Biodivers Conserv (2011) 20:119–139
DOI 10.1007/s10531-010-9950-6

ORIGINAL PAPER

Geographic sampling bias in the South African Frog Atlas Project: implications for conservation planning

Emily A. Botts · Barend F. N. Erasmus · Graham J. Alexander

Received: 23 September 2009 / Accepted: 18 November 2010 / Published online: 5 December 2010

Abstract Quality conservation planning requires quality input data. However, the broad scale sampling strategies typically employed to obtain primary species distribution data are prone to geographic bias in the form of errors of omission. This study provides a quantitative measure of sampling bias to inform accuracy assessment of conservation plans based on the South African Frog Atlas Project. Significantly higher sampling intensity near to cities and roads is likely to result in overstated conservation priority and heightened conservation conflicts in urban areas. Particularly well sampled protected areas will also erroneously appear to contribute highly to amphibian biodiversity targets. Conversely, targeted sampling in the arid northwest and along mountain ranges is needed to ensure that these under-sampled regions are not excluded from conservation plans. The South African Frog Atlas Project offers a reasonably accurate picture of the broad scale west-to-east increase in amphibian richness and abundance, but geographic bias may limit its applicability for fine scale conservation planning. The Global Amphibian Assessment species distribution data offered a less biased alternative, but only at the cost of inflated commission error.

Keywords Biodiversity · Biological atlas · Conservation planning · Geographic sampling bias · South African Frog Atlas Project

Abbreviations

SAFAP	South African Frog Atlas Project
QDGC	Quarter Degree Grid Cell
GAA	Global Amphibian Assessment

E. A. Botts (✉) · B. F. N. Erasmus · G. J. Alexander
School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, WITS, 2050 Johannesburg, South Africa
e-mail: Emily.Botts@students.wits.ac.za

Introduction

Recent concern over negative anthropogenic impacts on species and ecosystems has resulted in many studies detailing the necessity for large-scale, effective and timely conservation planning (Eken et al. 2004; Ferrier et al. 2004). Conservation planning requires primary input data, the optimal nature of which is still debated in the literature. In the absence of comprehensive species richness data, some authors have proposed the use of surrogates and proxies i.e. land cover types or 'indicator' species (see Funk and Richardson 2002; Pressey 2004; Ferrier et al. 2004; Larsen et al. 2009a). However, research has shown that surrogates perform poorly in regions where the spatial pattern of biodiversity differs among taxa (Van Jaarsveld et al. 1998a, b; Margules and Pressey 2000; Pressey 2004). Thus, most authors agree that robust conservation planning would benefit from accurate and current species distribution data for a wide variety of taxa (Donald and Fuller 1998; 1998a, b; Ferrier 2002; Pressey 2004; Larsen et al. 2009a; Boakes et al. 2010).

This has consequently given rise to an increasing number of new biological atlases (Elphick 1997; Donald and Fuller 1998; Rondinini et al. 2006). A biological atlas is usually a grid-based, presence-only record of species occurrence for a specific taxonomic group, region and time scale (Donald and Fuller 1998; Dennis et al. 1999; Dunn and Weston 2008; Robertson et al. 2010). The South African Frog Atlas Project (SAFAP) is one such atlas with potential for inclusion in conservation plans. Atlases often offer the only broad-scale, current and spatially contiguous data available for conservation assessments; but their limitations for these applications are still much emphasised (McCollin et al. 2000; Funk and Richardson 2002; Telfer et al. 2002; Reddy and Dávalos 2003; Tyre et al. 2003; Pressey 2004; Robertson et al. 2010). Atlases are particularly vulnerable to omission error, which in turn introduces bias to the datasets. False-negative error (or omission error) occurs when an organism or species is present but is not recorded (Williams et al. 2002; Tyre et al. 2003; Rondinini et al. 2006). Atlases should, in principle, be immune to such issues as they are systematic and aim to cover an area completely and thoroughly. However, in reality, atlases are subject to geographic bias due to logistical, budgetary and time constraints (Donald and Fuller 1998; Robertson et al. 2010). Geographic, or spatial, sampling bias refers to the case in which omission errors occur non-randomly in space. This arises when certain areas are favoured during sampling, thus receiving greater sampling intensity, and others are neglected (Funk and Richardson 2002; Robertson and Barker 2006; Rondinini et al. 2006).

The nature of false-negative error means that the bias in atlases is often predictable: higher sampling intensity in accessible areas or those favoured by researchers (Freitag et al. 1998; Funk and Richardson 2002; Parnell et al. 2003; Reddy and Dávalos 2003; Robertson and Barker 2006; Romo et al. 2006; Ferreira et al. 2007). It has been shown that conveniently located areas are better-sampled (Freitag et al. 1998; Parnell et al. 2003; Reddy and Dávalos 2003). In Thailand, there were many more plant records than expected within 4 km of populated places (Parnell et al. 2003). Records of passerine birds in sub-Saharan Africa were significantly closer to cities, rivers and roads than random (Reddy and Dávalos 2003). Areas within an easy-access distance of populated places or common travel routes are therefore predicted to have higher sampling intensity due to greater accessibility.

In addition to conveniently accessible areas, collectors tend to favour areas that they presume will be rewarding or valuable for research (Williams et al. 2002; Reddy and Dávalos 2003; Küper et al. 2006; Romo et al. 2006; Boakes et al. 2010). This includes protected areas and those areas already known for high biodiversity. This phenomenon has been termed 'diversity tracking' (Romo et al. 2006). On mainland Thailand, the three provinces with the highest plant collection density were those associated with national parks (Parnell et al. 2003). The correlation between one-degree grid cells containing protected areas and sampling of passerine birds in Africa was high ($R^2 = 0.74$; Reddy and Dávalos 2003). Both Iberian butterflies (Romo et al. 2006) and plants in Thailand (Parnell et al. 2003) were better sampled in mountainous regions as these are the preferred study areas of researchers in an otherwise transformed landscape.

Variation in sampling intensity in space can result in incorrect species richness and distribution measures. Higher sampling intensity results in a higher probability of species detection (Elphick 1997; Williams et al. 2002). Thus, well-sampled areas appear to be more species rich than poorly sampled areas (Reddy and Dávalos 2003). Remote, poorly sampled regions will appear to have low species richness and will consequently be erroneously excluded from conservation plans (Reddy and Dávalos 2003; Küper et al. 2006). These areas then appear of little scientific or conservation interest and continue to attract few researchers (Reddy and Dávalos 2003; Küper et al. 2006). Conversely, planners will consider the placement of existing reserves particularly effective since high sampling intensity in these areas results in over-estimated biodiversity (Freitag et al. 1998; Reddy and Dávalos 2003). Seemingly high diversity near to populated places also increases conflict between conservation and human development (Küper et al. 2006; Romo et al. 2006). Low sampling effort can result in smaller documented range sizes (Gaston and Rodrigues 2003; Küper et al. 2006; Ferreira et al. 2007). It is thus prudent to identify areas of low sampling intensity before performing any conservation planning exercise.

The biodiversity information captured in biological atlases is usually incorporated into systematic conservation plans in the form of species distribution maps. These distribution maps can take a number of forms (Rondinini et al. 2006). They can simply be the unprocessed grid cell maps that are commonly the product of atlas projects. Alternatively, they can be generalised into interpreted distribution maps that display the broad areas in which species occur within the study area. This manipulation aims to reduce the frequent omission errors found in atlas data (Rondinini et al. 2006). Modelled distribution maps are occasionally included, often also with the intention of managing uncertainty and reducing omission error (Margules and Pressey 2000; Funk and Richardson 2002). Sometimes only species of special interest are included: rare, threatened or endemic species (e.g. Rouget et al. 2004). Some conservation plans use expert consultation to combine this information into areas of concern for a specific taxon. Locations with high endemism or high numbers of endangered species are thus included as biodiversity features (e.g. Desmet et al. 2008).

Once the biodiversity features have been mapped, specific targets are set which are defensible and quantitative (Margules and Pressey 2000; Desmet and Cowling 2004). These targets can range from a single occurrence of each species (e.g. Rouget et al. 2004) to 100% of the distribution in cases of Critically Endangered species (e.g. Ferrar and Lötter 2007). Algorithms are then used to identify areas which can achieve the targets while minimising costs and maximising complementarity (Margules and Pressey 2000). The result is a map of irreplaceability, in which planning units that are very important for meeting the targets are highly irreplaceable (Margules and Pressey 2000).

Previous atlas projects in the southern African region have been extensively used in conservation planning exercises (Van Jaarsveld et al. 1998b; Larsen et al. 2009b), despite limitations of resolution, taxonomy, geographic coverage and bias. Indeed, conservation planning has been one of the major uses of the pioneering South African Bird Atlas Project, completed in 1997 (Dunn and Weston 2008; Harrison et al. 2008). Similarly, both national and provincial conservation planning efforts have already made use of the South African Frog Atlas Project (SAFAP) (Harrison et al. 2004). The South African National Spatial Biodiversity Assessment used the SAFAP for distribution data and conservation status of

11 threatened frog species (Rouget et al. 2004). This assessment regarded the SAFAP as one of the better datasets available and largely exempted it from acknowledged biases due to systematic sampling (Rouget et al. 2004). The SAFAP has also been used in several regional conservation assessments (e.g. GDACE 2004; Turner and De Villiers 2007), although in these cases the Quarter Degree Grid Cell scale is often too coarse a resolution and only those records with accurate GPS positions can be included (57% of the total records). Most of these conservation planning efforts concede that bias-associated data may be limiting, but rarely quantify the degree of bias.

The editors of the SAFAP clearly acknowledge that geographic bias was present within the atlas data. Two specific restrictions were identified: arid and mountainous regions (Minter et al. 2004). Dry weather in the arid northwestern part of South Africa meant that trips to this region resulted in few records. In some cases, a visit to a grid cell returned no frog records. Thus, trips to this part of the country were targeted at the larger Half Degree Grid Cell (HDGC) scale to save money and time (Minter et al. 2004). Difficulty in accessing mountain peaks resulted in limited sampling in the Cape Fold Mountains, Drakensberg Mountains and mountainous areas of Lesotho (Minter et al. 2004).

In this paper we provide a quantitative measure of geographic bias to inform accuracy assessment of conservation prioritisation efforts based on the SAFAP. We also include a parallel assessment of geographic bias within South African species distribution data available from the Global Amphibian Assessment (GAA), a possible alternative to the SA-FAP. We evaluate the advantages and limitations of these data for use in conservation planning.

Methods

The South African Frog Atlas data

The South African Frog Atlas Project (SAFAP) was a project managed by the University of Cape Town's Animal Demography Unit (Minter et al. 2004). The project was initiated to improve the quality of the distribution data available for South African amphibians (Minter et al. 2004). Twenty-eight universities and scientific institutions participated in the frog atlas project, either by supplying historical data or by contributing to active data collection (Minter et al. 2004).

The atlas data were collected between 1996 and 2003 by volunteers and herpetologists who entered species information onto report forms (Minter et al. 2004). The report forms and associated evidence were checked and processed by regional organisers before being sent to the University of Cape Town for data capture. The methodology involved the systematic survey of the majority the Quarter Degree Grid Cells (QDGCs) in South Africa, Lesotho and Swaziland. At each grid cell, presence records for frog species were acquired based on either visual or audio evidence (Minter et al. 2004). Frogs, like birds, have distinctive calls that are unique to a species and this greatly enhances the ability to atlas frogs. Data collection was usually timed to correspond with the breeding season to optimise the discovery and identification of species based on breeding calls (Minter et al. 2004). Historical museum and archive data were included to augment the data collected during the atlas survey. Museums, personal databases and literature records were used as sources for historical data that covered approximately 100 years of sampling (Minter et al. 2004). Data were corrected for updated taxonomic classifications and doubtful records were excluded. Ultimately, 16,983 historical records were included to support the 25,486 records collected during the atlas period.

Assessment of geographic sampling bias in SAFAP data

Two different methodological approaches were used to analyse geographic sampling bias in the SAFAP data. The first approach used numbers of atlas records as an indication of the spatial variation in sampling intensity (Freitag et al. 1998; Parnell et al. 2003; Reddy and Dávalos 2003). Relatively better-sampled areas were expected to have returned greater numbers of amphibian records. Thus, the collection density was computed and related to hypothesised sources of bias. The second approach employed species richness as a measure of sampling intensity. Higher sampling intensity in certain areas would have resulted in the detection of relatively more species. The number of species in each QDGC was used to identify gaps in sampling (Robertson and Barker 2006).

Biased collection density in relation to geographical features

Collection density is the number of records per unit area (Freitag et al. 1998; Parnell et al. 2003). It is a reflection on both the abundance of amphibians and the sampling intensity in an area. For comparison with previous studies (Freitag et al. 1998; Parnell et al. 2003), we calculated the collection density 'per 100 km^2' for the entire study area. The number of records per QDGC mapped the spatial variation in collection density over the country.

The distribution of records was then spatially related to features hypothesised to allow ease of access to researchers and thus act as sources of sampling bias: cities, roads and protected areas. Observed numbers of records were computed as the number of records falling within certain distance categories from these features.

Cities: The six major cities in South Africa were used in this analysis: Bloemfontein, Cape Town, Durban, Johannesburg, Port Elizabeth, and Pretoria. These cities also house the institutions that contributed the most records to the SAFAP. Thus, distance from these cities was presumed to be the distance that herpetologists would have to travel to survey frogs.

Roads: The largest national roads were used, since these are the main routes that connect the major cities. QDGCs in the countries of Swaziland and Lesotho were excluded from this analysis due to lack of roads data. Roads are much smaller than the QDGCs through which they transect, and there are often many roads within a single QDGC. This meant that QDGCs could not easily be assigned to a certain distance from roads. So, roads data were converted into distance image maps (ArcGIS Spatial Analyst, "Distance"), such that each QDGC contained 25 smaller pixels. Each of these smaller pixels was assigned the distance to the nearest road (Parnell et al. 2003). Thus, the distances used are the average distance (km) to any national road within each QDGC.

Protected areas: Since there is a large variation in the size of protected areas, some QDGCs are completely enclosed within reserves, while other reserves are completely enclosed within a single QDGC. Thus, the zero distance class used in this analysis contained QDGCs that were either completely or mostly covered by protected areas. The 1–10 km distance class included cells that either partially intersected reserves or were located within 10 km of the reserve boundary. The QDGCs in the remaining distance classes had no direct contact with any reserve, but were located within the indicated distance from the protected area.

In line with several previous studies, expected number of records was calculated as:

$$\text{Expected number of records} = \frac{\text{Number of QDGCs within category}}{\text{Total number of QDGCs}} \times \text{Total number of records}$$

Hence, the expected number of records became a function of the area covered by each distance category (Freitag et al. 1998; Parnell et al. 2003). Observed patterns in the distribution of records were compared to expected numbers and the differences indicated the level of sampling within a category (Freitag et al. 1998). Since observed and expected proportions were available per category, the appropriate statistical test was the χ^2 test (McDonald 2009). A separate χ^2 test was performed for cities, roads and protected areas to test for a significant difference between the observed and expected patterns. Since the high sample size resulted in an unsuitably high power for these tests, a power analysis was conducted (Quinn and Keough 2002; Lenth 2006). A minimum effect size of 5% (or $\approx$2,000 records) was chosen as this was approximately the size of the sampling bias effect found in previous studies (Quinn and Keough 2002; Reddy and Dávalos 2003; Parnell et al. 2003). Thus, sample size of observed numbers of records was reduced proportionally for each test to obtain a power of 0.8 ($\beta = 0.2$, a conventional choice for power; McDonald 2009; Quinn and Keough 2002). We corrected for multiple tests using both the step-down Bonferroni and False Discovery Rate methods (McDonald 2009). Only the corrected *P*-values are presented in the results section.

The location of cities, roads and protected areas are often correlated, since they are all manifestations of human presence (population density) within an area. Three correlations (cities vs. roads; cities vs. protected areas; roads vs. protected areas) were performed to test whether this covariance was evident at the QDGC resolution utilised here. Non-parametric (Spearman) correlations were used to account for possible non-normality in the measured distances from each feature (McDonald 2009). These distances from features were the same categorical distances assigned to each QDGC and used in the previous analyses. Step-down Bonferroni and False Discovery Rate corrections for multiple tests were performed.

Identifying sampling gaps based on species richness

Species richness is defined as the number of species per unit area. The term species richness is usually used to describe real patterns of species location in space. However, when there is significant geographic sampling bias, measured species richness can be a function of the degree of sampling intensity (Dennis et al. 1999; Reddy and Dávalos 2003; Ferreira et al. 2007). As sampling intensity increases, so measures of species richness become more dependent on the actual species richness and less dependent on the sampling intensity.

Funk and Richardson (2002) plotted the mean annual temperature and mean annual rainfall of all the QDGCs in their study area, thus creating a climate space in which they then plotted the rainfall and temperature variables of collecting sites. The gaps between collecting sites on the graph were then related back to a map of the study area to identify areas where conditions appear suitable, but which were not sampled. Robertson and Barker (2006) expanded on this method by using species richness values and adding heterogeneity variables to the analysis. The methods formulated by Robertson and Barker (2006) were followed here and are described below.

Mean annual rainfall and mean annual temperature were calculated for each QDGC from interpolated climate surfaces produced by the Climate Research Unit and based on weather station data (New et al. 2002). Each Quarter Degree Grid Cell was then plotted in a two-dimensional climate space defined by mean annual rainfall and mean annual temperature (Robertson and Barker 2006). A bubble-plot was used to show how species richness varied within this climate space. The size of each bubble represented the number of species within each QDGC. The climate space was then divided into units of 100 mm rainfall by 2°C temperature. The QDGC with the highest species richness within each climate unit was then identified. QDGCs with less than 50% of this highest value were 'possibly under-sampled' (Robertson and Barker (2006) used the more conservative value of 20%). In addition, if these 'possibly under-sampled' cells had a greater topographic heterogeneity (standard deviation of altitude per QDGC; 900 pixels per QDGC) than the cell with the highest species richness, these were considered to be 'probably under-sampled' (Robertson and Barker 2006). These categories were then mapped to illustrate which parts of South Africa were poorly sampled.

The proportions of well-sampled and under-sampled cells were then calculated for locations near to cities, roads and protected areas; and similarly for locations further from these features. The same distances that had significantly more amphibian records in the previous analysis were used as the threshold value for proximity to features. Results were presented as a stacked bar chart. Chi-square tests for independence were used to determine whether the proportion of 'well-sampled', 'possibly under-sampled' and 'probably under-sampled' QDGCs were the same at locations near to, or far from, geographic features (McDonald 2009). Again, only the *P*-values after correction for multiple tests are given.

Assessment of geographic sampling bias in the GAA species distribution data

The IUCN commissioned the Global Amphibian Assessment to evaluate the status of all amphibians described worldwide. This analysis was completed in 2004 and updated during the 2008 IUCN Red List categorisation (IUCN, Conservation International and NatureServe 2008). A product of this assessment was a set of geographic data that mapped the Extent of Occurrence of each amphibian species (IUCN 2009). Extent of Occurrence is a measure of species distribution defined as the area within the shortest boundary drawn around all presence records of a species (IUCN 2001). This measure reduces omission error, and geographic bias, in presence-only record data by encompassing poorly sampled areas within the boundary of known occurrences. It is therefore possible that GAA Extent of Occurrence data, while undoubtedly based on the SAFAP presence records, is a feasible bias-free alternative for use in conservation plans.

To test whether Extent of Occurrence measures reduce geographic bias, we obtained the GAA amphibian distribution data from the IUCN website (IUCN 2009). The species distribution polygons were spatially joined to the South African QDGCs, and the species richness of each grid cell was calculated based on the number of species polygons it intersected. Since this provided species richness information, we then performed the same species richness analysis of sampling bias that was described above. The same climate units as before were used. Grid cells were classified into "well sampled", "possibly under-sampled" or "probably under-sampled" and these categories were mapped.

Results

Sources of geographic sampling bias in the SAFAP

Collections from the entire dataset contributed an average collection density of 3.35 records per 100 km^2 (or 21 records per QDGC). However, the data were highly skewed, with many QDGCs containing low numbers of records (13.3% had no records) and few cells containing many records (742 specimens from a single cell in northern Swaziland). There was a clear gradient in number of records from the east to the west of the country with parts the northeastern region and eastern coastline having the highest number of records per cell (>101) (Fig. 1). The northwestern region and a few cells along the Drakensburg Mountains showed a paucity of records. The Cape coastline also contained many records, particularly in the surrounds of the city of Cape Town. Urban centres of Johannesburg, Pretoria, Bloemfontein and Durban also showed local increases in the number of records.

The results showed that the observed numbers of records were significantly different from expected in relation to cities ($\chi^2 = 91.452$; $P < 0.0001$), roads ($\chi^2 = 87.060$; $P < 0.0001$) and protected areas ($\chi^2 = 96.814$; $P < 0.0001$). There were substantially more records than expected within 100 km and even 200 km of cities (Fig. 2a). Areas furthest from cities (more than 401 km) were poorly sampled, with 6,697 fewer records than expected. For national roads, there were 4,966 more records than expected within an

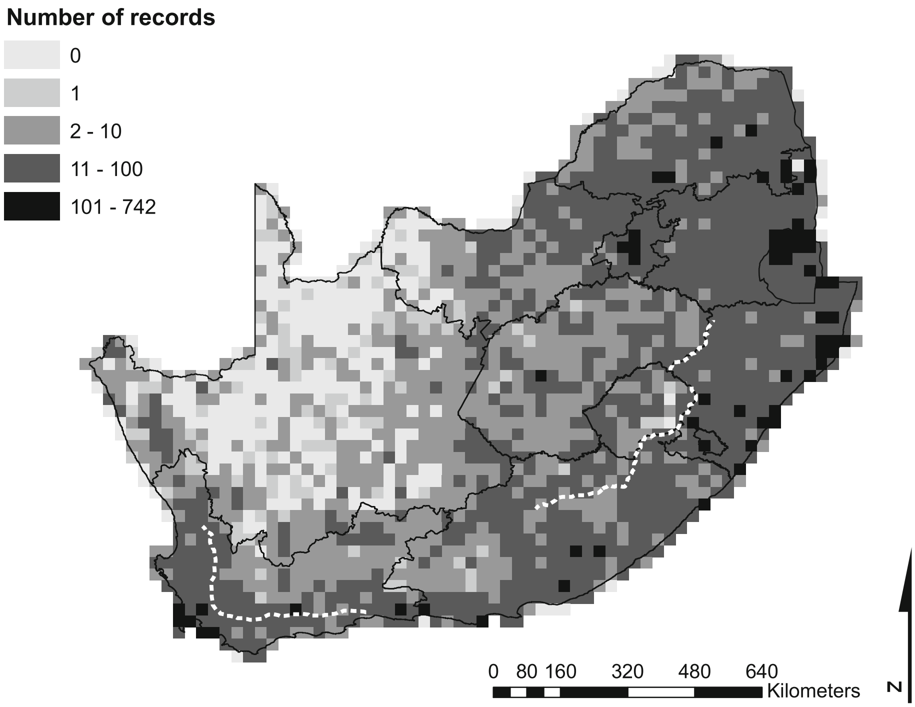

Fig. 1 The spatial distribution of South African Frog Atlas Project records by Quarter Degree Grid Cell. The *white dotted lines* show the locations of the Cape Fold Mountains in the west and the Drakensberg Mountains in the east

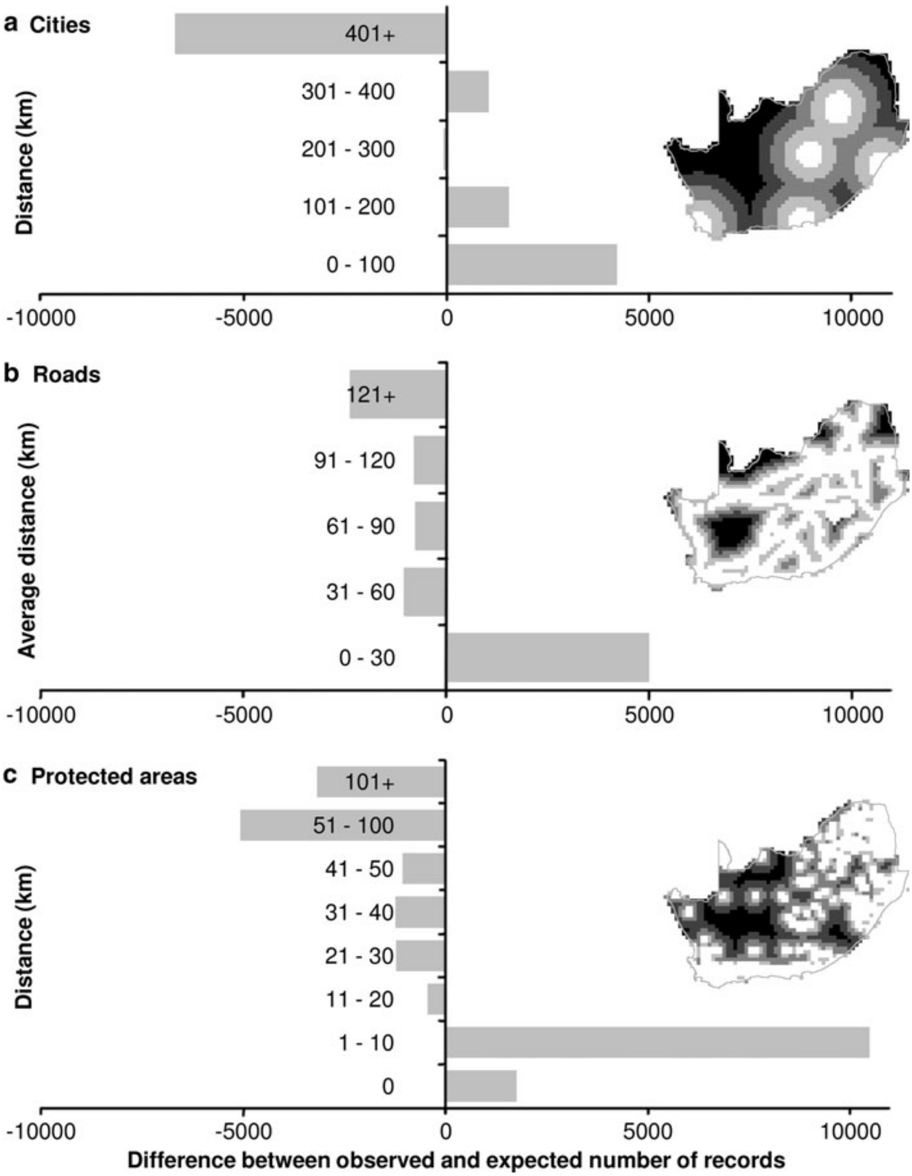

Fig. 2 The difference between the observed number of records and expected number of records within certain distance categories from: **a** cities, **b** national roads and **c** protected areas. The *insets* show maps of the same distance categories as the vertical axes, with progressively *darker shades* indicating increased distance from the features

average of 30 km from national roads (Fig. 2b). Further than 30 km from a national road, there were fewer records than expected. Similar to cities and roads, the degree of sampling showed a decreasing trend with distance from protected areas (Fig. 2c). QDGCs that fell entirely within, or less than 10 km from reserves were well sampled in comparison to other distance classes. The 1–10 km distance category was particularly well sampled, containing 10,482 more records than expected (Fig. 2c). The areas further from reserves, in particular those further than 50 km, had many fewer records than were expected. Most cities, roads and protected areas were located in the eastern half of the country and along the southern coast (see inset maps in Fig. 2). Spearman correlations confirmed the spatial associations between these features ($r = 0.21$–0.40, $P < 0.0001$). This reflects that fact that human population density and infrastructure development is higher in these areas.

Sampling gaps in the SAFAP based on species richness

Species richness showed a similar east–west gradient as collection density, with many more species per grid cell on the eastern side of South Africa, and few species per cell in the north-west (Fig. 3). The highest measure of species richness came from a grid cell on the northeastern coast, with a value of 49 species. The majority of the QDGCs in the country contained between two and ten species.

When species richness was plotted within a two dimensional climate space, the highest species richness values were found with a combination of high temperature and high precipitation (Fig. 4). QDGCs with lower than 400 mm mean annual precipitation had consistently lower measured species richness.

The results of this climate and heterogeneity assessment of sampling bias showed that the eastern and coastal areas were better sampled than the central and western interior (Fig. 5). Thirty-six percent of all QDGCs were well sampled. This analysis highlighted areas of high heterogeneity as the most likely to have been poorly sampled, in particular

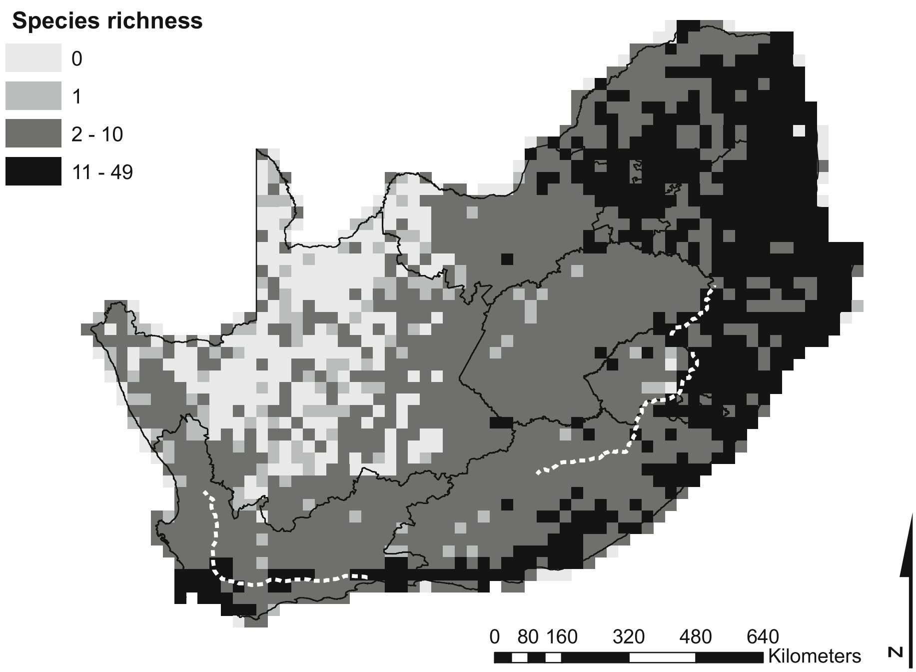

Fig. 3 Measured species richness of amphibians per Quarter Degree Grid Cell. The *white dotted lines* show the locations of the Cape Fold Mountains in the west and the Drakensberg Mountains in the east

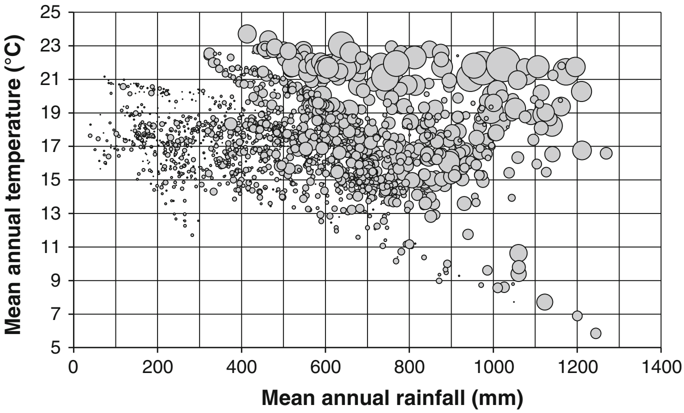

Fig. 4 Each Quarter Degree Grid Cell is a point within a two-dimensional climate space defined by mean annual temperature and mean annual rainfall. The size of the bubble indicates the species richness of the QDGC, such that the larger the bubble the greater the numbers of species recorded in that grid cell. Climate variables are calculated from interpolated climate surfaces produced by the Climate Research Unit and based on weather station data (New et al. 2002)

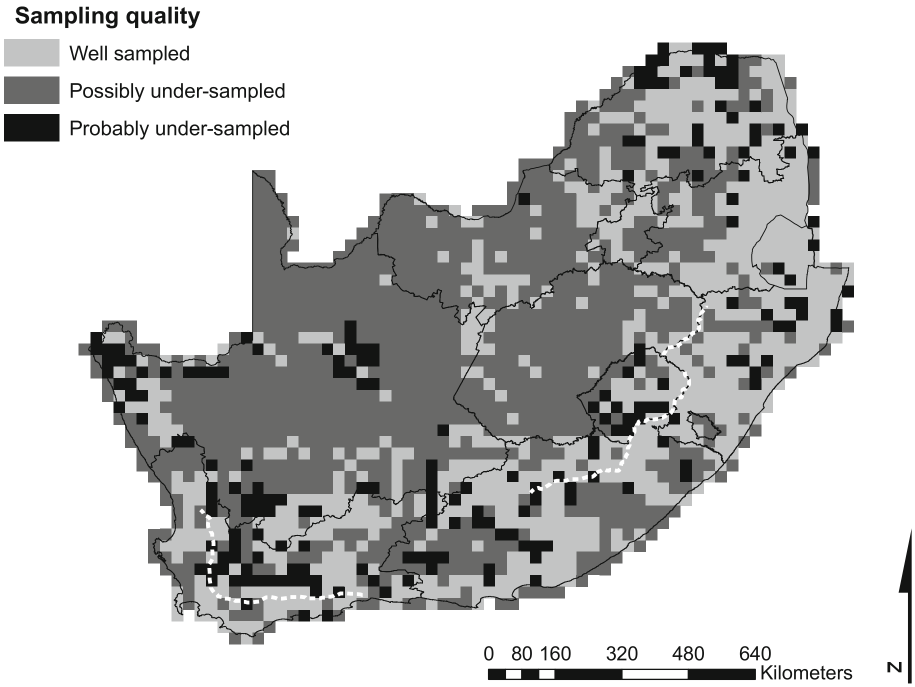

Fig. 5 An assessment of the quality of sampling in the SAFAP database for each Quarter Degree Grid Cell, based on climate and heterogeneity. The pale grey cells have more than 50% of the highest species richness in their climate unit, and are considered to be well sampled. The darker grey cells have lower than 50% of the highest species richness in their climate unit, but also have lower heterogeneity. These cells are possibly under-sampled. The black QDGCs have lower than 50% of the highest species richness in their climate unit, but higher heterogeneity, and are probably under-sampled. The *white dotted lines* show the locations of the Cape Fold Mountains in the west and the Drakensberg Mountains in the east

the cells which occur along the escarpment (Fig. 5). "Probably under-sampled" grid cells made up 10% of the total number of QDGCs.

This species richness assessment again revealed significant differences in the quality of sampling regarding proximity to cities, roads and protected areas. Areas near to these features contained more well-sampled cells and fewer 'probably under-sampled' cells than areas further away. Within 200 km of cities, only 7.0% of QDGCs were 'probably under-sampled' (Fig. 6), while further than 200 km from cities 11.6% of QDGCs fell into this category. This difference was significant according to a χ^2 test ($\chi^2 = 10.742$; $P > 0.01$). Proximity to roads presented a similarly significant difference in sampling quality ($\chi^2 = 26.170$; $P < 0.0001$). In the vicinity of national roads, 41.5% of QDGCs were well sampled. This percentage decreased to 30.4 further than 30 km from roads. Again, protected areas showed the most pronounced difference ($\chi^2 = 170.447$; $P < 0.0001$). Only 25.5% of cells more than 10 km from protected areas were well sampled. Near to protected areas, this measure doubled, with 52.7% well sampled QDGCs within 10 km of protected areas.

Geographic bias in the GAA species distribution data

The GAA polygons reduced much of the geographic sampling bias that was evident in the SAFAP presence data. The number of well-sampled grid cells increased to 86%, which covered the majority of the country including much of the arid western half (Fig. 7). Only a small number of northern QDGCs remain 'possibly under-sampled'. The numbers of "probably under-sampled" QDGCs dropped to less than 2%, but these were still centred on the Cape Fold and Drakensburg Mountains.

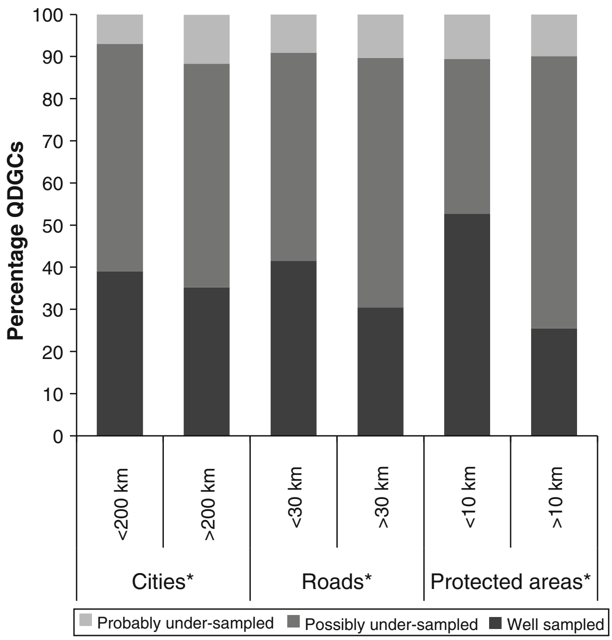

Fig. 6 The percentage of well-sampled and under-sampled Quarter Degree Grid Cells within certain distances of cities, roads and protected areas. An *asterisk* (*) indicates a significant difference according to a χ^2 test for independence ($P < 0.05$)

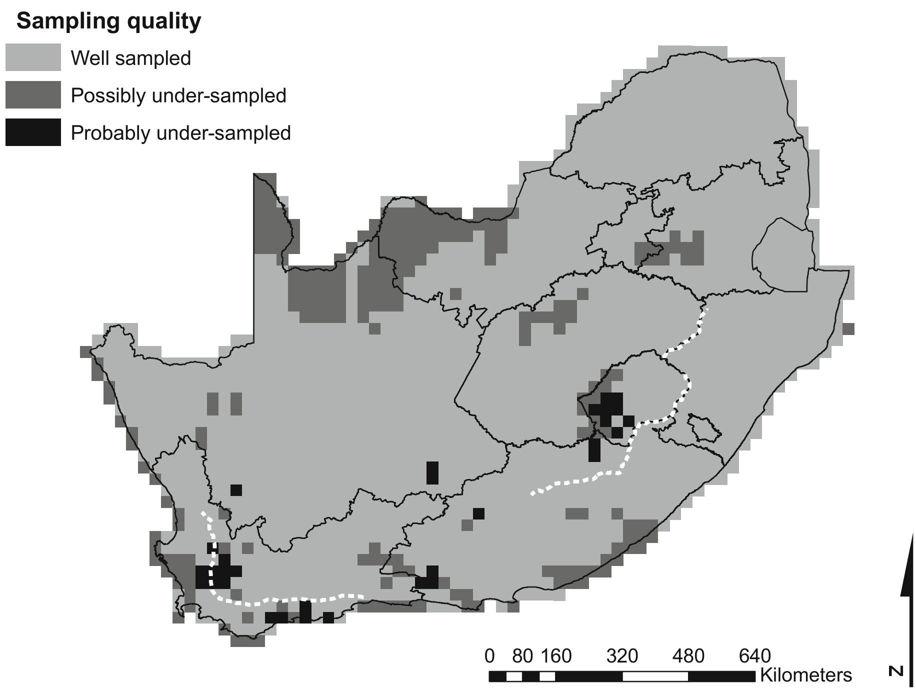

Fig. 7 An assessment of the quality of sampling of GAA species distributions for each Quarter Degree Grid Cell, based on climate and heterogeneity. The pale grey cells have more than 50% of the highest species richness in their climate unit, and are considered to be well sampled. The darker grey cells have lower than 50% of the highest species richness in their climate unit, but also have lower heterogeneity. These cells are possibly under-sampled. The black QDGCs have lower than 50% of the highest species richness in their climate unit, but higher heterogeneity, and are probably under-sampled. The *white dotted lines* show the locations of the Cape Fold Mountains in the west and the Drakensberg Mountains in the east

Discussion

The observed distribution of amphibian records in the SAFAP results from a combination of real biological pattern and geographic sampling bias. In deciding whether conservation planning can be reliably based on an atlas dataset, researchers must make a determination of which of these two processes has been foremost in producing the observed pattern (Funk et al. 1999). If variation in sampling intensity is exceedingly high, then sampling bias may completely obscure the real biological pattern. In this situation, use of the dataset for any form of conservation planning would be inappropriate. However, many authors believe that the general patterns of species richness and species range size can be reliably detected despite known geographical bias (Williams et al. 2002; Parnell et al. 2003). The acceptable level of geographic bias depends on many factors, including the objectives, scope, data formats and methodology employed within any particular conservation plan (Rouget et al. 2004; Rondinini et al. 2006; Robertson et al. 2010). This research provides both a quantitative and spatial assessment of geographic bias in the SAFAP, so that conservation planners can make an informed decision about whether the data are sufficiently complete to provide a reliable assessment.

In comparison to some other studies, the 3.35 amphibians recorded for every 100 km^2 represents relatively sparse sampling. Dunn and Weston (2008) conducted a review of bird atlases, and found that sampling intensity ranged from 2 to 8,480 records per 100 km^2 (with an average of 470 records per km^2). The first South African Bird Atlas amassed 7 million records of avian occurrence, a collection density of over 550 records per 100 km^2 (Harrison et al. 2008). This was due to the participation of over 5,000 members of the public, who added substantially to the database (Harrison et al. 2008). In contrast, only 420 volunteers submitted amphibian records to the SAFAP, and the majority of the records came from experienced herpetologists (Minter et al. 2004). This reflects not only the lower level of public interest in amphibians, but also some of the difficulties involved in frog atlasing: amphibians are nocturnal, they are generally restricted to water sources, are often only seasonally active and cryptic species require expert knowledge to identify with certainty. However, conservation plans that only include well-sampled ‘indicator’ taxa are not representative of all aspects of biodiversity (Larsen et al. 2009). The SAFAP is currently the most comprehensive dataset for amphibians in South Africa (Minter et al. 2004). It presents the only opportunity for this often-overlooked taxon to be included in conservation plans.

Collection density of the SAFAP records varied widely over the country. This variation followed a predictable pattern with a clear east-to-west gradient in collection density. The eastern parts of the country and the southern coastline had many more occurrence records than the arid central western interior. There was also a significant association between the location of amphibian records and human infrastructure. Cities, roads and protected areas are most commonly located within the eastern part of the country, where human population density is highest (Chown et al. 2003; Evans et al. 2006). The association between these features and a high collection density may be only a consequence of spatial congruence between amphibian distribution patterns and human activities. Evans et al. (2006) found that human population density in South Africa was well correlated with both bird and amphibian richness along this east-to-west gradient. Neither human populations nor amphibian species diversity were concentrated in desert areas. This relationship between species richness and human population density is also evident for other vertebrate taxa in Africa and is primarily attributed to similar use of environmental energy (Reddy and Dávalos 2003, Chown et al. 2003; Evans et al. 2006). Temperature, precipitation and net primary productivity all follow an analogous variation from high values in the eastern part of South Africa to lower values in the west (Evans et al. 2006; Chown et al. 2003). This finding complicates the separation between sampling bias and real biological pattern that is necessary to assess the worth of atlas data for conservation planning. Evans et al. (2006) conclude that bias in sampling effort plays only a minor part, if any, in this correlation. Conversely, Reddy and Dávalos (2003) suggest that sampling bias may actually be responsible for the correlation due to high sampling in populated areas.

A species richness approach, which accounted for climatic variables and topographical heterogeneity, provided an alternative measure of sampling quality to facilitate the resolution of this issue. Only cells with similar climatic conditions were compared and then those with lower topographic heterogeneity were held to a lower standard of expected species richness. A possible limitation of this method is that the choice of a threshold for poorly sampled areas is arbitrary. Robertson and Barker (2006) used a threshold of 20% of the value in the most species rich cell. Increasing the threshold value to 50% meant that our assessment of sampling quality was more rigorous and more cells were considered undersampled. There was a distinct relationship between the climate in a QDGC and its amphibian species richness. Had this not been the case, the use of this method would have been uninformative. High species richness was associated with a combination of high mean annual rainfall and high mean annual temperature. Cells with low rainfall (below 400 mm) all had low species richness values. Robertson and Barker (2006) found a similar relationship with rainfall when assessing species richness maps from the National Herbarium's plant database for southern Africa.

This method reiterates the fact that large parts of the central western interior were under-sampled. This result is comparable to that of Robertson and Barker (2006) who found that plants were also under-sampled in this region. Similarities in the locations of under-sampled areas between amphibians and plants suggest that factors determining geographic sampling bias may be consistent across taxa. 'Probably under-sampled' were many topographically heterogeneous cells along the Cape Fold Mountains and Drakensburg escarpment. The same areas of incomplete coverage acknowledged by the editors of the SAFAP are highlighted here, namely arid and montane regions (Minter et al. 2004). Moreover, the specific grid cells in which this under-sampling manifested were pinpointed.

Relating this species richness measure of sampling quality to cities, roads and protected areas confirmed that the higher numbers of amphibian records in close proximity to these features were the result of significantly more well sampled cells. Herpetologists involved in surveying frogs concentrated their efforts within a 100 km radius of their residences and places of work in the major cities: the syndrome of the easy day trip. When travelling further afield, they followed the major national roads and sampled preferably within 30 km of these convenient routes. In addition to bias towards easily accessed areas, our results indicate that researchers do focus on areas that they believe to have a higher biodiversity. There were significantly more well sampled QDGCs within 10 km of protected areas. Researchers likely target these areas due to their natural state and perceived higher biodiversity. Roads and protected areas may also provide sampling opportunities in publically accessible areas. This minimises potentially time consuming and complex negotiations with multiple private landowners. Thus, despite the attempt at systematic sampling on the part of the atlas researchers, cities, roads and protected areas received a disproportionate amount of sampling effort.

We have shown that the SAFAP data contains significant geographical bias and substantial sampling gaps, but is this bias extensive enough to mask the real pattern of biodiversity? There is some evidence that suggests that the observed patterns represent true amphibian species richness and abundance in South Africa. The broad scale east-to-west gradient is prominent in both the numbers of records and species richness. This gradient follows a real moisture gradient across the country and species richness was consistently low in low rainfall regions. Amphibian reproduction is highly dependent on the availability of standing water and the pattern of rainfall (Carey and Alexander 2003; McCarty 2001) and thus this relationship is theoretically valid. The observed species richness patterns are also consistent with biogeographical theories concerning the relative distributions of tropical and temperate species (see Alexander et al. 2004). Furthermore, this broad east-to-west gradient is not unique to amphibians. Species richness of birds recorded from the South African Bird Atlas Project (Fairbanks et al. 2002) and plants from the National Herbarium's plant database (PRECIS; Robertson and Barker 2006) showed a similar gradient. In the SAFAP, it is likely that both the real pattern in biodiversity and the pattern of sampling bias both follow an east-to-west gradient. Thus, the east-to-west gradient in amphibian richness is a real pattern, but it may well have been exaggerated by sampling bias. While the broadest scale patterns are generally discernable from the SAFAP, significant levels of geographic sampling bias mean that the dataset may not be sufficiently accurate at the finer scales optimal for systematic conservation planning (Rouget et al. 2004).

The GAA species distribution data generalised the pattern of amphibian species richness in South Africa, decreasing local heterogeneity in the number of species recorded. This reduced, but did not altogether eliminate, the geographic bias that was evident within the SAFAP. The same arid and mountainous regions still displayed some geographic bias. Whilst minimising much of the concern with bias associated data, these simplified species boundaries do not necessarily provide a superior alternative to biological atlases. The global nature of the GAA data means that they provide little more than the broad-scale amphibian distribution patterns already apparent within the SAFAP data. Despite decreased omission error, the data are now subject to increased commission error: the incorrect assumption of species presence (Rondinini et al. 2006). Commission error renders most alternative datasets ineffective for conservation planning at a scale suitable for implementation. Accurate, up-to-date and formally documented species occurrence data is generally a better option for including species distributions into fine scale conservation planning. Until a dataset is available that minimises both omission and commission error, conservationists must consider the implications of including flawed data in their plans.

Any biodiversity-based conservation planning requires the comparison of biodiversity between regions. In datasets that contain significant geographic sampling bias, these kinds of comparisons are ill-advised, since species richness can be both under- and over-estimated relative to regions with differing sampling intensity (Reddy and Dávalos 2003; Williams et al. 2002; Robertson et al. 2010; Boakes et al. 2010). Over-estimated biodiversity may seem an unlikely difficulty, since the better sampled an area, the closer the measured species richness is to the true value. However, when certain areas are preferentially sampled, the species richness appears elevated in relation to poorly sampled areas. Conservation plans will consequently afford greater priority to those areas for which sampling was more intense, rather than areas with real high biodiversity. If the distributions of important species are included in the conservation plans, geographic sampling bias will have a similar effect. These species will have had a higher probability of detection, and more presence records, in well-sampled grid cells and will therefore impart higher irreplaceability to these areas.

The association between high sampling intensity and human infrastructure is likely to increase conservation conflicts around metropolitan areas. Conservation conflicts occur when high biodiversity coincides spatially with high human activity, and land-use planners must make difficult decisions between conservation and development (Van Rensburg et al. 2004; Chown et al. 2003). Conservation conflicts may be unavoidable when species richness correlates with high human population density, as is the case in South Africa (Chown et al. 2003; Van Rensburg et al. 2004; Evans et al. 2006). However, it would be unfortunate if these problems were amplified by sampling bias. Geographic sampling bias towards cities and roads in the SAFAP means that the perception of species richness near to these features is inflated relative to other areas. The cells surrounding cities and roads will contain an apparently high number of species occurrences within a smaller area, increasing their value in complementary selection algorithms. Areas near to human infrastructure are generally highly transformed and priority for reserve selection in these areas should be avoided rather than enhanced.

The symptoms of this bias are evident in the irreplaceability map for endemic and threatened animals in the National Spatial Biodiversity Assessment. Seven datasets, including the SAFAP, formed the basis for this map (Rouget et al. 2004). Grid cells in the province with the highest human population density, Gauteng, showed high irreplaceability scores in this assessment. Other populous cities: Cape Town, Durban and Port Elizabeth also had high irreplaceability. This bias is also carried through to regional biodiversity assessments, such as the Gauteng State of the Environment Report, which asserts that the province "represents a relatively large proportion of South Africa's biodiversity in a small area" (GDACE 2004). In Gauteng and other provinces with high population density, the priority for conservation may be exaggerated due to high sampling intensity and this may be incompatible with essential social development.

Another important part of systematic conservation planning is the assessment of the current reserve system. This gap-analysis stage identifies how well current reserves meet the conservation targets (Margules and Pressey 2000). Historically, issues of biodiversity conservation received little attention during the placement of South African reserves. The current protected area system is thus an inadequate representation of biodiversity (Driver et al. 2005). However, if a biased dataset is used in the evaluation of current reserves, they may appear to conserve either high species richness or a high percentage of an individual species distribution.

In the SAFAP database, there were considerably more amphibian records and double the number of well-sampled QDGCs within 10 km of protected areas. This result was similar to that of Reddy and Dávalos (2003) who found that bird species where much better sampled within and surrounding the protected areas of sub-Saharan Africa. Since conservation of frog species is rarely the primary justification for the establishment of protected areas and amphibians are not restricted to reserves, this situation signifies a legitimate and substantive sampling bias. This bias will negatively affect evaluations of the effectiveness of the current reserve network for the conservation of amphibians. Current reserves will apparently contribute highly to amphibian biodiversity targets, when in reality these areas have just been preferentially sampled in comparison to other regions. Using the GAA alternative data will yield a similar problem. Commission error may mean that many species appear to be protected when in reality there are few actual occurrences within the current reserve system (Rondinini et al. 2006). Furthermore, this may result in future conservation plans assigning priority to areas in which the important biodiversity features do not occur (Rondinini et al. 2006).

Over-estimated species richness may falsely enhance the conservation priority of certain areas, or make protected areas seem particularly effective. Under-estimated species richness is an equally severe problem. It causes conservation plans to neglect areas that should have received high priority (Reddy and Dávalos 2003). Sampling of amphibians was poor in remote areas, which consequently had low measured species richness. In the SAFAP, large parts of the arid west had poor sampling intensity. These areas represent almost the entirety of the Succulent Karoo, Nama-Karoo and Desert biomes. While it is unlikely that these arid areas would ever achieve the high species richness of the tropical eastern parts of the country, additional records will possibly result in the discovery of further species occurrences. Thus, these areas may have an important contribution to make to biodiversity conservation in the country, but at present, they will be overlooked during the conservation planning process. Exchanging SAFAP presence records for species distribution polygons in poorly sampled areas would enhance the inclusion of these areas into conservation plans. However, when few occurrence data are available (as is the case for arid specialist amphibians in South Africa), the construction of species distributions becomes more speculative and less precise (Rondinini et al. 2006), with obvious implications for the accuracy of conservation plans.

Gaston and Rodrigues (2003) tested the effects of different data types on reserve selection, and found that even presence-absence datasets with low sampling effort can be effective in producing reasonable reserve networks. However, the scenario that they applied maintained equal sampling for every grid cell i.e. no geographic bias. Thus, if the geographic bias in the SAFAP could be minimised without concurrent increase in commission error, the dataset could become invaluable to conservation planning. To achieve this, sampling should be intensified in areas and for taxa known to have poor sampling effort in the past (Parnell et al. 2003; Reddy and Dávalos 2003; Küper et al. 2006). The SAFAP itself has been helpful in identifying the areas that require additional sampling (Donald and Fuller 1998; Dennis and Shreeve 2003; Graham et al. 2004). Dennis and Shreeve (2003) found that focused sampling as part of a new atlas of French butterflies addressed extensive under-sampling in previous atlases.

The geographic bias within the SAFAP could be minimised by additional targeted sampling. The arid northwestern region of the country held few well-sampled QDGCs and many grid cells with no amphibian records. This entire region would benefit from additional sampling. The distributions of common and widespread species undoubtedly extend further into this region than is currently documented (Minter et al. 2004). More records of little known arid specialists would improve knowledge of their distributions and life histories. Precipitation is scarce and unpredictable in this region. In these situations, amphibians are generally fossorial and breed opportunistically, awaiting good conditions before emerging (Skelly et al. 2003). It may take many sampling trips, specifically timed to brief periods of amphibian activity, before the bias in this region is completely reversed. Conversely, the under-sampled QDGCs in topographically heterogeneous mountainous areas require only that researchers overcome the logistical travel difficulties. These areas should then return additional records and species occurrences within a short timeframe. The Cape Fold Mountains, Lesotho Highlands, Limpopo Highlands and some QDGCs along the Eastern Escarpment should be the target of supplementary sampling.

It is unlikely that any projects harnessing ‘citizen science’ could be completely free from geographic sampling bias. Volunteers are increasingly vital contributors to broad-scale atlas datasets (Boakes et al. 2010), but systematic and standardised sampling cannot be demanded of them (Robertson et al. 2010). Since standardised surveys are seldom achievable, it is imperative that sampling effort is precisely recorded. Person-hours, the number of visits to an area, the distance travelled whilst observing, search methods or some combination of similar survey effort variables must be documented (Dennis et al. 1999; Gaston and Rodrigues 2003; Romo et al. 2006; Robertson et al. 2010). The second South African Bird Atlas Project attempts to achieve this by requiring a strict sampling protocol per datasheet submitted. A single datasheet provides for a maximum 5-day sampling period in a particular pentad. Sampling must begin with an initial continuous 2 h of observation, and hourly observations thereafter must be noted (www.sabap2.adu.org.za). Thus, a temporal unit of sampling has been created (Robertson et al. 2010). Furthermore, technological advances have increased the options for communication between project administrators and volunteers. Improved information dissemination means that participants can be notified of locations requiring additional sampling whilst the project is ongoing (Robertson et al. 2010). Again, the South African Bird Atlas 2 website provides a monthly updated gap analysis map and reports on the achievement of the project goals (www.sabap2.adu.org.za).

Improved surveys will no doubt amass far superior data in terms of coverage, comprehensiveness and quality. However, there is urgency in the requirement for data with which to analyse the anthropogenic effects on species ranges and for the prioritisation of conservation. Concern for certain taxa, especially amphibians, means that investigations aimed at identifying threats and creating conservation plans must proceed without unnecessary delay (Skelly et al. 2003; Tyre et al. 2003).

In the interim, the SAFAP presents the best available dataset of amphibian occurrence in South Africa. Conservation planners will have to employ methods that reduce the effects of geographic sampling bias. While using GAA species distributions presents an easily available opportunity to reduce geographic bias, the global scale of this data means that uncertainties arising from commission error are maximised. We have provided a simple score of sampling quality in the SAFAP based on climate and heterogeneity for each QDGC. This could be used to weight analyses in a manner similar to initiatives that have measured recorder effort (Elphick 1997). Other options include bioclimatic modelling, which can fill gaps in sampling and enhance incomplete datasets. Modelling predicts the probability of occurrence of species based on environmental relationships discovered from the primary data (Dennis and Shreve 2003; Reddy and Dávalos 2003; Graham et al. 2004; Segurado and Araújo 2004; Küper et al. 2006). Prior to this assessment of sampling bias, there were only vaguely stated concerns about the standard of SAFAP occurrence data and its potential biases. The applicability of the SAFAP as the primary data for niche modelling was indeterminate. It would now be possible to state the sampling quality of QDGCs included in particular species models, and hence give an assessment of model quality. Some conservation planners choose only to include adequately sampled species and regions. For this reason, only threatened and endemic amphibians were included in the National Spatial Biodiversity Assessment (Rouget et al. 2004). In areas with poor biodiversity data, expert knowledge can be used to define areas of high biodiversity (Berliner and Desmet 2007). All of these measures rely to a greater or lesser extent on the original survey data and geographic bias will influence their accuracy. We therefore recommend that a detailed description of geographic bias accompany any use of an atlas dataset. In addition, preliminary conservation plans constructed using biased data must be re-evaluated once improved data becomes available.

Acknowledgments We acknowledge the Animal Demography Unit of the University of Cape Town for the use of the South African Frog Atlas Project data. The Andrew W. Mellon Postgraduate Mentoring Programme at the University of the Witwatersrand provided financial support. We thank four anonymous reviewers for comments that improved the manuscript.

References

Alexander GJ, Harrison JA, Fairbanks DH, Navarro RA (2004) Biogeography of the frogs of South Africa, Lesotho and Swaziland. In: Minter LR, Burger M, Harrison JA, Braack HH, Bishop PJ, Kloepfer D (eds) Atlas and Red Data Book of the Frogs of South Africa, Lesotho and Swaziland. SI/MAB Series #9. Smithsonian Institution, Washington DC

Berliner D, Desmet P (2007) Eastern Cape Biodiversity Conservation Plan: Technical Report. Department of Water Affairs and Forestry Project No 2005-012, Pretoria

Boakes EH, McGowan PJK, Fuller RJ, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8:e1000385

Carey C, Alexander MA (2003) Climate change and amphibian declines: is there a link? Divers Distrib 9:111–121

Chown SL, Van Rensburg BJ, Gaston KJ, Rodrigues ASL, Van Jaarsveld AS (2003) Energy, species richness, and human population size: conservation implications at a national scale. Ecol Appl 13:1233–1241

Dennis RLH, Shreeve TG (2003) Gains and losses of French butterflies: tests of predictions, under-recording and regional extinction from data in a new atlas. Biol Conserv 111:131–139

Dennis RLH, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: the effects of sampling effort. J Insect Conserv 3:33–42

Desmet P, Cowling R (2004) Using the species-area relationship to set baseline targets for conservation. Ecol Soc 9:11 (online) http://www.ecologyandsociety.org/vol9/iss2/art11. Cited 30 July 2010

Desmet P, Schaller R, Skowno A (2008) North West Province Conservation Assessment Technical Report. North West Province Department of Agriculture, Conservation & Environment, Mafikeng

Donald PF, Fuller RJ (1998) Ornithological atlas data: a review of uses and limitations. Bird Study 45:129–145

Driver A, Maze K, Rouget M, Lombard AT, Nel J, Turpie JK, Cowling RM, Desmet P, Goodman P, Harris J, Jonas Z, Reyers B, Sink K, Strauss T (2005) National Spatial Biodiversity Assessment 2004: priorities for biodiversity conservation in South Africa. Strelitzia 17. South African National Biodiveristy Institute, Pretoria

Dunn AM, Weston MA (2008) A review of bird atlases of the world and their application. Emu 108:42–67

Eken G, Bennun L, Brooks TM, Darwall W, Fishpool LDC, Foster M, Knox D, Langhammer P, Matiku P, Radford E, Salaman P, Sechrest W, Smith ML, Spector S, Tordoff A (2004) Key biodiversity areas as site conservation targets. Bioscience 54:1110–1118

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

Evans KL, Van Rensburg BJ, Gaston KJ, Chown SL (2006) People, species richness and human population growth. Glob Ecol Biogeogr 15:625–636

Fairbanks DHK, Kshatriya M, Van Jaarsveld AS, Underhill LG (2002) Scales and consequences of human land transformation on South African avian diversity and structure. Anim Conserv 5:61–73

Ferrar AA, Lötter MC (2007) Mpumalanga Biodiversity Conservation Plan Handbook. Mpumalanga Tourism and Parks Agency, Nelspruit

Ferreira D, Malard F, Dole-Olivier M, Gilbert J (2007) Obligate groundwater fauna of France: diversity patterns and conservation implications. Biodivers Conserv 16:567–596

Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst Biol 51:331–363

Ferrier S, Powell GVN, Richardson KS, Mannion G, Overton JM, Allnutt TF, Cameron SE, Mantle K, Burgess ND, Faith DP, Lamoreux JF, Kier G, Hijmans RJ, Funk VA, Cassis GA, Fisher BL, Flemons P, Lees D, Lovett JC, Van Rompaey RSAR (2004) Mapping more of terrestrial biodiversity for global conservation assessment. Bioscience 54:1101–1109

Freitag S, Hobson C, Biggs HC, Van Jaarsveld AS (1998) Testing for potential survey bias: the effect of roads, urban areas and nature reserves on a southern African mammal data set. Anim Conserv 1:119–127

Funk VA, Richardson KS (2002) Systematic data in biodiversity studies: use it or lose it. Syst Biol 51:303–316

Funk VA, Zermoglio MF, Nasir N (1999) Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. Biodivers Conserv 8:727–751

Gaston K, Rodrigues ASL (2003) Reserve selection in regions with poor biological data. Conserv Biol 17:188–195

GDACE (2004) Chapter 8: Biodiversity. In: Gauteng Conservation State of the Environment Report. Gauteng Department of Agriculture, Conservation and the Environment, Gauteng Provincial Government

Graham CH, Ferrier S, Huettman F, Moritz C, Townsend Peterson A (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol 19:497–503

Harrison JA, Minter LG, Burger M (2004) Atlas and Red Data book for frogs completed. S Afr J Sci 100:11–13

Harrison JA, Underhill LG, Barnard P (2008) The seminal legacy of the South African Bird Atlas Project. S Afr J Sci 104:82–84

IUCN (2001) Categories & criteria (version 3.1). Prepared by the IUCN Species Survival Commission. IUCN, Gland. http://www.iucnredlist.org/static/categories_criteria_3_1. Cited 23 April 2009

IUCN (2009) IUCN red list of threatened species. Version 2009.1. http://www.iucnredlist.org. Cited 15 June 2009

IUCN, Conservation International and Nature Serve (2008) An analysis of amphibians on the 2008 IUCN Red List. http://www.iucnredlist.org/amphibians. Cited 21 April 2009

Küper W, Sommer JH, Lovett JC, Barthlott W (2006) Deficiency in African plant distribution data—missing pieces of the puzzle. Bot J Linn Soc 150:355–368

Larsen FW, Bladt J, Rahbek C (2009a) Indicator taxa revisited: useful for conservation planning? Divers Distrib 15:70–79

Larsen R, Homern T, Prager SD, Maliti H, Roskaft E (2009b) Using the extended quarter degree grid cell system to unify mapping and sharing of biodiversity data. Afr J Ecol 47:382–392

Lenth RV (2006) Java applets for power and sample size [Computer software]. http://www.stat.uiowa.edu/~rlenth/Power. Cited 26 Feb 2009

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–253

McCarty JP (2001) Ecological consequences of recent climate change. Conserv Biol 15:320–331

McCollin D, Moore L, Sparks T (2000) The Flora of a cultural landscape: environmental determinants of change revealed using archival sources. Biol Conserv 92:249–263
McDonald JH (2009) Handbook of Biological Statistics, 2nd edn. Sparky House Publishing, Baltimore, Maryland
Minter LR, Harrison JA, Burger M, Braack HH (2004) Introduction. In: Minter LR, Burger M, Harrison JA, Braack HH, Bishop PJ, Kloepfer D (eds) Atlas and Red Data Book of the Frogs of South Africa, Lesotho and Swaziland. SI/MAB Series #9. Smithsonian Institution, Washington DC
New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. Clim Res 21:1–25
Parnell JAN, Simpson DA, Moat J, Kirkup DW, Chantaranothai P, Boyce PC, Bygrave P, Dransfield S, Jebb MHP, Macklin J, Meade C, Middleton DJ, Muasya AM, Prajaksood A, Pendry CA, Pooma R, Suddee S, Wilkin P (2003) Plant collecting spread and densities: their potential impact on biogeographical studies in Thailand. J Biogeogr 30:193–209
Pressey RL (2004) Conservation planning and biodiversity: assembling the best data for the job. Conserv Biol 18:1677–1681
Quinn GP, Keough MJ (2002) Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge
Reddy S, Dávalos L (2003) Geographic sampling bias and its implications for conservation priorities in Africa. J Biogeogr 30:1719–1727
Robertson MP, Barker NP (2006) A technique for evaluating species richness maps generated from collections data. S Afr J Sci 102:78–84
Robertson MP, Cumming GS, Erasmus BFN (2010) Getting the most out of atlas data. Divers Distrib 16:363–375
Romo H, García-Barros E, Lobo JM (2006) Identifying recorded-induced geographic bias in an Iberian butterfly database. Ecography 29:873–885
Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP (2006) Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol Lett 9:1136–1145
Rouget M, Reyers B, Jonas Z, Desmet P, Driver A, Maze K, Egoh B, Cowling RM, Mucina L, Rutherford MC (2004) South African National Spatial Biodiversity Assessment 2004: Technical Report. Volume 1: Terrestrial Component. South African National Biodiversity Institute, Pretoria
Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. J Biogeogr 31:1555–1568
Skelly DK, Yurewicz KL, Werner EE, Relyea RA (2003) Estimating decline and distributional change in amphibians. Conserv Biol 17:744–751
South African Bird Atlas 2 website, http://sabap2.adu.org.za. Cited 30 July 2010
Telfer MG, Preston CD, Rothery P (2002) A general method for measuring relative change in range size from biological atlas data. Biol Conserv 107:99–109
Turner AA, De Villiers AL (2007) Chapter 4: amphibians. In: Western Cape State of Biodiversity Report 2007. Scientific Services, CapeNature
Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecol Appl 13:1790–1801
Van Jaarsveld AS, Freitag S, Chown SL, Muller C, Kock S, Hull H, Bellamy C, Krüger M, Endrody-Younga S, Mansell MW, Scholtz CH (1998a) Biodiversity assessment and conservation strategies. Science 279:2106–2108
Van Jaarsveld AS, Gaston KJ, Chown SL, Freitag S (1998b) Throwing biodiversity out with the binary data? S Afr J Sci 94:210–214
Van Rensburg BJ, Erasmus BFN, Van Jaarsveld AS, Gaston KJ, Chown SL (2004) Conservation during times of change: correlations between birds, climate and people in South Africa. S Afr J Sci 100: 256–272
Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area selection. J Bioscience 27:327–338