, 2007).
We used the procedure proposed by Phillips and co-workers
(Phillips, 2010; Elith et al., 2011) to evaluate the relative
importance of predictors in SEVM. First, we built the models
including each environmental variable in isolation. Subse-
quently, each variable (including both linear and quadratic
terms) was removed in turn, and the model run with the
remaining variables. The difference in explained variation
between the full model and the one excluding the variable of
interest provides an estimate of the independent contribution
of such variable, after taking into account other parameters
and spatial autocorrelation (Phillips, 2010; Elith et al., 2011).

BCA is a hierarchical Bayesian technique that incorporates
a spatial random effect, describing the neighbourhood of each
site, into Bayesian regression models. Simulations have shown
that BCA is one of the spatial techniques with the best per-
formance under a variety of conditions (Latimer et al., 2006;
Beale et al., 2010). We built BCA assuming a Poisson error
distribution. A model assuming negative binomial error had
a higher deviance information criterion, indicating that
overdispersion did not affect the BCA model. To ensure con-
vergence, we ran three different MCMC chains. For each
chain, we performed a burn-in of 15,000 iterations, followed
by 45,000 iterations thinned every 10 iterations. For all vari-
ables, Brooks–Gelman–Rubin diagnostics were approximately
1, indicating convergence (Kéry, 2010). We then calculated
the estimates of regression coefficients and their 95% highest
posterior density credible intervals (HPDI).

The results of ecogeographical studies may be affected by
the scale at which analyses are performed (Hawkins, 2012).
Furthermore, the large number of cells might increase the
chance of finding significant results. To confirm our results,
we therefore repeated our analyses using a grid with a resolu-
tion of $2^\circ \times 2^\circ$ (975 cells).

To reduce skewness and improve normality where needed,
variables were transformed using logarithms (cell surface, el-
vational range and accessibility), square-root (summed annual
precipitation) or square-root-arcsine (percentage of natural
vegetation). We calculated the proportion of explained devi-
ance ($R^2_D$) as a measure of the variation explained by species
richness models. For SEVM, values of $R^2_D$ were adjusted to
take into account the number of predictors in the model
(Borcard et al., 2011). When more than one modelling tech-
nique is used, differences among algorithms can determine
differences in performance and lead to non-identical results
(model-based uncertainty) (Marmion et al., 2009). To reduce
the model-based uncertainty for model predictions, we kept
among the best-performing techniques when using non-nor-
mal data: spatial eigenvector mapping (SEVM) and Bayesian
intrinsic conditional autoregressive models (BCA) (Dormann
et al., 2007; Beale et al., 2010). Our dependent variable was a
count, and so we assumed a Poisson error distribution rather
than more standard techniques such as generalized least
squares, simultaneous or conditional autoregressive models,
which are suitable for normally distributed data (Dormann
et al., 2007; Beale et al., 2010). Species richness may have a
maximum when certain environmental variables reach cer-
tain values; furthermore, previous studies suggested that rela-
tionships between species richness and ecogeographical
variables are often nonlinear (Whittaker et al., 2007; Qian,
2010). We therefore also included quadratic terms for cli-
matic variables, vegetation and accessibility in all models, to
take into account potential nonlinear relationships.

SEVM allows the translation of the spatial arrangement of
data points into explanatory variables (eigenvectors) that cap-
ture spatial effects (Dormann et al., 2007). Neighbouring cells
were linked using a king’s connection scheme and we gener-

Figure 2 Relationships between environmental variables and
reptile richness for the Western Palaearctic, as estimated by
spatial eigenvector mapping. Predictions are shown for (a) mean
annual temperature; (b) summed annual precipitation; (c)
normal annual evapotranspiration (annual AET); (d) proportion
of each cell occupied by natural vegetation (%); (e) mean
accessibility (travel time in hours from the nearest city using
land-based or water-based transport). The plots include variables
for which the model takes quadratic terms into account. Grey
dots represent observed species richness in cells. The predictions
for each variable are made while holding the other variables
fixed at their global median. Fig. S2 in Appendix S2 reports the
results from the Bayesian autoregressive model and the global
medians of variables.
the average consensus prediction of the two species richness models (Marmion et al., 2009). Using the individual results of either model yielded nearly identical results (see Results). Subsequently, we projected models using a new data set, with identical environmental variables and spatial structure, but assuming accessibility is 1 h for all cells. This allowed us to evaluate the potential species richness of each cell, if it received a surveying effort similar to the most accessible ones (‘projected richness’ hereafter).

Validation of model predictions

To test whether the projected richness is an accurate estimate of the actual species richness, we obtained recent data on reptile distributions (from recent field research or publications) in 80 grid cells that have received intense sampling (Fig. 3, and see Appendix S1 in Supporting Information). These recent data were not included in the 2008 data set used to build the species richness models. Independent validation data were obtained mostly from recent publications on reptile distributions (73 cells). In six cells, recent data were obtained both from extensive field surveys performed by the authors and from the literature. Sampling was performed from 2008 to 2010 through diurnal and nocturnal visual encounter surveys, trying to cover the widest range of environmental features (Heyer et al., 1994; Ficetola et al., 2010). In one additional cell, data were obtained through field surveys only (total: 80 cells with recent data) (Appendix S1). As these cells recently received an intense and focused surveying effort, they represent areas with reliable information for the validation of model predictions. We thus assumed that reptile richness recorded from recent, intensive surveys (‘validation richness’ hereafter) represents an improved estimate of the actual species richness in these cells. Validation cells were in multiple regions of the study area, with a wide range of accessibility (Fig. 3, Appendix S1).

In order to assess the ability of the projected richness model to correctly predict the actual species richness in the validation cells, we built three Poisson GLMs considering validation richness as the dependent variable in all cases. As potential predictors, we included (1) the projected richness (i.e. the richness predicted by the species richness model assuming accessibility is 1 h in all cells), and (2) the number of species recorded in the 2008 data set (i.e. the original data set of Sindoaco & Jeremčenko, 2008). In both these models, we assumed $B = 1$ and intercept $= 0$, i.e. a perfect overlap between best available value of the species richness and either (1) our projected richness or (2) the richness recorded until 2008. We considered the null model as a third GLM (3), assuming no relationships between the validation and projected species richness. We then used the Akaike information criterion (AIC) to compare the performance of the three GLMs.

Due to sampling constraints, remote cells were less frequent in our validation data. Furthermore, several validation cells were clustered, and this may affect the results of regression-based validation. As an additional metric of prediction performance, we compared the projected richness to the validation richness. For each cell, we calculated the absolute value of [(projection – validation)/validation] $\times$ 100%. We then subtracted that value from 100%, to obtain the ‘percentage accuracy’ for each cell. Subsequently, we repeated the same procedure with the richness recorded until 2008.

We used Spearman’s correlation to evaluate the relationship between projected reptile richness in each cell, and the cell percentage covered by protected areas. Because of spatial autocorrelation, significance was calculated using Dutilleul’s (1993) degrees of freedom. We performed statistical analyses in R 2.12 (R Development Core Team, 2010) using the package spdep (Bivand et al., 2010); we ran BCA in WinBUGS 1.4 (Spiegelhalter et al., 2008).

RESULTS

Average reptile richness in 1° x 1° cells ($\pm$ SD) was 4.3 $\pm$ 5.6 species (range: 0–41). The areas with the highest known richness were in the eastern Mediterranean and western Maghreb. Richness was irregularly distributed in wide areas of the Middle East, with species-rich cells near cells with very few known species (Fig. 1c). This pattern was most apparent in the area of Israel–Jordan–Sinai Peninsula, southwestern Yemen, and around Riyadh, Saudi Arabia (Fig. 1c). SEVM and BCA generally gave the same outcome, and their results were in strong agreement ($R^2 = 0.82$; Fig. S1 in Appendix S2). Both models included 12 predictors (seven variables with five quadratic terms); SEVM also included 91 eigenvectors representing spatial autocorrelation (Table 1). Both models explained a large amount of the variation in reptile species richness (SEVM: adjusted $R^2_{\text{adj}} = 0.69$; BCA: $R^2_D = 0.63$). When controlling for the effect of cell surface, species richness increased with elevational range, temperature and precipitation, and decreased with annual AET. Furthermore, richness was highest in cells with intermediate values of natural vegetation cover (Table 1, Fig. 2; see also Fig. S2 in Appendix S2). Temperature, natural vegetation, elevational range and accessibility explained 20–32% of variation, when used in isolation (Table 1). After taking into account spatial autocorrelation, accessibility was the variable with the highest independent contribution to the model. The independent contribution of vegetation cover and temperature was 2–4% while, when taking the other variables into account, the independent contribution of surface, elevational range and precipitation was limited (Table 1).
When taking the other variables into account, the highest richness was observed in easily accessible cells, and gradually decreased in more remote areas, following a nonlinear pattern. For instance, if environmental features were held constant, reptile richness was 4–5 species in cells with an average accessibility of 1 h, compared with 1–2 species in cells with an accessibility of 10 h (Fig. 2e, Fig. S2e in Appendix S2). Results were extremely similar according to SEVM and BCA, although the relationship between two climatic features (annual AET and annual precipitation) were significantly nonlinear in SEVM but not in BCA (Table 1). The analysis performed at 2° × 2° resolution yielded very similar results (Appendix S3). For accessibility, regression coefficients were slightly different from the model at 1° × 1° resolution, but the negative, nonlinear relationship was confirmed (Appendix S3).

We then projected species richness models, assuming that all cells are accessible in 1 h, while keeping the original values of environmental variables. This allowed us to estimate the potential richness patterns, if remote areas received the same surveying effort as the accessible ones. In several regions, the projected consensus model predicted a higher species richness than the currently known values, particularly in the western Maghreb, in the south of the Arabian peninsula, in the Irano-Anatolian region, and in the Central Asian mountains (Fig. 1d, Fig. S3 in Appendix S2). Cells with high projected richness tended to be those with less cover of protected areas (Fig. 1; correlation between projected richness and cover of protected areas: \( r_1 = -0.25 \), Dutilleul’s modified d.f. = 96.9, \( P = 0.015 \)). Furthermore, in the Western Palaearctic, the cover of protected areas was highest in easily accessible cells (\( r_1 = -0.24 \), d.f. = 145.3, \( P = 0.005 \)).

### Validation of biodiversity predictions

The recent distribution data updated the available information, and usually increased known richness in validation cells. The average number of newly recorded species per cell was 2.5 (range: 0–12; Appendix S1). The projected model predicted reptile richness in the recently surveyed cells well (GLM: 2008 data = 79%, 2008 data = 77%). The average accuracy of predictions in validation cells (n = 80) was 79%, and was higher than the accuracy calculated using 2008 data (77%). The good performance of predictions was particularly evident in remote cells (cells with accessibility > 3 h: n = 39 accuracy of prediction = 74%, accuracy of 2008 data = 70%; cells with accessibility > 6 h: n = 7, accuracy of prediction = 80%, accuracy of 2008 data = 55%).

### DISCUSSION

In remote areas, we only know part of the species richness (Nelson et al., 1990), but quantifying the magnitude of richness underestimation is extremely difficult. Our analyses suggest that, after taking key ecogeographical variables into account.
account, accessibility influences the known reptile biodiversity at the biogeographical scale. Thus, integrating accessibility into models helps to estimate reptile richness in remote, poorly studied areas (Fig. 4a). Obviously, accessibility is not a perfect proxy of sampling effort because, particularly at fine scales, other factors may influence the completeness of inventories, such as perceived attractiveness for researchers and the presence of institutions specializing in certain taxa or specific geographical areas (Nelson et al., 1990; Sánchez-Fernández et al., 2008; Boakes et al., 2010; Vale & Jenkins, 2012). Nonetheless, broad-scale measures of accessibility can provide information on the bias of field surveys, which may be used to improve the performance of distribution models and allows first assessments of the biodiversity in remote areas where observations are sparse (Bini et al., 2006).

Our analysis suggests that reptile richness in remote areas such as the Iranian Plateau (including Afghanistan and south-western Pakistan), as well as parts of Central Asia, may be higher than currently known (Fig. 1, Fig. S3 in Appendix S2). Although this region includes two of the world’s recognized biodiversity hotspots (the mountains of Central Asia and the Irano-Anatolian region), known reptile richness is still relatively low compared with other hotspots (Mittermeier et al., 2004). The remoteness of the region, the roughness of landscape and the lasting political instability may have limited systematic surveys, and the actual biodiversity is likely to be higher than currently known. Two major mechanisms may cause such underestimation of species richness (Lomolino, 2004; Mokany & Ferrier, 2011): (1) species distribution is poorly known (Wallacean shortfall), and (2) some species are yet to be described (Linnean shortfall). Recent surveys provide a measure of causes of richness underestimation: on average, surveys increased the known richness by 47% and, in 13 cells, the increase in richness was \( \geq 100\% \) (Appendix S1). Limited knowledge of species distribution was the major cause of richness underestimation, as many species are recorded for only one or very few localities, and most new records in the validation cells expanded the range of known species. The discovery of new species is much less frequent than distribution updates. Nevertheless, most of the recently described reptile species in the Western Palaearctic (see e.g. Masroor, 2008; Rösler et al., 2008; Nazarov et al., 2009; Torki, 2010; Busais & Joger, 2011; Torki et al., 2011) have been found in areas where our model suggests a higher richness than currently reported, such as the south of the Arabian Peninsula and the Irano-Anatolian region. Finally, in recent years, molecular tools are improving our definition of species identity, with the recognition of cryptic taxa and the raising of geographically isolated ‘subspecies’ to the rank of species. These studies tend to have a geopolitical bias (Harris & Froufe, 2005), but molecular tools and integrated taxonomy are nevertheless starting to expand our knowledge of biodiversity, even for remote or understudied regions (e.g. Barata et al., 2012). Overall, we conclude that both Wallacean and Linnean shortfalls affect known reptile richness and our models help to identify regions where these shortfalls are particularly strong.

Reptiles are the only terrestrial vertebrates for which, so far, no comprehensive maps of species distribution range are available at the global scale, thereby limiting large-scale analyses of biodiversity patterns. For instance, this is the first analysis of reptile richness covering the whole Western Palaearctic on the basis of a regular grid (see Rodríguez et al., 2005, for an analysis of Europe). Previous large-scale studies of reptile richness highlighted that the availability of energy is the major determinant of reptile richness (Rodríguez et al., 2005; Qian et al., 2007; Whittaker et al., 2007; Qian & Ricklefs, 2008; Powney et al., 2010; Qian, 2010). Our results were in agreement with these findings, as reptile richness was highest in areas characterized by warm climate and low annual AET (Fig. 2). Integrating proxies of sampling effort, such as accessibility, may improve the performance of biodiversity models, allowing better estimates of the relationships between species richness and environmental features (Phillips et al., 2009; Bonardi et al., 2011; Gómez-Rodríguez et al., 2012), and helping to improve our understanding of the impact of environmental changes (e.g. climate or land-use change) on species richness. This can be particularly valuable.
for regions or taxa where insufficient data are available for modelling each species individually (Mokany & Ferrier, 2011). More accurate estimates of alpha diversity can also be integrated into a metacommunity framework, to estimate community composition in poorly known areas (Mokany et al., 2011). Nevertheless, it should be noted that no single environmental variable explained richness well, and that spatial autocorrelation accounted for a substantial amount of the variation (Table 1), suggesting an important role of historical or other unmeasured factors.

Systematic conservation planning requires more complete knowledge of biodiversity distribution and the identification of relationships between biodiversity and environmental features (Margules & Pressey, 2000; Rondinini et al., 2006; Mokany & Ferrier, 2011). Efforts to fill this gap are a priority for conservation biogeography, and must be performed in all regions and environments. Unfortunately, large areas of the globe have received limited surveys so far. Our approach allows spatially explicit estimates of unknown species richness, and can therefore guide the allocation of funding for biodiversity surveying towards specific regions, in which actual biodiversity values may be much higher than currently known.

Conservation actions are usually based on known biodiversity, with protected areas targeted to regions with high species richness, endemism and threat (Hughes et al., 2002; Brooks et al., 2006; Rondinini et al., 2006). This may bias the allocation of conservation efforts towards the best-known areas, while regions with rich biodiversity, but that are poorly studied, may remain under-represented (Bini et al., 2006). In the Western Palaearctic, remote regions can host high biodiversity, but tend to receive limited protection (Fig. 1). Our study identifies areas where the true biodiversity is probably much higher than the current known estimates. An exhaustive assessment of the network of protected areas for the conservation of reptiles is beyond the aim of this study, and should be performed at a finer spatial scale, taking into account not only species richness, but also other parameters such as the degree of endemism, the irreplaceability of areas, the presence of endangered taxa, and threatening processes (Margules & Pressey, 2000; Bini et al., 2006; Brooks et al., 2006). Our improved estimates of species richness can be integrated with the available information on endemism and threat, to identify regions where new protected areas may be required, despite biodiversity data remaining scarce (Hughes et al., 2002; Bini et al., 2006; Rondinini et al., 2006). Furthermore, integrating our broad-scale analysis with high-resolution data on habitat availability and changes in land cover may allow finer-scale information to be developed (Rondinini et al., 2011), which could help to refine the identification of conservation priorities. Some remote regions with high reptile richness had a very limited cover of protected areas, or no protected areas at all. There is a notable scarcity of protected areas in the south-west of the Arabian Peninsula (particularly in Yemen), in the east of the Iranian Plateau and in the western Maghreb (Fig. 1b), even though these regions are recognized as global biodiversity hotspots, and have a very high proportion of endemism for both reptiles and other taxa, including several microendemic reptiles recorded only in one or very few localities (Mittermeier et al., 2004; Sindaco & Jeremčenko, 2008).

ACKNOWLEDGEMENTS

We thank R. Ambrosini for statistical suggestions. J. Veech, T. Garner, D. Fontaneto, A. Aswad, S. Meiri, A. Feldman, L. Signorile and three referees provided comments on early versions of the manuscript. G.F.F. was funded by a scholarship of Università degli Studi di Milano-Bicocca.

REFERENCES


Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, G. F. Ficetola


**SUPPORTING INFORMATION**

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

**Appendix S1** Validation grid cells.

**Appendix S2** Model predictions (Figs S1–S2) and difference between projected and known richness (Fig. S3).

**Appendix S3** Results of regression models performed at the resolution of 2° × 2°.

**BIOSKETCHES**

**G. Francesco Ficetola** is a post-doc specializing in conservation and ecology of amphibians and reptiles. He combines multiple approaches (landscape ecology, macroecology, evolutionary ecology and population genetics) to obtain information that can be used to improve conservation actions.

Author contributions: All authors jointly designed the study; R.S. gathered most of the data; G.F.F. and A.B. performed the analyses; G.F.F. wrote the first version of the paper; and all authors discussed the results and commented on the draft manuscript.

Editor: Joseph Veech