ORIGINAL PAPER

Evaluating the sampling bias in pattern of subterranean species richness: combining approaches

**Maja Zagmajster · David C. Culver · Mary C. Christman ·
Boris Sket**

Abstract We investigated the pattern of species richness of obligate subterranean (troglobiotic) beetles in caves in the northwestern Balkans, given unequal and biased sampling. On the regional scale, we modeled the relationship between species numbers and sampling intensity using an asymptotic Clench (Michaelis–Menten) function. On the local scale, we calculated Chao 2 species richness estimates for 20×20 km grid cells, and investigated the distribution of uniques, species found in only one cave within the grid cell. Cells having high positive residuals, those with above average species richness than expected according to the Clench function, can be considered true hotspots. They were nearly identical to the observed areas of highest species richness. As sampling intensity in a grid cell increases the expected number of uniques decreases for any fixed number of species in the grid cell. High positive residuals show above average species richness for a certain level of sampling intensity within a cell, so further sampling has the most potential for additional species. In some cells this was supported by high numbers of uniques, also indicating insufficient sampling. Cells with low negative residuals have fewer species than

---

M. Zagmajster (✉) · B. Sket
Oddelek za biologijo, Biotehniška fakulteta, Univerza v Ljubljani,
Večna pot 111, p.p. 2995, 1000 Ljubljana, Slovenia
e-mail: maja.zagmajster@bf.uni-lj.si

B. Sket
e-mail: boris.sket@bf.uni-lj.si

D. C. Culver
Department of Environmental Science, American University,
4400 Massachusetts Ave, NW, Washington, DC 20016, USA
e-mail: dculver@american.edu

M. C. Christman
Department of Statistics—IFAS, University of Florida, PO Box 110339,
Gainesville, FL 32611-0339, USA
e-mail: mcxman@ufl.edu

would be expected, and some of them also had a low number of uniques, both indicating sufficient sampling. By combining different analyses in a novel way we were able to evaluate observed species richness pattern as well as identify, where further sampling would be most beneficial. Approach we demonstrate is of broad interest to study of biota with high levels of endemism, small distribution ranges and low catchability.

Introduction

A first step in understanding spatial patterns of biodiversity is to map the distributions of taxonomic groups investigated. However, the observed pattern depends on the quality of data available, so when relating observed patterns to historical and present influences that have shaped them one has to consider the potential effect of differential sampling intensity (Williams et al. 2002; Magurran 2004). Incomplete sampling and sampling bias can have an important effect on conservation strategies, as the choice of priority conservation areas may be affected. Common examples of sampling bias are proximity to roads (Reddy and Dávalos 2003) or proximity to collector's residence (Dennis and Thomas 2000). Despite such impediments, in light of ongoing threats to biodiversity, it is important to make full use of existing data, even when it may be biased and incomplete (Williams et al. 2002).

One would aim to sample all areas completely or at least at the same rate (Magurran 2004), yet this is neither cost effective nor feasible for many biotas. For better studied groups, especially vertebrates, data deficiencies can probably be regarded as relatively uncommon, or can often be remedied by increased sampling. But, the problem of incomplete and biased sampling is especially acute for biota with high levels of endemism, small distribution ranges and low catchability. In such cases, techniques based on modeling species ranges (Graham and Hijmans 2006) have limited utility. Other methods that depend on the proportion of rare species, such as the well known non-parametric estimators (Colwell and Coddington 1994), need to be implemented carefully.

A model system where adjusting for sampling effect is particularly challenging is the obligate subterranean cave fauna. Subterranean species, especially those living exclusively in caves (troglobionts, Sket 2008), are limited in their dispersal abilities due to discontinuities in the appropriate subterranean habitat. As a consequence, they exhibit high levels of endemism (Jeannel 1924, 1928; Sket 1999; Christman et al. 2005), individual species often having small distribution ranges (Lamoreux 2004; Trontelj et al. 2009). Sampling of cave fauna is especially difficult. Entrances to caves can be difficult to access, presence of vertical pits of various sizes demands the use of more complicated caving techniques, and sampling can depend even on the physical abilities of the collector. The number of caves in a region can be quite large, for example, there are over 10,000 caves with more than 730 km length of corridors reported from about 8,800 km^2 of karst in Slovenia (Cave Cadastre of Slovenia, Karst Research Institute ZRC-SAZU, 2010). Additionally, obligate subterranean animals are of small size and capable of moving into unreachable crevices. As a consequence, catchability can be quite low and false negatives (species present but not sampled) are frequent (Tyre et al. 2003).

Nonetheless, the obligate subterranean cave fauna offers a number of opportunities for the study of sampling issues. First, obligate cave species usually exhibit low local species

richness compared to regional subterranean species richness (Malard et al. 2009), so they require extensive sampling. Second, the problem of vagrant (non-resident) species which can result in oversampling, that is erroneously counting species as resident even though it is just temporarily present in the habitat (Harrison and Martinez 1995), can be excluded. Obligate subterranean species can mostly be distinguished from non-specialized ones due to characteristic morphological features (termed troglomorphies): reduced eyes, absence of pigment, and elongated appendages (Sket 2008; Culver and Pipan 2009). Due to limited dispersal abilities, a troglobiotic species recorded in a cave can be regarded as resident and not as an accidental finding from a more distant population. Third, extensive databases on subterranean fauna of some regions developed recently enable implementation of spatial data in more detailed studies (Deharveng et al. 2009; Zagmajster et al. 2008).

Many of the various techniques to ameliorate the sampling effect have not been implemented on subterranean fauna, some due to limited appropriateness. The approach of using a predictive model of species richness as a function of environmental information based on well sampled locations and extrapolating it to less well sampled areas (e.g. Lobo and Martin-Piera 2002; Hortal et al. 2004) has limited utility. The relationships of surface bioclimatic variables and distribution of cave fauna also are not well understood (Culver et al. 2006). Variability of climatic variables in the interiors of caves that could affect the number of species they harbor are known insufficiently to be related to richness patterns. Lower environmental variability in subterranean environments compared to surface ones implies that historical development of the subterranean community, e.g. whether or not the region had been under an ice sheet in the past, or local small scale factors, are likely more important (Culver and Pipan 2009).

Species accumulation curves (Gotelli and Colwell 2001) have been used to compare subterranean diversity among different cave regions, with respect to sampling intensity (e.g. Culver et al. 2006; Dole-Olivier et al. 2009). Non-parametric estimates of species richness (Colwell and Coddington 1994; Magurran 2004) have been used for comparisons among cave fauna of karstic regions (Culver et al. 2006; Dole-Olivier et al. 2009) and also for predicting the expected additional species in the regions (Zagmajster et al. 2008; Deharveng et al. 2009). The approach of modeling the relationship of observed species numbers versus sampling intensity, and investigating the residuals, was used in bird studies (Elphick 1997) and parasite communities (Poulin 1995; Poulin and Rohde 1997). Any remaining structure of the residuals can contain biologically meaningful information (McIntire and Fajardo 2009).

In this study we investigated the biodiversity pattern of the obligate cave-dwelling fauna in the karst of northwestern Balkans (mainly Dinarides), ranging from northeast Italy to northern Albania. The region is a hotspot, with about 1,000 described terrestrial and aquatic species limited to the subterranean environment (Sket et al. 2004; Sket 2005), i.e. troglobionts. Distribution and biodiversity patterns of the Balkan subterranean fauna have been of interest to various investigators throughout the 20th century, but only recently have statistical methods been applied (Culver et al. 2004; Zagmajster et al. 2006, 2008). Beetles comprise about half of the terrestrial obligate subterranean fauna, and based on the relatively large number of publications and species described, are the best studied group of the subterranean animals in the region (Sket 2005; Zagmajster et al. 2008). The primary aim of this study is to evaluate the observed pattern of subterranean beetle species richness, adjusted for the effect of incomplete, biased sampling. We also investigate where additional species would be expected. To address these issues we propose the approach of combining different analytical methods.

Methods

Study area and data used

The study area comprises the karst regions of countries in the northwestern part of the Balkan Peninsula in Europe, Dinarides and eastern parts of Southern Calcareous Alps (Fig. 1). The approximately 700 km of mountain ridges run parallel to the Adriatic Sea coast on the southwest side and the Pannonian Plain on the northeast. The area is mainly comprised of carbonates (limestone, dolomite), with numerous caves. The exact number of caves in the whole study area is not known, but at least 25,000 caves are likely to be present (Zagmajster et al. 2008).

In this study, we have utilized the distribution records for 371 obligate subterranean-dwelling beetle species from five families: Cholevidae (226 spp.), Carabidae (113 ssp.), Pselaphidae (29 spp.), Ptilidae (1 spp.) and Scydmaenidae (2 spp). We used data published in 155 literature sources as of October 2008, which included most of the studies published by that time (the list of all references is available on request). We included only localities with geographic position determined with at least 6 km accuracy (depending on the quality of description in the literature as well as available maps to determine the coordinates; see Zagmajster et al. 2006, 2008 for further details). Following these criteria we included 1,857 caves in which troglobiotic beetles have been found. A further 328 caves, where only non-troglobiotic beetles were found, were included in the analysis of whole dataset. We used number of sites (within a grid cell; see next section) with beetles (troglobiotic or not) as a surrogate for sample intensity. This number is not the same as the total number of caves within the grid cell, which could be regarded a measure of habitat availability (Christman and Culver 2001). Number of sites with recorded species per cell is a relative measure of sampling intensity already used in analyses of subterranean fauna (Deharveng et al. 2009; Zagmajster et al. 2008). We assumed that the collecting methods used for beetles were such that had a troglobiotic beetle been in the cave there was a reasonable likelihood of it being collected. We do not use information such as the time spent in each cave, or how many times each cave was sampled, as such data are almost never reported. Additionally,

Fig. 1 Map of the study area, comprising Dinaric Alps and eastern parts of Southern Calcareous Alps (both mountainous regions being karstic). *Black dots* present sampled subterranean localities (caves) where beetles (troglobiotic or not) were found (Lambert Conformal Conical Projection)
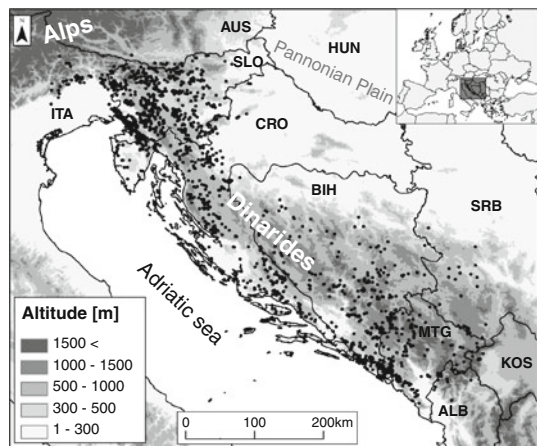
if no beetles were found, the results were typically not published and no record of the cave being visited or sampled by a biologist would exist.

Data analysis

The study site was overlain by a grid dividing the region into 20×20 km cells; this cell size is based on the results of a study of the best cell size for representation of cave biodiversity (Zagmajster et al. 2006, 2008).

First, we investigated the relationship between species numbers and sampling intensity using the number of caves in a grid cell with beetles (troglobiotic or not) as a surrogate for sampling intensity. We modeled the relationship using the equation based on the Michaelis–Menten model (Clench 1979, herewith referred to as the Clench function), with the parameterization given in Soberón and Llorente (1993):

$$S = \frac{aL}{(1 + bL)}$$

where S is observed species richness in a grid cell, L is number of caves in a grid cell (sampling intensity), and a and b are the model parameters. The model was fitted using the Gauss–Newton method for least squares nonlinear regression. This model is suitable for systems with many rare species, such as caves, where the chance of finding new species increases with field experience (Soberón and Llorente 1993). We mapped the residuals, separating the scale for positive and negative residuals, in order to determine if there was any remaining spatial pattern.

Next we investigated the sampling intensity on a local level. We estimated species richness within each grid cell using the non-parametric species richness estimator Chao 2 (Chao 1984):

$$S_{\text{Chao2}} = S_{\text{obs}} + \left[\frac{U^2}{2D}\right]$$

where S_{obs} is the number of distinct species observed in the grid cell, U is the number of species present in one cave only and D is the number of duplicates in the grid cell, i.e. the number of species present in exactly two caves. It is a commonly used estimator (Magurran 2004). We removed any grid cells with fewer than five caves with troglobiotic beetles to ensure that the results were not driven by the grid cells with very small numbers of caves. We included only caves where troglobiotic beetles were found, as the caves with non-troglobiotic beetles only are not part of the estimation. This smaller dataset of grid cells is referred to as truncated dataset.

To determine the ability of the adjustment factor (portion added to the observed number of species) in the Chao 2 estimator to reduce sampling bias, we investigated the uniques which affect the estimator most. We use the number of uniques in our study not as single cave endemics (sensu Christman et al. 2005), but as species known from only one cave per grid cell, which allows for species to be counted as a unique species in more than one cell. We modeled the relationship between uniques and number of caves in a grid cell adjusted for the total number of species observed in the grid cell using Poisson regression with a log link. Both explanatory variables were log-transformed. We tested for the interaction of the two explanatory variables but it was not statistically significant ($P > 0.15$) and so it was removed from the final model.

All mapping was done in ArcGIS 9.3 (ESRI). Statistical calculations and modeling were done in the statistical package programs in JMP® 8.0 and SAS (SAS Institute Inc., Cary, NC) and EstimateS 7.5 (Colwell 2005).

Results

Species richness

At least one cave with a beetle can be found in 265 20 × 20 km grid cells in the study area; in 233 grid cells, at least one troglobiotic beetle was recorded. In 115 grid cells at least 5 caves with troglobiotic beetles were found, constituting a truncated dataset (Fig. 2a). The cells that are excluded from the truncated dataset (having four or less caves with troglobiotic beetles) are located on the edges of the study area. All analyses were done on the truncated dataset, and when possible on whole dataset too. As no distinguishable differences were observed among the two datasets, we report here results of the truncated one (results for the whole dataset are in Supplementary material).

In the region, two areas of high observed species richness are apparent, one in the northwest in Slovenia and the other in the southeast near the border of Bosnia and Herzegovina (hereafter BiH) and Montenegro. In the northwest, the cells in the top two categories for richness are contiguous; conversely, in the southeastern region they tend to fall into three clusters (Fig. 2b).

Patterns of sampling intensity and species richness (Fig. 2a, b) differ. There are five highly sampled grid cells, all located in the northwestern region, and only one highly sampled cell is found in the southeastern part (BiH and Montenegro border). There are no grid cells sampled to the same extent in the central part of the region (Croatia and western BiH). There are only two cases where well sampled grid cells and species rich grid cells overlap. One is in Slovenia and one on the border of BiH and Montenegro. There are species rich grid cells that do have comparatively low sampling intensity—and vice versa. A highly sampled cell in northwestern part of the region, on the Italy-Slovenia border, does not result in the high species richness.

The fitted Clench function indicates that the number of species begins to level off as the number of caves per cell increases (Fig. 3). To investigate the effect of differences in sampling intensity, we mapped the differences between expected and observed species richness, i.e. the residuals from the fitted Clench function (Fig. 2c). The residuals show spatial clustering of similar values. Grid cells with high positive residuals occur in the northwestern and southeastern part of the study region. High positive residuals indicate areas where more species have been observed than would be expected based on sampling intensity; they can be considered true hotspots. They do overlap hotspots of observed species richness.

The Chao 2 estimator increases the observed number of species using the number of uniques and doubletons (Fig. 2d). All estimated species rich cells (those having at least 60% species of the richest cell) overlap with the observed species rich cells implying that the adjustment factor in the Chao 2 estimator also reinforces the general spatial pattern of species hot spots. Conversely, there are some observed species rich cells that are not classified similarly as the estimated values; these occur both in the northwest as in the southeast. In spite of that, overall the pattern of two areas of high species richness is confirmed by the Chao 2 estimator.
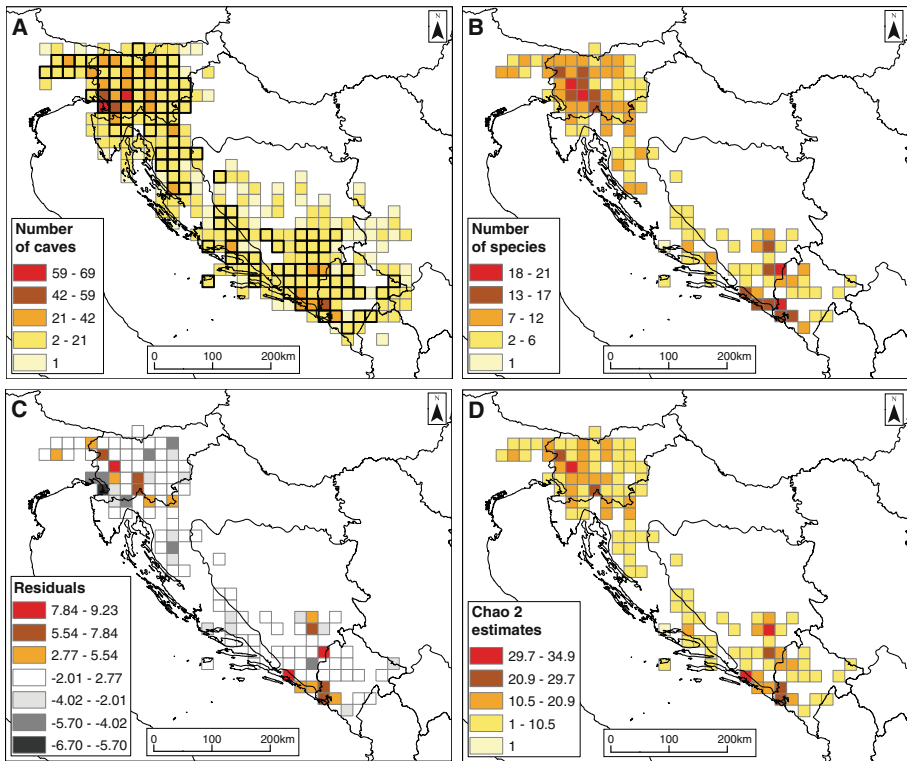
Fig. 2 The map of sampling intensity (**a**) measured with number of caves with beetles (all beetles, no matter whether they are troglobiotic or non-troglobiotic), and three representations of species richness patterns of troglobiotic beetles in the northwestern Balkans: **b**—observed numbers of species, **c**—residuals of the Clench function fit (Soberón and Llorente 1993) of the observed number of species to sampling intensity, **d**—Chao 2 estimates. The grid cells with thick *black outline* are those with five or more caves with troglobiotic beetles (termed truncated dataset). We used the following classes: (1) grid cells having at least 85% of the maximum observed in the richest grid cell; (2) grid cells having between 60 and 85% of the maximum; (3) grid cells with between 30 and 60% of the maximum; (4) grid cells with between two species and 30% of the maximum number of species; and, (5) grid cells with exactly one species. In the case of residuals, delimitation is presented separately for positive and negative ones, with the fourth and fifth classes merged in one (Lambert Conformal Conical Projection)

Recommendations for further sampling

We mapped the number of uniques in the grid cells (Fig. 4) in order to compare the spatial pattern to the observed species richness pattern. Some grid cells with the larger number of uniques also have the highest observed species richness (Figs. 2b, 4). High numbers of uniques are also found in less rich cells adjacent to the high richness cells, such as in southeastern BiH, southern parts of Slovenia, and near the Croatia–Slovenia border.

To understand this pattern, we investigated the relationship between number of uniques and sampling intensity, regulated for the number of species in the cell. The following equation best describes the relationship:
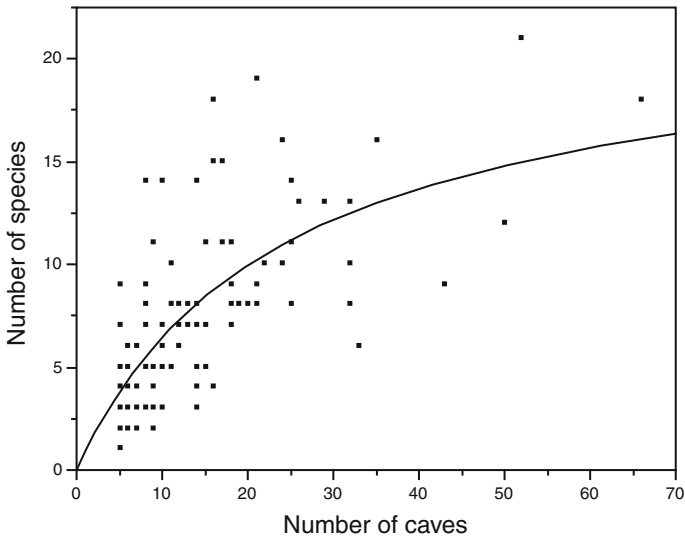
$$\ln(E[Y]) = -0.3790 - 0.6394X_1 + 1.4983X_2$$

Fig. 3 The Clench function fit (Soberón and Llorente 1993) of the observed number of species of obligate subterranean beetles to sampling intensity (measured as number of caves with troglobiotic beetles per 20×20 km grid cell) in the northwestern Balkans. The asymptote parameter estimate is 0.91315, the rate parameter estimate is 0.04159 (Clench function parameters a and b, see text), with RMSE 3.05

Fig. 4 The map of the number of uniques (species of obligate subterranean beetles present in one cave within the grid cell only) in the northwestern Balkans. Class delimitation as described in legend of Fig. 2 (Lambert Conformal Conical Projection)
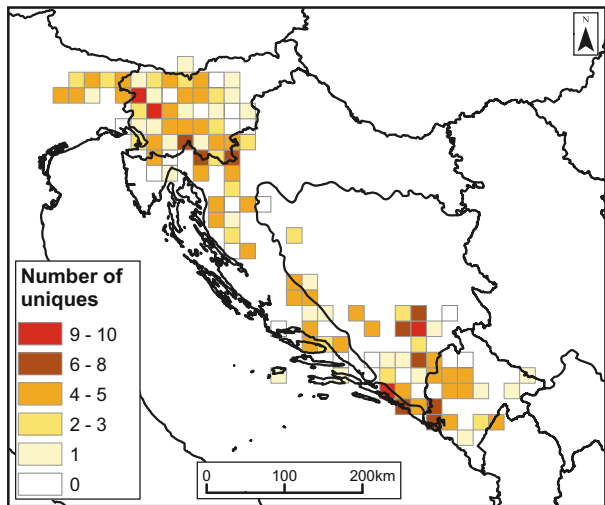
where the intercept is not statistically significant ($P = 0.1467$), and both the slope for $X_1 = \ln$(number of caves) and $X_2 = \ln$(number of species) were statistically significant ($P < 0.0001$). This confirms that as the number of caves increases the expected number of unique species decreases for any fixed number of species in the grid cell (Fig. 5), which implies that the number of uniques can be used as a measure of sampling completeness.

By comparing the residuals from the Clench function with the number of uniques we can infer where additional sampling would be necessary (Fig. 6a). High positive residuals
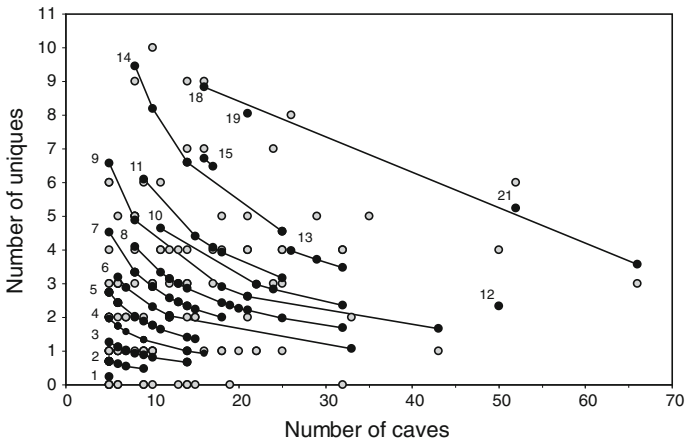
Fig. 5 Plot of the relationship between uniques (species of obligate subterranean beetles present in one cave within the grid cell only) and number of caves in a grid cell adjusted for the total number of species observed in the grid cell. Poisson regression with a log link was used. Observed numbers are presented with *grey circles*, and predicted values with *black circles*. Labels next to predicted values are the number of observed species and lines connect predicted values with the same number of observed species

show that for a given sampling intensity, species richness is higher than would be expected from sampling intensity alone. So, these cells have the highest potential to reveal additional species with additional sampling. It becomes apparent that many (though not all) of the highest positive residuals actually overlap with the areas of highest numbers of uniques within a grid cell. Such cells are distributed in northwestern (four) and southeastern part (five) of the region (Fig. 6b). Based on what is known from them at the moment, more than twice (two cells in southwest) or nearly twice (two cells in northeast) the observed number of species is estimated by Chao 2 estimator (Fig. 2d).

Negative residuals show less species than expected on the basis of sampling intensity, so additional sampling is less likely to reveal many additional species. This is further supported by low number of uniques in those cells (Fig. 6). More grid cells can be regarded as sufficiently sampled in the northwestern part of the study area.

Discussion

In the northwestern Balkans, there are two areas of high species richness for subterranean beetles. This pattern appears to be accurate despite the sampling bias present in the dataset. The inclusion of additional species (three more families, albeit with fewer species) and expanding the study area (including the Alpine karst in Slovenia) in the current study compared to previous ones (Zagmajster et al. 2008) did not greatly change the pattern. This shows it is robust to change in area of coverage and taxonomic scope. Similarly, for subterranean fauna in the Slovenian Dinaric karst, Culver et al. (2004) have shown that when the known species richness as of a particular date was mapped for different points in time the position of the richest grid cells did not change despite more data being added in later time periods. Similar results were reported also in some surface studies. For example, in a study of fern biodiversity in Bolivia, patterns of richness were fairly consistent along three time periods in the past despite increasing collection intensity (Soria-Auza and Kessler 2008).
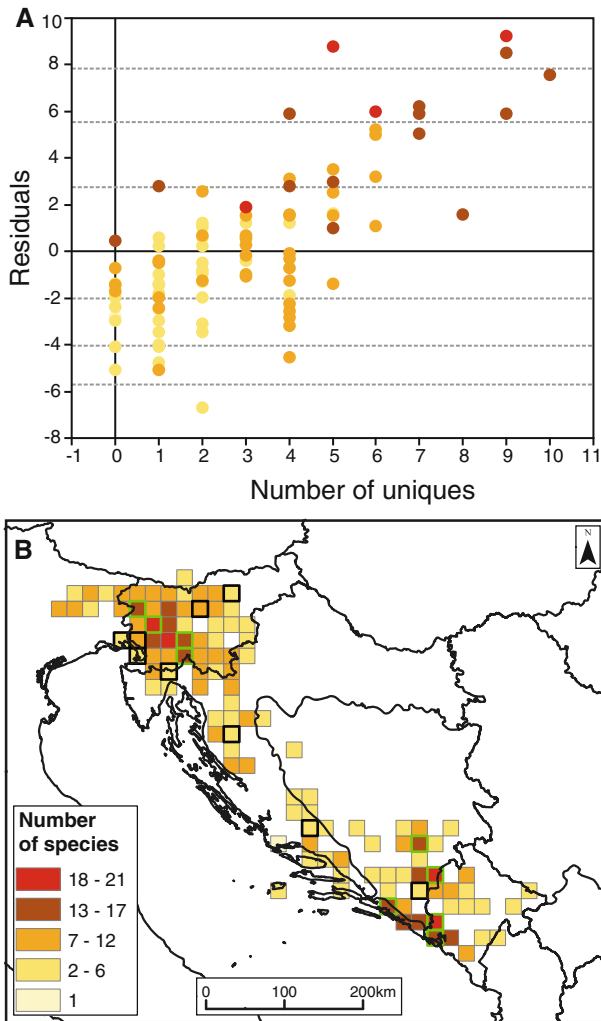
Fig. 6 The relationship of the residuals of the Clench function fit (Soberón and Llorente 1993) of the observed number of species to sampling intensity to number of uniques (**a**) and the map of species richness. On the map, grid cells with high residuals and high number of uniques are outlined with *green*, while the ones with low residuals and low number of uniques with *black*. Class delimitation as described in legend of Fig. 2 (Lambert Conformal Conical Projection) (Color figure online)

The study of biodiversity is a comparative discipline as a single species richness number for one area is not necessarily informative when not compared to richness in other area (Magurran 2004). Therefore, we were interested in patterns within a region rather than the exact species numbers within each grid cell. Unfortunately, we cannot do this directly when the sampling intensity is not homogeneous across the region. To address whether this is an issue for our study region, we used different approaches to ameliorate sampling effect and have confirmed the observed pattern of areas of rich fauna diversity. High positive residuals from fitting the observed number of species against sampling intensity, indicating locations of higher species richness than would be expected on sampling alone, confirmed

most of the observed hotspots. The overlap of hotspots predicted by positive residuals and observed hotspots show that the observed rich areas can be regarded as true hotspots and so the observed pattern is valid in spite of the likely unequal sampling intensity.

Estimators that reduce bias by accounting for sampling intensity are widely used in many biodiversity studies (Magurran 2004). We used the Chao 2 estimator, regarded as being a good choice for adjusting the observed richness (Colwell and Coddington 1994). The appropriateness of the use of this estimator in fauna with small distribution ranges should be considered in relation to size of the sampling unit. In an earlier study of species richness of two families of troglobiotic beetles, grid cells were used as samples for calculating the estimated number of species in the whole region (Zagmajster et al. 2008). Chao 2 was considered inappropriate, as the number of species found in one grid cell only (unique) was less likely to go to zero (Zagmajster et al. 2008). Cave species by nature have very small distribution ranges (Christman et al. 2005; Trontelj et al. 2009) so the distribution range of one species could fit within one grid cell and uniques would never diminish. But, in the current study, estimation is done within each grid cell, using individual caves as samples. On this scale, cave animals are less likely present in one cave (one sample) only; cave animals are small and able to cross very narrow crevices to neighboring caves (Jeannel 1924, 1928; Culver and Pipan 2009). Hence, uniques are likely an artifact of insufficient sampling and not species actually being limited to a single cave.

The number of unique species within a grid cell is informative for estimating sampling completeness within the cell. Some studies have used the ratio of observed and estimated species richness to evaluate sampling completeness (Soberón et al. 2007). Yet the estimator contains the number of observed species and so is correlated with the observed species richness. As a result the estimated species richness pattern will not be independent of the observed pattern. The number of uniques, being part of estimated values, can be regarded as an alternative measure of sampling completeness. If in fact the number of uniques contains independent information about the sampling intensity, as sampling improves, we expect that number of uniques should approach zero. We have shown that when adjusted for the observed species numbers, the number of uniques did decrease with increased sampling.

In both the northwestern and southeastern regions of the study area, there are grid cells that show high species richness but also insufficient sampling, i.e. a high number of uniques. The high positive residuals from fitting the Clench function generally supports the pattern observed from the distribution of uniques. The residuals show that if species richness is already above average, it is these areas that tend to have the potential to reveal the most additional species. This is true for some of the grid cells with large numbers of uniques. This interpretation is different from that of Elphick (1997). He interprets areas with negative residuals as the ones that would need further sampling since these are areas where less species were observed than expected. The issue is that they are low even after the sampling intensity is accounted for implying that increasing intensity will not necessarily increase species numbers. Our argument is further supported by low numbers of uniques in some of those same cells and so sufficient sampling is indicated. In our study, grid cells with sufficient sampling (and low species richness) tend to be distributed at the edges of the study region. This is not unreasonable as there is less suitable habitat available being at the edge of the karst region.

Lennon et al. (2004) have shown that in species richness maps of birds, common species, defined as having a wide distribution range, account for a bigger part of diversity than rare species. But in subterranean habitats species have very small distribution ranges (Trontelj et al. 2009) and common species as defined by Lennon are a minority.

So, exploring whether rare species are informative is a reasonable approach; we were able to show that the rare species (in terms of uniques) can be very informative in evaluating the species richness pattern for subterranean beetles. When further sampling is planned, it may be reasonable to question what the contribution of additional sampling to the hotspots would be in resolving the regional pattern. More intensive sampling of obligate ground-water animals over the last few years in six European regions in Europe already known as being species rich had little effect on the broad-scale pattern of the groundwater biodiversity in Europe (Deharveng et al. 2009).

Sampling bias is not always a problem if relative differences within a study region are important, and lower sampling intensity can give a satisfactory representation of biodiversity patterns (Brose 2002). Typically biologists have sampled in areas where they expect more species no matter what group of organisms is being studied. So, if they went to the species rich area and found new species they returned to that region again. And vice versa: if they went to the area and found no new species they may have returned a second time but then no more. Such adaptive sampling is a legitimate sampling method in cases dealing with rare species (Smith et al. 2004). In cave fauna, the adoption of the adaptive sampling strategy principles when applied to the number of unique species rather than total number of species may help to support the argument that the observed pattern reflects the underlying diversity. Also, as an adaptive sampling design is statistically defensible, further studies in this direction would be useful.

Large amounts of data on species distributions exist even though in many cases they are scattered throughout databases, published literature, unpublished collection lists, or even personal field notes. Such information can offer a valuable contribution toward addressing biodiversity but the effect of sampling bias should be evaluated. The approach we took in this study may be especially useful for fauna with species with small ranges, i.e. high endemism, especially when we have little other information to indicate specific levels of sampling intensity (such as number of visits) in a grid cell. After evaluating the sampling effect, other possible factors like natural drivers should be considered (proportion of habitat, habitat fragmentation, past and present bioclimatic variables etc.). Understanding species distribution and finding areas where the diversity of species is highest are important prerequisites for investigating the processes that have created them, for planning further sampling, and for creating efficient conservation priorities.

References

Brose U (2002) Estimating species richness of pitfall catches by non-parametric estimators. Pedobiologia 46:101–107

Chao A (1984) Non-parametric estimation of the number of classes in a population. Scand J Stat 11:265–270

Christman MC, Culver DC (2001) The relationship between cave biodiversity and available habitat. J Biogeogr 28:367–380

Christman MC, Culver DC, Madden M et al (2005) Patterns of endemism of the eastern North American cave fauna. J Biogeogr 32:1441–1452

Clench HK (1979) How to make regional lists of butterflies: some thoughts. J Lepid Soc 33:216–231

Colwell RK (2005) EstimateS: statistical estimation of species richness and shared species from samples. Version 7.5. User's guide and application. http://purl.oclc.org/estimates. Accessed 10 Dec 2009

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc B 345:101–118

Culver DC, Pipan T (2009) The biology of caves and other subterranean habitats. Oxford University Press, Oxford

Culver DC, Christman MC, Sket B et al (2004) Sampling adequacy in an extreme environment: species richness patterns in Slovenian caves. Biodivers Conserv 13:1209–1229

Culver DC, Deharveng L, Bedos A et al (2006) The mid-latitude biodiversity ridge in terrestrial cave fauna. Ecography 29:120–128

Deharveng L, Stoch F, Gibert J et al (2009) Groundwater biodiversity in Europe. Freshw Biol 54:709–726

Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. J Insect Conserv 4:73–77

Dole-Olivier MJ, Castellarini F, Coineau N et al (2009) Towards and optimal sampling strategy to assess groundwater biodiversity: comparison across six European regions. Freshw Biol 54:777–796

Elphick CS (1997) Correcting avian richness estimates for unequal sample effort in atlas studies. Ibis 139:189–190

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett 4:379–391

Graham CH, Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. Glob Ecol Biogeogr 15:578–587

Harrison JA, Martinez P (1995) Measurement and mapping of avian diversity in southern Africa: implications for conservation planning. Ibis 137:410–417

Hortal J, Garcia-Pereira P, Garcia-Barros E (2004) Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. Ecography 27:68–82

Jeannel R (1924) Monographie des Bathyscinae. Arch Zool Exp Gen 63:1–436

Jeannel R (1928) Monographie des Trechinae. Morphologie comparee et distribution geographique d'un groupe de Coleopteres. Troisieme Livraison. Les Trechini cavernicoles. J Entomol 35:1–800

Lamoreux J (2004) Stygobites are more wide-ranging than troglobites. J Cave Karst Stud 66:18–19

Lennon JL, Koleff P, Greenwood JJD et al (2004) Contribution of rarity and commonness to patterns of species richness. Ecol Lett 7:81–87

Lobo JM, Martin-Piera F (2002) For a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. Conserv Biol 16:158–173

Magurran AE (2004) Measuring biological diversity. Blackwell, London

Malard F, Boutin C, Camacho AI et al (2009) Diversity patterns of stygobiotic crustaceans across multiple spatial scales in Europe. Freshw Biol 54:756–776

McIntire EJB, Fajardo A (2009) Beyond description: the active and effective way to infer processes from spatial patterns. Ecology 90:46–56

Poulin R (1995) Phylogeny, ecology, and the richness of parasite communities in vertebrates. Ecol Monogr 65:283–302

Poulin R, Rohde K (1997) Comparing the richness of metazoan ectoparasite communities of marine fishes: controlling for host phylogeny. Oecologia 110:278–283

Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. J Biogeogr 30:1719–1727

Sket B (1999) The nature of biodiversity in hypogean waters and how it is endangered. Biodivers Conserv 8:1319–1338

Sket B (2005) Dinaric karst, diversity. In: Culver DC, White WB (eds) Encyclopedia of caves. Elsevier Academic Press, Amsterdam

Sket B (2008) Can we agree on an ecological classification of subterranean animals? J Nat Hist 42:1549–1563

Sket B, Paragamian K, Trontelj P (2004) A census of the obligate subterranean fauna of the Balkan peninsula. In: Griffiths HI, Kryštufek B, Reed JM (eds) Balkan biodiversity, pattern and process in Europe's biodiversity hotspot. Kluwer Academic Publishers, Dordrecht

Smith DR, Brown JA, Lo NCH (2004) Application of adaptive sampling to biological populations. In: Thomson WL (ed) Sampling rare or elusive species. Concepts, designs, and techniques for estimating population parameters. Island Press, Washington, DC

Soberón JM, Llorente JB (1993) The use of species accumulation functions for the prediction of species richness. Conserv Biol 7:480–488

Soberón J, Jimenez R, Golubov J et al (2007) Assessing completeness of biodiversity databases at different spatial scales. Ecography 30:152–160

Soria-Auza RW, Kessler M (2008) The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. Divers Distrib 14:123–130

Trontelj P, Douady CJ, Fišer C et al (2009) A molecular test for cryptic diversity in groundwater: how large
    are the ranges of macro-stygobionts? Freshw Biol 54:727–744

Tyre AJ, Tenhumberg B, Field SA et al (2003) Improving precision and reducing bias in biological surveys:
    estimating false-negative error rates. Ecol Appl 13:1790–1801

Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority
    area selection. J Biosci 27:327–338

Zagmajster M, Sket B, Podobnikar T (2006) Choosing a grid network for spatial analysis of subterranean
    biodiversity. In: Perko D, Nared J, Čeh M et al (eds) Geographical information systems in Slovenia
    2005–2006. ZRC SAZU Publishing, Ljubljana

Zagmajster M, Culver DC, Sket B (2008) Species richness patterns of obligate subterranean beetles (Insecta:
    Coleoptera) in a global biodiversity hotspot-effect of scale and sampling intensity. Divers Distrib
    14:95–105