MIGRATION WITHIN AND OUT OF AFRICA: IDENTIFYING KNOWLEDGE GAPS BY DATA-MINING THE GLOBAL REGISTER OF MIGRATORY SPECIES

Klaus Riede
Zoologisches Forschungsinstitut und Museum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany, Email: k.riede.zfmk@uni-bonn.de

Abstract: Knowledge gaps for birds migrating within Africa, or visiting the continent either for breeding or wintering, have been identified systematically by data-mining the Global Register of Migratory Species (GROMS). The Register contains a migratory vertebrate species reference list, GIS maps and a literature database with full-text passages on migration. Using the GROMS literature and accessory tables for data mining, knowledge gaps with respect to migration routes and seasonal timing were diagnosed for more than 150 bird species. This analysis was based on a search for text strings such as “migration unknown”, “poorly known”, “surprisingly little known”, etc. The species list of lesser-known migrants generated by data-mining has been published on the GROMS website (www.groms.de), and provides a starting point for future research. The underlying complex query uses Standard Query Language (SQL), requiring a fully documented relational database. Such documentation is often lacking for current biodiversity information systems, which limits their use for complex data mining.

Key words: Migratory birds; intra-African migrants; biodiversity informatics; data-mining

1. INTRODUCTION

Africa provides wintering habitats for billions of birds, mainly from the Palearctic. In addition, its marine shelf areas are important breeding and feeding grounds for marine mammals, seabirds and diadromous fishes. Besides this well-known significance of African ecosystems for inter-

B.A. Huber et al. (eds.), African Biodiversity, 245–252.

continental migrants, there are less well-known intra-African migration phenomena, which seem to depend on rainfall patterns and/or insect outbreaks, e.g. locusts.

The present article exemplifies the potential and problems of data-mining by querying GROMS for African migratory species, particularly searching for knowledge gaps on migration behaviour of non-passerine birds. The GROMS is a cross-sectional database based on a reference list of currently 4,400 vertebrate species and subspecies, combined with 1,100 global distribution maps in GIS format. Besides publication on the World Wide Web (www.groms.de), detailed documentation is available in book format, with a full off-line version of the relational database and the complete GIS map repository on CD-ROM (Riede 2001, 2004a). First results of a global meta-analysis of the entire GIS dataset have recently been published elsewhere (Riede, 2004b). In this paper I will show that such advanced data-mining is only possible by using Standard Query Language (SQL, see Taylor, 2001), together with full documentation of the data model. The result from this text-mining exercise is a document summarising knowledge gaps in African bird migration behaviour, which can be downloaded from the GROMS website (www.groms.de/groms/knowledge_gaps.pdf).

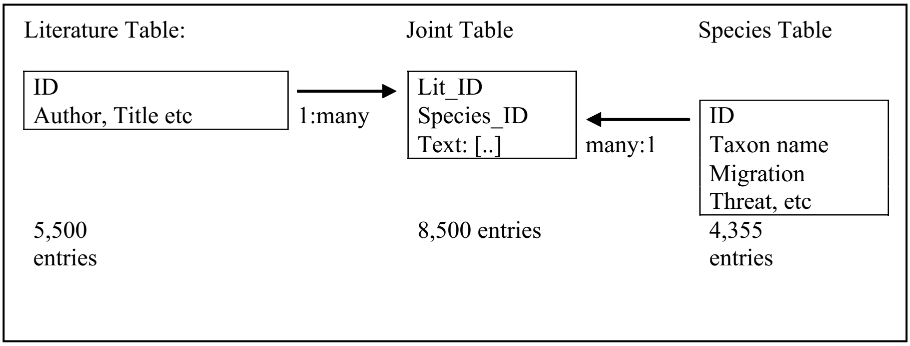
2. METHODS

The definition of migration phenomena, as used in this analysis, is based on cyclical, true migration of more than 100 km (for details, see Riede, 2001, p. 25, and Dingle, 1996).

Full-text citations about migration of all non-passerine birds identified as "migratory" have been entered into the GROMS database, mainly using the "movement" section within species accounts published in del Hoyo et al. (1992-2003). Figure 1 shows the many: many structure of the relational data model, which makes it possible to relate one species with many citations or text passages, or relate one text passage with many species, with several keywords (in this case: migration). Such a data model allows storage of contradictory information from different sources, referring to migration, threat, taxonomic status and other key themes. The data entry module is available on the GROMS CD-ROM (Riede, 2004a).

The recent off-line edition of the GROMS database (Riede, 2004) has been used to identify migratory species occurring in Africa, using the ACCESS (MS-Windows) database query interface to select species according to geographic criteria (GEO-Search: l.c., p. 94).

For text mining, a genuine SQL-statement was generated, using the ACCESS query generator (Fig. 1), searching for text strings such as




```
SELECT Tab_Arten.Latein, Tab_Arten.Englisch, Tab_Arten.Migration,
Jointab_Art_Lit.Lit_Bezug, Tab_Literatur.Autor_Name, Tab_Literatur.Autor_Vorname,
Tab_Literatur.Coautoren_Namen, Tab_Literatur.Jahr, Tab_Arten.animalgroup,
Tab_Arten.Familie
FROM Tab_Literatur RIGHT JOIN (Tab_Arten INNER JOIN Jointab_Art_Lit ON
Tab_Arten.ID = Jointab_Art_Lit.ID_Art) ON Tab_Literatur.ID = Jointab_Art_Lit.ID_Lit
WHERE (((Jointab_Art_Lit.Theme)=7) AND ((Tab_Arten.Animal_Class)=2) AND
((Jointab_Art_Lit.Lit_Bezug) Like "*   unknown*")) OR (((Jointab_Art_Lit.Theme)=7) AND
((Tab_Arten.Animal_Class)=2) AND ((Jointab_Art_Lit.Lit_Bezug) Like "*     perhaps*")) OR
(((Jointab_Art_Lit.Theme)=7) AND ((Tab_Arten.Animal_Class)=2) AND
((Jointab_Art_Lit.Lit_Bezug) Like "*   little*")) OR (((Jointab_Art_Lit.Theme)=7) AND
((Tab_Arten.Animal_Class)=2) AND ((Jointab_Art_Lit.Lit_Bezug) Like "*     poor*")) OR
(((Jointab_Art_Lit.Theme)=7) AND ((Tab_Arten.Animal_Class)=2) AND
((Jointab_Art_Lit.Lit_Bezug) Like "*   possib*"))
ORDER BY Tab_Arten.animalgroup, Tab_Arten.Familie;
```

Figure 1. Text-mining by using the query generator. Species (Tab_Arten) and literature (Tab_Literatur) table are connected with a many: many relation by a Jointab containing full-text passages. The latter are searched for relevant text strings characteristic for knowledge deficits, which are high-lighted in the statement.

"migration unknown", "poorly known", "surprisingly little known", "thought to be sedentary but...", etc.

3. RESULTS

Searching the GROMS for species migrating within Africa, or migrants visiting the continent either for breeding or wintering, resulted in 1,285 vertebrate species, with a surprisingly high number of migratory fish species from Sub-Saharan Africa (Table 1). The more complex SQL statement described in Figure 1 was necessary to search for birds with insufficiently known migration behaviour. A search for all text entries on bird migration containing text strings such as "unknown", "perhaps", "little", "poor" or "possib" resulted in a list of 349 birds with insufficiently known migration behaviour.

Apus caffer White-rumped swift	Migratory in northernmost and southernmost parts of range. Spanish population present early May to Aug-Oct, some recorded into early Dec, with autumn migration through Straits of Gibraltar mid-Aug to mid-Oct; S African population present Aug-May, mainly absent from S Cape and much reduced farther N within S breeding range Jun-Jul. **Poorly** understood wet-season movements into Sahel may be feature of N sub-Saharan populations. Otherwise resident. Migrates in flocks of up to 100. S African migrants may be transequatorial. Some degree of altitudinal migration in Natal. First record from rabia 1982, and seen at least once subsequently in Tihamah coastal plains, Saudi Arabia, in Mar 1989. Vagrant to Norway (May, Jun) and Finland (Nov).
Caprimulgus climacurus Long-tailed nightjar	**Poorly** known. Nominate race migratory and partially sedentary, some populations moving S after breeding season. Race sclateri **possibly** sedentary and partially migratory. Race nigricans **probably** sedentary. Outside breeding season, range also includes S Ivory Coast, SW Nigeria, S Cameroon, Equatorial Guinea, Gabon, SE Congo (lower Congo river valley), NE Angola (one record Luaco), SE Sudan, SW Ethiopia, W Kenya (sporadic in Turkana and Pokot region) and E Uganda.

Figure 2: Sample output for 2 out of 349 species found searching GROMS with the query listed in Fig. 1. Text particles matching search strings are highlighted.

Examples of two typical text passages are shown in Figure 2; the full list is available as pdf-document at www.groms.de/groms/knowledge_gaps.pdf. Filtering this list for species occurring in Africa still left 107 non-passerine birds from 15 orders (Table 2). Forty-one species are intercontinental migrants, with most of the population visiting the African continent for wintering (Example: Eurasian cuckoo - *Cuculus canorus*). A high number of 42 species shows partial migratory behaviour of a certain subspecies or population (Example: Little green bee-eater - *Merops orientalis*). In these cases, uncertainties with respect to migration mainly refer to difficulties in recognition and tracking of the migratory individuals in the field, where they often mix with the resident part of the population. The fraction of 12 intracontinental migrants contains intra-African migrants with poorly studied, but evidently complex migration patterns (Curry-Lindahl, 1981a, b; example: Pennant-winged nightjar - *Macrodypterix longipennis* [Gould, 1837]). Intracontinental migration often blends with other, less-well defined categories used within the GROMS database, such as "nomadising" (e.g. Tawny eagle - *Aquila rapax* [Temminck, 1828]), "range extension" (e.g. White-backed night heron - *Gorsachius leuconotus* [Wagler, 1827], or

Table 1. Migratory vertebrate species occuring in Africa, according to GROMS.

	Vertebrates	Mammals	Birds	Sea Turtles	Fish
North-	604	58	234	5	307
Sub-Saharan	1.174	79	388	5	702
AFRICA	**1.285**	**93**	**447**	**5**	**740**

Table 2. Number of species of non-passeriform African migrants with insufficiently known migration behaviour within each order (following the systematics used by del Hoyo et al., 1992)

Order	Number of species
Charadriiformes	36
Procellariiformes	13
Gruiformes	12
Falconiformes	10
Caprimulgiformes	9
Cuculiformes	8
Apodiformes	3
Columbiformes	3
Pelecaniformes	3
Anseriformes	2
Ciconiiformes	2
Coraciiformes	2
Strigiformes	2
Podicipediformes	1
Sphenisciformes	1

"possibly migratory" (e.g. Three-banded plover - *Charadrius tricollaris* Vieillot, 1818), altogether comprising an additional 12 insufficiently known migrants.

Table 2 shows that most (36) species with insufficiently known migration behaviour belong to the Charadriiformes, a speciose and highly diverse group comprising 300 migratory species – plovers, lapwings and sandpipers, but also seabirds such as skuas, gulls and auks. A much higher proportion of knowledge deficits is observed within Caprimulgiformes: movements are poorly understood for 9 out of 31 migratory nightjar species. With 8 out of 46, a similarly high proportion is observed in cuckoos (Cuculiformes).

A closer look at some text examples reveals a wide variety of knowledge gaps, related to timing, seasons, food availability and regional distribution. In the following, some typical examples are given. The monotypic white-rumped swift *Apus caffer* (Lichtenstein, 1823) exhibits "...poorly understood wet-season movements into Sahel may be feature of N sub-Saharan populations" (del Hoyo et al., 1999). The Sahel during the wet season is a huge, mostly inaccessible region, and even with better ornithological coverage most specialists will find it difficult to recognise swifts belonging to the northern sub-Saharan populations.

For the little ringed plover *Charadrius dubius* Scopoli, 1786, del Hoyo et al. (1996) noted: "Race *curonicus* migratory, but possibly resident in S breeding areas; W European population migrates across Sahara to tropics;[...]." This is a typical example of unclear delimitation of the frontier between resident and migratory individuals within one population, probably showing clinal variation along a North-South gradient, and possibly variation with annual weather conditions. Weather conditions, and in particular rain, as a trigger for migratory movements is hypothesised for the standard-winged nightjar *Macrodipteryx longipennis* (Shaw, 1796): "An intra-African migrant, movements protracted and possibly influenced by rains, not fully understood. Leaves breeding grounds in southern savannas of W & C Africa from mid-Apr to about Aug and moves N to spend wet season in savannas of Sahel and Sudan." (del Hoyo et al., 1999). In the case of nightjars, it is probably crypsis, together with their nocturnal habits, which hamper research. Even in well-studied regions, such as the Netherlands, recent investigations revealed 50% underestimation of populations by traditional monitoring methods, which could be overcome by application of new techniques, such as song playback (Bult, 2003).

4. DISCUSSION

In summary, most knowledge deficits can be attributed to poorly understood wet season movements. A recent search of the BIOSIS database revealed only a very few papers adding to our understanding of African intra-tropical migration. The extensive study of Anciaux (2000) on migration phenology of a wide range of African non-passerines adds to our knowledge, but is limited to a small study area of inland South Benin. Progress will only be made if the few data from satellite telemetry and ringing are pooled in a more efficient way, using advanced information technologies. Regular monitoring efforts are restricted to a few countries, such as the southern African region (Harrison, 1997), or certain groups, such as waterbirds (Dodman and Diagana, 2003).

Besides the biological results presented here, this small data-mining exercise illustrates that advanced data-mining requires access to entire, fully documented databases, using queries based on Standard Query Language (SQL). This had to be complemented by searching commercial literature databases, such as BIOSIS (www.biosis.org). At present, the Global Biodiversity Information Facility (www.gbif.org) endeavours to harmonise the numerous databases and information systems by harmonising taxonomic standards and designing portals, connecting different databases within "federated systems". However, it will be a challenge to provide user-friendly tools for complex data-mining, which requires deep-drilling into federated systems composed of several biodiversity databases, differing in structure and function.

ACKNOWLEDGEMENTS

Supported by the Secretariat of The Convention on the Conservation of Migratory Species of Wild Animals (UNEP-CMS), Bonn. The development of the GROMS database was supported by the German Federal Agency for Nature Conservation with funds from the German Federal Environment Ministry. Thanks to all members of the GROMS team, and in particular Thomas Lingen and Bedru Sherefa-Muzein, for patiently entering full-text passages on migration from del Hoyo et al. Thanks also to Lynx editions for allowing digital publication of maps and text passages from “Birds of the World”.

REFERENCES

Anciaux, M. R., 2000, Approche de la phenologie de la migration des migrateurs intra-africains de l'interieur des terres du sud-benin (plateau d'allada et sud de la depression de la lama). 1. Les non-passeriformes et les non-coraciiformes, *Alauda* **68:**311-320.

Bult, H., 2003, Nachtzwaluwen *Caprimulgus europaeus* onder de rook van Antwerpen, *Limosa* **75:**91-102.

Curry-Lindahl, K., 1981a, *Bird migration in Africa (Movements between six continents). Vol.1*, Academic Press, London.

Curry-Lindahl, K., 1981b, *Bird migration in Africa (Movements between six continents). Vol.2*, Academic Press, London.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 1992, *Handbook of the birds of the world. Vol.1: Ostrich to Ducks,* Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 1994, *Handbook of the birds of the world. Vol.2: New World Vultures to Guineafowl,*- Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 1996, *Handbook of the birds of the world. Vol.3*: *Hoatzin to Auks*, Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 1997, *Handbook of the birds of the world. Vol.4: Sandgrouse to Cuckoos,* Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 1999, *Handbook of the birds of the world. Vol.5: Barn Owls to Hummingbirds*. Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 2001, *Handbook of the birds of the world. Vol.6: Mousebirds to Hornbills.* Lynx Edicions, Barcelona.

del Hoyo, J., Elliott, A., and Sargatal, J. (eds), 2002, *Handbook of the birds of the world. Vol.7: Jacamars to Woodpeckers.* Lynx Edicions, Barcelona.

Dingle, H., 1996, *Migration - The biology of life on the move*, Oxford University Press, New York.

Dodman, T., and Diagana, C. H., 2003, *African Waterbird Census 1999, 2000 and 2001 - Les Dénombrements d'Oiseaux d'Eau en Afrique 1999, 2000 et 2001*, Wetlands International, Netherlands.

Harrison, J. A., Allan, D. G., Underhill, L. G., Herrmans, M., Tree, A. J., Parker, V., and Brown, C. J., 1997, *The Atlas of Southern African Birds*, Vol.1: Non-Passerines, 2, Johannesburg, South Africa.

Riede, K., 2001, *Global Register of Migratory Species - Database, GIS Maps and Threat Analysis. Weltregister wandernder Tierarten,* Landwirtschaftsverlag, Münster.

Riede, K., 2004a, *Global Register of Migratory Species: From global to regional scales,* Landwirtschaftsverlag, Münster.

Riede, K., 2004b, The "Global Register of Migratory Species" - first results of global GIS Analysis. In: Werner, Dietrich (ed.): *Biological Resources and Migration*. Springer-Verlag, Berlin.

Taylor, A. G., 2001, *SQL for dummies,* Books Worldwide, Indianapolis.