



Getting the most out of atlas data

M. P. Robertson^{1*}, G. S. Cumming² and B. F. N. Erasmus³

¹Centre for Invasion Biology, Department of Zoology & Entomology, University of Pretoria, Pretoria 0001, South Africa, ²Percy FitzPatrick Institute for African Ornithology, DST/NRF Center of Excellence, University of Cape Town, Rondebosch, Cape Town 7701, South Africa, ³School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private bag 3, WITS 2050, Johannesburg, South Africa

ABSTRACT

Aim To review some of the applications in ecology and conservation biogeography of datasets derived from atlas projects. We discuss data applications and data quality issues and suggest ways in which atlas data could be improved.

Location Southern Africa and worldwide.

Methods Atlas projects are broadly defined as collections or syntheses of original, spatially explicit data on species occurrences. We review uses of atlas datasets and discuss data quality issues using examples from atlas projects in southern Africa and worldwide.

Results Atlas projects must cope with tradeoffs between data quality and quantity, standardization of sampling methods, quantification of sampling effort, and mismatches in skills and expectations between data collectors and data users. The most useful atlases have a good measure of sampling effort; include data collected at a fine enough resolution to link to habitat variables of potential interest; have a sufficiently large sample size to work with in a multivariate context; and offer clear, quantitative indications of the quality of each record to allow for the needs of users who have specific demands for high-quality data.

Main conclusions Atlases have an important role to play in biodiversity conservation and ideally should aim to offer reliable, high quality data that can withstand public, scientific and legal scrutiny.

Keywords

Atlas projects, biodiversity databases, conservation biogeography, data quality, distributions, GIS.

*Correspondence: M. P. Robertson, Centre for Invasion Biology, Department of Zoology & Entomology, University of Pretoria, Pretoria, 0001 South Africa.
E-mail: mrobertson@zoology.up.ac.za

INTRODUCTION

Species distribution data are of central importance to documenting and conserving biodiversity. The newly emerging field of Conservation Biogeography is concerned with the distributional dynamics of species and how they relate to the conservation of biodiversity (Whittaker *et al.*, 2005). Conservation biogeography relies directly on distribution data to address a range of conservation problems. However, distributions of many taxa are poorly known, and data quality varies among taxa and regions (Whittaker *et al.*, 2005). The distributions of birds and large mammals appear to be reasonably well known, but for many other groups knowledge is poor (Donald & Fuller, 1998). Even for birds, some regions of high biodiversity have been poorly surveyed (Gibbons *et al.*, 2007;

Dunn & Weston, 2008). The quantity and quality of distributional data can have a profound influence on the quality of products that are then used to direct (or misdirect) conservation action. Atlas projects have an important role to play in collecting and managing high-quality distributional data that can be applied to a range of issues in the field of conservation biogeography (Harrison *et al.*, 2008). In addition, advances in the field of conservation biogeography, including the development and testing of new biogeographic theories, the application of biogeographic principles and the development of new analytical techniques, rely directly on large datasets of high quality distribution data that atlas projects can gather (Whittaker *et al.*, 2005).

Atlases are broadly defined as datasets of primary, spatially explicit data on species occurrences (Dunn & Weston, 2008).

They usually aim to collect occurrence data (by means of field observations) for a specific group of organisms, e.g. breeding birds or butterflies. They typically cover the full extent of a discrete and clearly defined geographic area (map region). Atlas projects are usually designed to collect data within a specific time period (usually several years), but they may be repeated to enable temporal comparisons. Most atlases use a predefined grid for sampling, employ a sampling protocol and have a minimum set of requirements for the submission of records. Although most atlases record the presence or abundance of a set of species in a grid cell (e.g. many bird atlases) they may also incorporate point observations (e.g. Southern African Plant Invaders Atlas). An important attribute of atlas projects is that they often rely on volunteers, who are usually amateurs rather than scientists, to collect the data. In addition, each atlas project has a coordinator (or project team) to direct and manage the project.

Atlas data can be distinguished from other types of spatially explicit biological data. Data taken from the labels of specimens that are housed in museums and herbaria have been referred to as 'collections data' (Funk & Richardson, 2002). Collections data are based exclusively on specimens collected by scientists, whereas atlas data are usually not specimen based. Although specimens are not usually collected with atlas data, in some cases, all available specimen data (from specimen labels or collections databases) are assembled prior to the commencement of the atlas project. Unlike atlas data, collections data are not usually gathered as part of a project with a specific aim, sampling protocol and time frame by a coordinated group of observers. Collections data are usually point records, whereas atlas data are most often presence or abundance records made within a grid cell. Brotons *et al.* (2007) made the distinction between atlas projects and long-term monitoring (LTM) programs. They defined LTM programs as those based on a network of sampling locations at which species occurrence and relative abundance are collected at given time steps to document temporal trends. An example is the Catalan common bird survey for which birds are recorded along a set of 3-km transects that are visited twice during the breeding season (Brotons *et al.*, 2007). In this article, we also distinguish atlas datasets from surveys of particular sites (e.g. vegetation plots or transect based datasets) by expert scientists, even if these do not involve repeated sampling.

Atlas projects are undertaken with a variety of different objectives in mind (Donald & Fuller, 1998; Gibbons *et al.*, 2007; Dunn & Weston, 2008). Most objectives will overlap to some extent, although particular kinds of analysis may demand different tradeoffs between data quality and data quantity. The uses of atlas data range from purely ecological explorations (such as biogeography or niche modelling) through assessments of anthropogenic impacts (climate change, urbanization) to conservation planning applications and environmental impact assessments (Donald & Fuller, 1998; Gibbons *et al.*, 2007; Dunn & Weston, 2008; Harrison *et al.*, 2008). There is a growing demand for data sets that can be used to detect the impacts of general environmental change (climate, land use) on species distributions in global analyses [e.g. Araújo *et al.*,

2005a,b; The North American Breeding Bird Survey (Table S1), EBCC Atlas of European Breeding Birds (Table S1)]. Atlases may also be developed to monitor ecosystem services that can be tied to specific species or groups of species, such as the pollination of plants that are economically important or of particular conservation concern, or plankton that indicate good fishing areas (Plankton Atlas of the North Atlantic Ocean, see Table S1). Directed atlassing efforts may try to find new species (Funk *et al.*, 2005) or resolve taxonomic issues, such as cryptic congeneric species (Tolley & Burger, 2004). Lastly, distributions for a representative suite of species are routinely used in conservation planning and species distribution modelling, even if the data are poor (Donald & Fuller, 1998; Gaston & Rodrigues, 2003). Atlases need not be spatially extensive to be useful; intensive regional surveys (e.g. Carolina Herp Atlas, Massachusetts Butterfly Atlas, Biodiversity Atlas of the Columbia River Basin, see Table S1 for details) frequently yield fine scale data that may complement broader scale efforts.

Regardless of the original intent of the developer, once a database has been assembled it is likely to be subjected to ever more complex analyses, many of which were not considered during data collection. As a result there is a high likelihood of users not taking into account the limitations of existing databases. For instance, one of the commonest forms of occurrence data is that of *ad hoc* datasets, which are based on an accumulation of records that have been collected with varying sampling effort in space and time. Most museum collections and checklist-based datasets are *ad hoc* data (Funk & Richardson, 2002). The people who are involved in collecting these data are seldom the same people who try to use them to draw general principles or describe broad scale patterns (Donald & Fuller, 1998; Dunn & Weston, 2008; Harrison *et al.*, 2008).

The aim of this article is to highlight the growing importance of atlas projects in ecology and conservation biogeography and to suggest ways in which these datasets could be improved. Although atlas projects can be defined quite broadly, our discussion focuses on atlas projects that collect data using a predefined grid and rely largely on a group of volunteers for data collection. We start by reviewing some of the applications of atlas data, then consider ideal qualities of biological data contained within atlases, discuss data quality issues and make recommendations for improving atlas datasets.

A huge diversity of atlas projects has been undertaken in different parts of the world. In this article, we pay particular attention to examples from southern Africa, for two reasons: first, because we are familiar with them and they illustrate our central arguments clearly and second, because they will be unfamiliar to many developed-country readers and offer some novel insights into atlassing efforts in a developing nation. These atlases are described in Tables S2 and S3 and include: the Southern African Bird Atlas Project (SABAP), the Tick Distributions Project (Tick DiP), the Southern African Frog Atlas Project (SAFAP), the Southern African Reptile Conservation Assessment (SARCA), the Protea Atlas, the Southern African

Plant Invaders Atlas (SAPIA), the South African National Survey of Arachnida (SANSA) and the South African Butterfly Conservation Assessment (SABCA).

While we do not aim to provide a comprehensive review of existing atlases elsewhere in the world, a few examples of typical atlas projects outside of southern Africa are summarized in Table S1. They range from global or national initiatives through to less ambitious state or provincial surveys and convey the general flavour of atlas efforts.

APPLICATIONS OF ATLAS DATA

Atlas data have been used internationally in documenting and understanding biological responses to climate change. Data collected by the New York State Amphibian and Reptile Atlas Project were used to establish that frogs are calling and breeding earlier than recorded for a baseline study (Gibbs & Breisch, 2001). Virkkala *et al.* (2008) made use of bird atlas data from Finland and Norway to predict likely distribution changes as a result of climate change. Atlas data are valuable in distribution modelling and for exploring factors that influence the performance of these models. For example, Marmion *et al.* (2009) used data from the Distribution Atlas of European Butterflies to compare prevalence, latitudinal range and spatial autocorrelation of species distribution patterns on the predictive accuracy of eight modelling techniques. Parsons *et al.* (2009) used atlas data to supplement survey data to model the distribution of a ground-dwelling bird in Australia. Atlas datasets have formed the basis of many large conservation-oriented initiatives, such as the development of a global map of plant diversity (Kier *et al.*, 2005), GAP analysis (see Jennings, 2000 and other articles in the same special issue of *Landscape Ecology*; also <http://gapanalysis.nbi.gov/portal/server.pt>), ecoregional planning (e.g. Bailey, 1983; TNC 2006) and the Millennium Ecosystem Assessment (Millennium Ecosystem Assessment 2005a,b). One of the core problems that the Millennium Ecosystem Assessment highlighted was that there is a critical lack of global biodiversity datasets that can be used in making broad-scale predictions about species loss and changes in ecosystem services (Cumming, 2007). It is also interesting to note that relatively few of these larger projects have introduced formal data screening techniques. GAP analysis, for example, appears to have used a wide range of datasets of varying quality in tandem with species models and expert opinion (Jennings, 2000). The likelihood that participants in these large initiatives will simply accept datasets without further screening or ground-truthing places an additional onus on atlas creators to provide quality indicators and to develop a high-quality product.

Data from atlas projects in southern Africa have seen many applications, offering insights that have relevance for prospective atlas developers, both within and outside of southern Africa. At the most basic level, the data collected by southern African atlas projects have provided data on distribution ranges of the species targeted. Most atlas projects have resulted in the publication of an atlas containing species distribution

maps, maps of species richness, hotspots, coldspots and collection intensity (e.g. the Frog atlas; Harrison *et al.*, 2004 and SABAP; Harrison *et al.*, 1997, 2008). These atlases are usually the basis for conservation assessments for the species (Barnes, 2000; Harrison *et al.*, 2004).

Atlas data for southern Africa have been used to produce distribution maps for identification guide books (Henderson, 2001; Hockey *et al.*, 2005) and scientific papers (Henderson, 1999; Olckers & Hill, 1999), and to inform studies on particular species (Dean, 2000a; Peacock *et al.*, 2007) or groups of species (Robertson *et al.*, 2003; Richardson & van Wilgen, 2004). Niche models that are used to predict potential ranges of species have been calibrated using atlas data (Osborne & Tigar, 1992; Cumming, 2000b; Robertson *et al.*, 2001, 2004; McPherson *et al.*, 2004; Rouget *et al.*, 2004). These models have been valuable for mapping tick species richness (Cumming, 2000b), understanding factors that limit tick distributions (Cumming, 1999, 2002), managing invasive alien plants (Nel *et al.*, 2004; Rouget *et al.*, 2004), testing theories about invasion biology (Thuiller *et al.*, 2006; Mgidi *et al.*, 2007; Wilson *et al.*, 2007), exploring the potential impacts of climate change (Erasmus *et al.*, 2002; Bomhard *et al.*, 2005; Cumming & Van Vuuren, 2006; Estrada-Pena *et al.*, 2006; Coetzee *et al.*, 2009), and investigating host–parasite relationships for ticks (Cumming, 1998, 1999, 2000c, 2004; Cumming & Guegan, 2006). Similarly, Thuiller *et al.* (2004) investigated species distributions in relation to plant traits using data from the Protea Atlas Project.

Atlas data for southern Africa have been very valuable for developing, refining and testing new distribution modelling techniques (Cumming, 2000a; Robertson *et al.*, 2001, 2004; Richardson & Thuiller, 2007) or their performance (McPherson *et al.*, 2004). The SABAP data have seen extensive use in conservation planning (Harrison *et al.*, 2008). Data from the frog atlas project were used in designing conservation plans (Harrison *et al.*, 2004). Several studies have investigated various aspects of conservation planning and conservation area selection (Lombard, 1995; Reyers *et al.*, 2000, 2002; Fairbanks *et al.*, 2001; Gaston *et al.*, 2001; Rodrigues & Gaston, 2001, 2002a,b,c; Bonn *et al.*, 2002; Gaston & Rodrigues, 2003; Lombard *et al.*, 2003; Bonn & Gaston, 2005; Grantham *et al.*, 2009). Williams *et al.* (2005) identified corridors to ensure connectivity of suitable habitat for Proteas under climate change using data from the Protea Atlas Project. A number of studies have investigated theories about species richness patterns or macroecological–environment relationships (Allan *et al.*, 1997; Dean, 1997, 2000b; Fairbanks *et al.*, 2002; Van Rensburg *et al.*, 2002, 2004; Chown *et al.*, 2003; Fairbanks, 2004; Lennon *et al.*, 2004).

IDEAL QUALITIES OF ATLAS DATASETS

Given the frequency and diversity of atlas efforts currently under way in different parts of the world, it seems that different partners in large atlas efforts would benefit from a stronger understanding of one another's agendas (Dunn &

Weston, 2008). From the perspective of a quantitative analyst or 'end user', atlassing efforts are not equal, and there are certain kinds of atlassing data that yield higher scientific return than others. Seven ideal qualities of such datasets are discussed in the next paragraph.

1. The sampling strategy should be informed by the amount of variation in the taxon under study. To obtain maximum information, data should be collected at spatial and temporal scales that are commensurate with those at which the study taxon varies. Many of the questions of interest to ecologists revolve around the ways in which organisms respond to environmental heterogeneity. If samples are collected at too coarse a scale, variation within a sampling unit can swamp variation between sampling units. For example, insect communities may vary through the course of an evening as well as through the year. Collections of light-trapping samples for annual analysis need to either be made over the same time period each evening or over a sufficiently wide range of times that hourly trends can be estimated and corrected for.

2. The spatial and temporal resolution of the dataset should be as high as possible to ensure maximum value of the dataset. The data should therefore have a high resolution and a broad extent in both space and time. Data can always be aggregated – for example, point data can be summarized by quarter-degree cell or daily data can be presented as monthly means. However, if the data are collected at a coarse (e.g. quarter-degree) resolution from the start then finer-scale analyses are ruled out. Similarly, if only part of a species range is covered by a survey, answering biogeographic questions becomes difficult; for example, it is harder to draw conclusions about environmental preferences or realized niches.

3. The taxonomic resolution of the dataset should be as high as possible. The units of analysis in most ecological studies are species. Reliable species-level data for poorly studied taxa may be hard to obtain, but working with families or genera (for example) is notably inferior. By contrast, for well-studied species like large mammals, a level of taxonomic resolution that takes into account subpopulations and gene flow may be ideal.

4. The demographic resolution of the dataset should be as high as possible. Population ecologists in particular, and ecologists in general, can pose and answer many relevant and interesting questions using the age and stage structure of populations. For example, distinguishing between larvae, nymphs and adults can be important for invertebrate studies where these individuals lead different lifestyles.

5. The sampling protocol should be standardized and each record should include a good measure of sampling effort. One of the fundamental aims of atlassing is to provide rigorous comparison between different locations, and/or tracking of change through time. Such comparisons are only possible if samples can be validly compared with one another. The number of species found in a location will increase logarithmically with sampling time and area sampled. More time spent searching or a greater area covered, will mean that more individuals of more species are observed. Providing an estimate

of sampling effort offers one way of standardizing results, assuming that the relationship between sampling effort and sample size is consistent. Comparison between samples is usually easiest when data are collected in identical ways and without any kind of systematic bias.

6. The sampling protocol should be described in detail, including potential sources of error and bias in the dataset. Whether or not this is the original intent, atlassing efforts should be repeatable so that environmental changes can be tracked. In addition, many end users will not have had hands-on familiarity with data collection. It is important that end users should know how to interpret the results and that any particular biases in an individual atlas dataset are openly noted and highlighted.

7. Sample size – i.e. the number of unique sampling units (e.g. grid cells) for which data are recorded – should be as large as is feasible (although increasing the size of the dataset should never come at the expense of data quality). Errors creep in to even the most carefully collected datasets. Large sample sizes provide the quantity of evidence that is needed to separate trends from errors or 'noise' in the data. Stratification and high quality coverage of a smaller total area are often more useful than extensive coverage and low quality data. Many atlassing efforts succumb to the temptation to 'fill in every grid cell'. This often leads to inadequate sampling in the majority of sampling locations. For the scientific user, high-quality data that have been collected at sufficiently many locations to adequately cover the full range of relevant variation in the environment can be more useful than low-quality data that 'fill in' more spaces.

These ideals must be interpreted relative to context. In many cases, it will not be possible to meet all of the 'ideal' criteria for an atlas dataset, and the demands of scientific users must generally be balanced against a set of real-world constraints. Foremost among these are constraints on time, funding and expertise. Many atlassing projects have small budgets and relatively few full-time personnel. For some taxa, such as insects and spiders, there may be few taxonomists who are capable of reliably identifying individual specimens to a species level; and these taxonomists are often not able to devote large amounts of time to a new atlassing effort. We would emphasize, however, that some aspects of data quality – particularly the development and maintenance of a consistent sampling protocol and the quantification of sampling effort – are so fundamental to the interpretation of atlassing data that they should not be compromised.

DATA QUALITY OF EXISTING ATLASES AND PROSPECTS FOR IMPROVEMENT

Data quality determines whether atlas data are appropriate for a particular purpose. Atlas projects should provide measures of data quality so that users of the data can make informed decisions about the appropriateness of the data for particular applications and can take limitations into account when analyzing the data. Aspects of data quality that we discuss

include spatial scale, temporal resolution, sampling bias, errors in the records and completeness of the records.

Spatial scale

Spatial scale has two components: extent and grain (Whittaker *et al.*, 2005). Spatial extent refers to the map region (geographical area) over which the data for the atlas project are collected. Grain (spatial resolution) refers to the size of the sampling unit over which a single observation is made. In most cases, a sampling unit is a grid cell of a particular size, e.g. 15 min. Extent is related to grain in that atlases with a large extent tend to have a coarser spatial resolution (Dunn & Weston, 2008). An important consideration is that different patterns of diversity can be observed with the same dataset by varying the spatial resolution (Whittaker *et al.*, 2005). If data are collected at fine spatial resolution then they can always be aggregated to coarse resolution but the converse is not true. The spatial resolution of the data can severely limit the types of questions that can be addressed. For some atlas projects (e.g. SABAP, SAPIA), data were collected at a fairly coarse spatial resolution such as Quarter Degree Squares (QDS; 15' × 15'). In most cases, data collected at QDS are too coarse to be used for selection of reserve networks in conservation planning (Pressey *et al.*, 2003; Driver *et al.*, 2005). Finer scale distribution data are also needed to understand the combined effects of climate and land use change (De Chazal & Rounsevell, 2009).

Increasing spatial resolution from atlases that collect data at 15 min (Quarter-Degree Square) to finer resolution (e.g. 5 min for SAPAB2; Harrison *et al.*, 2008) would be useful for conservation, especially conservation area selection. Increases in spatial resolution are, however, limited by the number of observers and the spatial extent of the map region (Gibbons *et al.*, 2007). It has also been suggested that surveys could be designed such that records at higher spatial resolution are collected in regions that are particularly vulnerable to development relative to the rest of the mapped region (Donald & Fuller, 1998). Certain atlas projects (e.g. SAPIA) are flexible in that they allow point records to be submitted in addition to grid-based data. This is one way of ensuring that high-resolution data are recorded. It may not be possible to sample every grid cell of a fine scale grid, especially when the extent is large. For the atlas of breeding birds of Britain and Ireland, the sampling unit was a 10-km grid but timed visits were also undertaken to a number of 2 × 2 km grid cells (tetrads) nested within the 10-km grid to map indices of relative abundance (Gillings, 2008).

Temporal resolution

Problems of sampling in space are closely allied to those of sampling in time. When sampling effort is a limiting factor, tradeoffs may arise between the benefits of repeating sampling at the same location versus spreading sampling effort in space. Temporal resolution can be considered at the level of the entire atlas dataset or at the level of individual sampling units. For

most atlas projects, the data are collected over a discrete time period, usually a few years. These data can be used as a baseline with which to compare future changes when another phase of the project is repeated at a later date (e.g. SABAP 1 & 2; Harrison *et al.*, 2008). Donald & Fuller (1998) give examples of studies that have documented range changes for birds by comparing atlas datasets from different years. It is thus important when developing sampling protocols to ensure that the sampling protocol can be repeated in the future so that temporal changes can be documented. Cases have been reported where methods between two sampling periods were so different that direct comparisons were impossible (Donald & Fuller, 1998). Individual sampling units may be sampled by different individual observers at different times and in this way provide repeat observations. However, the temporal resolution of individual sampling units in the map region will vary. If better temporal resolution data are required then timed visits to a limited number of specific sites will be better. For the Catalan common bird's survey, observers record all birds seen or heard along a set of 3-km transects that are visited twice during the breeding season (Brotons *et al.*, 2007). This type of data has been referred to as 'long-term monitoring data' (Brotons *et al.*, 2007), but it could easily be incorporated into an atlas project.

Sampling bias

Sampling bias is a major problem in atlas datasets (e.g. Dennis *et al.*, 1999) and in collections data (Funk & Richardson, 2002). It can include geographical (spatial) bias, taxonomic bias and temporal bias (Funk & Richardson, 2002). Geographical bias refers to uneven sampling effort across the map region. Taxonomic bias can include over or under-representation of certain species in the dataset. For example, species with cryptic coloration, fossorial species and species with low vagility may be under-represented (Robertson *et al.*, 1995). Dennis *et al.* (2006) reported bias in butterfly atlas datasets based on apparency of butterfly adults (defined using wing colour, size and behaviour).

Temporal bias occurs when records are collected in one season only or more often at certain times of the year (Funk & Richardson, 2002). This type of bias can also occur when species have very specific environmental triggers for activity periods, such as ectotherms. Dunn & Weston (2008) reported that for many bird atlases, data collection was limited to summer. The same is generally true in cold regions for invertebrates and ectotherms. Spatial bias in collections data and atlas data has received more attention in the literature than has temporal bias (Reddy & Davalos, 2003; Robertson & Barker, 2006). Robertson *et al.* (1995) suggested for SABAP that rare or endemic species may be over-represented in game reserves or national parks because people tend to actively search for these species in these areas. This may also apply to datasets for other organisms. Prior knowledge of grid cells may result in observers favouring particular areas and differences in levels of experience of observers could influence detection rates

of species (Robertson *et al.*, 1995). Low sampling effort may mean that rarer species are not recorded or are under-represented in certain cells (Robertson *et al.*, 1995). Sampling effort is usually quantified by examining the number of records submitted per grid cell (Robertson & Barker, 2006), and equal effort is assumed per record. However, sampling effort per record may vary as some recorders may spend more time searching and cover greater distances within each grid cell (spatial sampling unit) than others. The number of records per grid cell may thus not be a reliable measure of sampling effort. A further limitation is that accurate and useful measures of sampling effort such as time spent observing or distance travelled in a grid cell are generally not reported by recorders. The time spent observing has been referred to as the temporal unit of sampling (Dunn & Weston, 2008) and may vary considerably across atlases.

Atlas projects are at an advantage compared with collections data when it comes to addressing sampling bias. Atlas projects have a coordinator (or project team) who can identify biases in the dataset and communicate these to the data collectors so that sampling can be altered in response to the identified biases. In many cases, the data collectors consist of a fairly large group of volunteer observers that have the potential to collect a vast amount of data in a relatively short period of time. To reduce spatial bias, it is possible to generate maps that document grid cells that are considered to be poorly sampled and to encourage observers to sample these cells. It may also be necessary to highlight species or higher taxa for which data are lacking, as a means of addressing taxonomic bias. Temporal biases could be addressed in a similar manner by making observers aware of the trends in the dataset. Making the data freely available during the atlas project is likely to encourage scientists to undertake analyses that will reveal biases and other data quality issues in the dataset. In addition, a reliable measure of sampling effort (per record) would also help to overcome some of the problems associated with sampling bias when analyzing the data.

In an attempt to avoid taxonomic bias in sampling (under or over representation), it may be worth attempting to quantify the observability of each major group. For example, more conspicuous species (larger, brighter, louder) may be reported more frequently (Dennis *et al.*, 2006); and this bias, if consistent, may be corrected for by using a measure of observability based on comparisons between results of quick surveys and more extensive surveys. This could also be used to correct for sampling effort. Gillings (2008) assigned a detection score, ranging from 1 (easy to detect when present) to 4 (difficult to detect), to each species in a study of sampling effort in birds. He found that the likelihood of missing a species was significantly positively correlated to its detection score.

Errors and record completeness

Errors in, and completeness of, records are likely to be a problem with all atlas datasets, especially when citizen scientists are involved (Cohn, 2008). The most serious errors

are likely to be misidentification of species. This is likely to be influenced by the experience of the observer. Errors can also occur when recording the geographical position of the observation or the sampling unit in the map region. Problems with completeness of records include cases where not all the fields for the record are completed or insufficient data are provided for a record. For example in the SAPIA database, certain species that are difficult to identify have been identified only to the level of genus. The result is that records with the species name '*Eucalyptus* spp.' could include one of several species of *Eucalyptus* that have been introduced to South Africa.

Various systems can be put in place to minimize errors in the dataset. For example in SABAP2, observers who submit records to the online database will receive a notification if the species that they have reported is considered to be out of its range. The range data were collected during the first phase of the project. This system ensures that observers check their data for obvious misidentification errors as part of the submission process. Using multimedia electronic field guides may be another way of reducing misidentification errors. For example, recordings of calls provided with some bird guides may help to confirm the identity of the species. These field guides are now available for several taxa (Stevenson *et al.*, 2003) and are likely to become more popular with the increase in popularity of mobile devices. Similarly, the increasing use of GPS and the availability of free online maps and imagery (e.g. Google Earth) are likely to result in fewer spatial errors. The contribution of digital photos and specimens could allow the identification of the species to be checked and would increase confidence in the records. The use of an online virtual museum has proven to be very useful for SARCA; amateurs submit photographs according to specific guidelines, and experts check these records online. Expert agreement is used as a measure of identification reliability.

Observer skill and experience are usually important factors that can influence data quality. As a result, it may be sensible to implement a two-tier system for observers based on their experience and ability to identify species. This system would provide a means of distinguishing experienced observers from inexperienced observers. This could be useful for example when a new location for a species is reported as one would tend to have greater confidence in the record if it were submitted by an experienced observer than by an inexperienced observer who may have misidentified the species. As part of the two-tier system it would also be possible to accept only certain (easily quantifiable) data from inexperienced observers while accepting all possible fields of data from an experienced observer or scientist. The extent to which this will be necessary is likely to differ among groups of organisms. For certain atlas projects, there is less reliance on amateur observers to collect data. For example, Harrison *et al.* (2004) reported for the SAFAP that most of the data were collected by professional herpetologists, as fieldwork was more demanding than collecting data for SABAP. For certain bird atlas projects, observers are allowed to contribute data to different levels of

detail (Dunn & Weston, 2008). This means that potentially valuable data are not lost when data collectors either do not have all the required data or do not submit data because they feel that the data requirements for submitting a record are too time-consuming. In SABAP2, there is a facility for submitting incidental observations.

Improving data quality

Atlas projects often face a tradeoff between quality and quantity. Quality can be improved by increasing the minimum data requirements for the submission of records and by implementing various systems to reduce error (see above). Unfortunately, the likely consequence of increasing the minimum requirements (e.g. a specific sampling effort) for a record will result in fewer records being submitted, especially by amateurs. However, many atlas projects (e.g. SABAP, Protea atlas and SAFAP) specifically aim to involve amateurs in the project as a means of generating interest in the particular group of organisms (Harrison *et al.*, 2008). Many atlas projects rely on volunteers to collect data and would not be able to collect these data without the contribution of considerable time and resources by these people (Cohn, 2008). The needs of these volunteers should be balanced against scientific needs (Gillings, 2008).

In regions with complex biogeographical histories, the legacy of historical range limiting factors may prevail over present-day range limitations. This may well lead to the belief that every cell must be sampled, usually with lower quality data as a result. It is important to realize that it is not necessary to sample every cell to create a useful dataset. The key is to be explicit about survey objectives and to design the sampling protocol accordingly. Good precedents do exist for projects that have collected data only in priority grid cells and have not attempted to sample every cell (Dunn & Weston, 2008). Brotons *et al.* (2007) highlighted the value of data collected at a network of observation sites at regular intervals for modelling distributions, thereby complementing atlases with a coarser resolution that attempt to survey every cell. Although various measures can be taken to improve atlas data, it is often more important to document the quality of the data that have been collected.

Quantifying sampling effort

There is a clear need in atlassing for solid quantification of sampling effort by observers, especially the effort expended per record. The way in which this is carried out will depend on the type of organisms concerned. For noisy, mobile organisms such as birds, the number of hours spent observing is the primary measure of effort. For less mobile or less obvious taxa such as frogs or insects, time spent searching might need to be supplemented with information on search techniques and/or habitats searched (e.g. to what extent were both terrestrial and aquatic habitats sampled). In addition, the total distance travelled through the grid cell

(sampling unit) could be recorded. Quantification of sampling effort can be particularly useful for assessing the quality of the data. This is important when assessing the likelihood of detecting a rare species or the reliability of an assumption of absence for a species in a particular sampling unit, as occurs when range changes are assessed (Donald & Fuller, 1998).

Atlas data are often presence-only datasets with no reliable absence data, as is the case for most collections data (Funk & Richardson, 2002). A species may not be recorded in a particular grid cell either because it is genuinely not there or for a number of other reasons, such as low sampling effort, cryptic coloration, inappropriate search methods and so on. The rigorous evaluation of species range predictions generated from ecological niche models requires that accuracy measures incorporate errors of omission and errors of commission (Fielding & Bell, 1997; Araújo *et al.*, 2005b), although if the prevalence of positive records is high enough, assumptions of absences may be safely made in developing predictive models (Cumming, 2000a; Parsons *et al.*, 2009).

In several studies, where abundance was not specifically recorded, including SABAP, the relative abundance of species has been inferred using relative reporting rate (Dean, 1997; Bonn *et al.*, 2002; Gaston & Rodrigues, 2003; Bonn & Gaston, 2005). The relative reporting rate for a species is calculated as the proportion of the total number of cards for a grid cell on which that species is recorded. A study by Robertson *et al.* (1995) investigated the use of reporting rate to estimate population size of bird species and found significant relationships between reporting rate and observed abundance for three out of the four species studied and a marginally significant relationship for the fourth species. However, the generality of these relationships is yet to be demonstrated. Reporting rate is likely to be influenced by sampling effort and the skills of the observer (Robertson *et al.*, 1995). Abundance data (inferred from reporting rate) have been used as a measure of habitat quality or species performance for selecting conservation areas. Grid cells with peaks in abundance for species are assumed to represent areas where the species is most likely to persist in future (Bonn *et al.*, 2002; Gaston & Rodrigues, 2003).

Atlas projects that collect data for mobile species (such as birds and butterflies) could benefit from more accurately recording some measure of abundance or frequency of occurrence for species (Donald & Fuller, 1998). Gibbons *et al.* (2007) suggested that bird atlases that collected abundance data instead of presence were no more costly to undertake and produce in terms of number of observers required or time taken. Several methods are available for quantifying abundance of birds (Gibbons *et al.*, 2007), some of which will be applicable to other organisms. However, reporting rate (one of the proposed methods) is an indirect measure of abundance and direct measures (e.g. counts over a given time period or analyses of a standard quadrat size) is preferable. What has been learned from this aspect of bird atlassing efforts has broad general relevance to many other taxa. In SABAP2, observers are

required to list the order in which species were observed. This can be used as a rough index of abundance although it is clearly inferior to true abundance data and carries its own biases (for example, when arriving at a location such as a wetland where many species are present, birders may record rare species first because they are more likely to fly away during the observing period).

Managing successful atlas projects

Atlas development is generally a collaborative process that involves multiple interest groups. At least three important groups of people can be identified: the data collectors, the project manager or project team and the scientific users. The group of data collectors is usually amateur enthusiasts who volunteer their time and resources to collect data (Dunn & Weston, 2008; Harrison *et al.*, 2008). These data collectors can collect vast amounts of data at no cost to the atlas project (Harrison *et al.*, 2008). Data collectors participate primarily because they enjoy being able to put their skills to use in searching for and identifying species of the target group and may have relatively little interest in the final product or the resulting scientific conclusions. Communication between the project management team and data collectors is important to ensure success of the project. Regular updates on progress should be communicated using the project website and via newsletters to ensure that the collectors remain interested and feel that the data being collected are being used (e.g. SABAP2). Making collectors aware of sampling biases is important for ensuring the highest possible data quality of the atlas dataset.

The project team is responsible for maintaining the dataset, coordinating data capture, performing certain analyses, communicating with relevant parties and usually with publishing an atlas or conservation assessment of the taxon at the end of the project. The project team also has the responsibility of designing the sampling protocol with input from scientists and data collectors. Atlas efforts do not end with the first published analysis of the data or even with the provision of an on-line copy of the database (Harrison *et al.*, 2008). They often become a standard dataset that is used widely for a 5- to 10-year time horizon after collection. This general use period represents an important opportunity for value-adding to atlas projects. The atlas dataset should be made easily accessible to scientific users so that the maximum value can be derived from the dataset. Unfortunately, atlas databases occasionally end up as the preserve of a small number of people who claim ownership of the data and may even charge for access to these data. Such an approach is usually crippling to both the success of the current venture and the potential for future initiatives. Data ownership can also be maintained in less obvious ways, such as by making data superficially free 'on request' but delaying data provision, providing data in outdated or specialized formats, excluding the majority of fields that are not explicitly requested, or providing data at a lower spatial or temporal resolution than that at which they were originally collected.

A number of summary statistics should be routinely provided to users that request atlas data as this will help them to assess the fitness of the data for a particular application. Summary statistics per grid cell (sampling unit) such as number of species, number of observers, number of records and average effort per record can be useful for quantifying sampling effort in a particular grid cell. Date of first and last record can be used to assess temporal coverage. The difference between number of observers for the focal cell and average of number of observers for the neighbouring cells can give an indication of sampling effort.

Lastly, given the decay of data accuracy with time, atlas data will have maximum impact if they are published soon after collection and are made publicly available within a short period after atlas completion.

DISCUSSION

Developing a spatially comprehensive atlas for a specific taxon, let alone a number of representative taxa, requires a large investment of time and effort. Any survey will be a trade-off between available resources, sampling intensity (a measure of effort at a particular sampling site), sampling extent (the spatial configuration of sampling sites) and the detection probability of the taxon of interest. Analyses of bird atlases have found positive relationships between the number of observers and geographical extent, and a positive relationship between grain and extent (Gibbons *et al.*, 2007; Dunn & Weston, 2008).

With constant and comparable effort at a given sampling site, false positive errors will be the lowest at the time of sampling. With increasing time after sampling, the accuracy of the data decreases (the false positives increase) because of the intrinsic dynamic nature of species' ranges as well as to the effects of external stressors (e.g. fragmentation). Subsequent sampling reduces the level of error again. Quantifying decay in data accuracy, and teasing apart the relative influence of natural variation and externally induced stressors, is only feasible if sampling is at a finer spatiotemporal scale than that at which the taxon varies.

The reality is that many atlas efforts start by compiling *ad hoc* collections data and then seek to fill in the gaps to maximize the usefulness of existing data. This is a challenging problem since the level of undersampling is inconsistent and its spatial distribution is unknown. Even if subsequent sampling is designed to make maximum use of existing data, and is of a high quality, the final pooled dataset will always have the lowest quality data as a common denominator. Without knowing the shortcomings of the original data, the high quality data cannot be analysed to its full potential. If the quality of the original data is known (in terms of locational accuracy, date of sample, skill of collector, sampling effort, *ad hoc* or coordinated collection) then analysis limits, and subsequent impacts on atlas objectives, can be set. One potential solution for a transitional period is to include an indicator of data quality within the database, allowing users to

extract good data for applications in which data quality is particularly important.

In recent years, there has been an increase in the accuracy and reliability of methods for interpolating species distributions from presence-only records (Elith *et al.*, 2006). Presence-only data are easy to collect with relatively low sampling effort, and provided the locational and taxonomic information is reliable and sampling effort is known, they lend themselves to useful analyses by a large number of end-users. This type of data is also typical of existing *ad hoc* collections. A combination of well-selected representative intensively sampled sites and broader scale presence-only monitoring at spaced intervals across the landscape offers a good compromise between effort, extent and quality. It has been shown that detection probability can have large impacts on data quality (Dennis *et al.*, 2006), and therefore sampling effort should be appropriately high for rare or cryptic taxa. Uneven sampling effort has been shown to limit change detection ability at fine scales (Dennis & Hardy, 2001), therefore consistent and standardized sampling effort is of paramount importance. As Fielding & Bell (1997) point out, improved statistical analysis of datasets will never be adequate to compensate for a poorly designed sampling regime in which 'ecological' errors are high.

Conservation planning is increasingly incorporating consideration of processes into planning and the question arises as to whether or not surveys should also include measures of process-oriented indicators (e.g. water quality, extent of burned areas, reproductive success). Species-focused surveys may need to follow an indicator approach to monitor ecosystem processes as, at best, it is difficult to link ecosystem function to a particular species or groups of species. Mapping of ecosystem services can be performed independently of survey efforts, such as satellite-derived net primary productivity, water catchment, rates of green-up, soil erosion, etc. The real value for sustainability science lies in linking observed biodiversity patterns with data on the spatial distributions of relevant processes.

In cases where relatively low-quality datasets are of high historical value, some kind of working compromise between data inclusion and the overall quality of the analysis must be reached. The key in developing any such compromise will be to establish a reliable method for quantitatively testing data quality and removing records that are truly 'suspect'. Such methods will generally have to be based on either (1) a 'voting' procedure, where multiple records of the same species from the same locality and/or multiple adjacent records carry greater weight than single or isolated collection records or (2) some kind of calibration against current high quality data. Indeed, for many atlas projects, it will make sense to set aside a small portion of the total budget for the development of reference or calibration points in which high-quality sampling (i.e. frequent, standardized, exhaustive) is carried out on a regular basis by experts. If control points are carefully selected to cover major habitat types and environmental gradients, and if a consistent relationship is

obtained between such calibration points and the broader dataset, it will be possible to apply some kind of correction factor to areas in which lower quality sampling has been undertaken. Well-designed calibration data will also allow for a formal quantification of error rates and a quantitative assessment of the adequacy of the spatial and temporal grain of sampling.

In their review, Whittaker *et al.* (2005) discuss the sensitivity of conservation biogeography to assumptions in terms of four factors. These include scale dependency, the effects of model structure and parameterization, inadequacies in taxonomic and distributional data and inadequacies of theory. To make advances in conservation biogeography, further research is required into issues of scale and how the effects of model structure and parameterization influence the conservation decisions made. This research is dependent on high-quality distribution data for a range of taxa: the sort of data that atlas projects can provide. In addition, atlas projects can help to improve taxonomic and distributional data for many taxa by making use of volunteers. Advances in theory are often reliant on quantitative analyses of competing hypotheses. Atlas projects thus have a major role to play in the advancement of conservation biogeography.

CONCLUSION

Atlasing efforts should play an important role in biodiversity conservation by providing essential data on the occurrences of species. They should also make useful contributions to the development of ecological understanding. Although the emphasis of any given atlas exercise may be more on one or the other, these two goals are not incompatible.

The ultimate usefulness of atlas projects, from both an academic and a conservation perspective, is contingent on the quality and quantity of data collected, particularly in terms of the standardization of sampling methods, the appropriateness of the scale of sampling for the question being answered, and the potential for data calibration and the quantification of error rates. Statistical analyses can compensate for defects in some of these areas, but their ability to do so is limited by a few important constraints. In particular, a good measure of sampling effort is essential (for instance, to determine if absences are genuine or simply a consequence of inadequate sampling); the scale of data collection must be at a fine enough resolution to link to habitat variables of potential interest (because upscaling is possible but downscaling is not); the total sample size must be large enough to work with in a multivariate context; and some indication of data quality must be presented to allow for the needs of users who have specific demands for high-quality data. Atlas projects have an important role to play in the advancement of conservation biogeography by providing much needed distribution data that are essential for developing and testing new theories and analytical approaches.

ACKNOWLEDGEMENTS

We thank Simon Ferrier and three anonymous reviewers for their insightful comments on an earlier draft of the manuscript.

REFERENCES

- Allan, D.G., Harrison, J.A., Navarro, R.A., van Wilgen, B.W. & Thompson, M.W. (1997) The impact of commercial afforestation on bird populations in Mpumalanga province, South Africa: insights from the bird atlas data. *Biological Conservation*, **79**, 173–185.
- Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M. (2005a) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005b) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Bailey, R. (1983) Delineation of ecosystem regions. *Environmental Management*, **7**, 365–373.
- Barnes, K.N. (ed.) (2000) *The Eskom Red Data Book of birds of South Africa, Lesotho and Swaziland*. BirdLife South Africa, Johannesburg.
- Bomhard, B., Richardson, D.M., Donaldson, J.S., Hughes, G.O., Midgely, G.F., Raimondo, D.C., Rebelo, A.G., Rouget, M. & Thuiller, W. (2005) Potential impacts of future land use and climate change on the Red List status of the Proteaceae in the Cape Floristic Region, South Africa. *Global Change Biology*, **11**, 1452–1468.
- Bonn, A. & Gaston, K.J. (2005) Capturing biodiversity: selecting priority areas for conservation using different criteria. *Biodiversity and Conservation*, **14**, 1083–1100.
- Bonn, A., Rodrigues, A.S.L. & Gaston, K.J. (2002) Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecology Letters*, **5**, 733–741.
- Brotons, L., Herrando, S. & Pla, M. (2007) Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions*, **13**, 276–288.
- Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S.L. & van Jaarsveld, A.S. (2003) Species richness, human population size and energy: conservation implications at a national scale. *Ecological Applications*, **13**, 1233–1241.
- Coetzee, B.W.T., Robertson, M.P., Erasmus, B.F.N., van Rensburg, B.J. & Thuiller, W. (2009) The impact of climate change on the southern African Important Bird Areas network. *Global Ecology and Biogeography*, **18**, 701–710.
- Cohn, J.P. (2008) Citizen science: can volunteers do real research? *BioScience*, **58**, 192–197.
- Cumming, G.S. (1998) Host preference in African ticks: a quantitative data set. *Bulletin of Entomological Research*, **88**, 379–406.
- Cumming, G.S. (1999) Host distributions do not limit the species ranges of most African ticks (Acari: Ixodida). *Bulletin of Entomological Research*, **89**, 303–327.
- Cumming, G.S. (2000a) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27**, 441–455.
- Cumming, G.S. (2000b) Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). *Journal of Biogeography*, **27**, 425–440.
- Cumming, G.S. (2000c) Host use does not clarify the evolutionary history of African ticks (Acari: Ixodoidea). *African Zoology*, **35**, 43–50.
- Cumming, G.S. (2002) Comparing climate and vegetation as limiting factors for species ranges of African ticks. *Ecology*, **83**, 255–268.
- Cumming, G.S. (2004) On the relevance of abundance and spatial pattern for interpretations of host–parasite association data. *Bulletin of Entomological Research*, **94**, 401–409.
- Cumming, G.S. (2007) Global biodiversity scenarios and landscape ecology. *Landscape Ecology*, **22**, 671–685.
- Cumming, G.S. & Guegan, J.-F. (2006) Food webs and disease: is pathogen diversity limited by vector diversity? *EcoHealth*, **3**, 163–170.
- Cumming, G.S. & Van Vuuren, D. (2006) Will climate change affect ectoparasite species ranges? *Global Ecology and Biogeography*, **15**, 486–497.
- De Chazal, J. & Rounsevell, M.D.A. (2009) Land-use and climate change within assessments of biodiversity change: a review. *Global Environmental Change*, **19**, 306–315.
- Dean, W.R.J. (1997) The distribution and biology of nomadic birds in the Karoo, South Africa. *Journal of Biogeography*, **24**, 769–779.
- Dean, W.R.J. (2000a) Alien birds in southern Africa: what factors determine success? *South African Journal of Science*, **96**, 9–14.
- Dean, W.R.J. (2000b) Factors affecting bird diversity patterns in the Karoo, South Africa. *South African Journal of Science*, **96**, 609–616.
- Dennis, R.L.H. & Hardy, P.B. (2001) Loss rates of butterfly species with urban development. A test of atlas data and sampling artefacts at a fine scale. *Biodiversity and Conservation*, **10**, 1831–1837.
- Dennis, R.L.H., Sparks, T.H. & Hardy, P.B. (1999) Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation*, **3**, 33–42.
- Dennis, R.L.H., Shreeve, T.G., Isaac, N.J.B., Roy, D.B., Hardy, P.B., Fox, R. & Asher, J. (2006) The effects of visual apparency on bias in butterfly recording and monitoring. *Biological Conservation*, **128**, 486–492.
- Donald, P.F. & Fuller, R.J. (1998) Ornithological atlas data: a review of uses and limitations. *Bird Study*, **45**, 129–145.
- Driver, A., Maze, K., Rouget, M., Lombard, A.T., Nel, J., Turpie, J.K., Cowling, R.M., 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 Biodiversity Institute, Pretoria.
- Dunn, A.M. & Weston, M.A. (2008) A review of bird atlases of the world and their application. *The Emu*, **108**, 42–67.
- Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Erasmus, B.F.N., van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. & Wessels, K.J. (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology*, **8**, 679–693.
- Estrada-Pena, A., Corson, M., Venzal, J.M. *et al.* (2006) Changes in climate and habitat suitability for the cattle tick *Boophilus microplus* in its southern Neotropical distribution range. *Journal of Vector Ecology*, **31**, 158–167.
- Fairbanks, D.H.K. (2004) Regional land-use impacts affecting avian richness patterns in southern Africa – insights from historical avian atlas data. *Agriculture, Ecosystems and Environment*, **101**, 269–288.
- Fairbanks, D.H.K., Reyers, B. & van Jaarsveld, A.S. (2001) Species and environment representation: selecting reserves for the retention of avian diversity in KwaZulu-Natal, South Africa. *Biological Conservation*, **98**, 365–379.
- Fairbanks, D.H.K., Kshatriya, M., van Jaarsveld, A.S. & Underhill, L.G. (2002) Scales and consequences of human land transformation on South African avian diversity and structure. *Animal Conservation*, **8**, 61–73.
- Fielding, A.H. & Bell, J.F. (1997) A review of method for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Funk, V.A. & Richardson, K.S. (2002) Systematic data in biodiversity studies: use it or lose it. *Systematic Biology*, **51**, 303–316.
- Funk, V.A., Richardson, K.S. & Ferrier, S. (2005) Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society*, **85**, 549–567.
- Gaston, K.J. & Rodrigues, A.S.L. (2003) Reserve selection in regions with poor biological data. *Conservation Biology*, **17**, 188–195.
- Gaston, K.J., Rodrigues, A.S.L., van Rensburg, B.J., Koleff, P. & Chown, S.L. (2001) Complementary representation and zones of ecological transition. *Ecology Letters*, **4**, 4–9.
- Gibbons, D.W., Donald, P.F., Bauer, H., Fornasari, L. & Dawson, I.K. (2007) Mapping avian distributions: the evolution of bird atlases. *Bird Study*, **54**, 324–334.
- Gibbs, J.P. & Breisch, A.R. (2001) Climate warming and calling phenology of frogs near Ithaca, New York, 1900–1999. *Conservation Biology*, **15**, 1175–1178.
- Gillings, S. (2008) Designing a winter bird atlas field methodology: issues of time and space in sampling and interactions with habitat. *Journal of Ornithology*, **149**, 345–355.
- Grantham, H.S., Wilson, K.A., Moilanen, A., Rebelo, T. & Possingham, H.P. (2009) Delaying conservation actions for improved knowledge: how long should we wait? *Ecology Letters*, **12**, 293–301.
- Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. & Brown, C.J. (1997) *The atlas of southern African birds*. Birdlife South Africa, Johannesburg, South Africa.
- Harrison, J.A., Minter, L.G. & Burger, M. (2004) Atlas and red data book for frogs completed. *South African Journal of Science*, **100**, 11–13.
- Harrison, J.A., Underhill, L.G. & Barnard, P. (2008) The seminal legacy of the Southern African Bird Atlas Project. *South African Journal of Science*, **104**, 82–84.
- Henderson, L. (1999) The Southern African Plant Invaders Atlas (SAPIA) and its contribution to biological weed control. *African Entomology Memoir* no. 1. 159–163.
- Henderson, L. (2001) Alien weeds and invasive plants. *Plant Protection Research Institute handbook no. 12*. Agricultural Research Council, Pretoria, South Africa.
- Hockey, P.A.R., Dean, W.R.J. & Ryan, P.G. (2005) *Roberts – birds of Southern Africa*, 7th edn. The trustees of the John Voelcker Bird Book Fund, Cape Town.
- Jennings, M.D. (2000) GAP analysis: concepts, methods, and recent results. *Landscape Ecology*, **15**, 5–20.
- Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Kuper, W., Kreft, H. & Barthlott, W. (2005) Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, **32**, 1107–1116.
- Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2004) Contribution of rarity and commonness to patterns of species richness. *Ecology Letters*, **7**, 81–87.
- Lombard, A.T. (1995) The problems with multi-species conservation: do hotspots, ideal reserves and existing reserve coincide? *South African Journal of Zoology*, **30**, 145–163.
- Lombard, A.T., Cowling, R.M., Pressey, R.L. & Rebelo, A.G. (2003) Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region. *Biological Conservation*, **112**, 45–62.
- Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, **220**, 3512–3520.
- McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.
- Mgidi, T.N., Le Maitre, D.C., Schonegevel, L., Nel, J.L., Rouget, M. & Richardson, D.M. (2007) Alien plant invasions – incorporating emerging invaders in regional prioritization: a pragmatic approach for southern Africa. *Journal of Environmental Management*, **84**, 173–187.
- Millennium Ecosystem Assessment (2005a) *Ecosystems and human well-being: synthesis*. Island Press, Washington, DC.
- Millennium Ecosystem Assessment (2005b) *Ecosystems and human well-being: biodiversity synthesis*. World Resources Institute, Washington, DC.
- Nel, J.L., Richardson, D.M., Rouget, M., Mgidi, T., Mdzeke, N., Le Maitre, D.C., van Wilgen, B.W., Schonegevel, L., Henderson, L. & Naser, S. (2004) A proposed classification of

- invasive plant species in South Africa: towards prioritizing species and areas for management action. *South African Journal of Science*, **100**, 53–64.
- Olckers, T. & Hill, M.P. (eds) (1999) Biological control of weeds in South Africa (1990–1998). *African Entomology Memoir* no. 1.
- Osborne, P.E. & Tigar, B.J. (1992) Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of Applied Ecology*, **29**, 55–62.
- Parsons, B., Short, J. & Roberts, J.D. (2009) Using community observations to predict the occurrence of malleefowl (*Leipoa ocellata*) in the Western Australian wheatbelt. *Biological Conservation*, **142**, 364–374.
- Peacock, D.S., van Rensburg, B.J. & Robertson, M.P. (2007) The distribution and spread of the invasive alien Common Myna *Acridotheres tristis*, L (Aves: Sturnidae) in southern Africa. *South African Journal of Science*, **103**, 465–473.
- Pressey, R.L., Cowling, R.M. & Rouget, M. (2003) Formulating conservation targets for biodiversity pattern and process in the Cape Floristic Region, South Africa. *Biological Conservation*, **112**, 99–127.
- Reddy, S. & Davalos, M. (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.
- Reyers, B., van Jaarsveld, A.S. & Kruger, M. (2000) Complementarity as a biodiversity indicator strategy. *Proceedings of the Royal Society of London B*, **267**, 505–513.
- Reyers, B., Fairbanks, D.H.K., Wessels, K.J. & van Jaarsveld, A.S. (2002) A multicriteria approach to reserve selection: addressing long-term biodiversity maintenance. *Biodiversity and Conservation*, **11**, 769–793.
- Richardson, D.M. & Thuiller, W. (2007) Home away from home — objective mapping of high-risk source areas for plant introductions. *Diversity and Distributions*, **13**, 299–312.
- Richardson, D.M. & van Wilgen, B.W. (2004) Invasive alien plants in South Africa: how well do we understand the ecological impacts? *South African Journal of Science*, **100**, 45–52.
- Robertson, M.P. & Barker, N.P. (2006) A technique for evaluating species richness maps generated from collections data. *South African Journal of Science*, **102**, 77–84.
- Robertson, A., Simmons, R.E., Jarvis, A.M. & Brown, C.J. (1995) Can bird atlas data be used to estimate population size? A case study using Namibian endemics. *Biological Conservation*, **71**, 87–95.
- Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions*, **7**, 15–27.
- Robertson, M.P., Villet, M.H., Palmer, A.R., Fairbanks, D.H.K., Henderson, L., Higgins, S., Hoffmann, J.H., Le Maitre, D.M., Riggs, I., Shackleton, C.M. & Zimmermann, H.G. (2003) A proposed prioritization system for the management of invasive alien plants in South Africa. *South African Journal of Science*, **99**, 37–43.
- Robertson, M.P., Villet, M.H. & Palmer, A.R. (2004) A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. *Diversity and Distributions*, **10**, 461–474.
- Rodrigues, A.S.L. & Gaston, K.J. (2001) How large do reserve networks need to be? *Ecology Letters*, **4**, 602–609.
- Rodrigues, A.S.L. & Gaston, K.J. (2002a) Rarity and conservation planning across geopolitical units. *Conservation Biology*, **16**, 674–682.
- Rodrigues, A.S.L. & Gaston, K.J. (2002b) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biological Conservation*, **105**, 103–111.
- Rodrigues, A.S.L. & Gaston, K.J. (2002c) Optimisation in reserve selection procedures – why not? *Biological Conservation*, **107**, 123–129.
- Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Egoh, B. & Mgidi, T. (2004) Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, **10**, 475–484.
- Stevenson, R.D., Haber, W.A. & Morris, R.A. (2003) Electronic field guides and user communities in the eco-informatics revolution. *Conservation Ecology*, **7**, 1–17.
- Thuiller, W., Lavorel, S., Midgley, G., Laverigne, S. & Rebelo, T. (2004) Relating plant traits and species distributions along bioclimatic gradients for 88 *leucadendron* taxa. *Ecology*, **85**, 1688–1699.
- Thuiller, W., Richardson, D.M., Rouget, M., Procheş, S. & Wilson, J.R.U. (2006) Interactions between environment, species traits, and human use describe patterns of plant invasions. *Ecology*, **87**, 1755–1769.
- TNC (2006) Conservation by design: a strategic framework for mission success. Available at: <http://www.nature.org/aboutus/howwework/cbd/files/cbd.pdf> (accessed 20 March 2009).
- Tolley, K.A. & Burger, M. (2004) Distribution of *Bradypodion taeniabronchum* (Smith 1831) and other dwarf chameleons in the eastern Cape Floristic Region of South Africa. *African Journal of Herpetology*, **53**, 123–133.
- Van Rensburg, B.J., Chown, S.L. & Gaston, K.J. (2002) Species richness, environmental correlates, and spatial scale; a test using South African birds. *American Naturalist*, **159**, 566–577.
- Van Rensburg, B.J., Koleff, P., Gaston, K.J. & Chown, S.L. (2004) Spatial congruence of ecological transition at the regional scale in South Africa. *Journal of Biogeography*, **31**, 843–854.
- Virkkala, R., Heikkinen, R.K., Leikola, N. & Luoto, M. (2008) Projected large-scale range reductions of northern-boreal land bird species due to climate change. *Biological Conservation*, **141**, 1343–1353.
- Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.M.E. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.
- Williams, P., Hannah, L., Anelman, S., Midgley, G., Araújo, M., Hughes, G., Manne, L., Martinez-Meyer, E. & Pearson, R. (2005) Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. *Conservation Biology*, **19**, 1063–1074.

Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, S., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, **13**, 11–22.

SUPPORTING INFORMATION

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

Table S1 Examples of atlas data sets, illustrating some of their variety.

Table S2 Summary of atlas projects focused on the southern African region.

Table S3 Details of four example atlas projects, Southern African Bird Atlas Project (SABAP), Southern African Plant Invaders Atlas (SAPIA), Tick Distributions Project (TickDip), South African Reptile Conservation Assessment (SARCA).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Mark Robertson is interested in the distributions of species and the application of ecological niche modelling to understanding potential distributions, particularly for invasive alien species.

Graeme Cumming is a spatial ecologist and an interdisciplinary theorist who has worked on a wide variety of systems and problems, mostly relating to the role of spatial variation in ecological and social–ecological complexity. He has published over 60 peer-reviewed articles and a coedited book, ‘Complexity Theory for a Sustainable Future’.

Barend Erasmus is a spatial ecologist with a strong interest in the spatiotemporal impacts of global change on elements of biodiversity at different scales. His current research focuses on quantifying and understanding patterns of long-term change in savannas of southern Africa by combining fine-scale field studies with broader scale landscape analyses.

Editor: Simon Ferrier