Use Of Databases For Research and Conservation of the Eastern Arc Mountains

Authors: Burgess, N. D., Fjeldså, J., and Botterweg, R.

Source: Journal of East African Natural History, 87(1) : 159-180

Published By: Nature Kenya/East African Natural History Society

URL: https://doi.org/10.2982/0012-8317(1998)87[159:UODFRA]2.0.CO;2

Journal of East African Natural History **87**: 159–180 (1998)

USE OF DATABASES FOR RESEARCH AND CONSERVATION OF THE EASTERN ARC MOUNTAINS

N.D. Burgess, J. Fjeldså
Danish Centre for Tropical Biodiversity
Universitetsparken 15, DK-2100, Copenhagen, Denmark

R. Botterweg
Tanzanian Biodiversity Database, Department of Zoology
P.O. Box 3560, University of Dar es Salaam, Dar es Salaam, Tanzania

ABSTRACT

The establishment and use of computer databases to hold biological records is common in many countries. Such systems can accommodate a large volume of data with many potential uses, but many databases have failed due to poor design or programming, or to the lack of a proper user-need analysis. In all biological databases there are problems with the data being used. The most common problems are uneven collection effort, taxonomic confusion leading to unreliable records, difficulties with finding exact collection localities, differences in map projections leading to inaccurate record positions and more.

Data in databases, either the raw biological records, or interpreted/modelled range distributions, can be used for analyses of value to both conservationists and to academic biogeographers. However, all such analyses are highly influenced by the analytical scale, and results from one scale cannot be used at another. This is a fundamental problem with analyses using computerised data, especially for conservation planning. Data at different scales can be used to illustrate areas of high species richness, or areas where species of narrow distributional range congregate, although the results from one scale may not be applicable at another.

Computer programs can be used to select areas so that all species in the database are covered. The last method, using the principles of complementarity, is the most efficient way to select 'ideal' conservation areas. With such an analysis the theoretical minimum number of areas required to conserve all the species within a given database (*e.g.* all the birds in Sub-Saharan Africa) can be chosen. However, all such analyses are only indicative, as they do not take into consideration population viability, threats to the areas selected, or other 'real-world' variables that are important when conservation plans are being formulated.

For academics the patterns of species richness and range-restriction (*i.e.* endemism) can be used in large-scale models that can develop and test hypotheses to explain why species are distributed as they are and how evolution/extinction may have operated over time in order to produce the patterns observed. Such studies can have relevance to the development of conservation plans at the broad scale.

INTRODUCTION

'Database' or 'Databank' are terms describing systems that hold information in an organised and retrievable form. However, the terms have increasingly come to mean computerised systems that can rapidly extract, display, and analyse large volumes of data.

More powerful computers, at affordable prices, have resulted in many biologists becoming interested in computer databases which are able to a) store information on many sites in a country (or larger area) where certain vegetation types, habitats, or species are found; b) update taxonomy for species and records as these change in the literature (which is very hard with paper-based systems); and, c) undertake complex analyses that cannot be done by hand. For example, biological data can be overlain with vegetation maps, climate maps, topographical maps, population maps, protected areas maps etc., and various types of analyses can be undertaken (for discussion see: Scott *et al.*, 1993; Stoms & Estes, 1993; Turpie & Crowe, 1994; Lombard, 1995; Tushabe *et al.*, in press; Reynolds *et al.*, in press). The increasing availability and types of satellite derived data (see Olivieri *et al.*, 1995 for review) means that there is a ready supply of digital information, in vast quantities, on various physical and biological parameters of the earth, which can be used to compare against biological data sources (*e.g.* Erhlich & Lambin, 1996; Lambin & Erhlich, 1996; Nøhr & Jørgensen, 1996; Jørgensen & Nøhr, 1996; Fjeldså *et al.*, 1997). With the addition of more and more data, and the better interfacing between computers, the possibilities for interesting analyses increase every year.

The computerisation of biological and physical data has also allowed the 'true' distribution of species to be modelled using physical data which are readily available (*e.g.* climate and soil data), even if the species is rare and infrequently recorded (Walker, 1990; Stoms, 1992; Scott *et al.*, 1993; Margules & Austin, 1994; Margules & Redhead, 1996; Tushabe *et al.*, in press). Searches on the ground can confirm how accurate these predictions are, allowing the model in the computer to be refined, or confidence limits to be placed on the accuracy of the model.

The aim of this paper is to review the type of databases/databanks available for tropical conservation planning and to assess their strengths and weaknesses. We will illustrate the fundamental importance of geographical scale to the design and use of databases, including two existing systems that hold data on the Eastern Arc area of eastern Africa. We review factors that cause problems with the data added to biodiversity databases, and look at how data in these systems can be analysed to detect areas important for biodiversity. The relevance of these systems to research and conservation in the Eastern Arc is stressed.

DATABASE SYSTEMS

A number of non-computerised database systems have existed for decades. Collectively these systems hold a lot of data, but typically the data cannot easily be retrieved or manipulated. In recent years there has been an explosion of computerised database systems. These have the same general functions as the traditional databases, but can hold far more data, manipulate it, and undertake analyses that were previously impossible.

Human mind

Strengths

People who have life-long experience of a biological group have great accumulated knowledge on species distribution, habitat preferences etc. The minds of such people can tell if data are flawed in some way, or assess if a result is ecologically reasonable. Their knowledge and skills are still essential for a computerised database program if it is to produce meaningful results.

Weaknesses

The human mind makes mistakes, especially when it comes to remembering precise details. There are also limits to the types and quantity of data that can be stored. The human mind is thus a poor system to save large numbers of geographical co-ordinates etc. A further weakness is that people die and their accumulated information is then lost.

Museum collections

Strengths

There are probably tens of millions of biological specimens in museum collections around the world, most of which have details of the collection site, habitat and collection date. Most collections of African flora and fauna (except in Southern Africa) are found outside Africa, particularly in the old colonial powers of Britain, France, and Belgium (and to a lesser extent Germany, Portugal and Spain) (table 1). These collections are a vast potential source of distributional data. Eastern Arc specimens are mainly found in Britain, although there are also old German collections, and smaller samples in other European countries (especially Scandinavia) and the USA.

Table 1. Numbers of bird specimens held in some of the world's major museums, with an indication of the countries from Africa that are represented in the collection

Museum	No. Specimens in collection (not all Africa)	African focal areas
Europe		
Bonn (Germany)	75,000	Afrotropics, North Africa
Edinburgh (Scotland)	60,000	Worldwide
Frankfurt (Germany)	90,000	North Africa
Copenhagen (Denmark)	110,000	Tanzania, Uganda, Kenya
Leiden (Netherlands)	300,000	Liberia, Tanzania, Kenya
Milan (Italy)	34,000	NE Africa
Oxford (England)	19,000	South Africa
Tervuren (Belgium)	134,000	Central Africa and Afrotropics
Florence (Italy)	21,000	NE Africa, Libya
Tring (England)	1,000,000	Former British Colonies in Africa (Gambia, Sierra Leone, Ghana, Nigeria, Cameroon, Kenya, Uganda, Tanzania, Malawi, Zambia, Zimbabwe, Botswana, South Africa, Lesotho, Swaziland)
Vienna (Austria)	95,000	E. Africa

Others

Large collections of African birds also exist in various museums in the USA, and in South Africa, Zimbabwe and Kenya.

Weaknesses

The data contained in museum collections are generally very difficult to access. Collections of European museums are largely or even totally non-computerised, largely because funding levels have been declining for many years and routine collection management (prevention of infestations by pests) is now the main activity carried out. Many European museums also now have data access policies that are designed to ensure that museum data are not obtained for free, and then used for profit (scientific or financial) by others. However, these policies can also stifle beneficial collaborative endeavours with those more familiar with computerisation procedures. Conversely, most museum collections in the USA are computerised, and the Government's open data-access policy means that data must be given to those requesting it. South African museum collections are also typically computerised, but hold little data of relevance to eastern Africa.

Some data and specimens from these overseas museums have been repatriated to eastern African countries. More is likely in the future, but the management of specimen collections in eastern Africa (particular in the hot and humid coastal areas) is difficult and expensive. The management of computerised data, on the other hand, represents a smaller problem and is considerably cheaper.

Paper-based systems

Strengths

Paper-based systems can hold information that is not standardised and thus cannot easily be entered into a computer. For example, a paper-based file on a particular species could contain lists of localities, references on the species, and distribution maps at different scales. Once organised they can provide a convenient way to hold these kinds of data, which may be the basis of a programme of data computerisation, or the source of verification for computerised data.

Weaknesses

This form of data is very static. It cannot easily be manipulated, cannot be easily provided to other parties, and its use requires considerable additional work, particularly if they are to be put into a standardised computer database.

Point-locality computer databases

Strengths

The main advantage of such systems is that they can hold precise locality details for a very large number of biological specimens. They also hold various other attributes of the specimens: dates, season, weather, collector name, collection method, habitat etc. When sufficient data are compiled then many different analyses are possible by using different combinations of datafields. Point locality data for different specimens can also be extracted and presented on maps using the many different GIS systems now commercially available. Computerised databases in Kenya and Tanzania hold a lot of data on the species of the Eastern Arc, but the coverage is far from complete, and a lot of further data are available elsewhere in the world, but have not been entered on the systems in eastern Africa.

Weaknesses

The design and establishment of a functional point-locality database is much more complex than initial discussions with computer scientists might suggest. Considerable investments of time and money have been made around the world to establish biological database systems that have never worked. Common reasons for failure are: a) over-complication, b) poor programming, c) use of software which cannot transfer data to newer systems, d) unreasonable expectations of the possibilities of computers, e) lack of a clear purpose for the database, f) lack of data-entry capacity (people), g) entering large volumes of poor quality data, h) insufficient data checking, and so loss of database credibility. In eastern Africa the point-locality biodiversity databases in Uganda, Kenya and Tanzania are functioning quite well using the commercially available Microsoft ACCESS programme, a small number of staff, and normal PC computers (Reynolds *et al.*, in press). Inputting new data into these systems and checking existing data are, however, extremely time consuming. Experts are needed to ensure that data-typists do not enter large numbers of errors. There are no shortcuts here and unregulated entering of data will result in an unusable database.

Geographic information systems and related programmes

Geographic Information Systems (GIS) are computerised tools designed to assemble and analyse spatial data. These data can be on species, vegetation, human population, or any number of other mappable variables. There are two main types of systems, and some derivations or specialised applications of the main themes. Vector-based GIS systems hold data in the form of x-y co-ordinates, which make up polygons, and thus link closely to the database systems described above. There are several commercially available systems, including the market-leader: Arc Info/Arc View. Raster-based GIS systems hold data as squares, with the number of squares needed to represent a mapped feature being dependent on the size the feature and the size of the squares used to map it. A commercially available system of this type is IDRISI. The WORLDMAP software, developed by Paul Williams of the Natural History Museum in London, stores maps of species distributions on the basis of presence/absence records within grids, and is thus similar to a raster-based GIS.

Strengths

All types of GIS systems (and WORLDMAP) can be linked to point-locality databases, and thus data from other sources can be imported and analysed. The grid structure of raster-based GIS and WORLDMAP is advantageous when comparing with other grid-based data sources, for example satellite-derived weather, topographical, and image data. WORLDMAP software is specifically developed for fast interactive analyses of biodiversity data using sophisticated analytical procedures, but without the need for expensive computers and GIS training (Williams, 1995). It is thus within the reach of almost all biodiversity scientists with a modern computer. Vector-based GIS systems, however, offer the greatest flexibility to work at different scales and high degrees of complexity.

Weaknesses

Raster-based GIS systems and WORLDMAP have the general disadvantage of a pre-set data structure (grid), although this is more serious in WORLDMAP where data cannot be transferred between grid sizes. It is also difficult to export WORLDMAP maps to other software and thus the mapping and flexibility advantages of GIS systems cannot be realised easily. Vector-based GIS systems (especially Arc Info) are expensive in terms of machinery and training before they can be used, although such problems are declining as simpler programmes are produced. All these systems have a general problem with data-availability and quality, and a further problem is that poor quality data can easily be 'hidden' beneath sophisticated analysis and attractive graphical presentation.

The use of distributional modelling

Biological data have not been gathered evenly from all parts of the world, and indeed there are substantial biases towards some places opposed to others. Moreover, it is not possible to conduct extensive surveys of the vast unknown areas that exist, for example in the Central African swamp forests, or war zones. Thus some biodiversity studies have 'modelled' potential species distributions using available (incomplete) data. Modelling approaches are most useful for species that, although poorly known, have a reasonable number of collection sites and whose habitat preference is known.

Advantages

Modelling approaches may reduce collection biases in point-locality databases, and thus may provide species distribution patterns that are closer to the 'truth' (Scott *et al.*, 1993; Margules & Redhead, 1996; Tushabe *et al.*, in press). Computer-based modelling can even give a statistical confidence limit to the estimated occurrence of a species (Margules & Redhead, 1996). It is also repeatable and testable as the data are retained and the analyses can be re-run using the same, or another statistical routine. Modelling potential species ranges can also be done using the accumulated expertise in the brains of world experts (here termed 'mental modelling'), assisted by detailed habitat, climatic and topographic maps.

Disadvantages

Computer modelling requires adequate biological data points and the availability of physical data at the same level of resolution. In the Eastern Arc there would be problems to obtain adequate climate data from which to construct the models, although smoothed and computer-modelled climatic maps, derived from satellite information, are available (*e.g.* CRES, 1995). Vegetation data are often available, but may date from the colonial period and thus do not reflect the current situation on the ground. Older biological data may also be a problem as suitable habitat, and thus the relevant species, may have disappeared. Mental modelling approaches have been used in the construction of many range maps in textbooks, and can also give an assessment of the likely distributions for species that have not previously been mapped in this way. Regardless of what system is employed, there is no known way to model the distribution of a species known only from one locality, of which there are quite a number within Africa, including in the Eastern Arc.

ISSUES RELATED TO COMPUTERISED DATABASES

The remainder of this paper will be devoted to those issues that are particularly important for the establishment of computer databases, including the types of analytical procedures that are most commonly undertaken with computerised data. Strengths and weaknesses are brought out where possible, and links are made to the situation in the Eastern Arc.

The importance of scale

The scale of an analysis, the data used, and the required results are all closely interlinked and have a fundamental influence on decisions concerning computerised databases. In general, scale of design in a computerised database has similar divisions to that recognised in community and landscape ecology (*e.g.* Forman, 1995; Stoms & Estes, 1993; Whittaker, 1977). Thus, databases can be constructed for analyses at the following levels of scale: global, continental (epsilon), regional (delta), landscape (gamma) and within-community (alpha). In this situation 'beta' diversity is not relevant, as it relates to environmental gradients, and so can be analysed using an alpha level database. Also, the 'point' level is not relevant, as this is the unit of data gathering for most types of databases.

In general, the larger the scale of the analytical area and the 'coarser' the analytical filters, the larger the scale at which data can be stored in the database. When looking at theoretical questions at the global scale (*e.g.* Levin, 1992; Williams & Gaston, 1997), a 5 or even a 10 degree square grid might be adequate, but such a grid would not provide any useful information for conservation planning at the national scale (figure 1). Conversely, for a continental scale analyses of biodiversity patterns and processes, the gathering and inputting of all available data from the literature and museum collections as data-points would be so time consuming that it would require several lifetimes to complete. Useful results can in this case be obtained from presence/absence gridded systems (Burgess *et al.*, in press; Williams *et al.*, in press; figure 1). Some of the problems of using different scales on the results of a species-richness analyses have been presented elsewhere (Levin, 1992; Stoms, 1992).

Perhaps the most urgent challenge facing all distributional databases in the Afrotropical region (and elsewhere) is to either improve the quality of the distributional data, or make use of the available data by following some generally agreed rules (Frietag & van Jaarsveld, 1995; Stoms & Estes, 1993). Precise distributional maps at a high resolution using recently gathered point records from field studies covering the entire potential range of a species would solve most database problems, but these are generally not available for even the well-studied groups like birds. Such data could only be created through a major programme of intense survey work, but at a time when the amount of fieldwork is probably declining in most parts of Africa, such an aim is unrealistic. However, even if such survey programmes were completed data would still remain incomplete and it is thus extremely relevant to ask "how complete the data must be to change the qualitative results and inferences that are drawn from them" (Kodric-Brown & Brown, 1993).

A partial solution to these dilemmas is to use detailed point records only for small-scale (alpha level) investigations, and rely on range maps or modelled distributions for gamma (landscape) or epsilon (regional) studies. Both approaches, however, rely on the advice of experts, who in the African context are declining in number, ageing and often living outside Africa. They are typically not being replaced by nationals from countries within Africa (unlike in South America and SE Asia), and even where they do exist these people are often severely hampered by a lack of funds and have little access to reference literature needed to do their work.

Considerations at the global scale

At the scale of the globe, analyses often aim to test general hypotheses or provide a general indication of the distribution of biodiversity. Such analyses have often relied on extremely large grids (5 x 5 degree or 10 x 10 degree) as the analytical unit (*e.g.* Gaston & Williams, 1993; Williams & Gaston, 1994), or whole countries (*e.g.* Mittermeier, 1988). However, such a scale of analysis tends to hide much of the variation in the data, and thus cannot be used for conservation planning.

Regional

Analyses at the regional scale can provide both a regional overview of biodiversity, and also allow theoretical questions on the reasons behind this distribution to be addressed. Within Africa there are two biogeographical regions (Palaearctic and Afrotropical—or three if the Cape Region is also recognised), and various offshore islands (*e.g.* Socotra, São Tomé,

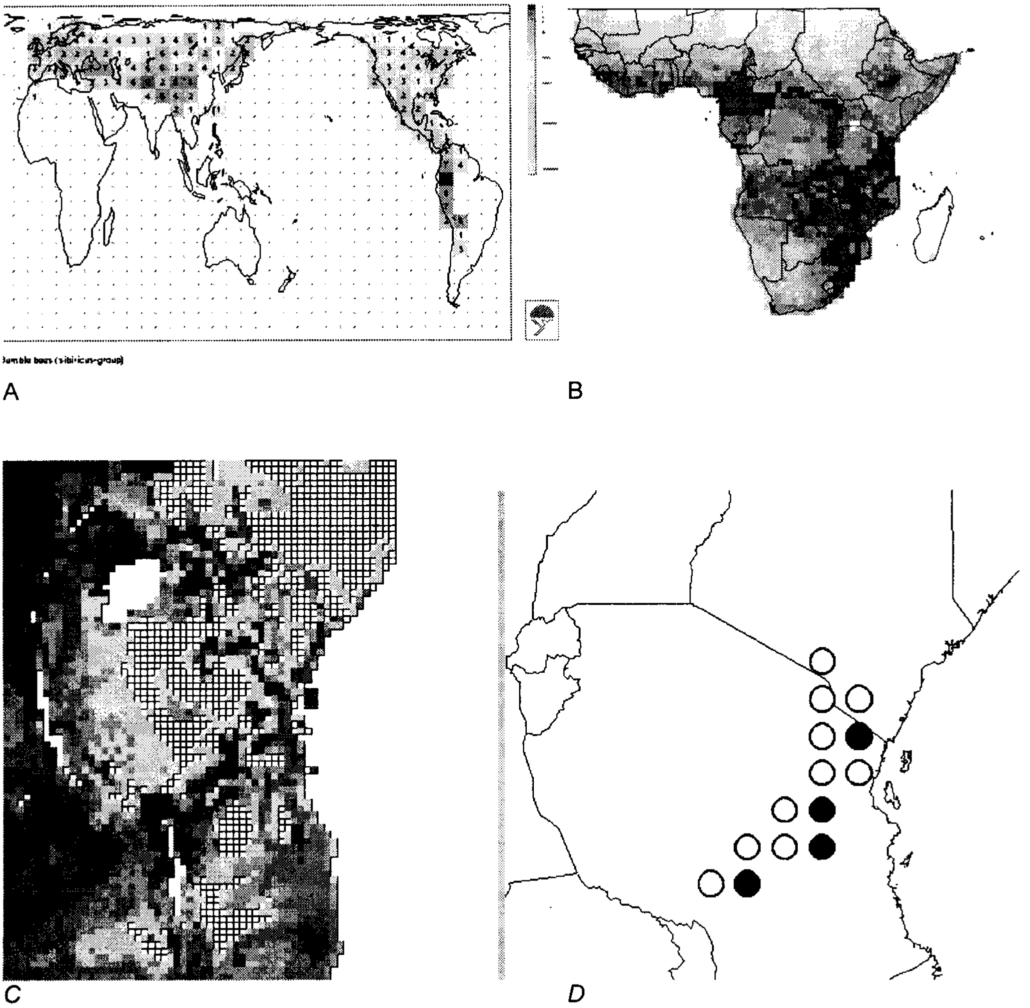

Figure 1. Biodiversity results at the different scales of analyses: A) Global scale distributional data—WORLDMAP analysis of bumblebee distribution; B) Regional scale distributional data—WORLDMAP analysis of Afrotropical amphibians: C) Landscape/country scale distributional data—WORLDMAP analysis of eastern African forest birds, and D) Point locality scale of distribution data—records of Afrixalus uluguruensis*—in eastern African forests. Black dots are records and open dots are interpreted range.*

Principé etc.) which have their own unique biodiversity. It is thus important to decide the limits of the study area before data compilation begins, and selecting one region (*e.g.* the Afrotropics) might be reasonable (*e.g.* Burgess *et al.*, 1997).

Because a region like the Afrotropics is so large the data compilation exercise and the type of database used needs to be considered. It is important to ascertain whether most data are in the form of range maps, point locality maps, a mixture of types, or in the form of already computerised specimen data. For this region (apart from South Africa), the available sources of biological data are varied, and are often old. Computerised specimen databases are few and not easily accessible. Some species are still known only from one locality, often as the holotype specimen. The scale of the analytical region, and the types of data available thus suggest the use of a relatively simple database, which does not rely on specimen localities alone, and which accommodates the data in quite large units. Raster-based GIS or the WORLDMAP software could be suitable for such a study (see Burgess *et al.*, 1997; in press) (figure 1; table 2).

Table 2. Factors believed to affect species richness at each ecological scale and of relevance to database design and operation

Scale	Main Factor	Other factors
Epsilon (continental)	Physiography	Productivity, climatic zones, historical, deforestation
Gamma (landscape/country)	Habitat and topographical diversity	Productivity, climate (rainfall etc.), habitat fragmentation, land-use changes etc., fire
Alpha (within habitat)	Structural complexity	Productivity, climate (rainfall etc.), fire, floods, pests, land-use changes, competition etc.
Point (sampling position)	Microhabitat factors	Competition etc., predation, parasitism

An example of a database that has been constructed to work at this scale, is the African Vertebrates database of the Danish Centre for Tropical Biodiversity, the Percy FitzPatrick Institute in Cape Town, and the Natural History Museum and York University in Britain. The primary aim of this project is to present and understand the reasons behind the distribution of species in the Afrotropical Region. Species distribution maps use a 1 x 1 degree grid as the mapping unit. A one-degree square may not be ideal, but using a finer scale would give a false impression of accuracy given the generalised nature of much available distributional data, and a coarser scale would loose the ability to assess variation in the distribution patterns.

The data gathering process started by creating a species list for the taxa to be mapped. This required consultation with standard taxonomical lists (*e.g.* see Burgess *et al.* (this volume) for references), which were updated to include newly described species. Distributional data for the species are then gathered from various sources to make the maps and these data are input into WORLDMAP, which is specially developed for rapid interactive analysis of biodiversity data (Williams, 1995). For specimen data considerable checking of localities was required and standard national gazetteers were extensively used. Experts were also consulted extensively. Supporting material on how the species maps were made is stored in paper-based filing systems and also in a computer spreadsheet format. There are problems with this approach as there are with all others, but it allowed results to be produced within the time-scale of a single funding proposal (three years).

Some of the main problems are outlined below. Firstly, species data are often patchy and, for many species, sparse. Extrapolations therefore have to be made, most notably to try and produce reasonable distributional maps for Afrotropical species that are known from only a few locality records. It also has to be accepted that only the vertebrates (and probably not all the reptiles, small mammals and amphibians), butterflies and some plant groups have been studied in sufficient detail to allow them to be mapped with any confidence. The majority of the invertebrates (the majority of biodiversity) are so poorly known that they cannot be mapped at all.

Secondly a decision had to be made on whether to map species distributions as they used to be, or as they are now after man has altered habitats and reduced species ranges. The best example of this is black rhinoceros *Diceros bicornis,* which used to be widely distributed but is now extremely rare and found in a few well-protected sites. In our database such species are mapped according to their distribution around the 1939–1945 war and hence ranges are too generous for a number of species, especially the larger mammals. Decisions also have to be made over the best ways of mapping species that are rare and only known from a few localities; for example if a dense forest specialist species may be known from 25 specimens collected over 100 years from a large area in the Congolian forest—should the species be assumed to occur throughout the area, or should only the confirmed records be used in analysis and the distribution regarded as patchy or disjunct? Moreover, in the case of taxonomically unstable species, decisions have to be made on where to split, or where to lump, or in the worst cases where to omit data from whole genera as unusable. These and other decisions have had to be made, at a species-by-species (or genus by genus) level for the c. 4000 species contained in this database. Expert taxonomists, documentation, and some rules are essential in this work, if it is to be defended scientifically.

Landscape

The Landscape scale includes those analyses within a single country, which is generally the scale at which national biodiversity databases are constructed. Individual countries will need to know how many species they have, which ones are endemic, what their populations and habitat preferences are, and where they are found. This is particularly important for political aspects of conservation, such as the implementation of the Convention on Biological Diversity. At this scale, and for these types of analyses, species record (point) locality databases are to be preferred over range maps and other estimation methods (Frietag & van Jaarsveld, 1995; Frietag *et al.*, 1996), although modelling of distributions (*e.g.* Tushabe *et al.*, in press) may also be important where locality records are sparse. However, one of the criticisms of range maps—that they tend to reflect historical rather than current distributional ranges—is also true of point data, as these are compiled from many years of exploration, and these areas may now be a large town, or the species may have been lost due to the activities of man (*e.g.* Happold, 1995). A further problem with points is that they are biased towards well-collected sites, such as biological field stations and along roads and rivers. Modelling species records against habitat and other attributes may assist in the creation of more accurate distributional data at the landscape scale (*e.g.* Margules & Austin, 1994). However, there will still be problems where the models are based on either old localities, or old vegetation maps, as the predictions will include areas where the species habitat is no longer found. A further problem is that data for groups other than vertebrates and butterflies (*i.e.* almost all invertebrates and most plants) are not comprehensive enough to allow them to be modelled.

In Tanzania, a database system has been designed to work principally at the national or landscape scale, holding point locality records. It aims to hold detailed and precise locality records for all species in Tanzania, especially those which are endemic to the country, or globally threatened by extinction. The purpose is to use these data for conservation planning, decision-making, and to prevent developments being undertaken in places where there is a valuable component of species. The database is situated at the Zoology Department of the University of Dar es Salaam and holds mainly zoological data, although a similar database has also been started in the Botany Department to contain plant records. The first task of database development was to select hardware and software, and then start to enter data. Data are entered using the commercially available Microsoft Access database system, which is robust and simple to use. The database is linked to the commercially available MapInfo and IDRISI mapping software, which allows the data held to be displayed and maps produced together with lists of the most important species and where they are found. Some analyses can also be undertaken.

The structure of this database (figure 2) shows the general arrangement of the data and the links to other software programs. The data in the databases are held according to a standard list of the species in Tanzania, checked through the standard scientific literature mentioned above. Data are then compiled for these species, especially those of the greatest conservation interest. Data have been obtained from studies completed at the University of Dar es Salaam, and from collaborating scientists working in Tanzania, and from the older data already held in foreign museums.

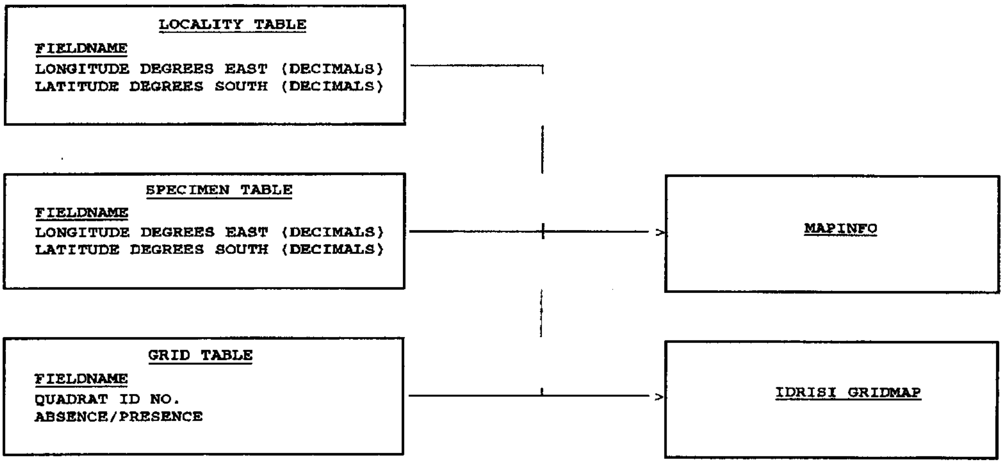

Figure 2. Data structure of the Tanzanian Biodiversity Database (University of Dar es Salaam)

Within habitat (Alpha)

For the Eastern Arc Mountains, the within-habitat (alpha) and the associated 'between similar habitat' (beta) scales are most often used for analytical purposes (*e.g.* Burgess *et al.*, this volume; Cordeiro, this volume; Fjeldså & Rabøl, 1995). At this scale point locality data are used, with no extrapolations. Such data are best held in point locality databases, and once entered they can be exported and mapped for the whole of the Eastern Arc (figure 1).

Gathering data for these analyses typically involves both new fieldwork and the critical appraisal of older studies. New work is very important as, for example in the Eastern Arc, there are still forested areas that have never been explored. At this scale it is often possible to use distributional data from biological groups that are poorly known at the level of the whole country, but where there have been detailed studies more locally (*e.g.* the spiders and millipedes of the Eastern Arc are reasonably known and this data can be compiled—see Burgess *et al.*, this volume). It is at this scale that detailed management decisions are made. Conservation projects in the Eastern Arc are eager to obtain relevant biodiversity data, and increasingly they support field data collection using standardised methods (see Johansson *et al.*, this volume) to provide the data they require for management planning purposes.

Data gathering for management purposes is well advanced in the East Usambara Mountains (Johansson *et al.*, this volume) and will start during 1999 in parts of the Udzungwa Mountains. In the future it is hoped to complete such programmes for each Eastern Arc forest, as has been achieved for all the Ugandan forest reserves. Data collected in the East Usambaras are stored within a module of the Tanzania Biodiversity Database at the University of Dar es Salaam, but specialised data-entry forms have been designed for use in the field, which are then linked to the fields on the data-entry screen in the computer and thus to the Access database (figure 2).

The importance of data-quality

The quality of the data in a database is extremely important when it comes to presenting results and drawing conclusions. The old proverb of 'rubbish in and rubbish out' is highly relevant to computerised databases, especially because the use of graphical and statistical programs can to a large extent hide the fact that the underlying data are bad. In database systems there are a number of potentially serious data problems. These are outlined below with examples from the Eastern Arc Mountains.

Survey effort

This may be the most serious problem of all in biological databases. Biological field stations illustrate the problem quite clearly; they are places where specialist researchers visit and while they are there they discover new species. These may remain known from only this site if further fieldwork is not carried out, making the field station seem very important in terms of species richness and the number of narrowly distributed endemics. Such areas can be described as 'over-collected' in comparison to the average. The opposite—gross 'under-collection' with regard to the average—is also found in Africa. Reasons for relative under-collection are that some areas are so remote (*e.g.* centre of the Congo Basin forests; within the extensive Miombo woodlands and deserts) or so dangerous (*e.g.* areas with prolonged wars such as Mozambique and Angola) that there have been few investigations of their flora and fauna. These areas will always score poorly for richness and generally also for endemism. Smoothing the effects of survey coverage is becoming an important scientific discussion (*e.g.* Tushabe *et al.*, in press). Even in the relatively well-collected South Africa, there are still serious problems of variations in survey effort (*e.g.* Lombard *et al.*, 1995).

In the Eastern Arc, the East and West Usambaras have been sampled more than other areas because of the Mazumbai Field Station in the West Usambaras, and the Amani Research Station in the East Usambaras. The lack of a similar station further south in Tanzania is perhaps the major reason that detailed exploration of the Udzungwa Range did not start until the later 1970s (Rodgers & Homewood, 1982). The lower coverage of the Udzungwas and other smaller ranges (*e.g.* Ukaguru and Nguru) influences the results of analyses using data held in biological databases. Moreover, local effects, such as the ease of access to Morningside and Bagilo in the Uluguru Mountains, as opposed to an area near a munitions factory on the same mountain, results in clusters of records from the same few sites.

Localities

The exact position of a particular locality is extremely important in a point locality database. It may seem like a simple matter to locate the place where a collection was made and then add the co-ordinates to the computer system in use. However, this is not always the case. Firstly, for old specimens, some localities were written as the port of shipping, which in Tanzania was often Zanzibar, Lindi, or Tanga. These specimens may have come from the interior. More recent collections are often located to a village near a forest, because (before GPS systems) there was no possibility to know precisely where you were in a dense forest or extensive savannah woodland. This may mean that specimens are mapped outside the habitat from where they actually came. A further problem is that many villages in Africa (especially in Tanzania under the Ujamaa programme) have moved, been re-established in a different location, or changed their name. This makes the use of modern maps to locate the position of old specimens difficult, and makes the precision of some records in point-locality databases spurious.

An example of a locality problem in the Eastern Arc is the use of the name Bagilo as the locality for numerous specimens from the northern Uluguru Mountains. This locality is a village outside the forest, but the specimens given this collection name were almost certainly collected inside the forest. Similar problems are found with the localities 'Amani' in the East Usambaras and 'Mazumbai' in the West Usambaras, which are the sites of biological stations and specimens bearing this name will have often come from nearby forests. Uncritical use of these data will tend to locate the areas of highest biological importance in odd places (towns, villages and field stations).

Taxonomists

The taxonomy of all organisms in the world is still unstable, and different taxonomists often disagree about the way that a particular group should be organised. As taxonomic units (normally species) are the basis of a biological database, constant changing of names and redefinition of species makes database design and maintenance very difficult. The level of activity and 'style' of a particular taxonomist can also be a problem. For example, if there has been one taxonomist working for 30 years on the amphibians of one country there will probably be a lot of species recorded, probably including a number of endemics described by that taxonomist. If the adjacent country (which may have similar habitats) has never had such a person, then the number of species known will be fewer and the number of endemics will probably be less. These differences may not be real, but cannot be estimated from the available data. This can cause some areas to have inflated species-richness and species endemism figures. Alternatively there could be two taxonomists in adjacent countries and one was a 'splitter' who recognised many species on small morphological differences, and the other a 'lumper' who did not. Data for these two countries would be difficult to compare, even if the effort they had expended was similar.

The birds of the Eastern Arc Mountains have been studied by several taxonomists over a long period of time. However, there have been long-standing disagreements on whether forms of birds on some of the mountains should be regarded as full species, or sub-species. Recent analysis of the DNA of these birds (*e.g.* Roy *et al.*, 1997) tends to support the sub-species being regarded as full species, but the controversy is likely to continue making it very difficult to organise Eastern Arc bird records in a database that will satisfy all people.

As another example, the presence of West African plant species in the Eastern Arc is not disputed (see Lovett, this volume), but there are some complications in that much of the taxonomy of the Congo Basin has been undertaken by French and Belgian botanists and most of Tanzanian plant taxonomy has been done by German and British botanists. Some plant species currently regarded as endemic to the Eastern Arc could actually be the same as species which are already described, but the relevant specimens have not been compared in different herbaria.

Maps

Most distribution maps are created according to the latitude longitude system, using a variety of projections to take account of the curvature of the earth. In such maps a single equally sized square on the map will cover a different land area on the ground. The grids are largest close to the equator and become smaller towards the poles, for example at the equator a 1 x 1 degree square will have a land area of 12,308 km^2, while at the southern part of South Africa, a 1 x 1 degree grid will cover 10,188 km^2. Such variations can affect the results of analyses as the larger grid close to the equator could hold more species simply due to the species-area relationship (*e.g.* Rosenzweig, 1995).

These problems are not confined to grid area, as projection problems can also cause the misplacement of different features on a map. For example, it is quite common to see maps of Africa where mapped boundaries of a forest reserve are not coincident with the boundaries of the forest (*e.g.* 1997 Tanzanian Land Cover maps). This is not because they do not agree on the ground, but because there have been problems with overlaying mapped data from different projections. Even if GPS co-ordinates are used, the same problems can occur because the GPS data are related to different 'datum points' and these can vary from those used to construct available paper or GIS maps of an area. Such problems can easily lead to biodiversity data from one system being mapping in the wrong place by another system.

Errors

There are errors inherent in many aspects of biological science. In biological databases, co-ordinates can be written down incorrectly, specimens can be incorrectly identified, and names can be transcribed wrongly. If data for biodiversity databases are taken from museum collections the rates of error are probably higher than in the published literature, and an identification error rate of 5 % for smaller mammal species, for example, is regarded as quite good. Adding these data to computerised databases might tend to lend an aspect of precision to data, which actually contain a fair number of errors. Such problems are increasingly serious in difficult to identify groups where the taxonomy is confused.

ANALYSES OF DATA IN BIODIVERSITY DATABASES

As outlined earlier the aims of a database, related to the scale of the area being covered, are linked to the types of analyses that can be undertaken. Despite these issues there are a number of commonly applied analyses of biodiversity data, most of which have been developed over the past 10 years. Many of these analytical methods are under constant development and there is considerable scientific debate on the merits and disadvantages of the various analytical approaches and what they can be used for (*e.g.* Pressey & Nicholls, 1989; Vane-Wright *et al.*, 1991; Pressey *et al.*, 1993; van Jaarsveld, 1995).

Species richness

This is a simple measure of the number of species found in a certain area. Analyses are normally undertaken within a gridded system and thus species richness relates to the number of species found in that particular square (figure 3). This number is related to the grid size. For example in a 1 x 1 degree square database each grid might be expected to contain more species than in a 15' x 15' square database. However, this relationship is not linear, some grid squares may cover only one habitat and thus the difference in species number between the 1 x 1 degree and 15' x 15' grids will be less than in squares which cover a wide altitudinal range or other forms of habitat heterogeneity.

The main strength of species richness information is that it is simple to calculate and to understand. It can show areas where there are a lot of species and those where there are few, but provides very little information on how important the species recorded are. Species richness scores have been used to generate lists of priorities for conservation. A general disadvantage of using species richness for conservation planning is that it will tend to select ecotones, areas with high heterogeneity, and thus not the most unique places (see Fjeldså & Lovett, 1997a). Moreover, tests of the efficiency of species richness versus other methods for selecting a set of areas to represent all species in a particular database, show that species richness is poor for this purpose, often no better than selecting areas at random (Williams *et al.*, in press) (table 3).

Range size rarity

This is a measure of the rarity of a particular species within the database. The main purpose of all such methods is to highlight areas that possess a high number of species with a small distribution and thus target conservation action to the prevention of species extinction. Williams *et al.* (in press) have shown that range size rarity is better at representing all the mammal species in a Afrotropical database than species richness (table 3), and that it is considerably better than selecting grids randomly. This is an important indication that conservation priorities selected using rarity are more efficient than those selected using richness.

There are many ways to calculate rarity values, using cut off approaches (e.g. selecting species with ranges of smaller than 50,000 sq. km, selecting species found in less than 10 grid squares etc.), or with more continuous approaches (e.g. of calculating the inverse range sizes for all species in a database and then adding together the inverse range size scores for all species in a particular grid to get a score for that grid (Figure 3)). The cut-off approaches have the advantage of being simple to implement but lack flexibility and cannot easily deal with species that fall marginally outside the pre-selected range. The continuous approach, using inverse range scores added up within grids, is also quite easy to understand in that grids with many species with small ranges will have higher scores than grids with many species with large ranges. However, there is a problem in that the range restrictedness scores are partially related to species richness, because values for grids with large number of species with small scores (large range) can be the same as scores for grids with one species with a large score (small range). Calculating the mean endemism scores for grids can help solve this problem. A disadvantage is that range size rarity approaches will tend to locate areas where the habitat patches are small and the species they contain can only possibly have small ranges.

Hotspots

The hotspots approach was originally devised to select areas of the world that possessed 5 % of the worlds species of plants within a reasonably constrained geographical area (Myers, 1988; 1990). The term has subsequently been used in a variety of ways to describe areas with high biodiversity values (either species richness or endemism), and has a generally understood but not very precise meaning. Data in biodiversity databases can be used to

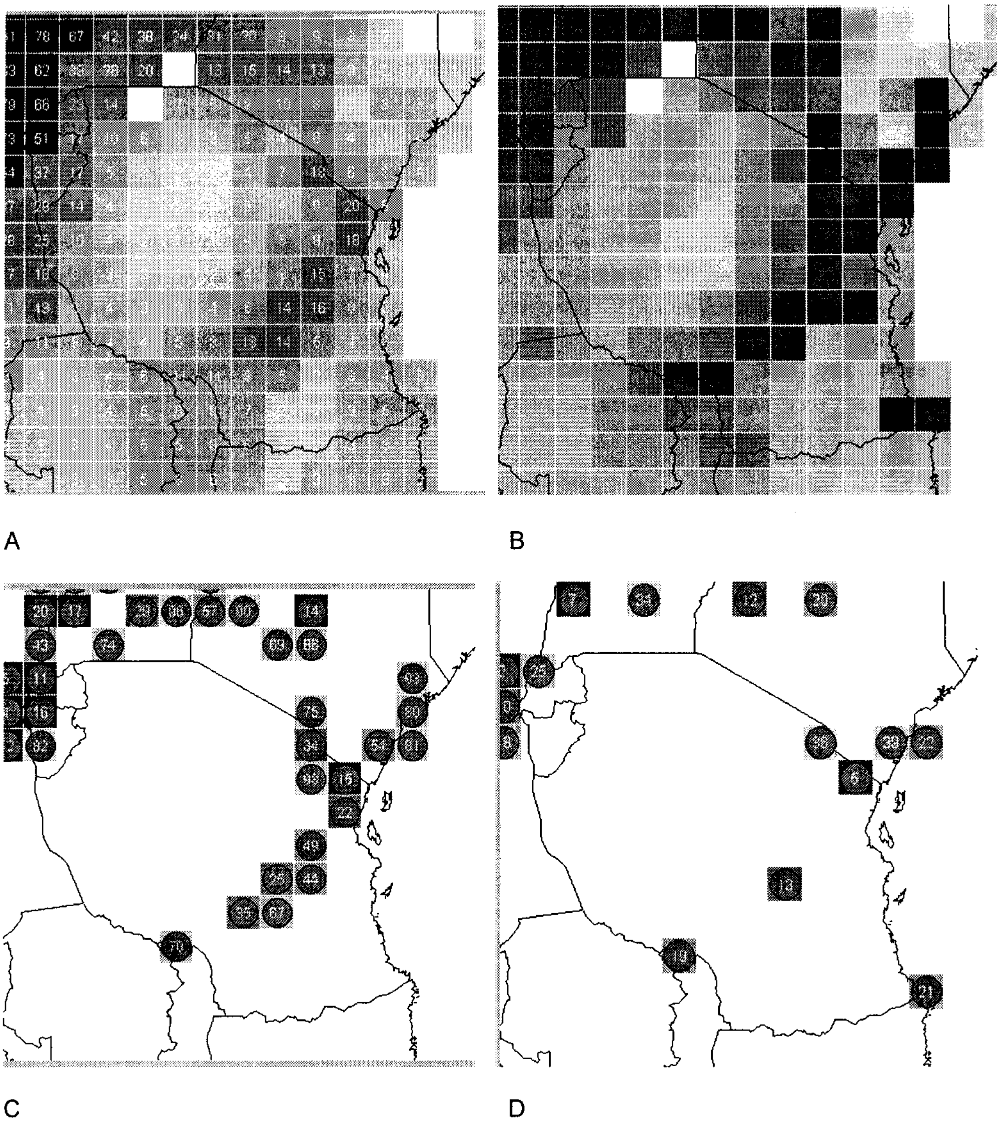

Figure 3. Analytical procedures to select priority areas. a) species richness—WORLDMAP analysis of species richness of forest mammals in eastern Africa; b) range size rarity (=endemism)—WORLDMAP analysis of forest mammals in eastern Africa; c) part of hot-spots—WORLDMAP analysis of endemism hotspots for forest mammals in the Afrotropical Region (top 100 grids; 9.5%); d) complementarity—WORLDMAP analysis of minimum set of areas (identified using the Greedy Area complementarity algorithm) to represent forest mammals in the Afrotropical Region at least once. Numbers indicate ranked importance, with low numbers having higher rank. Analysis shows only eastern Africa.

*Table 3. Three quantitative methods of selecting 50 priority areas for representing species applied to three questions in sub-Saharan biodiversity conservation and assessed against the consequences of choosing 50 areas at random. For each question, random draws of 50 areas are simulated 1000 times to calculate the mean percentage of species expected to be represented by chance. The percentage threshold score to the top 5% of randomly drawn scores (single-tailed, in parenthesis) shows the maximum percentage that would be expected by chance (significantly higher results in the same row shown with *) (Table from Williams* et al., *in press).*

		Method A Richness hotspots	Method B Endemism hotspots	Method C Comple-mentary areas	Random Mean, estimated from 1000 random draws (with estimated threshold to the 5% upper tail of the random distribution)
Question 1: - Which areas can represent the greatest diversity of sub-Saharan mammals?	% all mammal species (from 883) represented in 50 priority grid cells selected using all mammals	58	76*	90*	57 (61)
Question 2: - How well do areas chosen for mammals represent the diversity of birds?	% bird species (from 1911) represented in 50 priority grid cells selected using all mammals	73	85*	92*	79 (83)
Question 3: - How well do areas chosen for large mammals represent the diversity of small mammals?	% small mammal species (from 635) represented in 50 priority grid cells selected using large mammals (from 226)	47	47	59*	47 (52)

identify hotspots, based on a range of pre-defined criteria (for example the top 5 % grids according to species-richness or range size rarity scores). This approach shows a clear pattern of hotspots in the Afrotropics (figure 3), and for birds at least the range size rarity hotspots coincide closely with many of the Endemic Bird Areas selected by BirdLife International (ICBP, 1992). A general disadvantage of the hotspot approach is that although the top scoring grids are selected, these might all possess a similar species complement (more so with richness than endemism). Some other areas possessing fewer range-restricted species, which might be unique to that area, would tend to not be represented in this approach.

The hotspots approach (*e.g.* figure 3) brings out the importance of the Eastern Arc Mountains for all groups so far analysed (birds, mammals, snakes, amphibians and plants). It is suspected that the hotspot methodology would also result in the Eastern Arc being selected for most other groups as well, especially invertebrates (see Burgess *et al.*, this volume).

Complementarity

Complementarity works by first choosing the area that contains the most species in a database, and secondly an area which adds the largest number of species not already selected in the first area, and so on until all the species are covered. The computer can be told to represent each species once, twice or more times in the set of areas chosen, or to rank the areas in terms of their species richness or range size rarity scores. Such an analysis selects a more scattered selection of grids (figure 3), but these grids include all the species, which is not the case when areas are selected using either species richness or range-size rarity (Williams *et al.*, in press; table 3). Various algorithms have been proposed to do complementarity analyses, and there has been a considerable debate in the literature on which are the most efficient (*i.e.* take the smallest area of land (or numbers of grids) to achieve a pre-defined objective) (see Williams *et al.*, 1996 and Williams *et al,* in press for discussion).

Complementarity is important because it selects priority areas by looking at the full species and distributional content of the database, with the aim of achieving the objective set by the investigator. There are at least two problems though. One problem is that places where a species of animal or plant has been described and never found again are automatically selected. Some of these species may be scientifically dubious and perhaps should be omitted from a database before this type of analysis is run. A further (and potentially more serious) problem is that a grid with two or more narrowly overlapping species will be selected as the more efficient solution to selecting a 'minimum set' of areas, when in fact it might be biologically better to choose areas within the core ranges of both species (Larsen, 1997). Some of the areas selected by the complementarity approach may therefore represent ecotones and border areas that fulfil the computational objective, but that contain marginal and perhaps non-viable populations (Gaston, 1994). Despite the problems, on the regional or continental scale minimum sets identified using complementarity provide useful guidance on areas where conservation actions could be considered (Williams *et al.*, 1996). In all cases there needs to be a detailed consideration of the situation on the ground if the results are being used for conservation planning, and problems caused by the scale of data (see figure 4) should never be underestimated.

DISCUSSION AND CONCLUSIONS

Computerised databases are here to stay and will become increasingly useful to biologists working on both theoretical and conservation-related subjects. The scale of the problem to be tackled using the database, and the sheer quantity of data involved and the resources available will to some extent determine the type of database used. For rapid results at a large scale within the lifetime of a single funding proposal relatively crude map distribution databases may be acceptable. However, in the longer time frame, and for single countries and smaller scales than this, point locality databases are essential. These allow data to be moved between scales of analysis and provide the full details on every specimen or record, which are not available in map database systems. It is possible that all available data on the biodiversity of Africa could be added to point locality databases within one lifetime, if there were reasonable resources devoted to the project. However, at present it is much more likely that the databasing will be undertaken in a piece-meal fashion by countries and projects. It can only be hoped that these efforts can be joined up at some point in the future.

There is a general and serious problem of very uneven survey coverage in Africa. Drawing defensible conclusions from biodiversity data, when some areas have been studied in detail for more than 100 years by scientists from all over the world and from many different disciplines, and other areas have hardly even been even visited by a biologist, is not easy. Modelling potential or likely distributions using the available biological data and other available data sources such as climate, topography etc., offers a partial solution to the problem (*e.g.* Tushabe *et al.*, in press), although there are still problems with old data (for example the vegetation map used by Tushabe *et al.*, in press was from 1964). At least this solution is realistic and the methods exist to do it. The alternative—that of mass collection of many areas and the identification of everything that is collected is probably not realistic, particularly because of the differences of opinion between different taxonomists on the validity of many species, and the political instability of many countries in Africa today.

ACKNOWLEDGEMENTS

This paper owes a great debt to the discussions held with many people over the past three years, particularly Helen de Klerk of the FitzPatrick Institute in South Africa, Paul Williams of the Natural History Museum in London, and Carsten Rahbek, Bent Otto Poulsen and Poul Andersen of the Danish Centre for Tropical Biodiversity in the Zoological Museum of the University of Copenhagen, Denmark. It also incorporates much of the discussion received when this paper was presented orally to the Eastern Arc Conference in December 1997 in Morogoro, Tanzania.

The Danish Centre for Tropical Biodiversity thanks the many people who have assisted in the development of databases for Afrotropical Biodiversity, especially the Percy FitzPatrick Institute in South Africa, and the taxonomists mentioned in the paper in this conference volume on the fauna of the Eastern Arc Mountains. The Biodiversity database of the University of Dar es Salaam thanks the GEF Project on Developing the Capacity for Biodiversity Conservation in East Africa for the financial support to its establishment, and the Zoology Department of the University of Dar es Salaam for office space and technical assistance. It also thanks all those who have provided data used here, especially Dr Alan Rodgers (larger mammals of Tanzania), the FINNIDA-Catchment Forestry-Frontier Tanzania biodiversity survey project in the East Usambara Mountains, and several museum databases in USA. Paul Williams provided WORLDMAP and altered it for our requirements. He also allowed us to use the figure on the large-scale distribution of bumblebees. We thank Dr Alan Rodgers and Professor Kim Howell for their comments on the manuscript.

REFERENCES

Burgess, N.D., H. de Klerk, J. Fjeldså, C. Rahbek & T. Crowe (1997). Mapping Afrotropical birds: links between atlas studies and conservation priority analyses. *African Bird Club* 4: 93–98.

Burgess, N.D., H. de Klerk, J. Fjeldså, T. Crowe & C. Rahbek (in press). A preliminary assessment of congruence between biodiversity patterns in Afrotropical forest mammals and birds. *Ostrich*

Burgess, N.D., J. Fjeldså & R. Botterweg (this volume). The faunal importance of the Eastern Arc Mountains of Kenya and Tanzania. *Journal of the East African Natural History* **87**: 37–58.

Cordeiro, N.J. (this volume). Preliminary analysis of the nestedness patterns of montane forest birds of the Eastern Arc Mountains. *Journal of the East African Natural History* **87**: 101–118.

CRES (1995). *An African data sampler*. CD-ROM, CRES, Australia.

Ehrlich, D. & E.F. Lambin (1996). Broad scale land-cover classification and interannual climate variability. *International Journal of Remote Sensing* **17**: 845–862.

Fjeldså, J. & J.C. Lovett (1997a). Biodiversity and Environmental Stability: Guest editorial. *Biodiversity and Conservation* **6**: 315–323.

Fjeldså, J. & J.C. Lovett (1997b). Geographical patterns of phylogenetic relicts and phylogenetically subordinate species in tropical Africa forest biota. *Biodiversity and Conservation* **6**: 323–344.

Fjeldså, J. & J. Rabøl (1995). Variation in avian communities between isolated units of the Eastern Arc mountain forests, Tanzania. *Gerfaut* **85**: 3–18.

Fjeldså, J., D. Ehrlich, E. Lambin & E. Prins (1997). Are biodiversity 'hotspots' correlated with current ecological stability? A pilot study using the NOAA-AVHRR remote sensing data. *Biodiversity and Conservation* **6**: 399–420.

Forman, R.T.T. (1995). *Land mosaics: the ecology of landscapes and regions*. Cambridge University Press, Cambridge.

Frietag, S. & A.S. van Jaarsveld (1995). Towards conserving regional mammalian species diversity: a case study and data critique. *South African Journal of Zoology* **30**: 136–144.

Frietag, S., A.O. Nicholls & A.S. van Jaarsveld (1996). Nature reserve selection in the Transvaal, South Africa: what data should we be using. *Biodiversity and Conservation* **5**: 685–698.

Gaston, K.J. (1994). *Rarity*. Chapman and Hall, London.

Gaston, K.J. & P.H. Williams (1993). Mapping the world's species: the higher taxon approach. *Biodiversity Letters* **1**: 2–8.

Happold, D.C.D. (1995). The interaction between humans and mammals in Africa in relation to conservation: a review. *Biodiversity and Conservation* **4**: 395–414.

ICBP (1992). *Putting biodiversity on the map: priority areas for global conservation*. ICBP, Cambridge.

Jaarsveld van, A.S. (1995). Where to with reserve selection and conservation planning in South Africa? *South African Journal of Zoology* **30**: 164–168.

Johansson, S.G., P. Cunneyworth, N. Doggart & R. Botterweg (this volume). Biodiversity surveys in the East Usambara Mountains: preliminary findings and management implications. *Journal of the East African Natural History* **87**: 139–158.

Jørgensen, A.F. & H. Nøhr (1996). The use of satellite images for mapping of landscape and biological diversity in the Sahel. *International Journal of Remote Sensing* **17**: 91–109.

Kodric-Brown, A. & J.H. Brown (1993). Incomplete data sets in community ecology and biogeography: a cautionary tale. *Ecological Applications* **3**: 736–742.

Lambin, E.F. & D. Ehrlich (1996). The surface temperature-vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing* **17**: 463–487.

Larsen, J.B. (1997). Testing some possible ways to establish conservation priorities for Afrotropical mammals, using WORLDMAP. Unpublished Masters thesis: University of Copenhagen.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology* **73**: 1943–1967.

Lombard, A.T. (1995). The problems with multi-species conservation: do hotspots, ideal reserves and existing reserves coincide? *South African Journal of Zoology* **30:** 145–163.

Lombard, A.T., A.O. Nicholls & P.V. August (1995). Where should nature reserves be located in South Africa? A snakes perspective. *Conservation Biology* **9**: 363–372.

Lovett, J.C. (this volume). Importance of the Eastern Arc Mountains for vascular plants. *Journal of East African Natural History* **87**: 59–74.

Margules, C.R. & M.P. Austin (1994). Biological models for monitoring species decline: the construction and use of data-bases. *Philosophical Transactions of the Royal Society, London B* **344**: 69–75.

Margules, C.R. & T.D. Redhead (1996). *Guidelines for using the BioRap Methodology and Tools.* CSIRO for the World Bank, Australia.

Mittermeier, R.A. (1988). Primate diversity and the tropical forest. Case Studies from Brazil and Madagascar and the importance of the megadiversity countries. In E.O. Wilson (ed.), *Biodiversity*. National Academy Press, Washington D.C. Pp 145–154.

Myers, N. (1988). Threatened biotas: 'hot spots' in tropical forests. *The Environmentalist* **8:** 187–208.

Myers, N. (1990). The biological challenge: extended hot-spots analysis. *The Environmentalist* **10**: 243–256.

Nøhr, H. & A.F. Jørgensen (1996). Mapping of biological diversity in Sahel by means of satellite image analyses and ornithological surveys. *Biodiversity and Conservation* **6**: 545–566.

Olivieri, S., J. Harrison & J.R. Busby (1995). Data and information management and communication. In V.H. Heywood & R.T. Watson (eds), *Global Biodiversity Assessment*. Cambridge University Press, Cambridge. Pp 607–671.

Pressey, R.L. & A.O. Nicholls (1989). Efficiency in conservation evaluation: scoring versus iterative approaches. *Biological Conservation* **50**: 199–218.

Pressey, R.L., C.R. Humphries, C.R. Margules, R.I. Vane-Wright & P.H. Williams (1993). Beyond opportunism: key principles for systematic reserve selection. *Trends in Ecology and Evolution* **8**: 124–128.

Reynolds, J., H. Tushabe & P. Kasoma (1999). The establishment of a National Biodiversity Data Bank for Uganda. *EANHS Bulletin* **29**(3): 6–9.

Rodgers, W.A. & K.M. Homewood (1982). Biological values and conservation prospects for the forests and primate populations of the Udzungwa Mountains, Tanzania. *Biological Conservation* **24**: 285–304.

Rosenzweig, M.L. (1995). *Species diversity in space and time*. Cambridge University Press, Cambridge.

Roy, S.M., J.C.M. da Silva, P. Arctander, J. Garcia-Moreno & J. Fjeldså (1997). The role of montane regions in the speciation of South American and African birds. In D.P. Mindell (ed.), *Avian Molecular Evolution and Systematics*. Academic Press, London and New York. Pp 325–343.

Scott, J.M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, J. Ulliman & R.G. Wright (1993). Gap analysis: a geographical approach to protection of biological diversity. *Wildlife Monographs* **123**: 1–41.

Stoms, D.M. (1992). Effect of habitat map generalisation in biodiversity assessment. *Programmetric Engineering and Remote Sensing* **58**: 1587–1591.

Stoms, D.M. & J.E. Estes (1993). A remote sensing research agenda for mapping and monitoring biodiversity. *International Journal for Remote Sensing* **14**: 1839–1860.

Turpie, J.K. & T.M. Crowe (1994). Patterns of distribution, diversity and endemism of larger African mammals. *South African Journal of Zoology* **29**: 19–32.

Tushabe, H., D. Pomeroy & J. Reynolds (in press). Predicting bird species occurrence for conservation planning in Uganda. *Ostrich*

Vane-Wright, R.I., C.J. Humphries & P.H. Williams (1991). What to protect: systematics and the agony of choice. *Biological Conservation* **55**: 255–268.

Walker, P.A. (1990). Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography* **17**: 279–289.

Whittaker, R.H. (1977). Evolution of species diversity in land communities. *Evolutionary Biology* **10**: 1–67.

Williams, P.H. (1995). Using WORLDMAP for Windows 95: Priority Areas for Biodiversity, version IV. The Natural History Museum: London.

Williams, P.H., N.D. Burgess & C. Rahbek (in press). Hotspots of richness, hotspots of endemism, and complementarity areas: their success in representing the diversity of sub-Saharan mammals, their consequences for birds, and their application to large 'flagship' species mammals for representing small mammals. In A. Entwistle, & N. Dunstone (eds), *Future priorities for the conservation of mammalian biodiversity*. Cambridge University Press, Cambridge.

Williams, P.H. & K.J. Gaston (1994). Measuring more of biodiversity: can higher-taxon richness predict wholesale species richness? *Biological Conservation* **67**: 211–217.

Williams, P.H., D. Gibbon, C. Margules, A. Rebelo, C.J. Humphries & R. Pressey (1996). A comparison of richness hotspots, rarity hotspots and complementarity areas for conserving the diversity of British Birds. *Conservation Biology* **10**: 155–174.