



Sampling bias inverts ecogeographical relationships in island reptiles

Gentile Francesco Ficetola^{1,2*}, Massimo Cagnetta¹, Emilio Padoa-Schioppa¹, Anita Quas³, Edoardo Razzetti⁴, Roberto Sindaco⁵ and Anna Bonardi¹

¹Dipartimento di Scienze dell'Ambiente e del Territorio, e di Scienze della Terra, Università degli Studi di Milano-Bicocca, Piazza della Scienza 1, 20126 Milano, Italy, ²Laboratoire d'Ecologie Alpine (LECA), CNRS UMR 5553, Université Grenoble Alpes, BP 53, 38041 Grenoble Cedex 9, France, ³Department of Economics, Finance and Control, EMLYON Business School, 23 Avenue Guy de Collongue, 69134 ECULLY Cedex, France, ⁴Museo di Storia Naturale, Università degli Studi di Pavia, Piazza Botta 9, 27100 Pavia, Italy, ⁵Museo Civico di Storia Naturale, 10022 Carmagnola, Italy

ABSTRACT

Aim Species richness is one of the commonest measures of biodiversity, and is a basis for analyses at multiple scales. Data quality may affect estimations of species richness, but most broad-scale studies do not take sampling biases into account. We analysed reptile richness on islands that have received different sampling efforts, and assessed how inventory completeness affects the results of ecogeographical analyses. We also used simulations to evaluate under what circumstances insufficient sampling can bias the outcome of biodiversity analyses.

Location Mediterranean islands.

Methods We gathered data on reptile richness from 974 islands, assuming better sampling in islands with specific inventories. We used Moran's eigenvector mapping to analyse the factors that determine whether an island has been surveyed, and to identify the relationships between reptile richness, geographical parameters and anthropic parameters. We simulated islands, mimicking patterns of true data, and sampled them with varying effort. Simulated richness was analysed using the same approach used for real-world data.

Results The probability that islands were sampled for reptiles was higher in large, human-populated islands. The relationship between human impact and reptile richness was negative in well-surveyed islands, but positive in islands that had not been systematically surveyed, because densely populated and accessible islands receive better sampling. In simulations, analyses successfully retrieved the relationships between species richness and human presence only if the average species detection probability was $\geq 75\%$. Poorer sampling resulted in biased regression results.

Main conclusions Human activities may strongly affect biodiversity, but human presence and accessibility improve sampling effort and thus the quality of biodiversity information. Therefore, regressing known species richness on parameters representing human presence may result in apparent positive relationships. These two facets of human presence (positive on biodiversity knowledge, negative on actual biodiversity) represent a major challenge for ecogeographical studies, as not taking them into account would bias analyses and underestimate human impact.

Keywords

Accessibility, biodiversity surveys, detection probability, herpetological atlas, island biogeography, virtual ecologist, Wallacean shortfall, zero-inflated models.

*Correspondence: Gentile Francesco Ficetola, Laboratoire d'Ecologie Alpine (LECA), CNRS UMR 5553, Université Grenoble Alpes, BP 53, 38041 Grenoble Cedex 9, France.
E-mail: francesco.ficetola@gmail.com

INTRODUCTION

The analysis of species richness over broad spatial scales is a central topic in ecology and biogeography. These analyses can

help to understand the environmental factors that drive biodiversity, to evaluate the ongoing effect of human activities, and to forecast the potential impact of future environmental changes (Kerr *et al.*, 2007). Analyses of species richness over broad

geographical scales are increasingly popular due to technical advancements (e.g. geographical information systems, statistical tools suited for the analysis of spatially explicit data), but also due to the growing availability of environmental layers and distribution data, particularly in faunistic/floristic atlases and online databases (Yang *et al.*, 2013; Ficetola *et al.*, 2014). Our knowledge of species distribution remains, however, largely incomplete, even for the best-studied taxa (Wallacean shortfall) (Lomolino, 2004). Even if we assume that all the species detected in a given area are correctly identified and reported, the known species richness may be a severe underestimation of the true richness, because many species remain undetected. Analyses on plants and reptiles suggest that, in broad-scale databases, the underestimation of the true richness in a given area can exceed 30%, and underestimation is likely to be stronger in tropical regions and in less-studied, species-rich taxa (Ficetola *et al.*, 2013; Yang *et al.*, 2013).

Incomplete sampling can influence the outcome of analyses in multiple ways. In some cases, some species remain undetected at certain sites, but imperfect detection is randomly distributed across environmental patches. Under these circumstances, imperfect detection would moderately bias the results of analyses, for instance by reducing the absolute values of regression coefficients (Gu & Swihart, 2004), but random misdetection is not considered to have a major impact on the interpretation of results (Van Buskirk, 2005; but see also Foody, 2011). Unfortunately, it is unlikely that imperfect sampling would be random over broad spatial scales. If sampling is systematically biased toward areas with specific environmental features, the impact on ecogeographical analyses can be severe. For instance, regions with higher richness (e.g. strong elevation gradients) often attract more research interest (Sánchez-Fernández *et al.*, 2008; Vale & Jenkins, 2012; Yang *et al.*, 2013). More intensive sampling in mountainous regions is expected to overestimate the importance of topographic gradients on species richness, while neglecting other, relevant variables (Yang *et al.*, 2013).

Human presence can have contrasting effects on known biodiversity. On the one hand, human impact often has negative effects on many native species, reducing the richness of species that avoid human habitats. On the other hand, areas with high human presence and accessibility are most likely to receive accurate sampling and can therefore show more known species, regardless of the true species richness (Sánchez-Fernández *et al.*, 2008; Ficetola *et al.*, 2013). This sampling artefact may create misleading patterns of known biodiversity, determining underestimation of the human impact on biodiversity. In recent years, we have increasingly come to appreciate the importance of sampling accuracy on the analysis of species distribution and richness. A growing body of literature explored the effects of incomplete sampling, to identify new approaches for the analysis of biodiversity data (MacKenzie *et al.*, 2006; Miller *et al.*, 2011; Gómez-Rodríguez *et al.*, 2012). Most of these studies, however, were performed at rather fine spatial scales (e.g. regional or landscape), while less attention has been devoted to broad-scale analyses of data quality, and to the impacts on macroecological analyses (but see Sánchez-Fernández *et al.*,

2011; Foody, 2011; Kéry, 2011; Comte & Grenouillet, 2013; Yang *et al.*, 2013; Lahoz-Monfort *et al.*, 2014).

Islands are an excellent study system in which to assess the impact of data quality on species-richness analyses. Islands have attracted the attention of ecologists and biogeographers for more than 100 years, and many studies have identified the ecological, geographical and human factors that determine species richness (McArthur & Wilson, 1967; Whittaker & Fernández-Palacios, 2007). The availability of faunistic and floristic checklists for many islands has boosted biogeographical analyses, even over broad spatial scales (e.g. Kalmar & Currie, 2006; Kreft *et al.*, 2008; Triantis *et al.*, 2012). These sources often rely on a variety of surveys (Triantis *et al.*, 2012), but limited information is available on their completeness, and on the potential consequences of their quality for analyses.

The aim of this study was to evaluate the contrasting effects of human impact on species-richness data for islands. We focused on Mediterranean reptiles; reptiles are abundant vertebrates in small islands, and many data are available for this group in faunistic atlases and other publications, although the data quality is uneven. First, we identified the geographical and human factors related to the availability and quality of biodiversity data. Second, we evaluated whether different levels of sampling quality may affect the observed pattern of species richness. Specifically, we predicted a negative relationship between human impact and the true reptile richness on islands, but non-negative or positive relationships between richness and human impact if sampling is not evenly distributed, because islands that are easily accessible or with greater human presence are better sampled. Finally, we used the ‘virtual ecologist’ approach (Zurell *et al.*, 2010) to simulate islands suffering different human impacts. We evaluated how human impact may interact with sampling quality, and identified the conditions under which poor sampling may bias the estimation of the relationship between species richness and human impact.

MATERIALS AND METHODS

Study area and datasets

We analysed the islands of the Mediterranean basin and the adjacent Macaronesia (Fig. 1). We considered three geographical features that are known to affect reptile richness on islands – area, isolation and maximum elevation (McArthur & Wilson, 1967; Ficetola & Padoa-Schioppa, 2009) – and three variables that represent human impact – total human population, the presence of an airport, and naval connections. Airport presence is strongly related to the accessibility of islands for people, and is also a good indicator of economic development, trading exchanges and tourism (Green, 2007). Isolation was measured as the distance from the continent or from the nearest large island (Corsica, Sardinia, Sicily or Crete). Area, elevation, human population (in the period 2003–2007), naval connections and airport presence were mostly obtained from Arnold’s (2008) database of Mediterranean islands, and from Ficetola & Padoa-Schioppa (2009). Additional data were obtained from

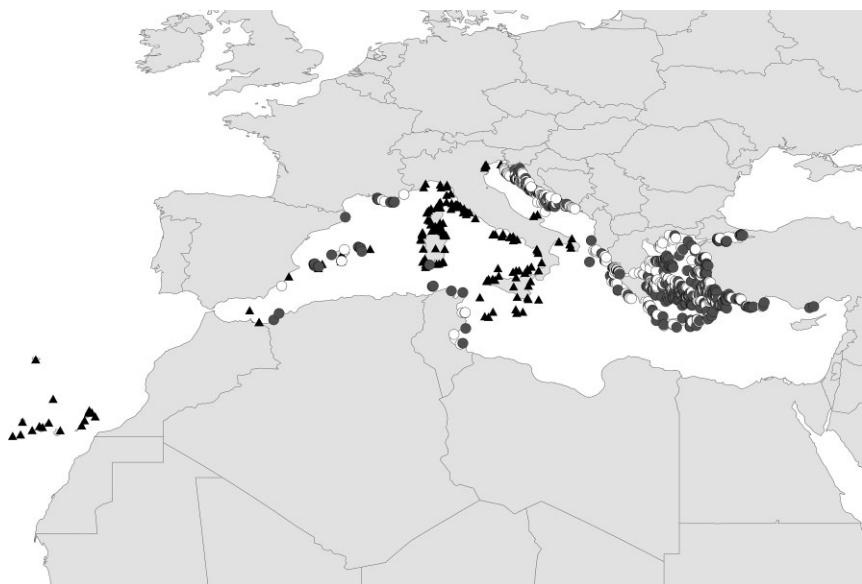


Figure 1 Study area and distribution of the 974 analysed Mediterranean and Macaronesian islands. Black triangles, islands in the ‘atlas’ dataset; open circles, islands in the ‘no-atlas’ dataset for which we did not find reptile records; grey circles, islands of the ‘no-atlas’ dataset with at least one reptile record.

national statistical databases, atlases and geographical publications. Naval connections were coded as follows: 0, no regularly scheduled connections with the mainland or with other islands; 1, ≤ 7 connections; 2, > 7 connections per week (Arnold, 2008). We only considered islands with surface area greater than 1 ha; islands with this area can host large populations of reptiles (e.g. Pérez-Mellado *et al.*, 2008).

Land use and climate are additional and potentially important parameters. We calculated the percentage cover of non-natural habitats (urban and cropland) as an additional measure of human impact, on the basis of the 300-m resolution GlobCover land-cover map (Bicheron *et al.*, 2008). Potential evapotranspiration (PET) is a climatic parameter that represents the amount of atmospheric energy, and is strongly related to reptile species richness (Rodríguez *et al.*, 2005). We extracted PET from Trabucco & Zomer (2009). Land-use and PET were only available for a subset of islands with relatively large area, because of their relatively coarse resolution.

We considered two different sources of data for reptile distributions, resulting in two distinct datasets. First, herpetological atlases have been published for several countries/regions (Appendix S1 in Supporting Information). These sources include specific sections on island biodiversity, representing the summary of multiple, intensive surveys, and provide detailed accounts of reptile communities on a number of islands. We thus assume that the islands described in detail in these atlases (‘atlas’ dataset) represent islands for which high-quality data are available.

For the remaining islands (‘no-atlas’ islands), we searched for information in a variety of sources, including published papers, the grey literature, publications summarizing the grey literature, and direct surveys by the authors (Sindaco & Jeremčenko, 2008; Sindaco *et al.*, 2013, and references in Appendix S1). Although a herpetological guide of Greece is available (Valakos *et al.*, 2008), it does not detail the species composition on the islands, and

Greek islands were therefore included in the ‘no-atlas’ dataset. The islands in this dataset probably received variable survey effort and, for many of them (about 50%; see Results), we found no information on reptile communities. For each island, we then calculated the number of known native reptile species (see Appendix S1).

Analysis of reptile data

In both datasets, environmental features (area, isolation, elevation, human population, naval connections and PET) and species richness showed spatial autocorrelation. Autocorrelation may be related to different biogeographical processes occurring among archipelagos and may affect regression analyses. In preliminary analyses, the residuals of ordinary least-squares regressions were significantly autocorrelated ($P < 0.05$; Fig. S1 in Appendix S2), so we integrated autocorrelation into the models, and analysed the data using Moran’s eigenvector maps (MEM). MEM allows the spatial arrangement of data points to be translated into explanatory variables (eigenvectors) that capture spatial effects (Dray *et al.*, 2006; Thayn & Simanis, 2013). We used a permutation test to evaluate the autocorrelation of residuals for all models. We generated MEMs on the basis of geographical distances, using the minimum distance that keeps all sites connected (3° ; Borcard & Legendre, 2002). We then selected the eigenvectors that best reduce the spatial autocorrelation of residuals, and included them as spatial predictors in generalized linear models, until residual autocorrelation was not significant any more ($P > 0.05$ after a permutation test) (Griffith & Peres-Neto, 2006; Dormann *et al.*, 2007). Comparisons among statistical methods showed that this implementation of MEM is flexible and efficient, even when analysing non-normal data, and provides results that are congruent with other spatially explicit techniques (Dormann *et al.*, 2007; Siesa *et al.*, 2011; Ficetola *et al.*, 2013; Thayn & Simanis, 2013). We performed the MEM

analysis using the ‘ME’ function in the R package SPDEP (Griffith & Peres-Neto, 2006; Bivand *et al.*, 2013).

If surveyed, even the most isolated islands of the Mediterranean, and all the islands with an area greater than 1 ha revealed the presence of some reptile species (e.g. Pérez-Mellado *et al.*, 2008). We assumed that an island received some sampling if it has at least one observation of either native or non-native species. We used MEM (binomial error distribution) to identify the factors that are most likely to determine whether an island has been surveyed. For the ‘no-atlas’ dataset, the presence or absence of reptile records was considered as the dependent variable, whereas the six environmental variables (area, isolation, elevation, human population, airport presence and naval connections) were the independent variables. We used Akaike’s information criterion (AIC) to compare candidate models, including all the possible combinations of independent variables. AIC trades off explanatory power against the number of predictors; parsimonious models explaining more variation have low AIC values and are considered to be the ‘best AIC’ models (Symonds & Moussalli, 2011). AIC may select overly complex models; we therefore considered a complex model only if it showed a lower AIC than all its simpler nested models (Richards *et al.*, 2011). For each candidate model, we also calculated the AIC weight, which measures the relative support of models (Symonds & Moussalli, 2011).

We then used MEM to relate the observed richness of native reptiles to the geographical and human variables. To compare the outcome of analyses performed on high-quality data, with analyses performed on data with varying quality, this analysis was repeated four times: (1) on all the islands that received some sampling, i.e. for which at least one species (either native or non-native) was recorded; (2) to test whether the relationship between human impact and observed richness was different between the ‘atlas’ and ‘no-atlas’ datasets, for which we repeated the analysis of the full dataset, including the interaction between data quality (‘atlas’/‘no-atlas’) and variables representing human impact; (3) on the ‘atlas’ dataset only; (4) on the islands of the ‘no-atlas’ dataset which had received some sampling.

Human population was strongly correlated to naval connections (Pearson’s $r = 0.84$). As highly correlated variables may bias regression results, we first ran analyses using human population and excluding naval connections, and then repeated them including naval connections and excluding human population. Finally, to evaluate the potential role of climate and land use, we repeated analyses for the subset of islands for which these data were available. Models were built with Poisson-distributed errors or, if the residual deviance was strongly different from residual degrees of freedom, with quasi-Poisson errors.

When necessary, variables were transformed using logarithm (island area, isolation, elevation and human population), square-root (PET) or arcsine-square-root (% land-use) transformations. We tested significance using likelihood-ratio tests. We detected no collinearity issues (variance-inflation factor always < 5). Models were run in R 3.0; partial regression plots were built using the package VISREG (R Core Team, 2013).

Analysis of simulated data

The analysis of virtual data that simulate ecological processes and sampling is a powerful framework, allowing the effect of sampling and analytical methods to be assessed (Zurell *et al.*, 2010). We simulated data on species richness in islands, assuming that the true richness was influenced by biogeographical variables and negatively affected by human impact. Species richness was then virtually sampled with varying effort, assuming a better sampling in human-dominated islands. Simulated data of observed richness were then analysed using the same approach used for real-world data. This allowed us to evaluate the consequences of poor sampling: under what circumstances does better sampling in human-dominated islands outweigh the negative effects of human impact? How good should sampling be in order to obtain reliable estimates of the true relationships between species richness and environmental variables?

In each simulation run, we considered a set of 300 islands. For each island, we generated variables representing environmental features, true species richness and the number of species detected. With regard to environmental features, we focused on area, isolation and presence/absence of airport, as these were the most important independent variables in explaining reptile distribution in the analysis of real-world data. Airport presence was modelled as a function of island area, as larger real islands are more likely to have an airport. The simulated number of species effectively present in each island, S_{Ns} , was generated from a Poisson distribution with parameter equal to the lambda value predicted by the regression model for the islands in the ‘atlas’ dataset (Table 1c).

We then generated the number of species detected in each island, S_{Nd} . To simulate different sampling intensities across islands, we computed for each island the variable Q , representing the probability of species detection, and therefore sampling quality. Q is related to environmental variables, as shown in eqn. 1, built from real-world results:

$$Q = \frac{1}{e^{-(K + \beta_1 \times S_{\text{airport}} + \beta_2 \times S_{\log.\text{area}})} + 1} \quad (1)$$

β_1 and β_2 are the coefficients of the regression model that relate the presence of reptile observations to environmental variables (Table 2a). Sampling quality, Q , varies between 0 and 1, and is better in large islands and in islands hosting an airport (see Results). K is a constant which differs across simulation cycles, allowing variation in the average sampling effectiveness across all the islands to be represented; higher K determines better sampling. In real data, the presence of a herpetological atlas leads to higher K . In an island hosting S_{Ns} species, each species has a different probability of being detected. The detection probability of each species (p) in an island was drawn from a beta distribution with mean of Q , and variance of 0.01. S_{Nd} was the number of species detected in an island hosting S_{Ns} species, and was calculated as the sum of S_{Ns} Bernoulli trials, each with probability of success p .

Table 1 Poisson regression models relating species richness of native reptiles to geographical and human variables of Mediterranean and Macaronesian islands. (a) Model built for all the islands for which at least one species was recorded. (b) Model built for the same islands used in (a), also including the interaction between data quality ('atlas'/'no-atlas') and the variables representing human impact (human population and airport). (c) Model built for the islands of the 'atlas' dataset. (d) Model built for the islands of the 'no-atlas' dataset for which at least one species was recorded. Correlograms of residual spatial autocorrelation are shown in Appendix S2, Fig. S1.

Independent	B	χ^2	d.f.	P
(a) all islands with data				
$R^2_D = 0.61$; residual autocorrelation: $P = 0.12$				
Area	0.29	41.82	1	< 0.001
Isolation	-0.21	51.71	1	< 0.001
Elevation	0.18	8.15	1	0.004
Human population	0.09	13.74	1	< 0.001
Airport	0.02	0.05	1	0.829
MEM eigenvectors		42.97	2	< 0.001
(b) all islands; interactions between data quality and human impact				
$R^2_D = 0.68$; residual autocorrelation: $P = 0.07$				
Area	0.36	72.41	1	< 0.001
Isolation	-0.12	18.47	1	< 0.001
Elevation	0.09	2.57	1	0.109
Human population	0.13	12.29	1	< 0.001
Airport	0.08	0.02	1	0.889
Human population × data quality	-0.16	30.21	1	< 0.001
Airport × data quality	-0.53	9.45	1	0.002
Data quality*	0.27	16.79	1	< 0.001
MEM eigenvectors		21.95	1	< 0.001
(c) Islands on atlas [†]				
$R^2_D = 0.60$; residual autocorrelation: $P = 0.17$				
Area	0.43	80.74	1	< 0.001
Isolation	-0.11	5.03	1	0.025
Elevation	0.12	3.00	1	0.083
Human population	-0.08	5.80	1	0.016
Airport	-0.48	11.83	1	< 0.001
MEM eigenvectors		58.18	3	< 0.001
(d) Islands not on atlas				
$R^2_D = 0.71$; residual autocorrelation: $P = 0.20$				
Area	0.26	16.31	1	< 0.001
Isolation	-0.13	12.95	1	< 0.001
Elevation	0.11	1.56	1	0.212
Human population	0.18	31.14	1	< 0.001
Airport	0.11	2.33	1	0.127
MEM eigenvectors		8.46	1	0.003

R^2_D : proportion of null-deviance explained by the model.

*Coded as follows: 0, 'no-atlas' dataset; 1, 'atlas' dataset.

†Quasi-Poisson error distribution.

We subsequently evaluated whether regression can successfully estimate the relationships between species richness and the three environmental predictors (area, isolation and airport presence), by relating the detected species richness S_{Nd} to the predictors. For each simulation, we built the regression models, and recorded the model coefficients as well as the average values

of Q and S_{Nd}/S_{Ns} , which represents the proportion of species successfully detected. We ran simulations with five values of K (-2, -1, 0, 1 and 2), and 300 replicates per each value of K . A complete description of the simulation methods is reported in Appendix S2; the R code of simulations is available in Appendix S3.

RESULTS

Mediterranean island data

We obtained data from 974 islands in 11 countries (Fig. 1). The 'atlas' dataset included 217 islands, and the 'no-atlas' dataset included 757 islands. The 'no-atlas' islands were mostly concentrated in the Eastern Mediterranean basin (72%), although several small islands of the Western Mediterranean were not detailed in herpetological atlases (Fig. 1). Atlas islands tended to be less isolated, smaller, with lower maximum elevation and larger human populations than no-atlas islands (*t*-test, all $P \leq 0.01$), whereas the frequency of ferry connections was similar between the two datasets (Mann–Whitney test, $P = 0.35$). There was nonetheless a wide overlap in geographical features and human impact between the two datasets, and the atlas dataset included the largest and most isolated islands (Appendix S2, Fig. S2).

Which islands are more likely to receive reptile surveys?

Out of the 757 islands of the 'no-atlas' dataset, we found at least one species record in 372 islands (see Appendix S4 for a list of islands for which we found no reptile records). The best AIC model indicates that the probability of having at least one recorded species was positively related to island area, airport presence and human population (Table 2). Three eigenvectors were also included into the model. They represent significant spatial autocorrelation (Table 2), and reduced residual autocorrelation to non-significant values (Moran's $I = 0.005$, $P = 0.76$). Some models with lower AIC weights did not include human population, and included a negative relationship with elevation (Table 2b). Area and airport presence were included in all candidate models with high support.

Analysis of observed species richness

When all the islands with at least one record were analysed (589 islands), species richness clearly increased with island area and elevation, and was negatively related to isolation. The observed richness was not related to airport presence, but was positively related to human population. Two eigenvectors were also included into the model, to take spatial autocorrelation into account (Table 1a).

We detected significant interactions between human population, airport presence and dataset of origin ('atlas' vs. 'no-atlas'); the relationship between reptile richness and the variables representing human impact was significantly more negative in the 'atlas' dataset than in the 'no-atlas' dataset (significant

Independent variables	B	χ^2	d.f.	P	K	AIC	w	R^2_D
(a) best-AIC model					7	811.5	0.63	0.24
Area	0.55*	12.20	1	< 0.001				
Human population	0.27	5.48	1	0.020				
Airport	15.37†	7.65	1	0.006				
MEM eigenvectors	147.64	3		< 0.001				
(b) Alternative models								
Area (+), airport presence (+), elevation (-)			7		814.1	0.17	0.24	
Area (+), airport presence (+)			6		815.0	0.11	0.23	
Area (+), human population (+), elevation (-)			7		817.1	0.04	0.23	
Area (+), human population (+)			6		817.1	0.04	0.23	
Airport presence (+), isolation (+), human population (+)			7		821.6	0.004	0.23	
Airport presence (+), human population (+)			6		821.7	0.004	0.23	
Area (+), elevation (-)			6		822.3	0.003	0.23	
Area (+)			5		823.6	0.001	0.22	
Isolation (+), human population (+)			6		828.4	< 0.001	0.22	
Human population (+)			5		828.6	< 0.001	0.22	
Airport presence (+), elevation (+), naval connections (+)			7		840.0	< 0.001	0.21	

*Used as β_2 in eqn. 1.

†Used as β_1 in eqn. 1.

Table 2 Binomial spatial eigenvector models relating presence/absence of reptile records on Mediterranean and Macaronesian islands to human and geographical variables. (a) Best model, selected on the basis of Akaike's information criterion (AIC); (b) environmental variables included in the other candidate models. Only the 12 models with lowest AIC are reported here; all these models also include MEM eigenvectors. K, number of parameters in the model; w, AIC weight; R^2_D , proportion of null deviance explained by the model.

interactions in Table 1b). When only the 'atlas' islands were analysed, richness was negatively related to human population and airport presence. Parameters representing human impact therefore showed opposite regression coefficients, compared to the results of the full dataset (Table 1c, Fig. 2). Conversely, when we analysed only the 'no-atlas' islands with at least one presence record (372 islands), the relationships between species richness and parameters representing human impact were similar to the analysis of the full dataset. The relationship between reptile richness and human population was positive, whereas the relationship with airport presence was not significant (Table 1d, Fig. 2). Spatial eigenvectors were included in all the models, and successfully reduced residual spatial autocorrelation to non-significant values (Table 1; Appendix S2, Fig. S1).

We obtained analogous results for models using naval connections instead of human population (Table S1 in Appendix S2). The relationships between richness and variables representing human impact (naval connections and airport presence) were generally positive both in the analysis of the full dataset, and in the analysis of the 'no-atlas' islands, but the interaction between naval connections and dataset was strongly significant (Table S1b), because the relationship between reptile richness and naval connections was weakly negative for the islands of the 'atlas' dataset (Table S1).

We obtained data on land use and climate (PET) for 307 islands ('atlas', 78 islands; 'no-atlas', 229 islands). After taking into account the other variables, reptile richness was not significantly related to either land use or PET (Appendix S2, Table S2). It should nonetheless be remarked that data on land use and climate were only available for the largest islands. For instance, in this dataset the median area was 9.11 km^2 , whereas the median area of the global dataset was 0.65 km^2 .

Simulations

The five values of K considered allow potential variation in sampling quality (Fig. 3a) to be well represented. When $K = -2$, only 21% of present species were detected; the proportion of detected species rose to 54% with $K = 0$, and to 87% with $K = 2$ (Fig. 3a). The results of regression models were strongly affected by sampling quality. For both island area and isolation, coefficients were close to the actual values when $K \geq 1$ (i.e. high sampling quality and c. 75% of species or more detected), whereas coefficient estimates were less accurate with lower values of K (Fig. 3b,c), although the sign of the coefficient remained unchanged for both area and isolation. The coefficient of the relationship between airport presence and species richness was strongly affected by changes to K. The relationship between airport presence and richness was positive if $K = -2$, and approached the real, negative value only when $K \geq 1$ (Fig. 3d). Poisson regression coefficients were therefore strongly affected by sampling bias, and the relationship with variables representing human presence may be inverted when bias was strong.

DISCUSSION

Human presence, data quality and known richness

Species richness is one of the commonest measures of biodiversity, and is a basis of ecological analyses at multiple scales. However, the outcome of species-richness analyses may be severely affected by data quality. Human presence, accessibility and attractiveness of areas are strongly related to sampling effort and the quality of biodiversity information (Sánchez-Fernández

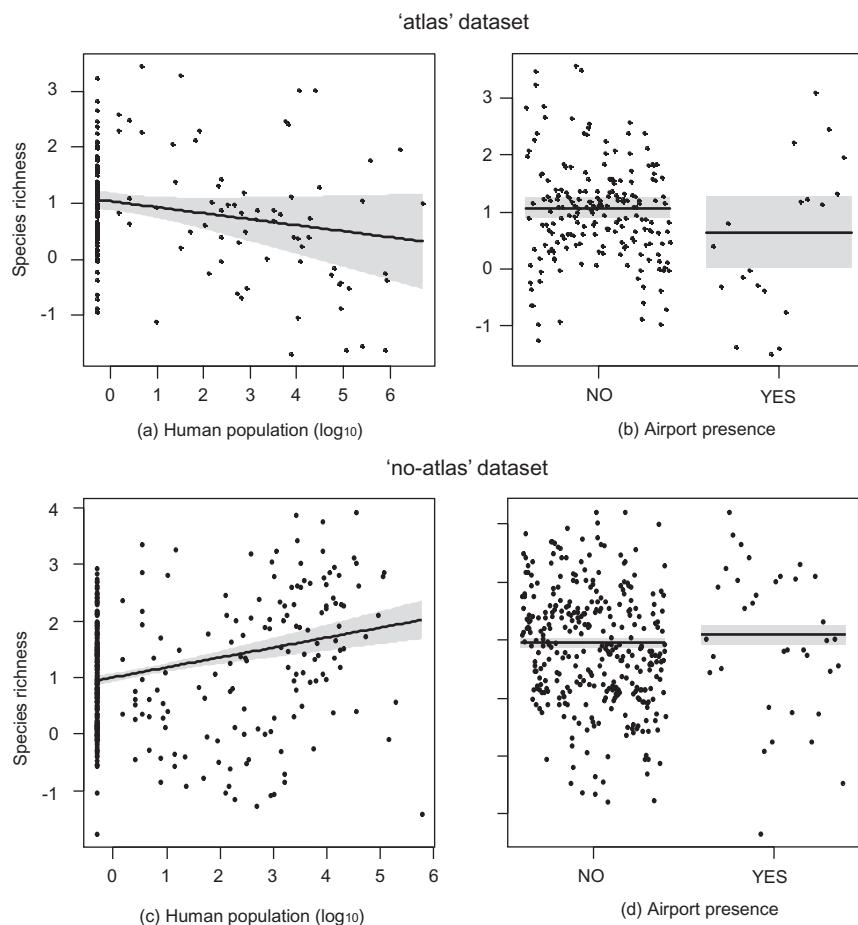


Figure 2 Partial residual plots showing the relationship between species richness of reptiles and variables representing human impact (total human population and airport presence) for Mediterranean and Macaronesian islands included ('atlas' dataset) and not included ('no-atlas' dataset) in faunistic/herpetological atlases. Shaded areas represent 95% confidence bands.

et al., 2008; Kent & Carmel, 2011; Ficetola *et al.*, 2013; Yang *et al.*, 2013). Across the Mediterranean, sampling is clearly better in islands with good transport connections and housing large human populations. As a result, regressing species richness against parameters representing human presence (e.g. human population, airport presence and naval connections) may result in apparently positive relationships (Table 1, Fig. 2c–d), even though human impact is likely to have negative effects (Table 1c, Fig. 2a–b). These two facets of human impact (positive on biodiversity knowledge; likely to be negative on actual biodiversity) represent a major challenge for ecogeographical studies, as not taking them into account would result in biased analyses and misunderstanding of the true human impact.

This interpretation of our results is supported by simulation analyses. Low sampling quality determines less accurate estimates for all relationships (e.g. inflated error in the relationship between isolation and richness; Fig. 3c), but the bias is particularly relevant for parameters related to both richness and detection probability, such as airport presence (Fig. 3d). Actually, the negative effects of human impact in our simulations could only be successfully assessed if island communities are well-sampled, and at least 75% of species are detected (Fig. 3). We do not know the actual level of completeness of surveys in the Mediterranean islands of our datasets. Richness underestimations of more than

50% are not infrequent in poorly-surveyed localities (Ficetola *et al.*, 2013; Yang *et al.*, 2013), but survey quality was not evenly distributed across islands in our real-world dataset. Even for the 'no-atlas' dataset, some islands received repeated focused surveys, whereas others have only been occasionally sampled.

Some studies suggested that positive relationships between biodiversity and human presence may arise because certain climates (e.g. warmer, with more precipitation) favour both human activities and biodiversity (Stohlgren *et al.*, 2006). However, this effect is unlikely to influence our results, because islands in our set span a limited latitudinal range (Fig. 1), and share a similar Mediterranean climate. In fact, observed species richness was not related to evapotranspiration, and the interactions between sampling quality and human impact remained evident when taking climate into account (Table S2). A positive correlation between reptile richness and human presence might also arise because humans settle on islands where water is available and with more habitats, and these parameters can also positively affect species richness (Kadmon & Allouche, 2007). Our models nonetheless considered island area and elevation, which strongly correlate with habitat availability and diversity (Davidar *et al.*, 2002; Kreft *et al.*, 2008). Furthermore, the positive correlation was only observed for the 'no-atlas' islands, supporting the role of sampling quality.

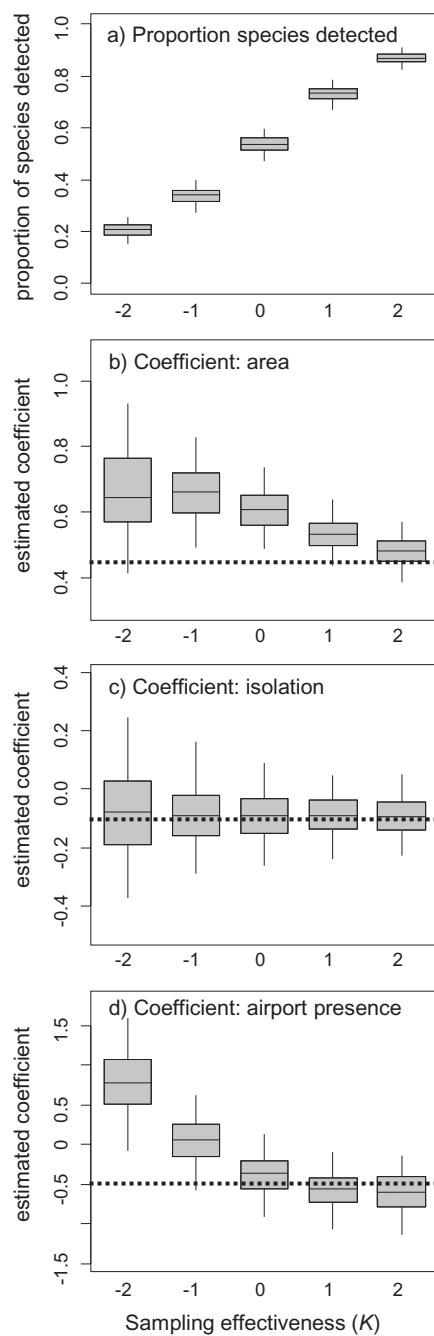


Figure 3 Results of simulations: effect of different values of sampling effectiveness (K) on (a) average proportion of reptile species detected in Mediterranean and Macaronesian islands; coefficients of Poisson regression relating observed species richness to (b) $\log(\text{area})$, (c) $\log(\text{isolation})$ of the islands and (d) airport presence. Box include the second and the third quartile of simulation results; error bars are 95% CI, the bold dotted lines represent the true parameters used to generate the data.

Human impact and reptile richness

For the ‘atlas’ islands, the relationships between species richness and both human population and airport presence were clearly

negative, suggesting a strong impact on reptile communities (Table 2, Fig. 2a–b). Human activities in the Mediterranean basin date back to the Neolithic, and their effects extend to most of the islands. The introduction of alien species is likely to be the factor with the strongest negative effects on native reptiles (Corti *et al.*, 1999). Many alien species have been introduced to Mediterranean islands in historical times, but the rate of introductions is quickly rising due to increasing commercial exchanges (Pinya & Carretero, 2011). Exhaustive data on trade intensity were not available, so we used airport presence and naval connections as proxies of tourism and trade between the islands and the continents. Airport traffic is a major predictor of economic development (Green, 2007), and airport presence is one of the strongest predictors of the richness of alien reptiles on islands (Ficetola & Padoa-Schioppa, 2009). Other human impacts may have weaker effects on reptiles. Many Mediterranean reptiles are well-adapted to traditional agriculture, and may therefore tolerate the partial loss of natural habitats. Nevertheless, recent urbanization poses the greatest threat, and some endemic reptiles are actually endangered by habitat loss and degradation. Finally, persecution by humans may have locally important effects, particularly on snakes (Mullin & Seigel, 2009).

How can we improve broad-scale ecogeographical analyses?

Given the heterogeneous quality of broad-scale distribution data, how can we improve their use for ecological analysis? One potential approach is to identify localities for which data quality is sufficiently high, and to focus analyses on well-surveyed sites (Yang *et al.*, 2013). This is the approach we used here, as we assumed an adequate quality for the ‘atlas’ dataset. Nevertheless, this approach remains subjective, and we did not have direct control of the actual quality for many of the islands in the ‘atlas’ dataset. Even if these islands generally received repeated and intensive sampling, information could be incomplete for some of them. Alternative and more objective approaches include using accumulation curves to estimate the completeness of data, but this would require access to the original data, which is rarely available for broad-scale databases. Furthermore, estimates obtained through accumulation curves may not be consistently reliable (Sánchez-Fernández *et al.*, 2011; Yang *et al.*, 2013).

A different approach would be to obtain estimates of sampling effort (e.g. accessibility to researchers, number of known surveys performed), and integrate them into models to improve the results of analyses (Sánchez-Fernández *et al.*, 2008; Gómez-Rodríguez *et al.*, 2012; Ficetola *et al.*, 2013). These measures were not available for all the islands in our dataset – for instance, accessibility maps have a resolution of 1 km² (Nelson, 2008), and many islands are smaller than this – but could be successfully integrated into studies that focus on larger spatial units. Finally, some analytical techniques have been proposed for taking the complex effects of variables on detection probability and richness into account, such as zero-inflated Poisson models (ZIP; Zuur *et al.*, 2009). Even with ZIP, however,

we were unable to successfully capture the negative relationships between human presence and species richness. Indeed, the ZIP results were no better than those of standard regression models (Appendix S2, Table S3, Fig. S3).

Even if there is general awareness that species distribution data are rarely complete (Lomolino, 2004), the impact of this underestimation on biodiversity analysis remains underappreciated (Kéry, 2011). Uneven geographical sampling can, however, strongly affect patterns of species richness and may even invert ecogeographical relationships if the same factors is related to both sampling quality and actual species richness. Insufficient sampling can have consequences both for ecological understanding and practical conservation planning. For instance, misjudging the effects of human impact may result in a lack of management actions. In the last decade, a growing body of literature has investigated the effects of imperfect detection on local-scale studies, although this issue has attracted less attention in the macro-ecological literature. More research on this theme will allow a better understanding of broad-scale ecological patterns, with potential consequences on biodiversity conservation.

ACKNOWLEDGEMENTS

We thank J. A. Diniz-Filho and two reviewers for comments on the manuscript. C. Arnold provided his database of Mediterranean Islands. G.F.F. was funded by a scholarship of University Milano-Bicocca.

REFERENCES

- Arnold, C. (2008) *Mediterranean islands*. Survival Books, London.
- Bicheron, P., Huc, M., Henry, C., Bontemps, S. & GLOBCOVER partners (2008) *GLOBCOVER products description manual*. Medias France, European Space Agency, Toulouse.
- Bivand, R., Altman, M., Anselin, L. et al. (2013) *spdep: spatial dependence: weighting schemes, statistics and models*. R package version 0.5-62. Available at: <http://cran.r-project.org/web/packages/spdep/>
- Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Comte, L. & Grenouillet, G. (2013) Species distribution modelling an imperfect detection: comparing occupancy versus consensus methods. *Diversity and Distributions*, **19**, 996–1007.
- Corti, C., Massetti, M., Delfino, M. & Pérez-Mellado, V. (1999) Man and herpetofauna of the mediterranean islands. *Revista Española de Herpetología*, **13**, 83–100.
- Davidar, P., Yoganandan, K., Ganesh, T. & Devy, S. (2002) Distribution of forest birds and butterflies in Andaman islands, Bay of Bengal: nested patterns and process. *Ecography*, **25**, 5–16.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Dray, S., Legendre, L. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.
- Ficetola, G.F. & Padoa-Schioppa, E. (2009) Human activities alter biogeographical patterns of reptiles on Mediterranean islands. *Global Ecology and Biogeography*, **18**, 214–222.
- Ficetola, G.F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. (2013) Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography*, **40**, 1202–1211.
- Ficetola, G.F., Rondinini, C., Bonardi, A., Katariya, V., Padoa-Schioppa, E. & Angulo, A. (2014) An evaluation of the robustness of global amphibian range maps. *Journal of Biogeography*, **41**, 211–221.
- Foody, G.M. (2011) Impacts of imperfect reference data on the apparent accuracy of species presence-absence models and their predictions. *Global Ecology and Biogeography*, **20**, 498–508.
- Gómez-Rodríguez, C., Bustamante, J., Díaz-Paniagua, C. & Guisan, A. (2012) Integrating detection probabilities in species distribution models of amphibians breeding in Mediterranean temporary ponds. *Diversity and Distributions*, **18**, 260–272.
- Green, R.K. (2007) Airports and economic development. *Real Estate Economics*, **35**, 91–112.
- Griffith, D.A. & Peres-Neto, P.R. (2006) Spatial modelling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, 2603–2613.
- Gu, W.-D. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195–203.
- Kadmon, R. & Allouche, O. (2007) Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: a unification of island biogeography and niche theory. *The American Naturalist*, **170**, 443–454.
- Kalmar, A. & Currie, D.J. (2006) A global model for island biogeography. *Global Ecology and Biogeography*, **15**, 72–81.
- Kent, R. & Carmel, Y. (2011) Presence-only versus presence-absence data in species composition determinant analyses. *Diversity and Distributions*, **17**, 474–479.
- Kerr, J.T., Kharouba, H.M. & Currie, D.J. (2007) The macroecological contribution to global change solutions. *Science*, **316**, 1581–1584.
- Kéry, M. (2011) Towards the modelling of true species distributions. *Journal of Biogeography*, **38**, 617–618.
- Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. (2008) Global diversity of island floras from a macroecological perspective. *Ecology Letters*, **11**, 116–127.
- Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014) Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.

- Lomolino, M.V. (2004) Conservation biogeography. *Frontiers of biogeography: new directions in the geography of nature* (ed. by M.V. Lomolino and L.R. Heaney), pp. 293–296. Sinauer Associates, Sunderland, MA.
- McArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.A. & Hines, J.E. (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press, Burlington, MA.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Campbell Grant, E.H., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.
- Mullin, S.J. & Seigel, R.A. (eds) (2009) *Snakes: ecology and conservation*. Cornell University Press, Ithaca, NY.
- Nelson, A. (2008) *Travel time to major cities: a global map of accessibility*. Office for Official Publications of the European Communities, Luxembourg.
- Pérez-Mellado, V., Hernández-Estévez, J., García-Díez, T., Terrassa, B., Misericordia Ramón, M., Castro, J., Picornell, A., Martín-Vallejo, J. & Brown, R. (2008) Population density in *Podarcis lilfordi* (Squamata, Lacertidae), a lizard species endemic to small islets in the Balearic Islands (Spain). *Amphibia-Reptilia*, **29**, 49–60.
- Pinya, S. & Carretero, M.A. (2011) The Balearic herpetofauna: a species update and a review on the evidence. *Acta Herpetologica*, **6**, 59–80.
- R Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Richards, S.A., Whittingham, M.J. & Stephens, P.A. (2011) Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, **65**, 77–89.
- Rodríguez, M., Belmontes, J.A. & Hawkins, B.A. (2005) Energy, water and large-scale patterns of reptile and amphibian species richness in Europe. *Acta Oecologica*, **28**, 65–70.
- Sánchez-Fernández, D., Lobo, J.M., Abellán, P., Ribera, I. & Millán, A. (2008) Bias in freshwater biodiversity sampling: the case of Iberian water beetles. *Diversity and Distributions*, **14**, 754–762.
- Sánchez-Fernández, D., Lobo, J.M., Abellán, P. & Millán, A. (2011) How to identify future sampling areas when information is biased and scarce: an example using predictive models for species richness of Iberian water beetles. *Journal for Nature Conservation*, **19**, 54–59.
- Siesa, M.E., Manenti, R., Padoa-Schioppa, E., De Bernardi, F. & Ficetola, G.F. (2011) Spatial autocorrelation and the analysis of invasion processes from distribution data: a study with the crayfish *Procambarus clarkii*. *Biological Invasions*, **13**, 2147–2160.
- Sindaco, R. & Jeremčenko, V.K. (2008) *The reptiles of the Western Palearctic, vol. 1. Annotated checklist and distributional atlas of the turtles, crocodiles, amphisbaenians and lizards of Europe, North Africa, Middle East and Central Asia*. Belvedere, Latina, Italy.
- Sindaco, R., Venchi, A. & Grieco, C. (2013) *The reptiles of the Western Palearctic, vol. 2. Annotated checklist and distributional atlas of the snakes of Europe, North Africa, Middle East and Central Asia*. Belvedere, Latina, Italy.
- Stohlgren, T.J., Barnett, D.T., Flather, C., Fuller, P., Peterjohn, B., Kartesz, J. & Master, L.L. (2006) Species richness and pattern of invasion in plants, birds, and fishes in the United States. *Biological Invasions*, **8**, 427–447.
- Symonds, M.R.E. & Moussalli, A. (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, **65**, 13–21.
- Thayn, J.B. & Simanis, J.M. (2013) Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. *Annals of the Association of American Geographers*, **103**, 47–66.
- Trabucco, A. & Zomer, R.J. (2009) *Global potential evapotranspiration (Global-PET) and global aridity index (Global-Aridity) geo-database*. CGIAR Consortium for Spatial Information, Washington, DC. Available at: <http://www.cgiar.org/>
- Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species-area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215–231.
- Valakos, E.D., Pafilis, P., Sotiropoulos, K., Lymberakis, P., Maragou, P. & Foufopoulos, J. (2008) *The amphibians and reptiles of Greece*. Chimaira, Frankfurt am Main.
- Vale, M.M. & Jenkins, C.N. (2012) Across-taxa incongruence in patterns of collecting bias. *Journal of Biogeography*, **39**, 1744–1748.
- Van Buskirk, J.V. (2005) Local and landscape influence on amphibian occurrence and abundance. *Ecology*, **86**, 1936–1947.
- Whittaker, R.J. & Fernández-Palacios, J.M. (2007) *Island biogeography*. Oxford University Press, Oxford.
- Yang, W.-Y., Ma, K. & Kreft, H. (2013) Geographical sampling bias in a large distributional database and its effects on species richness–environment models. *Journal of Biogeography*, **40**, 1415–1426.
- Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 References used as source for species distribution in the ‘atlas’ and in the ‘no-atlas’ datasets, and for the identification of non-native species.

Appendix S2 Additional analyses: correlograms of residual autocorrelation; complete description of methods used for simulations; results of regression models considering naval connections, land-use and climate; analyses using zero-inflated Poisson regression.

Appendix S3 R code used to run simulations.

Appendix S4 List of islands for which no reptile records were found.

BIOSKETCH

G. Francesco Ficetola is a researcher combining multiple approaches (landscape ecology, macroecology, evolutionary ecology) to improve knowledge of the ecology and conservation of amphibians and reptiles.

Author contributions: G.F.F., A.B., E.P.-S and E.R. designed the study; A.B., M.C., E.R., R.S and G.F.F. gathered species data; G.F.F., A.B., M.C and A.Q. performed the analyses; G.F.F. wrote the first draft of the manuscript. All authors commented on and approved the final manuscript version.

Editor: José A. F. Diniz-Filho