IEA ISTITUTO ECOLOGIA APPLICATA

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR DEVELOPMENT
DIVISION VIII/A/1

AMD
African Mammals Databank

A databank for the conservation and management of the African mammals

Project n° B7.6200-94-15/VIII/ENV/1994/67

A GUIDE TO THE METHODS AND THE DATA

January 1998

IEA – Institute of Applied Ecology

No-profit Association
for the assessment and management
of natural resources

Main Office:
Via L. Luciani, 41
00197 Roma (Italy)

VAT 02154781005

Technical Office:
Via L. Spallanzani, 32
00161 Roma (Italy)

tel/fax: (int)39-6-4403315
E-mail: md2605@mclink.it

Participants

Luigi Boitani
Co-ordinator

Fabio Corsi
Alessio De Biase
Ilaria D'Inzillo Carranza
Maria Ravagli
Gabriella Reggiani
Iacopo Sinibaldi
Patrizia Trapanese

Field Work Teams

Botswana Team
Marks Dithlogo
Tebogo R. Molefhe
Benjamin H. Raseroka
Gabotsewe B. Sekgororoane
Ditshupo B. Selepeng
Iacopo Sinibaldi

Cameroon Team
John T. Banser
Alessio De Biase
Giuseppe Gabbi
Andrea Ghiurghi
François Kameni

Morocco Team
Brahim Haddane
El Ayachi Sehhar
Patrizia Trapanese

Uganda Team
Mathias Behangana
Gabriella Reggiani

IEA ISTITUTO ECOLOGIA APPLICATA

The Institute of Applied Ecology is a non-profit organization dedicated to the research, assessment and management of natural resources.
Via L. Spallanzani n° 32, I-00161 Rome

Extract from:

IEA 1998, AMD African Mammals Databank - A Databank for the Conservation and Management of the African Mammals Vol 1 and 2.
Report to the Directorate-General for Development (DGVIII/A/1) of the European Commission. Project No. B7-6200/94-15/VIII/ENV. Bruxelles 1998. pp 1174

1. Introduction

Ecologists and conservationists have traditionally focused their attention on narrow scales, both temporal and spatial. However, they have gradually learned that factors affecting local communities can hardly be understood without full reference to broader spatial and temporal scales (Pimm, 1991). Broad scale projects are rare for a variety of reasons (funding, long-term commitment of programmes and researchers, academic pressures, etc.) and yet they are of paramount importance if we are to understand the intricacies of the functioning of ecological processes and be able to manage them effectively. A broad scale approach to conservation has also been clearly advocated as the new and necessary way of dealing with the current conservation crisis (May, 1994).

In this work we used a broad scale approach to analyse the geographical ranges of the medium and large mammals of Africa, considering the continent in its entirety: we are aware that by maintaining the focus on broad scale trends and patterns, the particularities of local situations are basically lost. Moreover, the conservation of several African species is today restricted to the small remnant populations inside the protected areas and a study of their global distribution range may appear as an academic exercise. It is likely that the conservationist or manager who is working on the solution of species problems at local level will consider these analyses of little use: and yet we believe that no local action should be taken without a clear understanding of the global conditions of the species and their habitats. We believe this approach to be not an option for conservation, but a must. And it is even more important for international organisations that have to decide on priority funding and international programmes: indeed much conservation effort is wasted for lack of proper understanding of global patterns.

This project intends to contribute to the conservation of the African mammals by providing a first study of their global distribution patterns and by nurturing the new chapter of conservation biology on broad scale analyses. The databank also intends to make available to the conservation and scientific community the raw and semi-processed data that is needed to develop further analyses on distribution trends and patterns.

1.1 General background on the project

Conservation at the local scale has traditionally been more effective than programmes at broad scale involving regions or entire sub-continents. A main reason for this has been the almost complete lack of a broad scale approach in analysing the status and trends of the natural resources, particularly animal species. Broad scale analyses are becoming a fundamental step toward consistent assessment of conservation needs within a region: actions and funding priorities must be decided against the background of a comprehensive assessment of the status of the resources at broad scale, and even local projects should be set within the framework of a broader assessment. Thus, conservation programmes at the local scale could become much more effective when their general context is clearly understood.

Conservation priorities for animal species have long been based on the traditional Threat Categories adopted by the IUCN Red Data Books. However, although the IUCN lists have been fundamental in identifying the species most in need of conservation support, a tool is required to place this information into the context of spatial dynamics and to enable the comparison of priorities on a continental scale - beyond the limits of national priorities and opportunities. Conservation biology, the modern and most scientific approach to conservation, calls for data of better quality and provides new conceptual bases for the assessment of threats and the evaluation of conservation actions. It also demands more frequent analysis of broad scale spatial and temporal trends, inviting scientists and conservationists to expand their views to the global patterns that affect the status of natural resources.

Moreover, in 1995 the IUCN/SSC (Baillie & Groombridge, 1996) adopted a new system of threat categories largely based on data on population biology, ecology and distribution patterns of the species. Information on the latter is currently very poor for most species, particularly in Africa: maps of the range of the most charismatic species are available but it is surprising how few species have been mapped and how poor the general quality of the mapping is. At national scale, some countries have an excellent tradition of research/monitoring and, for them, species distribution patterns are well known, but in order for the new IUCN/SSC Threat Category system to fully exploit its capabilities the global distribution range of each species must first be analysed.

This is where the GIS (Geographical Information System) technology comes in, providing a powerful tool to store and analyse data for the identification and evaluation of conservation actions. The results of the present project aim to provide such a tool for the medium and large mammals of Africa. The project was designed to collect, store, organise and pre-analyse data for distribution to the community of institutions and individuals worldwide concerned with the design and implementation of conservation projects in Africa: as such, its ultimate goal is to provide background data and a service to the conservation community.

In order to be in a position to take informed decisions on priority actions, conservation agencies should have access to background information and data in a format that is suitable for project identification and feasibility studies. Yet insufficient appraisal of the actual status and exploitation of natural resources has been a recurrent weakness in several programmes. This is even more evident for agencies working at international level where priorities and programmes are spread over an entire continent such as Africa. The lack of reliable means of predicting potential scenarios has also strongly constrained the development of more adequate conservation responses: when these scenarios have involved several countries, the tools for handling them have been extremely poor.

Data banks allowing for use of GIS analysis potentialities were not available for African fauna before this project was implemented. Various GIS efforts are being developed at national and continental level (Botswana, Ivory Coast, Kenya, Tanzania, Uganda, UNEP/GRID at its Nairobi Headquarters, JRC in Ispra, Varese – Italy, etc.), but focus on wild fauna is either rare or restricted to small (one country) areas.

A new international project launched by IUCN and its Species Survival Commission aims at building a new system for sharing environmental data and information: the Biodiversity Conservation Information Service is a new consortium of 12 international organisations (the IUCN Commissions, WCMC, Birdlife, Conservation International, The Nature Conservancy, etc.) having the objective of building a meta-data bank of conservation-related information worldwide. Within this new context, the IUCN/SSC is building its own species data bank which draws on its network of more than 7 000 experts around the world. This data bank is now being designed to manage all data on endangered species, including evaluation of population sizes in the various fragments of the species' ranges; it will include both alpha-numeric information and a full range of georeferenced data on species distribution.

The same team that worked on the present project was among the inspirers of the IUCN/SSC effort. Thus the implementation of this project has followed a path parallel to that of the IUCN/SSC data bank to ensure the full viability and utilisation of its output.

1.2 Terms of Reference: the objectives

The project was approved with the following objectives representing the terms of reference for its completion:

- to make an inventory of existing data bases on environmental components from European, international and national organisations; to define the limitations of existing material in terms of utility for African counterparts, and to assess what would be needed for research output to be utilised for planning and decision making by African networks,
- to build and implement a data bank especially designed to gather, process and analyse spatial and temporal information relevant to conservation and management of African mammal species,
- to store all relevant data on the status, distribution, and relevant ecological parameters of medium and large African mammals,
- to carry out a series of conservation assessments of the endangered species (actual and potential ranges, population fragmentation, role of protected areas, potential reintroduction areas, etc.),
- to make the data and the analyses available to international and national agencies, research centres, decision makers at all technical levels, also through the growing network of electronic distribution,
- to establish links with African universities and the international community for the exchange of information and for training in the application of GIS to natural resources and conservation.

2. The approach, in synthesis

Geographical Information Systems technologies are now widely available and provide the opportunity for a renewed approach to building and managing environmental data banks. The possibility of linking series of alpha-numeric data with geographical data at any scale, and the capabilities of a GIS to work out the relations between them, has significantly modified the scope and the management techniques of environmental data banks. It is now possible to use these data banks not only to store data, but also as a powerful research tool and for decision making in managing biodiversity. This approach is even more important for fauna, as GIS makes it possible to evaluate the animal species' distributions in relation to several environmental factors which may concur in defining a habitat's suitability for a species.

Of all environmental components, fauna is one of the most difficult to show on a map because of its mobility, and this limits our capability of locating individuals of a species. At any given time, we can only have a certain probability level of finding an individual of a given species at a given location. Conversely, detailed knowledge of the distribution patterns of the animal species is indeed one of the most important aspects to be taken into account in their conservation, in establishing and managing protected areas, and in many programmes of environmental evaluation and management. As mentioned above, evaluation of the new threat categories for animal species is also based, along with the requirements of conservation biology, upon analysis of the fragmentation patterns within the surviving meta-populations.

The classic presentation of the distribution ranges of animal species drawn on maps implicitly relies on a probability assumption: the boundaries of the polygon that describes the distribution range enclose the area in which an observer has a chance of finding individuals of the species. This is generally termed the Extent of Occurrence (EO) of the species. Traditional Extent of Occurrence is of little use today, as it generally lacks an explicit indication of the probability threshold used to draw the boundaries, while it has a resolution that is normally not sufficient to define the species' real Area of Occupancy (AO). Thus it fails to convey important information concerning the distribution pattern within the species' range in relation to the environmental factors, and regarding its historical evolution. The concept of Area of Occupancy represents the areas really occupied by the species within the Extent of Occurrence (Gaston, 1991), which is indeed the most crucial piece of information needed for implementing an effective conservation plan. In practice the concept of Area of Occupancy represents a better approximation of the EO, and the robustness of its representation depends on the biology of the species and the quality and scale of the data used. Nevertheless it is an interesting and useful development in the representation of animal species distribution and future improvement will certainly build on it.

The use of GIS in conjunction with adequate knowledge of the ecology of a species can be a first step in this direction. And this data bank represents an example of this approach applied to the identification of the species' Area of Occupancy. Different pieces of information have been assembled to achieve this goal, each of which has a specific meaning and represents one of the many components contributing to a species' distribution pattern.

The first piece of information included in the process of AO identification is the Extent of Occurrence. This data layer is used to discriminate between expected and possible presence of the species. In fact, models based on simple environmental suitability generally do not account for driving forces such as historical constraints (see Morrison et al., 1992 for a review) and species' behavioral patterns. Thus their results tend to create an image of suitability that is far from the true extent of the Area of Occupancy. In the process used in this project, the EO is assumed to represent all those driving forces for which an explicit descriptor could not be found in a form suitable for its inclusion in the models. Thus with regard to the boundaries of the Extent of Occurrence the results of the models should be read as expected (either presence or absence) inside of them and potential (in this case, obviously referring only to presence) outside of them.
The second piece of information is, in fact, composed of a set of layers that describes the spatial variation of the environmental conditions within the area under analysis (here the entire African continent). The basic requirements of the environmental layers included in this data set are some level of correlation with the driving forces of the species' distribution pattern and complete coverage of the entire area under analysis. It is important to note that the models are not meant to identify the variables that determine the species' distribution and only use those available to maximise the representation of the species' distribution.

Further information is needed to define an index of suitability based on the environmental variables selected. This calls for a way of defining how the available environmental variables are linked to the species' presence. In this project, two approaches were followed: a deductive approach and an inductive one (Corsi et al., in press).

The former describes the species' environmental preferences, as derived from available literature, in terms of environmental variables in the database. Based on this description, one or more experts can derive a ranking of suitability for each different combination of environmental layers observed within the area under analysis. Thus the Area of Occupancy appears as a patchwork of more or less suitable areas (Categorical-Discrete distribution model, CD).

The inductive approach uses the information which is implicitly available in the Extent of Occurrence to build a function that is capable of ranking the entire study area according to a continuous suitability index (Probabilistic-Continuous distribution model, PC). Even though its results are more objective, this is the first time that it was used on a continental scale in Africa, and it is thus presented only as a parameter for discussion and comparison with the distribution model derived from the deductive approach.

3. The species

The objectives of this project loosely referred to all medium and large African mammals. Selection of the species to be included was adapted to suit various criteria more than a rigorous prescription of minimum size, although the size of a rabbit was taken as the lower threshold.

The final list was compiled using the following rationale:

- all species belonging to the orders of Primates, Carnivora, Perissodactyla (except rhinos, see below), Hyracoidea, Tubulidentata, Artiodactyla, Pholidota and Lagomorpha were included, irrespective of their size, completely accounting for these taxonomic groups;
- although small, Macroscelidea are considered of particular conservation concern, and were thus added to the list;
- among Rodentia, 7 species (belonging to Dipodidae, Pedetidae and Hystricidae) were included, as they are of medium size and/or particular conservation or ecological interest. Following the same criteria, the three species of Tenrecidae not endemic to Madagascar and the entire family of Erinaceidae were included for Insectivora;
- possibly requiring a different environmental data set and separate analyses, all Chiroptera, Cetacea and Sirenia were excluded, except for the African manatee (*Trichechus senegalensis*);
- three notable species were excluded, the rhinoceros (*Diceros bicornis* and *Ceratotherium simum*) and the elephant (*Loxodonta africana*): the two species of rhinos were excluded because data on the last few areas in which they are found are being kept from the public and we did not want to interfere with this important decision; the elephant because an excellent and detailed database in a format very similar to the one proposed here is kept in Nairobi by the Elephant Specialist Group of the SSC/IUCN. For a similar reason, all Madagascar species were also excluded, as there is an extensive database produced by the World Bank;
- newly described species (i.e. *Pseudopotto martini*) were excluded due to an almost total lack of information, especially as concerns their distribution range;
- finally, species thought in recent years (the last 20-50 years) to be extinct were excluded, along with those that occur in Africa only as recently introduced populations (i.e. the fallow deer, *Dama dama*) or those that are totally domesticated (i.e. the dromedary, *Camelus dromedarius*).

A total of 281 species, belonging to 12 orders and 28 families were included in the data bank.

Species nomenclature generally follows “Mammals species of the world” (Wilson & Reeder, 1993). Even though this reference provides the most comprehensive and up-to-date taxonomic treatise on African mammals, a different classification was adopted for some groups. This was particularly true when a different systematic arrangement was recommended by the relevant IUCN/SSC Specialist Group or supported by recent evidence. For instance, the way in which the species *Procolobus badius* is considered in this data bank follows the indication of the IUCN/SSC Primate Specialist Group, which suggests retaining the different forms of red colobus monkeys (recognised as different species in Wilson and Reeder, 1993) in a single species classification until a thorough re-analysis of their relationships is undertaken (Oates, 1996).

The arrangement adopted does not pretend to be a final solution, given the continuous changes in and refining of mammalian phylogenetic information, the related systematic relationships and naming process will most probably change.

3.1 Literature review

A comprehensive review of the literature was carried out to collect and store all relevant and available information on the status, distribution and biological and ecological features of each mammal species. This search covered the literature published over the past 5-10 years, even though for some poorly known species it was extended to include most of the literature ever published.

The literature review was achieved by means of an exhaustive keyword-based search covering taxonomy, ecology and geographical distribution within the geographical and administrative bounds of the study area.

General and specialised publications were used as introductory literature for collecting references: i.e. "Keyword Index of Wildlife Research" (1990-1995) "Zoological Records" (1990-1996) and "Current Advances in Ecological and Environmental Sciences" (Pergamon Press, 1990-1995). Computer based searches were carried out on the following databases: Medline, Cab Abstracts, Life Sciences, Nisc Disc; either through direct access to the CD-Rom or through public library access.

The references listed in the most recent and comprehensive monographs concerning African mammals were also examined (Estes, 1991, Skinner & Smithers, 1990, Macdonald, 1984, Mammalian Species of the American Society of Mammalogy, IUCN/SSC Specialist Group Series, C. Helm Mammal Series, IUCN Red List Series, etc.).

The quantity and quality of information available for each species vary significantly; therefore although the same analyses were carried out on all species, the reliability of the results is strictly dependent on these factors..

3.2 Distribution data

Distribution maps were chosen as the best way to represent the available knowledge on each species' occurrence on the continent. Only in a few cases were point locations considered and the limits of the EO extended to include single known records of a species; this was generally restricted to those cases for which the existing distribution maps were not updated or precise enough to include some confirmed records of a species. In general, however, distribution maps were retained as the basic source of information on the limits of the distribution of each species.

Operationally all species' EO were obtained from published distribution maps, thus the quality of distribution data varies according to the information available for each species. As stated earlier, distribution maps are generally a way of summarising and graphically representing distribution data. Often they merely provide an explanatory support to the description of one characteristic of a species (its occurrence in space). Furthermore, delimitation of the area of occurrence is the result of an extrapolation from point locations, mostly done on the basis of a subjective interpretation by the author of the map (generally a specialist on that species) of the species' distribution pattern and other characteristics of its biology, such as the habitat preferences, hence it is subject to a certain degree of approximation.

3.2.1 Digitalisation

To make the data from published distribution maps usable for analysis, they had to be converted into Arc/Info polygon coverage. The methodology used in this project followed these steps:

- rasterisation
- georeferencing
- vectorialisation.

Distribution maps were initially acquired as digital raster images by means of a scanner. The acquisition was performed at either 300 or 600 dpi, depending on the quality and dimension of the original figure.

The next step involved the conversion of the image into a georeferenced raster, in which each point corresponds to real world coordinates. As most of the original distribution maps were based on explanatory figures of published papers, action plans, and field guides, which generally lack precise references to projection parameters, scale and control points, it was not possible to georeference them through simple Arc/Info procedures of transformation or projection. Georeferencing had to be performed by means of "rubber-sheeting", which stretches the raster by making the geographic features identifiable on the map correspond to the same features present in a template coverage which is obviously already georeferenced. The template was obtained merging into a single coverage the main geographic features of the continent derived from the available baseline geographical data set.
For each map scanned, a minimum of 10 to a maximum of 40 control points were used for registration. The points were always selected to match well identifiable geographic features such as rivers, administrative borders, etc.

The result of the georeferencing process was visually checked and repeated (adding and/or deleting control points) until the geographic features on the map corresponded sufficiently to those on the template. This process involved the introduction of a varying amount of inaccuracy, especially in those cases in which the original maps were at a very small scale or when the geographic features were poorly or unreliably depicted. To reduce the overall error, up to 10 different distribution maps from different sources were acquired for a single species. Of these, only the ones that performed best in the georeferencing process were retained for further analysis.

Once georeferenced, the raster image was used to video-digitise the Arc/Info polygon coverages. In this phase, the arcs defining the extent of occurrence were manually digitised on the screen by an operator, following the distribution boundaries shown on the raster image of the map. Each polygon in the coverage was assigned a code which discriminates between portions of the range that are certain and those that are possible or where the species has been introduced/reintroduced. The same code allows for ready retrieval of the map's source in literature from the bibliographic database.

For most of the species, more than one distribution map were available in the literature. In these cases, the ones best representing the species' distribution were chosen. Maps were evaluated in terms of the reliability of the literature source from which they were taken, updatedness, accuracy, resolution and performance in the georeferencing process. In several cases, a single map could be used to define the extent of occurrence of a species. Sometimes, however, more than one map had to be merged to obtain the final coverage. In such cases, each map was first converted into a polygon coverage, after which the various maps were merged into a single final coverage using the Arc/Info editing tools.

Taxonomic uncertainties regarding some of the species considered in the project were also taken into account in selecting the distribution map to be used for the final EO. Whenever possible, only the distribution maps reflecting the taxonomic status of the species as adopted in this project were used.
In some cases (e.g. *Genetta rubiginosa* and *G. pardina*), when no map following the taxonomic arrangement used in this project could be found, the EO was obtained from the maps available. When a map represented the distribution of a taxon including populations of several subspecies, we used only the ranges of the populations that could be assigned to the species as we had defined it.

3.3 Expert support

A preliminary map showing the known Extent of Occurrence of each species as determined from the distribution data obtained from the literature survey was sent to the chairmen of the IUCN/SSC Specialist Groups for review. The chairmen revised the distribution maps either personally or forwarding our request to other experts. Of 24 experts contacted, 12 replied in useful terms. The referees corrected or commented the maps, often sending further literature, especially the most recent publications. The experts' comments were taken into account in drawing the final distribution map.
In a few cases, the entire acquisition process was repeated using the new distribution map as suggested by the referee, but in most cases it was enough to simply edit the coverages already acquired.

For instance, whenever an existing geographical feature (e.g. river, lake, etc.) was known to act as the limit of a species' distribution, it was taken as the boundary of the polygon in the Arc/Info coverage. In such cases, the corresponding arcs were retrieved from the coverages in the geographic data bank (e.g. the coverage of the river network) and substituted into the distribution coverage. In other cases, when the exact limit of the distribution for a particular portion of the EO was not shown in any of the available distribution maps, it was placed along the boundary of the administrative or conventional entity (e.g. country, protected areas) in which the species' presence was reported either in the literature or by the specialists. Again, the arcs representing these boundaries were derived from the corresponding data sets available in the GIS data bank.

In some cases, it was not possible to follow the instructions of the experts exactly: a few recent maps were lacking valid geographical references, while the merging of too many very detailed country maps was considered unfeasable; though, in this last case, the country maps were used to verify at least the external boundary of the known extent of occurrence.

4. Baseline habitat data

No new environmental data set was produced as a part of the activities carried out during implementation of this data bank. Among the goals of the data bank was the identification and evaluation of the available environmental data sets with continental coverage of Africa. Of major importance, within the evaluation process, was the possibility to use an environmental parameter to produce species' distribution models. Therefore, we retained only those data sets that could help in describing the species' environmental preferences.

In addition to the acquisition of the available commercial data sets (Digital Chart of the World - DCW, Esri 1993; Africa Data Sampler - ADS, WRI 1995), an investigation was carried out among the major international organisations dealing with environmental issues in Africa. Though most of them rely on the two previously mentioned products as baseline environmental data sets, the FAO has an extensive information system which was made available to this project. Some very specific data sets (e.g. protected areas, UNESCO's White Vegetation Map, etc.) were contributed by the World Conservation Monitoring Centre. Not all of the data sets acquired were used for the project, though most of them provided an important source of information with which to check and correct the data sets used to derive the distribution models.

Data quality was a major issue throughout the setting up of the data bank. Errors were found in the data on both the attributes and the locations in the layers used for the models. Both attribute and location data were corrected, whenever possible (e.g. for the DCW, the ADS, and White's Vegetation Map), using the original data source**.** When the original data source did not allow for an adequate level of reliability (e.g. between the charts M4 and N4 of the Operational Navigational Charts, which are the base of the DCW data set) the original digital data was retained.

Standardisation of geographic projections was another major issue. All data sets were stored in a geographic projection using decimal degree units. Calculation of areas was done using the Lambert Azimuthal Equal Area Projection which allows for the best compromise among characteristics of the projections supported by Arc/Info.

We are quite confident of the level of final reliability of the individual layers. Nevertheless a shift of up to 5 km was observed between two of the main reference data sets coming from different sources used in the data bank. No real correction was attempted for those errors as it would have implied extensive use of rubber sheeting, thereby decreasing the reliability of the attribute data linked to locational data. Thus the CD model is based essentially on the projection parameters of the DCW-ADS while the PC model is based on those of the Global Land Cover Characterization (GLCC) project.

The following table contains a complete list of the data sets available and of the sources from which they are available. As mentioned, not all of them were used to produce the distribution models.

Coverage name	*Description*	*Source*
Admlim	administrative boundaries	FAO
Popden	similar to admlim + '94 pop. data	FAO
rainfll	mean annual rainfall	FAO
Wetday	number of wet days per year	FAO
Wndvel	mean annual wind velocity	FAO
Geomph	geomorphological regions	FAO
itu	integrated terrain units	FAO
Lnduse	general land use	FAO
phyuni	physiographic division	FAO
surfrm	land surface forms	FAO
vegeta	simplified vegetation	FAO
vegfao	FAO vegetation map	FAO
whites	White's Vegetation Map	WCMC
landcov	Seasonal Land Cover	GLCC
efz	ecofloristic zones	WCMC
geolog	geology	FAO
lndfrm	land forms	FAO
soils	FAO soils	FAO
surmod	surface modifiers	FAO
basins	river basins	FAO
bsnmaj	major basins	FAO
pop	Continuous population density	NCGIA
parkpoly	protected areas boundaries	WCMC/ADS
parkpnts	protected areas location	WCMC/ADS
dtm	Digital Terrain Model	USGS/GLCC
ndviXX	NDVI monthly averages	FAO/GLCC
aepoint	Aeronautical landmarks	DCW/ADS
clline	Cultural landmarks lines	DCW/ADS
clpoint	Cultural landmarks points	DCW/ADS
clpoly	Cultural landmarks	DCW/ADS
dnnet	Dreinage network	DCW/ADS
dnpoint	Dreinage points	DCW/ADS
dqnet	Data quality	DCW/ADS
dspoint	Supplemental dreinage point	DCW/ADS
hsline	Supplemental hypsography	DCW/ADS
hspoint	Supplemental hypsography points	DCW/ADS
hynet	Hypsography	DCW/ADS
hypoint	Hypsography points	DCW/ADS
lcpoint	Land cover points	DCW/ADS
lcpoly	Land cover	DCW/ADS
ofline	Ocean features	DCW/ADS
ofpoint	Ocean features points	DCW/ADS
phline	Physiography	DCW/ADS
ponet	Political and ocean boundaries	DCW/ADS
popoint	Political and ocean boundaries	DCW/ADS
pppoint	Populated places - points	DCW/ADS
pppoly	Populated places - polygons	DCW/ADS
rdline	Roads	DCW/ADS
rrline	Railroads	DCW/ADS
tsline	Transportation structure	DCW/ADS
tspoint	Transportation structure points	DCW/ADS
utline	Utility lines	DCW/ADS

The following is a description of the data sets actually used. They consist of two types: the first contains topographic data concerning geographic features (administrative boundaries, rivers, lakes, etc.) used mainly during the phase of acquisition of the EO of the different species, and as a topographic basis and reference for the standardisation of the environmental layers employed in modelling the AO. The second includes a series of layers containing data on the spatial distribution of environmental variables was later used for modelling the AO.

4.1 Topographic data

The data used were taken mainly from the Digital Chart of the World (ESRI, 1993); this data set is basically the digital version of the US Defense Mapping Agency Operational Navigation Chart (ONC) series, generated from base maps at a 1: 1 000 000 scale. We also used data from the Africa Data Sample (ADS) (WRI, 1995), which are essentially an updated version of the DCW.

- Administrative boundaries and national boundaries were taken from the DCW - Ponet coverage. Available maps were used to draw the border between Ethiopia and Eritrea.
- Hydrological network. This coverage was taken from the ADS, as this source appeared to be more updated than the DCW version. As the data in the ADS are divided into several sets covering individual countries, the various coverages where merged and the resulting coverage checked manually for continuity of features along the national borders; errors were corrected on the basis of the source coverages and the ONC maps. A subset of this coverage was obtained by selecting all polygon features enclosing permanent water bodies; this was used as a mask to superimpose on environmental data sets to cut out the areas covered by inland water bodies.

4.2 Environmental variables

The environmental data sets used to produce this data bank can be grouped under two general headlines: discrete sets and continuous sets. While the former group contains most of the information available and was used to produce the CD models, the latter is made up of a small fraction of the information available and was essentially used for the PC models.

- White's vegetation map; this vector coverage, provided as Arc/Info export file by the WCMC of Cambridge, is the digital version of the UNESCO/AETFAT/UNSO vegetation map of Africa (White, 1983). The source map is at 1:5 000 000 scale and covers the whole African continent. The coverage was also converted into an Arc-Info grid with resolution and origin identical to those of the GLCC Land Cover map (see below). See paragraph 5.1.1. for other details on this map's features.
- Seasonal Land Cover regions (Global Land Cover Characterization - GLCC project); this data set was downloaded from the Eros Data Center ftp site (edcftp.cr.usgs.gov) as a raster image, and then converted into an Arc/Info grid, with a 1 km pixel resolution. It is derived from an unsupervised classification of AVHRR data spanning one year, followed by manual refinement, and is the preliminary release of a global land cover assessment implemented on a continent by continent basis by the US Geological Survey's Earth Resources Observation System in co-operation with the University of Nebraska-Lincoln and the Joint Research Centre of the European Commission (a detailed description of the data set and the methodology can be found at the following World Wide Web site: http://edcwww.cr.usgs.gov/landdaac/glcc). The version that was used for this project dates to the beginning of 1997, although a more recent version was made available during the final phase of the project (October 1997) when most of the modelling had already been done. Even though no specific comparison between the data used here and the new release has been attempted, the preliminary release possibly contains inconsistencies in the classification of some areas and this suggests caution in its interpretation. See paragraph 5.1.1. for other details on this coverage's features.
- Protected Areas: the coverage used was obtained from the Africa Data Sampler (WRI, 1995), which basically contains information from the WCMC Protected Areas Digital Map Database. A version of the Protected Areas Digital Map Database coverage, directly supplied by the WCMC was also available, but the Africa Data Sampler data set was chosen, as it appeared to be more updated for some areas; in addition, polygon features in the ADS data set had already been processed to fit the administrative boundaries and water bodies as reported in the DCW, and were thus more consistent with the topographic set used as a reference in the project. As the data from the ADS are furnished at country level, all the coverages pertaining to the different countries had to be merged into a single one covering the entire continent. The resulting coverage was then checked and compared with the WCMC data.
- Population density data set: two data sets were used for the project, both distributed by the National Center for Geographic Information and Analysis, University of California: a coverage with population density estimates by administrative district was used as one of the stratification parameters to derive the validation sampling scheme (see Validation paragraph), while a grid representing estimated population densities over a continuous surface was used for production of the PC models. The former data set is derived from a variety of sources; one of its main limitations is that it is quite disparate in terms of data resolution (e.g. for some countries data are available only for very large units). The continuous data set was generated by interpolating the vector data for the administrative boundaries, which take the

accessibility of each point into account; in particular, while the totals for each administrative unit are maintained, the population density for each point is estimated on the basic assumption that it is directly correlated to its accessibility, which is estimated on the basis of the road, railway and river network, as well as the distribution of the major urban centres. (Further details on the methodology and limitations of these two data sets can be found at the following Internet address: http://grid2.cr.usgs.gov/globalpop/).

- DEM; the Digital Elevation Model used is essentially the one provided with the Land Cover data set (1 km^2 pixel size) except for the areas below sea level which were absent in that data set. For those areas, the altitude was estimated using the DEM provided by USGS.
- NDVI; both the ten year series provided by the FAO on CD-Rom and produced by the ARTEMIS project (7.6 x 7.6 km pixel size), and the monthly average images (1 x 1 km pixel size) which accompanied the Land Cover data set were used in this project.

4.2.1 Derived data sets

Some layers used in the modelling procedures were obtained by elaborating on or combining available data sets:

- Distance from water : a raster of the minimum distance from permanent water sources was derived from the ADS coverages of drainage features. After all permanent streams, shorelines or permanent water bodies were selected from the ADS coverages of linear and point drainage features, the minimum distance from each pixel to the closest water source was calculated. This raster was generated with parameters of resolution and extension identical to those of the GLCC land cover data set, in particular a 1 km^2 pixel size. The raster was used in the elaboration of both the CD and the PC models for all species whose presence is affected by distance from water sources. In interpreting the results, it must be considered that this data set is derived from data taken from the ONC maps and consequently suffers from all the possible shortcomings in resolution, updating and accuracy that can be found in maps produced for navigational purposes at a 1:1 000 000 scale and over a period of several years.
- Land Cover map / White's vegetation map intersection: the two data sets on vegetation and land cover were combined into a unique layer, with resolution identical to that of the GLCC Land Cover map, in which a univocal code was assigned to each pixel on the basis of the combination of land cover class and White's vegetation type. This data set was used in the generation of the CD models.
- Seasonality map: derived from the FAO-ARTEMIS data set. The raster was produced as the corrected coefficient of variation of temporal variability of the 10 year NDVI data set (Baird, 1996), and represents the degree to which environmental conditions, as measured by NDVI, vary seasonally or stay constant throughout the year. This data set, which was kindly made available to the project directly by its author, was used for the generation of the PC model.
- Mean Annual NDVI: extracted as the first Principal Component of the monthly averages of the NDVI (Estman, 1992), it was used for generation of the PC model.

5. Modelling the Area of Occupancy

This project relied on two different approaches for identification of the Area of Occupancy.
The deductive approach uses the species' known ecological requirements to extrapolate suitable areas from the environmental variable layers available in the GIS database. Analysis of the species-environment relationship is delegated to the synthesising capabilities and wide experience of one or more specialists who decide, to the best of their knowledge, which environmental conditions are the most favourable for the existence of the species.

Generally, some sort of logical (e.g. Breininger et al., 1991; Jensen et al., 1992) or arithmetic map overlay operation (e.g. Donovan et al., 1987; Congalton et al., 1993) is used to merge the different GIS environmental layers to yield the combined effect of all environmental variables. This approach was chosen as the central modelling effort of the project, and was used to produce the Categorical-Discrete (CD) distribution models.

To provide a means of comparison an inductive approach was also followed. Given that the species-environment relationships are not known *a priori*, this approach was used to derive the ecological requirements of the species from the information implicit in the Extent of Occurrence. Assuming that a species' environmental preferences can be derived from this information in the form of a function linking the environmental variables to the species' presence; then, the function is used to extrapolate the distribution model with a process very similar to the one used in deductive modelling (e.g. Pereira & Itami, 1991;

Aspinall & Matthews, 1994). Few applications of these kinds of models had been attempted at a continental scale (e.g. Busby, 1991; Walker, 1990; Skidmore et al., 1996) and none for Africa. Thus this project was seen as a good opportunity to produce a first assessment of their potential.

5.1 CD model

In the CD model approach, the AO is described using the species' known environmental preferences. A suitability score is assigned to each different combination of environmental variables found within the species' known EO. Three categories of suitability were considered, 1 being the highest. Suitability scores were deduced from literature analysis; when sufficient references were available, they were assigned according to relative abundance within a given habitat type. A conservative approach was maintained, emphasising the species' possible occurrence rather than its absence.

5.1.1 Analyses

The following environmental layers were evaluated and combined:

- White's vegetation map (White, 1983)

White's vegetation map of Africa was considered adequate as a representation of the vegetation types selected by mammals: floristic, physiognomic and spatial information (zonation) is synthesised and organised in a two-rank classification (17 and 80 categories, respectively), permitting two different levels of analyses; the accompanying description provides further details useful in identifying specific habitat features. However, White's vegetation map scarcely depicts cultivated or degraded vegetation, and the nomenclature used is rarely (though increasingly) found in mammals' habitat descriptions.

- Seasonal Land Cover regions map (GLCC project)

The Land Cover map provides a recent and detailed description of mainly physiognomic features (195 categories), integrating White's vegetation map; in fact, it includes a detailed classification of cultivated and degraded vegetation, and the nomenclature used is easily related to mammals' habitat descriptions. Moreover, its detail was considered sufficient to represent habitat fragmentation at the scale required by this study.

For every species, a suitability score was assigned separately to each legend category of the two maps.

In order to provide as objective a method as possible for scoring, a data file was compiled with the species' habitat descriptions from different literature sources. The criteria used are summarised in the following table together with scores and definitions.

Score	**Definition**	**Criteria**
1	suitable	- frequently cited as main or preferred habitat - presence of major populations - higher population densities
2	moderately suitable	- frequently cited as secondary habitat - presence of minor populations - lower population densities - patchy distribution
3	unsuitable	- cited as unsuitable - not specifically cited, but unsuitability easily inferred from its ecology
9	undefined	- not cited, possibly suitable or moderately suitable as inferred from its ecology - data deficient, possibly suitable or moderately suitable

The following rules for the assessment of mosaic categories (when they are not preferred habitat types) were established:

suitable/unsuitable ⇒ moderately suitable

suitable/moderately suitable ⇒ suitable

unsuitable/moderately suitable ⇒ moderately suitable

suitable/unknown ⇒ suitable

unsuitable/unknown ⇒ moderately suitable

Subsequently, for each combination of categories derived from the intersection of the two maps, a total score was calculated, according to the following matrix:

Land Cover map \ White's vegetation map	1	2	3
1	1	1	2
2	2	2	3
3	3	3	3

The matrix gives priority to the Land Cover classification, which is more recent and detailed, but the score decreases when the Land Cover units fall in unsuitable vegetation contexts as indicated by White's map. As a result, suitable habitat patches within unsuitable vegetation zones are maintained and scored moderately suitable (i.e. a Land Cover forest unit within a White's woodland zone is scored moderately suitable for a forest species), while moderately suitable patches in the same zone are disqualified. On the other hand, unsuitable habitat patches within suitable and moderately suitable vegetation zones are identified by means of the Land Cover map (i.e. cultivated areas within White's forest zone are scored unsuitable for a forest species).

Although the matrix rules were applied to all species, the resulting scores have been modified for some specific combinations, particularly when White's classification better described the specific vegetation types (e.g. montane and desert vegetation).
In general, the intersection of the two maps made it possible to assess over 3 500 combinations attributed to the following main vegetation classes:

Forests
Forest mosaics
Swamp forest
Mediterranean sclerophyllous forest and their mosaics
Mangroves
Woodlands
Woodland mosaics
Bushlands, thickets and shrublands
Bushland, thicket and shrubland mosaics
Fynbos
Savanna
Savanna mosaics
Grasslands
Grassland mosaics
Semi-desert vegetation
Montane forests and their mosaics
Montane shrublands and shrubland mosaics
Montane grasslands
Saharomontane vegetation
Altimontane vegetation
Desert vegetation
Absolute desert and desert dunes
Rocky and gravel desert

- River network

To stress the relation between the distribution of some species and the distance from permanent water, the river network coverage was used. Buffer areas of 1, 5 and 10 km around permanent water courses or lakes were calculated. The suitability scores obtained from the intersection of White's vegetation map and the Land Cover map were decreased outside and increased inside these buffer areas. The degree of modification of the scores was decided according to the specific requisites of each species, as was the dimension of the buffer areas chosen to represent the threshold.

In particular, water-bound species were grouped as follows:

<u>Species occurring in or near water</u>: this group includes species inhabiting rivers, lakes or adjacent areas or spending a considerable part of their daily time in or near water (e.g. otters, *Aonyx* spp. and *Lutra*; hippo, *Hippopotamus amphibius*; waterbuck, *Kobus ellipsiprymnus*). For these species, all vegetation types occurring outside the buffer area were considered unsuitable. The buffer size assigned was either 1 km or 10 km, depending on the species' mobility. Actually, most species ascribed to this category usually concentrate their activity within 500 m from water (i.e. otters), but 1 km is the minimum resolution of the data set. The threshold of 10 km for hippos and waterbuck was used as a standard distance, as the maximum distance from water given in literature for these species is comparable in magnitude (Rowe-Rowe, 1994; Spinage, 1982).

Species occurring in riverine or gallery forests: some species (many primates and some ungulates) show preference for gallery or riverine forests, or extend their forest range in more arid zones through gallery forest. This vegetation type is not represented in the maps used. To underline this preference, it was assumed that riparian forests occur inside a 1 km buffer around permanent water in a number of White's vegetation map classes, as indicated by White (1983). Thus, assigned preference scores were increased in this buffer area for species such as the blue duiker (*Cephalophus monticola*), the Diana monkey (*Cercopithecus diana*), the black and white colobus monkey (*Colobus guereza*), and a number of others. Only for *Cercocebus galeritus* was a different buffer size (5 km) used, as deduced from Homewood (1975).

Water-dependent species: this group includes all those species (mainly ungulates) whose density decreases away from permanent water, at least during the dry season. For these species, the suitability scores obtained from the intersection between White's vegetation map and the Land Cover map were decreased outside a 10 km buffer area around permanent rivers. In this way the suitability of areas that can support high numbers of a species on a year-round basis were enhanced as compared to areas that are suitable only during the wet season. Some support for the choice of the threshold distance of 10 km was found in Western (1975), who showed that in the Amboseli ecosystem the biomass density of large mammals beyond 15 km from water, a radius which encloses 99.5% of the biomass of water-dependent species during the dry season, is extremely low compared to areas within 10 km of water. This fixed threshold, however, must be considered with great caution, as stressed by East (in litteris) who revised and commented the list of species included in this group: probably each species has its own threshold distance, and no particular threshold distance from permanent water can be applied uniformly across Africa. Nevertheless, interestingly enough, the described procedure improved the model's results for the selected species and can easily be repeated as soon as new or more detailed information is available, or could be applied with more efficacy using regional or local data sets.

- Elevation

For some species restricted to well defined elevation zones (e.g. *Lepus starki, Theropithecus gelada*), all vegetation types occurring outside known elevation limits were considered unsuitable. These areas were identified through the intersection of the EO and the Digital Elevation Model (DEM).

5.1.2 Interpretation of the model

The Area of Occupancy (AO) was defined as the sum of suitable and moderately suitable areas; presence was considered "expected" within and "potential" outside the boundaries of the known Extent of Occurrence (EO).

The reliability of the results of the CD model is strictly dependent on the reliability of the maps used to describe the species' habitat and the level of knowledge of the habitat requirements of each species (see paragraph 6.). In all areas, combining White's vegetation map with the Land Cover map improved the potential of the two maps taken singularly.

Some basic statistics that could be of help in interpreting the deductive model in terms of its use for conservation were also calculated: essentially a few indexes of fragmentation and of the efficacy for each species of the protected areas.

5.1.3 Fragmentation

Fragmentation is one of the fundamental pieces of information needed to implement an effective conservation strategy. Conservation status can vary greatly depending on the level of fragmentation of the area of occupancy. Thus it seemed mandatory to include fragmentation indexes among the tools for interpreting the models.

Before calculating the fragmentation statistics, the overall resolution of the models was downgraded to a pixel size of 5x5 km in order to average the errors due to misclassification and uneven registration of the layers used to produce the CD model. The decrease in pixel size is reflected in the discrepancies in area figures in the various tables presented in the species account.

Of the many fragmentation indexes available, it was decided to rely on simple and easily interpretable figures., the FRAGSTAT package (McGarigal & Marks, 1994) was chosen and modified so that it could be compiled with Microsoft Visual C++ and could read the native Arc/Info GRID format using the GRID I/O software library provided with ArcView. For both the "suitable" and the "moderately suitable" classes and for the overall AO, the following indexes were calculated:

- **Number of patches** (**NP**): the number of patches within the EO
- **Mean patch size** (**MPS**); the average dimension of the patches
- **Patch size standard deviation** (**PSSD**): the standard deviation of the previous index
- **Largest patch index** (**LPI**): the percentage of the total area of the class that falls within the largest patch
- **Mean shape index** (**MSI**): the shape index of a patch varies between 1 (when the patch is square) to infinity. The mean value is calculated as the average index over all the patches belonging to the class.
- **Area-weighted mean shape index** (**AWMSI**): same as above, but bigger patches contribute more to the result of this index.

It is important to note that none of the above indexes alone can give a picture of the fragmentation of the species' AO. Thus they should always be read together, each one supporting the other.

The number of patches of the two classes compared with that of the overall Area of Occupancy can provide a key for interpreting the connectivity of the different patches. If the NP of the two suitability classes is very high compared with that of the total AO, the AO can be expected to be substantially homogeneous even though highly interspersed with suitable and moderately suitable patches. Moreover, comparison of the two shape indexes with the total number of patches can give an idea of the real fragmentation of the AO at the scale examined.

Since the shape index varies from one (for a patch that is perfectly square) to infinity, an MSI near 1 indicates either that all of the patches, both big and small (up to patches of a single pixel), are square or that most of them are one pixel size. Nevertheless comparing the MSI to the AWMSI can give further insight into the patchiness of the AO. In fact if the AWMSI is not equal to one, this means that there is at least one big patch (of varying complexity) of the given class within the AO.
This outcome can be strengthened by observing the value of the LPI for the same class; if the LPI is high, this further supports evidence of one single patch in the AO.

5.1.4 Intersection with protected areas

Based on the information available on protected areas, the efficacy of the protected network was tested against the expected Area of Occupancy as derived from the CD model.

After converting the protected areas coverage into a raster with the same parameters as the GLCC Land Cover and overlaying this raster on the raster of the CD model, the percentage of area pertaining to each suitability class within the EO falling inside and outside existing protected areas was calculated for each species. It must be stressed however that these percentages do not necessarily reflect the extension of areas truly protected. This is mainly due to the fact that not all existing protected areas are mapped by the coverage used in this project Several protected areas in different countries are not mapped at all by the polygon coverage and some are recorded only as point locations in a different coverage included in the ADS. As it was not possible to calculate the percentages of those areas that fall into the different suitability classes, they were excluded from calculations. Hence the result of this calculation could be biased, especially for some species with distribution restricted to certain areas or countries (for example South Africa, for which protected areas are available only as point data), and the percentage of the Area of Occupancy included in protected areas could be severely underestimated.

5.1.5 Data validation

The analysis of species distribution in relation to predetermined environmental parameters (vegetation, land use, altitude etc.) provides a projection based on a theoretical model of the relationship between the presence of the species, its ecology and the environmental variables chosen. This model would have no validity if it were not compared to data collected directly in the field: in fact, estimating precisely the reliability of models worked out only on theoretical bases is a major concern.

Even though it may be presumed that the reliability of models will be different for different species, in order for the outcomes to be applied for the purposes of management and conservation, an estimate of the error of the model proposed must be provided.

Thus, four sample areas were chosen (each encompassing at least one country) and 100 points were allocated in each, following a random stratified sampling design. The various strata were obtained by overlapping a) the vegetation layers defined by White's vegetation map of Africa (1983); b) the layers of human population density; c) the administrative borders of the four selected countries. Where necessary, points were added inside protected areas to provide more data for analysis of the influence of this variable.

Work in the field involved verifying the relationship between the presence of the species and the different classes/values of the environmental variables considered. To this end, the 100 plots defined in each country were not meant to produce an independent estimate of the distribution of the species, but to provide a statistical sample of locations with specific environmental conditions in which a species is expected to be present or absent and is found to be either present or absent.

The presence (or absence) of the species at each of the 100 predetermined points was verified by direct observation, *in loco* collection of publications and scientific reports, interviews with local authorities and inhabitants. The latter were carried out according to previously agreed standard procedures, using special checklists and pictures to help with recognition of individual species. Each item of information was judged for quality/reliability. Field verification was also aimed, by means of a similar approach, at establishing the reliability of the vegetation classification set down in White's vegetation map and at collecting data on other environmental variables that could be useful for analysis.

Not all points selected through this semi-random method were actually accessible. There were various causes hindering access to a predetermined point; these included security considerations due to civil disorders, untransitable roads or paths, military restrictions, distance from the road, prohibitive land conditions (desert, dense tropical forest, etc.). In such cases, points were shifted to more accessible areas within the same layer with environmental conditions similar to the chosen point.

5.1.5.1 Selected countries

The countries originally selected were Botswana, Cameroon, and Uganda; Morocco was added later.

The criteria behind this selection were:

- a maximum of environmental diversity in each country;
- general representativeness of conditions in East, West and South Africa;
- medium-large size with respect to the rest of the African countries;
- feasibility of collaborating with local universities or ministerial institutes;
- interest on the part of locals in participating in the project or in receiving specific training in GIS;
- availability of detailed and good quality data on natural resources in each country.

The fourth area, Morocco, was added to take account of the peculiar environmental and faunistic conditions found in North Africa, which would otherwise have been excluded. Even though Morocco is the most outlying of North African countries, it has the greatest biodiversity and is the most accessible for field work.

In each of these sample areas, an agreement was reached with a local institution for collaboration involving supply of services and personnel, specific training, participation in field work and analysis of results.

5.1.5.2 Data analysis

After a validation phase in which plots were checked for errors, a total of 427 plots were earmarked for analysis. For each species the plots were considered valid either if they were included within the Extent of Occurrence of the species or if the species was observed in that plot. Validation analysis was performed only for those species for which at least 1% of the known Extent of Occurrence is included in the sample areas.

A simple Index of Accordance was calculated for each species. The Index is the percentage of valid plots that agree with the results of the CD model: a plot agrees with the result of the CD model if it falls within the predicted Area of Occupancy and the species is found during field work or if it falls outside of the AO and the species is not reported during field work.

Caution concerning the reliability of the validation process was expressed for those species with less than 15 valid points. The threshold for model reliability was set at 50% accordance with the results of field work. Where relevant, the Index of Accordance was referred to, in the discussion, for possible explanation of excessively high or low figures in the species account sheets.

A few general guidelines for interpreting the Index of Accordance: values just below the threshold were found for species with habitat specificities that do not appear on the maps because they are probably too small for the mapping resolution of the data bank (e.g. species associated to rocky outcrops or gallery forest, etc.); high accordance was found for those species whose ecology is better known, allowing for better identification of the preferred habitat requirements.

5.2 Probabilistic models

The general methodology used for these models is based on a data set of known locations of the species (either known territories, observations or radio-locations). These are used to define the species' ecological profile which is obtained through characterisation of the known locations with the available environmental variables. The characterisation is then used to calculate the ecological distance of each location within the study area from the species' preferred ecological conditions. This produces a model of the species' distribution in the form of an environmental suitability area, within which smaller distances indicate higher environmental quality for the species.

To use this methodology within this project the major modification required was due to the lack of a data set comparable to that of the "known locations". In order to build the species' ecological profile, we had to rely only on its EO. But by assuming that most of the area within the EO is of good quality for the species, we were able to use the average ecological quality of the EO as the ecological profile of the species. Thus we can map the ecological variability in terms of distance from the average ecological quality of the EO. Following this assumption, we can expect the quality and reliability of the results to depend on the homogeneity of distribution throughout the EO of both the species and the environmental variables (see below).

5.2.1 Analysis

The input variables for these models were the following continuous environmental layers: the first Principal Component of the 12 monthly averages of the NDVI index (essentially equal to the annual average NDVI) (Eastman, 1992), a seasonality map derived from the same NDVI data set (Baird, 1996), the population densities raster and the DTM. All of these data sets were resampled to a 1 km pixel size, To allow a better representation of the output. Nevertheless interpretation should be carried out at a coarser resolution (namely 5 km^2).

For each species, its ecological characterisation was derived from the vector of the means and from the variance-covariance matrix of all pixels falling within the known Extent of Occurrence.
Species specifically linked to water were treated in a slightly different way in that the Extent of Occurrence used to calculate the vector of means and the variance-covariance matrix are previously limited to the portion within a 2 km buffer around the permanent water network.

The vector of the means and the variance-covariance matrix were then used to calculate the Mahalanobis distance of each pixel on the African continent from the species' ecological characterisation. Values near zero indicate a high similitude with the average ecological condition of the Extent of Occurrence, while higher values indicate an increasing difference from the species' ecological characterisation. A limit of 150 units (the Mahalanobis distance is dimensionless) was set to allow an adequate graphical representation of the models results.

5.2.2 Interpretation of the models

The first warning that should be kept in mind when analysing the output of the PC models is that the colours used have a relative value. That is, darker shades of green indicate more suitable (less distant) areas for the species; but the same level of green does not indicate equally suitable areas for two different species.

Moreover, while it seems important to underline that these models should be read as speculative projections, their importance resides in the possibilities opened up by use of GIS in species distribution modelling in the conservation arena.
In fact, probabilistic-continuous models can be used to address part of the stochasticity inherent in locating an individual of a species, as well as problems of corridor design and meta-population modelling (Akçakaya, 1993) by introducing the geographic dimension in analysis of species viability.

It was decided here to apply these models to all the species included in the data bank. Although the results of the PC models are discussed on a species by species basis, the assumptions introduced to modify the methodology suggest great care in the general interpretation of the results. This warning holds especially true for opportunistic species with wide distributions and for species highly specialised that use only limited portions of their distribution range; probably, in the case of opportunistic species, increased reliability could be achieved through a sound stratification of the study area. Another possible evolution of these models would be to use the output of the CD model to calculate the species' ecological profile. None of these possible enhancements was attempted to permit objective comparison of the applicability of this method to the different species.

6. Comments on distribution and conservation

Biological populations are not evenly distributed across their geographical ranges and conservation biologists are aware that this fragmentation is due to many different causes acting at different scales.

Maurer and Heywood (1993) pointed out that a geographical range is subject to at least two types of fragmentation: the first is the fragmentation of its perimeter, reflecting the level of smoothness of the boundary between the species' range and the external areas, the second is the fragmentation due to uneven density of the species within its range. Both are important for the conservationist but studying them requires two different sets of methodologies: investigation of population densities has traditionally been carried out using census techniques or extrapolation models (Maurer, 1994), while inquiry into geographical ranges has just started to receive more attention by ecologists. In particular, studies on the broad scale patterns of actual and potential ranges, their correlation with habitat factors and their time series have only been approached recently since GIS techniques have become more easily available.

The weaknesses and caveats of GIS applications to animal distributions are well known to experts and scientists (for a review see Corsi et al., in press) but less so to conservationists and lay people. The greatest danger is misunderstanding the true meaning of the models and the graphic representation of animal distribution, and hence the possibility of basing management action on insufficient ground data.

The output of GIS models is at best an indication of true trends and patterns; it is not a true representation of reality: although this may sound redundant to many and a truism to many others, it is always useful to repeat it.

There are obvious problems with the interpretation of maps of even the simplest type of range representation: the global geographic range of a species. The line marking the boundary of the geographic range of a species is usually drawn to separate the areas in which the species occur from those in which it does not occur. Instead of a line, this boundary should be a strip of decreasing densities toward the external areas, as the species' distribution rarely stops abruptly along an obvious geographic feature. Individuals of that species will be found increasingly rarely as one moves from the peripheral areas of the range to the areas immediately outside the boundary line. This is one of the problems that make these boundary lines inaccurate: there can be no standard method for drawing the line at a precise value of decrease in density.

There is another problem relating to the boundary lines of geographic ranges: their accuracy varies with scale. The boundary line drawn on a small scale map will include/exclude large areas that can be accurately identified by the a line drawn on a detailed map. The finer the scale, the greater the accuracy of the boundary.

All these problems are reflected in our models and the reader is warned to appraise tables and figures correctly: the models use environmental parameters that have their own errors and approximations and the GIS has the ability of masking these errors from the ingenuous reader by displaying final outputs in the form of appealing maps.

In particular, our models identify various level of suitability within the range: they are to be understood as broad categories of habitat quality averaged over large areas. It is evident that a finer scale analysis would reveal much higher habitat complexity. Such finer analyses are indeed the natural next step after this initial global overlook. A similar comment can be made for the degree of fragmentation deduced from the figures and fragmentation indexes: these fragments refer only to geographical patterns of habitat suitability at broad scale and have the same conceptual and practical limitations as the data upon which they are based. Moreover, the fragmentation patterns indicated by our models and indexes provide no information on another important type of fragmentation, i.e. the demographic fragmentation occurring naturally (uneven densities) or artificially (direct and indirect human disturbance). For several large mammals in Africa, this last type of fragmentation is often much more important than geographical fragmentation. However, the geographical fragmentation encompasses the demographic and our models therefore provide a better level of approximation to the analysis of population status.

We felt it particularly useful to use the two terms Extent of Occurrence and Area of Occupancy, even though both have a strong biological and ecological significance (Gaston, 1991) and it is very difficult, in practice, to apply them appropriately. It is obvious that the Area of Occupancy is nothing more than a higher level of approximation of the species' true distribution with respect to the Extent of Occurrence. Moreover, it is obtained through a modelling process and not from true field data, although it has been partly validated. In spite of this constraints, we feel that the overall method provides a useful means of obtaining broad scale results that cannot be gathered in any other way; and if the final results offer a distribution range whose resolution is even only a fraction better than what was previously available, this is believed to be a significant improvement for ecologists and conservation planners.

The data bank offers the opportunity of continuing to refine the levels of approximation: starting from the current output, the models can be loaded with other environmental parameters (covering the whole range or just a portion of it) and/or data at finer scale to narrow the definition of suitable areas.

Keeping in mind all the caveats just mentioned, we also undertook measurement of the overlap of the Extent of Occurrence and Area of Occupancy of each species with protected areas: regardless of the level of approximation, what was of interest here was the order of magnitude of the areas being protected for each species. The results are self-explanatory and require no comment, as the protection for most species is almost negligible and yet much of these species' future conservation depends on these tiny fractions of their ranges. Further analyses could show how the various sub-populations (and subspecies) are protected by these areas and a full GAP analysis could be applied to identify the need for other strategically placed areas.

Finally, when the level of approximation is felt to be adequate and when population densities are sufficiently known to be extrapolated, the output of the models can be used to evaluate population numbers: this may prove a useful tool for getting insight into small remnant populations in the various geographic fragments.

7. Output structure

The project's output is expected to add significantly to the understanding of the status, trends, and threats of African mammals, both the large charismatic and the smaller, lesser known species. The project's output will benefit all institutions dealing with research and conservation of animal species at national and international level.

Two complementary sets of output have been produced to maximise the diffusion of results. A printed volume provides a way of distributing the project's results to those offices and institutions that do not yet have the technical facilities to handle electronic data. Conversely, the magnetic data bank remains available largely as a reference body of data for other applications. The technical capacity to handle georeferenced data and GIS is spreading rapidly, also in African countries: costs are very reasonable today and the software is becoming increasingly user-friendly, becoming accessible to lower levels of university instruction.

7.1 The printed data bank

Orders and families are presented in a phylogenetic sequence; within each of these, the species (or the subfamilies - if present - within the families) are arranged in alphabetical order, as in the main reference source (i.e. Wilson & Reeder, 1993).

Each of the 281 species was given a separate account, while subgroups, infra-groups and super-groups were purposely ignored to avoid unnecessary repetitions and to maintain the overall scheme. Each species was given an identification code (ID code) and a hierarchical number for identification of its family and order; both are indicated at the top left of each sheet (i.e., Order 2: Primates, Family 1: Loridae, Species 2.1.1: *Arctocebus aureus*, Species 2.1.2: *Arctocebus calabarensis*, etc.).

The species accounts have neither a biology nor an ecology section, which are easily found in many good sources among those listed in the bibliography and the related FoxPro file.
The information presented in the species account is arranged under the following headings:

Nomenclature (***Scientific name***). In order to keep an international standard format, species nomenclature follows the taxonomy set out in the second edition of the volume by Wilson & Reeder (1993), except when important literature sources or IUCN/SSC specialists expressed a strong preference for another nomenclature, suggesting a valid alternative name and justifying the choice (for example: *Parahyaena brunnea* sensu Wilson & Reeder = *Hyaena brunnea*; *Herpestes naso* sensu Wilson & Reeder = *Xenogale naso; Genetta maculata* sensu Wilson & Reeder = *Genetta pardina* + *Genetta rubiginosa*). In the case of discrepancies between the Wilson & Reeder system of classification and available distribution maps, the reference text was not necessarily always given priority (i.e.: *Procolobus badius* sensu Wilson & Reeder + *Procolobus preussi* sensu Wilson & Reeder + *Procolobus pennanti* sensu Wilson & Reeder + *Procolobus rufomitratus* sensu Wilson & Reeder = *Procolobus badius* sensu IUCN).

Describer. The name of the species' describer, written just below the scientific name, was also taken from Wilson & Reeder (1993).

Vernacular names. On each sheet, species' names are given in English (Eng) and French (Fr); these vernacular names were supplied as a convenient and useful tool for species identification, even if, unlike the Latin names, they are neither precise nor univocal and there is no way of dictating a generally accepted standardisation. As a matter of fact, there may be many different names for the same species in general use and deeply entrenched in the language within even a very restricted area. For English common names, the main reference texts were the IUCN/SSC Action Plans, while for French names, the main source was the publication by Kingdon (1997). Wherever a species was found to have more than one colloquial name, at least two were reported to facilitate consultation.

Taxonomic notes. This section is meant to provide no more than a comment on special or significant features, not a listing of the entire species' or genus' taxonomic history. Closely related species, systematic uncertainties and well-defined forms, if any, are noted here. Subspecies are generally ignored, as the project focuses on the species taxonomic level; moreover, the validity and precise distribution of many subspecies are still uncertain, debated or pending review. Only in a few cases are subspecies listed: when they are classified at a different threat level by the IUCN or are of special conservation interest; when the named subspecies is considered a full species by other zoology texts; when there is no general agreement on the taxon classification; or in special cases (i.e. the gorilla, *Gorilla gorilla*), when the species' subdivision is accepted worldwide. The decision to include or exclude one or more subspecies in this section was arbitrary.

No morphological description is given, as that would be beyond the scope of this work. Information in this section is based mainly on data from Wilson & Reeder (1993) and the Action Plans available for different taxa, but sometimes, and especially in case of discrepancy, reference was made to the publication by Meester & Setzer (1971) and the more recent data contained in Kingdon (1997).

IUCN threat category. This section gives the species' recent threat assessment according to the 1996 IUCN Red List of Threatened Animals (Baillie & Groombridge, 1996). The status category is fully listed for each taxon (at the species and, when described, subspecies level) together with its symbol and criteria in parentheses. Countries of occurrence are often listed close to the relative species (or subspecies), together with some of the following symbols (in particular for the antelopes, as also indicated by East, 1996): «?» indicates that the taxon may occur, but its presence (former or current) is not confirmed; «ex» indicates that it is now extinct and «ex?» indicates that it is now probably extinct.

The main categories used are: Critically Endangered (CR); Endangered (EN); Vulnerable (VU); Lower Risk: conservation dependent (LR: cd), near threatened (LR: nt), least concern (LR: lc); and Data Deficient (DD) (see Baillie & Groombridge, 1996, for definitions of categories and criteria).

Available information. This section gives a brief report of the information available in the scientific and gray literature on the species. The amount of information included here varies from species to species, depending on the complexity of each animal's data, how much it has been studied and the main focus of the research done. The longest accounts are obviously those referring to the most popular and/or well-studied mammals (such as the impala, *Aepycerus melampus*, the gorilla, *Gorilla gorilla*, the cheetah, *Acinonyx jubatus*, and the lion, *Panthera leo*). In these cases (and in any case wherever possible) the description is broken down into North, South, East and West African subregions.

A major effort was made to include in the bibliography the most important sources of information about each species, excluding minor sources and papers dealing with topics not relevant to our project.

Known Extent of Occurrence. This section summarises the species' known distribution range. It also provides the source(s) of the map, the name of the specialist (if any) who revised it (and the date), and a list of other publications used to update the output.

Comments are given on the distribution maps that were available for each species and the one (or ones) thought to be most recent/reliable and retained for modelling. The maps show the limits within which - to the best of our knowledge - the animal occurs, and underline special geographic features such as "introduced or reintroduced" areas, and the distinction between «possible» and "certain" presence. Each species' geographical range has been described in a standard form to make the results as uniform and essential as possible: the text either lists the countries, or the parts of them, in which the species occurs**,** or makes reference to geographic coordinates. The most recent names have been used as much as possible (however, the Democratic Republic of Congo is still referred to as former Zaire).

Categorical-discrete (CD) distribution model. This section presents the ecological information on the species used for the model. It also includes a table with the preference scores for the vegetation types occurring inside the EO (1 for suitable; 2 for moderately suitable; 3 for unsuitable; 9 for unknown).

Two tables show the quantitative results of the model in terms of percentage of the various suitability classes within the EO and the 6 indexes of fragmentation.

Probabilistic-continuous (PC) distribution model. The PC distribution model represents distribution in the form of an environmental suitability surface, where the ecological variability has been mapped as distance from the average ecological quality from the EO.

Validation. The validation parameters of the categorical-discrete (CD) distribution model were calculated only for those species for which at least 1% of the known Extent of Occurrence (EO) is included in the sample countries. This table (if present) shows the percent of the species' Extent of Occurrence (EO) within the sample areas; the number of valid plots and the Index of Accordance. A plot was analysed if it was included in the species' EO or if the species was observed in it. The Index of Accordance is the percentage of valid plots which agree with the CD model's results: it is considered sufficient if it exceeds a 50% threshold.

Comments and conservation issues. This specific section of the species account contains comments on models, fragmentation analysis and protected areas overlap. These comments provide an initial interpretation of the output. First, the two models' output is compared with the known Extent of Occurrence to underline the inconsistency, on a large scale, between the blotch shapes and the models' suitability areas. The section also indicates differences in the output of the two models and tries to interpret these differences. The shape and size of the Extents of Occurrence are compared with the shape and size of Areas of Occupancy, also using the first fragmentation indexes such as the patch numbers, their average dimension and their Standard Deviation. Attention is drawn to small fragments distant from the core of the species' distribution. The other fragmentation indexes are used to confirm the connectivity between the different suitability areas. The fragmentation indexes should be used only as support for the maps.

This is followed by a discussion of the EO and AO percentages included in the protected areas: this information was influenced by the completeness and accuracy of protected areas coverage, which does not correspond to reality on a detailed scale, but is nevertheless a valid general indication of the degree of protection received. It is important to underline the differences between the EO percentages (in the table) and the AO percentages (in the comment section).

These comments are not intended as an indication of the species' conservation action, as analysis in this project was on a global scale and was not concerned with local situations.

References. The bibliography listed in this section includes the main publications used to compile each sheet and tends to be selective rather than exhaustive. Sources are listed by author and year in the text and are fully cited at the end of each account. Unless otherwise specified, information found in the Taxonomic notes and IUCN threat categories sections were compiled directly from Wilson & Reeder (1993) and Baillie & Groombridge (1996), respectively.

Tables and Figures. Tables and Figures bear the same numbers as the species to which they refer (e.g. 8.6.30 *Ourebia ourebi*: Fig. 8.6.30.a, Fig. 8.6.30.b, Tab.8.6.30.a, etc.).

Indexes. The index contains all the scientific names of the species considered in taxonomic and alphabetical order with reference to the ID code, the complete taxonomic code and the species account. Each species is then inserted into the respective family and order, both of which are ordered numerically.

7.2 The magnetic data bank

In accordance with the Terms of Reference a meta-data bank was set up containing the information needed for the user to be able to retrieve all the data relative to each of the species included in the project.

7.2.1 The data bank structure

A DBF structure contains all information for each species. The following table gives the structure of the DBF file (mammals.dbf) which contains 281 records, one for each species considered in the project.

Field N°	Field Name	Field Type	Field Width	Description
1	IDSYS	CHAR	8	Hierarchic systematic code
2	ORDER	CHAR	20	Order name
3	FAMILY	CHAR	20	Family name
4	GENUS	CHAR	20	Genus name
5	SPECIES	CHAR	20	Species name
6	AUTHOR	CHAR	40	Author's name and date of description
7	IDCOD	CHAR	6	Unique identification code
8	STATUS	CHAR	30	IUCN status
9	SUMMARY	CHAR	254	Full path and filename of the species' account sheet
10	EO	CHAR	254	Full path and filename of the species' distribution coverage
11	HABITAT	CHAR	254	Full path and filename of the species-habitat relationships
12	CDMODEL	CHAR	254	Full path and filename of the CD-model image
13	PCMODEL	CHAR	254	Full path and filename of the PC-model image
14	VDS	CHAR	254	Full path and filename of the validation data set

The user can browse the data bank file for available data on a taxon with any DBF compatible software (e.g. Dbase III, Dbase IV, FoxPro for Windows etc.). The data bank will provide, along with data regarding the species' taxonomy and the IUCN status (record fields : IDSYS, ORDER, FAMILY, GENUS, SPECIES, AUTHOR, IDCOD, STATUS), a list of files storing the species' data as reported in the printed data bank, some additional data used to model the areas of occupancy and the results of the field work validation analysis (record fields: SUMMARY, EO, HABITAT, CDMODEL, PCMODEL, VDS). In this way the user can retrieve the files of interest, copy them and use them with common software to manage texts, images and spreadsheets. The list below explains each field in more detail.

1. **IDSYS** is the hierarchic code, in which the first, second and third numbers refer respectively to the order, the family and the species (genus and species). The arrangement of orders and families follows Wilson & Reeder (1993); within these, the species are ordered alphabetically.
2. **ORDER**, **FAMILY**, **GENUS**, **SPECIES** and **AUTHOR** refer to the taxonomy of a given species.
3. **IDCOD** is the unique sequential code used throughout to identify a species; it acts as a link between the various pieces of information stored in distinct files, i.e. the numeric part of the IDCOD is common to all filenames dealing with a given species. For instance, the filenames amd001.doc and blo001.cgm refer to the file storing the species account sheet and the file storing the same species' Extent of Occurrence, respectively.
4. **STATUS** is the IUCN threat category, if applicable. The user will find the threat category along with the criteria adopted to assign it.
5. **SUMMARY** contains the filename and the full path of the species account sheet. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\amd001\amd001.pdf).
6. **EO** contains the filename and the full path of the species Extent of Occurrence coverage file. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\amd001 \blo001.e00).
7. **HABITAT** contains the filename and the full path of the species-habitat relationships matrix. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\ amd001\hab001.dbf).
8. **CDMODEL** contains the filename and the full path of the species CD-model image file. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\amd001\ cdm001.tif) [1].

[1] The user will find in the same directory two files with extention tfw (e.g. wat001.tfw and pro001.tfw) which were generated by Arc/Info during the conversion from the grid data set to the TIFF file. These are text files which store the information required by Arc/Info to georeference the image by the conversion from the TIFF file to a new grid data set.

9. **PCMODEL** contains the filename and the full path of the species PC-model image file. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\amd001\ pcm001.tif) [1].
10. **VDS** contains the filename and the full path of the fieldwork validation data set. The user will find the full path in the format «CdRom number\directory\filename» (e.g. CdRom01\amd001\vds001.txt). If the validation analysis was not performed for a given species, the field is left

7.2.2 Storage supports

All data were stored on CD-Rom; each CD-Rom contains several directories, each of which pertains to a single species and is named with the species' unique ID code (see the field IDCOD; e.g. amd001, amd255 etc.). The directories contain files with the species' data as specified below:

1. the species account sheet (see the field SUMMARY; e.g. amd001.pdf in Portable Document Format);
2. the coverage of the species Extent of Occurrence (see the field EO; e.g. blo001.e00 in Arc/Info Ms-Dos export format) [2];
3. the species-habitat relationships matrix (see the field HABITAT; e.g. hab001.dbf in Xbase compatible format);
4. the CD-model for the given species (see the field CDMODEL; e.g. cdm001.tif in TIFF PackBits compressed format) [1];
5. the PC-model for the given species (see the field PCMODEL; e.g. pcm001.tif in TIFF PackBits compressed format) [1];
6. the validation data set (see the field VDS; e.g. vds001.txt in text delimited format).

7.2.3 Information structure

The information stored in the files cited above is structured as follows.

7.2.3.1 Species account sheet

The data stored in the species account sheet are mainly texts describing the taxonomy, the extent of occurrence, and the results of the analyses performed. For more information on the structure of the sheet, see in the previous section "Printed data bank".

7.2.3.2 The Extent of Occurrence coverage

The boundaries of the EO of the given species were stored in a vectorial Arc/Info coverage and converted into the Arc/Info export Ms-Dos format[2]. The Lambert geographical projection was used with the following parameters:

Projection:	Lambert Azimuthal
Units:	meters
Radius of the sphere of reference:	6370997.00000
Longitude of centre of projection:	20° 00' 0.000
Latitude of centre of projection:	5° 00' 0.000

The polygons in the coverage have a 4- or 5-digit code with the following meaning:

Digit	Value	Description
1-4	nnnn	Bibliographic code of the source reference of the distribution map **certain presence**
1-5	9nnnn	Polygon enclosing areas for which the presence is **uncertain**
1-5	90000	Polygon enclosing areas of **certain presence** on the basis of expert support and without bibliographic support
1-5	99991	Polygon enclosing areas in which the species is reported to be **Introduced** or **re-introduced**
1-5	99999	Polygon completely enclosed within the EO of the species but in which the species is **absent**

The codification of polygons enables the user to retrieve readily the source of the map from the bibliographic database and to ascertain the quality of the information on presence associated with the various polygons.

[2] Require dos2unix conversion if used on Unix platform.

7.2.3.3 The species-habitat relationships matrix

For each species a DBF file, containing the suitability scores for each Land Cover class, White's vegetation type and the combination of the two falling inside the EO, is provided. The scores listed are those used to produce the CD model; for the species for which the distance from water or the elevation was taken into account, the suitability scores for each combination of vegetation and land cover inside and outside the specified threshold are also listed, together with the specified thresholds considered. Each file contains at least the following fields:

COD	Numeric code corresponding to the unique combination of White's vegetation type and Land Cover class;
VEG	White's classification second rank (see legend);
CODE2	White’s classification first rank (see legend);
LAND_CODE	Land Cover class (see legend);
SCOREW	Suitability score assigned to the vegetation type;
SCOREL	Suitability score assigned to the Land Cover class;
SCOREWLC	Suitability score assigned to the White's vegetation type/ Land Cover class combination;

For the species for which the distance from water was taken into account, the following fields were added:

DISTR	Distance from permanent water (in meters) considered as threshold;
IN	Suitability score assigned to the White's vegetation type / Land Cover class combination at distances below the specified threshold;
OUT	Suitability score assigned to the White's vegetation type / Land Cover class combination at distances greater than the specified threshold.

For the species for which elevation was taken into account, two elevation thresholds were defined, and scores were assigned to each combination depending on whether they fall either below the lower threshold, between the two thresholds or above the upper threshold; the following fields are thus present in the file:

ELEV1	Lower elevation threshold in meters above sea level.;
ELEV2	Upper elevation threshold in meters above sea level;
SCOREBE	Suitability score assigned to the White's vegetation type / Land Cover class combination at elevations below the specified lower threshold;
SCOREIN	Suitability score assigned to the White's vegetation type / Land Cover class combination at elevations between the two thresholds;
SCOREOV	Suitability score assigned to the White's vegetation type / Land Cover class combination at elevations over the specified upper threshold.

7.2.3.4 The Categorical-Discrete (CD) model image

The file stores the raster file of the CD model. The results of the analyses carried out with the software developed under the Arc/Info environment were stored in the Arc/Info grid data set format. These were then converted to a TIFF graphic file format (with PackBits compression) to make them readable with any image processing software package. Along with the TIFF file the user will find in the same directory two files with the extension tfw (e.g. cdm001.tfw and pcm001.tfw) which were generated by Arc/Info during conversion of the grid data set into the TIFF file. The files store the information required by Arc/Info for georeferencing the TIFF image when converting it into a new grid data set.
The image pixel values are the outcome of the intersect procedure between the species’ EO and the GLCC Land Cover data set. The values range from 1 to 8 as shown in the following table.

Pixel value	Description	Map Color
1	Suitable	green
2	Moderately suitable	pale green
3	Unsuitable	red
4	Undefined	gold
5	Environmental classes not found inside the EO	pale yellow
8	Water	cyan

7.2.3.5 The Probabilistic-Continuous (PC) model image

The file stores the raster file of the PC model. The results of the analyses carried out with the specific software we developed (using the Microsoft Visual C++ compiler under the Windows NT operating system) were stored in the Arc/Info grid data set format. We then converted them to the TIFF graphic file format (with PackBits compression) to make them readable with any image processing software package. Along with the TIFF file the user will find, in the same directory, two files with extension tfw (e.g. cdm001.tfw and pcm001.tfw) which were generated by Arc/Info during the conversion of the grid data set into the TIFF file. The files store the information required by Arc/Info for georeferencing the TIFF image when converting it into a new grid data set.

The image pixel values are the outcome of the analysis procedure described at point 5.2.1 to calculate the Mahalanobis distance. The values range from 0 to 1501, being the integer of the Mahalanobis distance times 10 and 1501 represents all values above a distance of 150, which was set as the upper limit distance. The false colours used for the map range continuously from a darker to a paler shade of green, indicating a more suitable (less distant) or less suitable (more distant) area for the species, respectively.

7.2.3.6 The validation data set

Not all the species have a validation data set. As explained in the paragraph concerning validation analysis, only the species that have at least 1% of their EO within any of the four sample areas were included in the validation process. Thus the validation data set is only available for those species.

When present, this information can be found in a text file stored under the species' directory and named vdsXXX.txt (where XXX stands for the species code used throughout this project). This text files are tab delimited and can easily be imported into the most popular spreadsheet programs, and if desired, DMBS packages.

The first line of the file contains the species scientific name while the rest of the file is organised according to the following table:

Point_code	unique identification code
CD_score	score of the location according to the CD model
Inside_EO	set to "in" if the location falls inside EO
Found	set to 1 if the specie was observed during field work
Country	country code of the location (B = Botswana, C = Cameroon, M = Morocco, U = Uganda)
White_Cod	White's Vegetation Map class code bserved at the location
Latitude	Latitude of location
Longitude	Longitude of location

7.2.4 Bibliographic data

To allow for more standard use of the bibliographic data, the literature database is stored in ASCII text format. The maximum length of a CHAR Xbase field is 254 byte, enabling storage of a 254 character string: yet several reference titles are longer than that. Use of the MEMO field type would not have solved the problem because management of these kinds of fields is not standard and changes according to the DBMS used. The ASCII text format enables the user to import the data easily into any DBMS structure. The structure of the literature database is organised as shown in the following table. Each field is quoted and comma delimited. (e.g. "1499","Wilson & Reeder (1993)","Wilson D.E., Reeder D.M. (Eds.) (1993).","…","Smithsonian Institution Press, Washington D.C.").

Field	Field Type	Field Width	Description
1	CHAR	free text length	Unique identification code
2	CHAR	free text length	Author/s and date in short form
3	CHAR	free text length	Author/s and date in full form
4	CHAR	free text length	Title
5	CHAR	free text length	Reference
6	CHAR	free text length	Keywords (comma delimited)

8. Updates, bugs and comments (a request for)

The data bank we present provides the first set of data on the medium and large African mammals organised and stored in a digital format. As such, it also provides the first of a time series that it is hoped will be updated and maintained throughout the years: time series are the necessary tool to monitor changes and trends and therefore to assess conservation needs. It is hoped that the standard formats used in the data bank will help in updating it. The viability of these data will also be a function of future work on the evolution of species and population status.

As mentioned before, the data bank is intended to provide raw and semi-processed data in a way that further analyses can be carried out at local scale: the present data should be taken as the starting point for ad-hoc research into smaller scale status of regions, nations, subspecies and local populations. These applications will be the best way to use the present data. The utilisation of these data is free: we will appreciate any communication on the use of these data and the kind of results that will be obtained.

As for any other new electronic data bank, bugs are likely to be found by the users and we would be most grateful for any comment that may help us in eliminating them.

9. Acknowledgements

A great number of people and organizations supported the projecy throughout its implementation. It would be impossible to name them all, and we like to mention at least the following who provided data, collaboration, criticism, review and general support:
G. Amori, N. Burgess, J-P. d'Huart, R. East, R. Emslie, H. Hoeck, P. Jackson, H. Klingel, Maria Jadot, M. G. L. Mills, Marcel van Opstal, H. Van Rompaey, D. Shackleton, C. Sillero Zubiri, the University of Botswana (UB, Gaborone), the Institut de Recherches Zootechniques et Veterinaires (IRZV, Yaoundé), the Association Marocaine pour la Protection de l'Environnement (ASMAPE, Rabat), and the Makerere University Institute of Environment and Natural Resources (MUIENR, Kampala), Center for Tropical Diversity - Copenhagen, FAO, WCMC, SSC.

10. Acronyms

AO	Area of Occupancy
ADS	Africa Data Sampler
AETFAT	Association pour l' Étude Taxonomique de la Flore de l'Afrique Tropicale
ARTEMIS	Africa Real Time Environmental Monitoring Information System
AVHRR	Advanced Very High Resolution Radiometer
AWMSI	Area-weighted Mean Shape Index
CD	Categorical – discrete (distribution model)
DBMS	Data Base Management System
DCW	Digital Chart of the World
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EO	Extent of Occurrence
FAO	Food and Agriculture Organisation of the United Nations
GIS	Geographical Information System
GLCC	Global Land Cover Characterization
GRID	Global Resource Information Database
IDCODE	Unique Identification Code
I/O	Input/Output
IUCN	International Union for the Conservation of Nature and Natural Resources
LPI	Largest Patch Index
MPS	Mean Patch Size
MSI	Mean Shape Index
NDVI	Normalized Difference Vegetation Index
NCGIA	National Center for Geographic Information and Analysis
NP	Number of Patches
ONC	Operational Navigation Chart
PC	Probabilistic – continuous (distribution model)
PSSD	Patch Size Standard Deviation
SSC	Species Survival Commission
UNDP	United Nations Development Programme.

UNESCO	United Nations Educational, Scientific and Cultural Organisation.
UNSO	United Nations Sudano-Sahelian Office
USGS	United States Global Survay
WCMC	World Conservation Monitoring Centre
WRI	World Resource Institute
WWF	World Wide Fund for Nature

11. References

Akçakaya H.R. (1993). RAMAS/GIS: linking landscape data with population viability analysis. Applied Biomathemathics, New York.
Aspinall R., Matthews K. (1994). Climate change impact on distribution and abundance of wildlife species: an analytical approach using GIS. Environmental Pollution 86: 217-223.
Baillie J., Groombridge B. (Eds) (1996). 1996 IUCN Red List of Threatened Animals. IUCN Species Survival Commission.
Baird D.R. (1996). Predicting East African ungulate diversity from satellite sensor imagery. Master of Research in the Natural Environment. University of Edinburgh.
Breininger D.R., Provancha M.J., Smith R.B. (1991). Mapping Florida Scrub Jay habitat for purposes of land-use management. Photogrammetric Engineering & Remote Sensing 57(11): 1467-1474.
Busby J.R. (1991). BIOCLIM - a bioclimatic analysis and prediction system. In: Nature Conservation: cost efective biological surveys and data analysis, C.R. Margules and M.P. Austin (eds), CSIRO: melbourne, pp 64-68.
Congalton R.G., Stenback J.M., Barrett R.H. (1993). Mapping Deer habitat suitability using Remote Sensing and Geographic Information Systems. Geocarto International 3: 23-33.
Corsi F., De Leeuw I., Skidmore A. Species distribution modelling with GIS In: Boitani L. and T.K. Fuller (Eds) Research Techniques in Animal Ecology. Columbia University Press, New York. In press.
Donovan M.L., Rabe D.L., Olson C.E. (1987). Use of Geographic Information Systems to develop habitat suitability models. Wildlife Society Bullettin, 15: 574-579.
East R. (Ed.) (1996). Antelope survey update. IUCN/SSC Antelope Specialist Group: n° 2.
Eastman J.R. (1992). Time series map analysis using standardised principal components. Proceedings, ASPRS/ACSM/RT '92 Convention: Mapping and Monitoring Global Change. Bethesda: ASPRS/ACSM.
ESRI (1993). Digital Chart of the World for use with Arc/Info software. Environmental Systems Research Institute, Inc. Redlands, CA.
Estes R.D. (1991). The Behaviour Guide to African Mammals. Including Hoofed Mammals, Carnivores, Primates. The University of California Press.
Gaston K.J. (1991). How large is a species' geographic range? Oikos 61: 434-438
Homewood K.M. (1975). Can Tana River Mangabey survive?. Oryx 13: 53-59.
Jensen J.R., Narumalani S., Weatherbee O., Morris K.S. Jr. (1992). Predictive modelling of Cattail and Waterlily distribution in a South Carolina reservoir using GIS. Photogrammetric Engineering and Remote Sensing 58(11): 1561-1568.
Kingdon J. (1997). The Kingdon field guide to African Mammals. Academic Press, London and New York: Natural World.
Macdonald D.W. (Ed.) (1984). The Encyclopaedia of Mammals. Facts on File, New York.
Maurer B.A. (1994). Geographical Population Analysis: Tools for the Analysis of Biodiversity. Blackwell Science, Oxford.
Maurer B.A., Heywood S.G. (1993). Geographic range fragmentation and abundance in neotropical migratory birds. Conservation Biology 7: 501-509.
May R.M. (1994). The effects of spatial scale on ecological questions and answers. In: Large-scale Ecology and Conservation Biology. (Ed. by P.J. Edwards, R.M. May and N.R.Webb), pp. 1-17. Blackwell Scientific Publications, Oxford.
McGarigal K., Marks B.J. (1994) FRAGSTATS Spatial Pattern Analysis Program for Quantifying Landscape Structure. Version 2.0. Oregon State University, Corvallis.
Meester J., Setzer H.W. (Eds) (1971). The Mammals of Africa, an Identification Manual. Smithsonian Institution Press, Washington D.C.
Morrison M.L., Marcot B.G., Mannan R.W. (1992). Wildlife-habitat relationships: concepts and applications. Univ. Wisconsin Press..
Oates J.F. (1996). African Primates Status Survey and Conservation Action plan. IUCN/SSC Primate Specialist Group.
Pereira J.M.C., Itami R.M. (1991). GIS-base habitat modeling using loistic multiple regression: a study of the Mt. Graham Red Squirrel. Photogrammetric Engineering & Remote Sensing. 57(11): 1475-1486.
Pimm S.L. (1991). The Balance of Nature? University of Chicago Press, Chicago.
Rowe-Rowe D.T. (1994). The ungulates of Natal. Natal Parks, Game & Fish Preservation Board, Pietermaritzburg, Natal, R.S.A.
Skidmore A.K., Gauld A., Walker P.W. (1996). A comparison of GIS predictive models for mapping kangaroo habitat.: International Journal of Geographical Information Systems, v. 10, p. 441-454.
Skinner J.D., Smithers R.H.N. (1990). The mammals of the Southern African subregion. University of Pretoria, Pretoria.
Spinage C.A. (1982). A territorial antelope: the Uganda waterbuck. Academic Press, London and New York.
Walker P.A. (1990). Modelling wildlife distributions using geographic information system: kangaroos in relation to climate. Journal of Biogeography 17: 279-289.
Western D. (1975). Water availability and its influence on the structure and dynamics of a savannah large mammal community. E. Afr. Wildl. J. 13: 265-286.
White F.J. (1983). The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. UNESCO, Paris.
Wilson D.E., Reeder D.M. (Ed.) (1993). Mammals species of the World: A Taxonomic and Geographic Reference. Second edition. Smithsonian Institution Press, Washington D.C.
WRI (1995). Africa Data Sample. World Resource Institute, Washington, D.C.