Long-term monitoring of Serengeti bird species occurrence, abundance, and habitat

L. Francisco Henao Diaz [1, 2]* and Anthony R. E. Sinclair[1, 2]

[1]Biodiversity Research Centre, University of British Columbia, 2212 Main Mall, Vancouver, BC V6T 1Z4 Canada

[2]Department of Zoology, University of British Columbia, 4200 – 6270 University Blvd., Vancouver, BC V6T 1Z4 Canada

***Corresponding author:** L. Francisco Henao Diaz, email: fhenaod@zoology.ubc.ca

Abstract: The Serengeti ecosystem contains one of the most diverse bird assemblages in Africa. We present here a data set consisting of abundances of bird species in different habitats of the Serengeti ecosystem over a 87-year timeframe. This data set comprises 66,643 georeferenced occurrences for 568 species from 1929 to 2017. Most records contain feeding location, food source, distribution status and observation locality. The records originate from three different but complementary methodologies: points, sites and transects. The point method (bird species records: 1929–2017) is based on *ad hoc* observations and includes rare species or those in special habitats. These points came from published records as well from the research program of A.R.E. Sinclair and colleagues. The site method (1966–2017) is based on structured observations at sites selected to represent specific habitats, and replicated within habitats and over time. At each site, birds were recorded by sight and sound over a radius of 50 m for 10 minutes. The transect method (1997–2011) is based on road transects covering different areas of the ecosystem. Road transects were traversed using a vehicle with observers travelling at 30 km/h. Bird species were those easily seen from a vehicle out to 50 m either side. As most transects were traversed multiple times, this method provides information on temporal change in abundance for a select set of species. No copyright restrictions apply to the use of this data set other than citing this publication.

Key words: Africa, Tanzania, occurrences, feeding, ecology, birding, long-term survey.

METADATA

CLASS I. DATA SET DESCRIPTORS

A. Data set identity: Long-term monitoring of Serengeti bird species habitat choice and changes in abundance.

B. Data set identification code: Henao Diaz and Sinclair - serengeti_birds_dataset.txt

C. Data set description

1. Originators: Anthony R. E. Sinclair[1, 2]

[1]Biodiversity Research Centre, University of British Columbia, Canada

[2]Department of Zoology, University of British Columbia, Canada

2. Abstract: The Serengeti ecosystem contains one of the most diverse bird assemblages in Africa. We present here a data set consisting of abundances of bird species in different habitats of the Serengeti ecosystem over a 87-year timeframe. This data set comprises 66,643georeferenced occurrences for 568 species from 1929 to 2017. Most records contain feeding location, food source, distribution status and observation locality. The records originate from three different but complementary methodologies: points, sites and transects. The point method (bird species records: 1929–2017) is based on *ad hoc* observations and includes rare species or those in special habitats. These points came from published records as well from the research program of A.R.E. Sinclair and colleagues. The site method (1966–2017) is based on structured observations at sites selected to represent specific habitats, and replicated within habitats and over time. At each site, birds were recorded by sight and sound over a radius of 50 m for 10 minutes. The transect method (1997–2011) is based on road transects covering different areas of the ecosystem. Road transects were traversed using a vehicle with observers travelling at 30 km/h. Bird species were those easily seen from a vehicle out to 50 m either side. As most transects were traversed multiple times, this method provides information on temporal change in abundance for a select set of species. No copyright restrictions apply to the use of this data set other than citing this publication.

D. Key words: Africa, Tanzania, occurrences, feeding, ecology, birding, long-term survey.

CLASS II. RESEARCH ORIGIN DESCRIPTORS

A. Overall project description

1. Identity: Long-term monitoring of Serengeti bird species habitat choice and changes in abundance.

2. Originators: Anthony R.E. Sinclair

3. Period of Study:

Points data (*ad hoc* records) started in 1929 but the majority began in 1996. The data collection is ongoing, reported here until 2017.

Site data began in 1966 and terminated largely in 2011, although a few sites continue until 2017.

Transect data began in 1997 and terminated in 2011.

4. Objectives:

1. To record the abundance of bird species in different habitats of the Serengeti ecosystem.
2. To monitor changes in abundance, using three road transects, of a selected group of bird species over time.

5. Abstract: Same as above

6. Sources of funding:

Canadian Natural Sciences and Engineering Resources Council.

Frankfurt Zoological Society.

B. "Specific subproject" description

1. Site description:

The Serengeti ecosystem (34–36° E, 1–4° S) in Tanzania, East Africa has been protected as a national reserve or park since the 1920s. In general, there is a wet season from approximately November to June and a dry season July to October. Nearly all species breed in response to rain, although the start of breeding varies with the feeding guild (Sinclair 1978). Weather responds to the effects of El Niño Southern Oscillation with either floods or droughts occurring at intervals of 4–6 years (Sinclair et al. 2013).

We will use the common name 'African Acacias' to because these are generally recognized in common usage but will enter the Latin name that has been changed. The predominant

vegetation of the protected area in Serengeti is a savanna dominated by fine-leaved African Acacia and broad-leaved *Terminalia* trees. The African Acacia savanna is composed of different species that form effectively monospecific stands in patches about 200 m across and that separate along a soil gradient called 'catena'. At the top of low ridges the soil is shallow and rocky; these ridges support *Vachellia* (*Acacia) tortilis*, *Senegalia* (*Acacia) senegal* and *V. hockii* together with two species of *Commiphora*. In mid-slope, with deeper soils, the dominant species *V. robusta* occurs, the most frequent large tree species in the ecosystem. On sandy washouts, at the base of hills, the wait-a-bit thorn, *S. mellifera,* occurs. These areas collectively compose the 'upper catena'. At the bottom of the slope, with deep silt soils that impede drainage, are the small gall African Acacias, *V. drepanolobium* and *V. seyal*, and the tall tree, *Balanites aegyptiaca*. These areas collectively make up the 'lower catena'. There are also patches of open grassland in largely impeded drainage areas; most are small of 200–300 m across, but a few, such as Musabi and Ndoho in the west and Togoro in the center, span 2–3 km. In general, the 'African Acacia savanna' is composed of a fine-scale mosaic of different monospecific stands of trees creating patches of 100–300 m across. The grass layer over most of the catena is dominated by red oats grass (*Themeda triandra*).

The *Terminalia* woodland, a type of broad-leaved miombo woodland characteristic of southern and central Africa, occurs in the far northwest of the ecosystem with a grass layer of tall (1m) perennial *Hyparrhenia* species. It occurs on granitic rocky ridges which give way downslope to either open *Themeda* grassland or grassland with *V. gerrardii* trees; these trees replace *V. robusta* in the African Acacia savanna.

A number of smaller but distinct habitats all occur within the savanna. Rocky hills that rise steeply 200–500 m above the surrounding savanna are found along the eastern boundary and through the middle of the western corridor. These hills support *Combretum* woodland on the lower stony slopes, a subset of the *Terminalia* woodlands further north, with similar grasses and herbs. The main rivers support dense riverine forests that depend on groundwater. These forests are of two sorts: those along the Mara River in the north are montane (Loita Hills) in origin; whereas those along the Grumeti and Mbalageti Rivers to

the west are lowland Congo forest in origin (see Sinclair et al. 2015). Numerous small seasonal streams dissect the savanna and support a thin strip of bushes and riverine trees; the larger streams have large African Acacias, *V. xanthoploea* and *V. kirkii*. Smaller drainages occur as wet grasslands with small bushes and rushes. There are three lakes, all very shallow and highly alkaline: Lakes Lagarja and Masek , which form the top end of the Olduvai Gorge on the southeast plains, and Lake Magadi on the western edge of the plains. Freshwater is confined to the rivers, a few springs seeping out of the hills, and Lake Victoria at Speke Gulf in the far west.

The southeastern part of the ecosystem is open, treeless grassland with a gradient of long grass (similar to that of the *Acacia* savanna) in the northwest of the plains grading to intermediate grasslands of *Pennisetum* (20–50 cm) in the center, and short grasslands (5–15 cm) in the far east and south. Detailed descriptions of these habitats are given in Sinclair 2008.

In the past, similar natural savanna extended west of the present park borders covering about 2050 km^2 until agriculture, smallholdings with cereal and root crops, took over in the mid-19^{th} century; these land cover types extended eastwards in the 1950s and now about the western border of the natural ecosystem. The present agricultural areas were originally similar in flora and fauna, geology, soil nutrients and other ecological features to those of the native savanna. Agriculture, which forms an abrupt boundary with the savanna on the western border, has removed most trees. Many small native shrubs surround crops of millet and cassava.

2. Experimental or sampling design

Bird Surveys

Data on the bird communities were obtained by one of three methods: transects by vehicle along tracks and roads (“transects”); sites located in selected habitats (“sites”); and ad hoc points (“points”). These data have been collected over several decades, beginning in 1929 and continuing to the present (Figure 1).

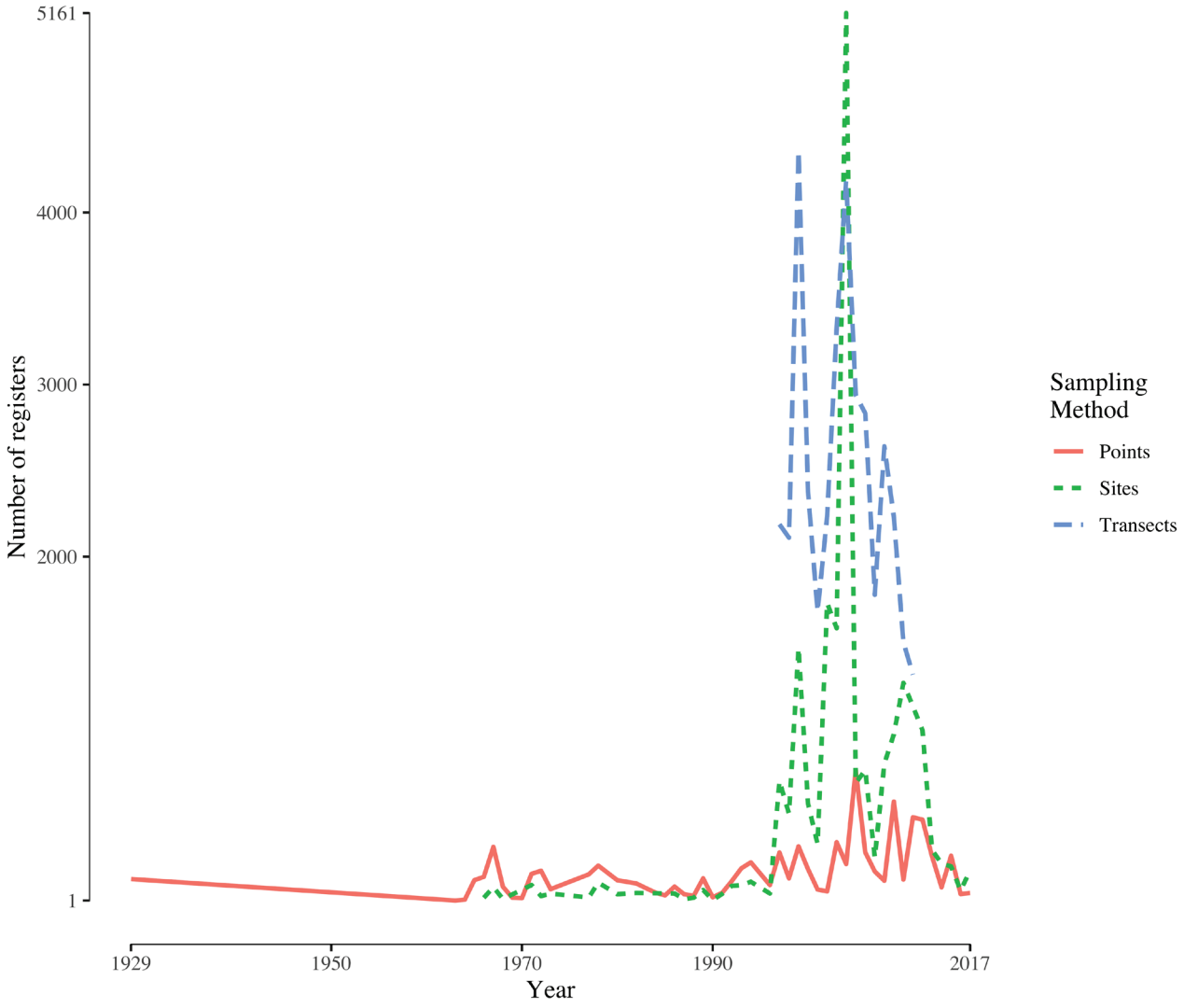

Figure 1. Number of registers through time per sampling scheme.

Transects. Thirty transects of lengths varying from 15 km to 100 km were located over the majority of the ecosystem along pre-existing tracks. Transects were driven slowly (30–50 km/h), stopping where necessary, using two observers, a recorder and a driver. All species, their numerical abundance, and habitat were recorded up to 50 m either side of the track. Each transect was divided into 5 km segments and records for species, as well as information on occurrence of habitat types, were pooled for each segment. The bird groups counted were medium and large sized insectivores, and all granivorous, frugivores and raptors. These focal species were easily detectable in the open grass and savanna habitats

that were surveyed. Forests were not sampled by this method. In this case the georeferenced points are at the start of each transect.

Three transects were traversed on a regular basis twice a year, at the end of the short rains in December–January and at the end of the long rains in May–June. Transect 1 crossed the plains from Seronera to Olduvai Gorge (75 km), and Transect 4 ran from Seronera to Kirawira and sampled the savanna of the western corridor (100 km). Transect 7 ran from Banagi homestead to Schaller's spring ca. 5 km south of Bologonja gate (Figure 2). The remaining transects were run on an *ad hoc* basis mainly to sample different areas of the Serengeti ecosystem, and were usually repeated only a few times, sometimes only once.

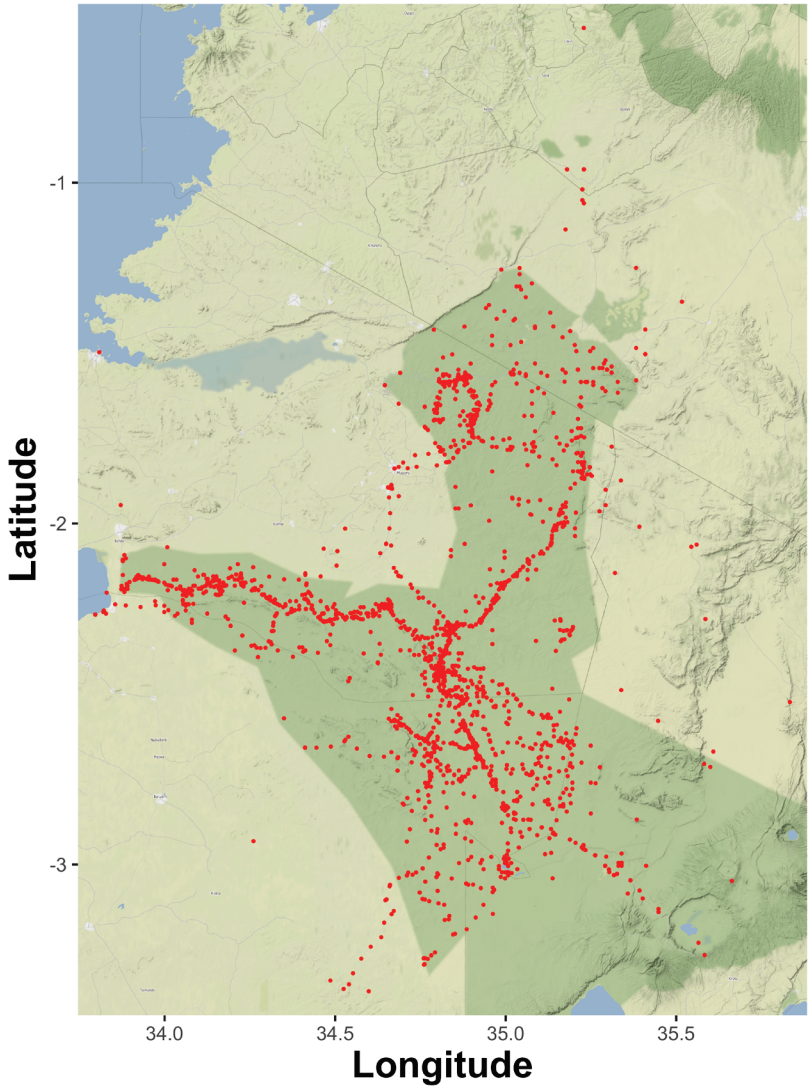

Figure 2. Bird sampling localities in Serengeti.

Sites. Sites were chosen so that they were homogeneous for a specific habitat (e.g., a stand of *V. robusta* trees, a wet grassland drainage line, or a thicket of *S. mellifera*). A total of 213 sites covering all habitats were established, with multiple replicates of sites within each habitat type. Within a habitat, sites were separated by at least 1 km. Surveys at sites include all birds seen or heard within a 50-m radius during a ten-minute count. Sites were surveyed between 7.00 and 10.00 h, and the order of site visits was rotated so that each site was surveyed at different times within this period. This is the standard method as it is described in (Bibby et al. 1992).

Points. In addition to transect and site surveys, *ad hoc* sightings of birds were recorded, usually when an unusual species appeared, or at a rare habitat such as a spring or rocky hillside. Information on location and habitat was recorded whenever appropriate Point surveys do not represent a systematic means of data collection. Point survey data are instead used to supplement the comprehensive species list of Serengeti birds.

3. Research methods

Bird taxonomy and common names are under the IOC World Bird List (v 6.4) (Gill and Donsker 2016) taxonomic nomenclature.

CLASS III. DATA SET STATUS AND ACESSIBILITY

A. Status

1. Latest update: 2019

2. Latest archive data: 2019

3. Metadata status: metadata is complete

4. Data verification: all data was quality checked (see CLASS V section B)

B. Accessibility

1. Storage location and medium: Publication as Supporting Information to this Data Paper in *Ecology*

2. Contact person: L. Francisco Henao Diaz, Biodiversity Research Centre and Department of Zoology, University of British Columbia, Vancouver, Canada fhenaod@zoology.ubc.ca

3. Copyright restrictions: None

4. Proprietary restrictions: None

CLASS IV. DATA STRUCTURAL DESCRIPTORS

A. Data set file

1. Identity: Henao Diaz and Sinclair - serengeti_birds_dataset.txt

2. Size: 66,643rows + header, 21 columns **(**17,7MB)

3. Format and storage: Text, values separated by tabs (.txt)

B. Variable information

Each row of the data set represents one observation and shows information about its abundance, taxonomy, habitat occurrence, feeding and location.

Variable Name	Variable definition	Units
transectID	Transect identification number	
year	Observation year	
month	Observation month	
day	Observation day	
individualCount	Number of observed individuals	individuals
method	Survey method	
observedDistance_meters	Observed distance	meters
macroHabitat	Macrohabitat	
IOCEnglishName	International Ornithological Congress Names in English	
family	Taxonomic family	
genus	Taxonomic genus	
specificEpithet	Taxonomic specific epithet	
scientificName	Scientific Name (Genus and specific epithet)	

scientificNameAuthorship	Scientific Name Authorship or Authority	
feedingLocation	Location of feeding	
foodType	Type of food	
residential_status*	Residential status	
locality	Distribution across habitats‡	
observer	Record observer or source	
decimalLongitude	Longitude coordinate	decimal
decimalLatitude	Latitude coordinate	decimal

* Residential status describes whether birds are residents or migrants. Residents are those present year-round, palearctic migrants come from the north and stay in Serengeti for the northern winter. Passage migrants also come from the palearctic but move through Serengeti on their way to areas further south for the winter, they are present for a short time. Some birds migrate within Africa, both from South Africa into Serengeti and northern Africa into Serengeti. These are the intra-African migrants and some stay to breed (intra-African breeders) while other come in the non-breeding season (intra-African non-breeders). Some species show more local movements changing in habitat and elevation, usually reducing elevation during the cold dry season, these are the seasonal movers. A few species occur very occasionally as wanders and these are called vagrants.

Feeding location provides a general description on where birds find their food. So, the category 'tree' means that the species would find its food in trees in general, and that entry applies to all observations of species A.

‡Ubiquitous: found everywhere or in all habitats

CLASS V. SUPPLEMENTAL DESCRIPTORS

A. Quality assurance/ quality control procedures

Independent data sets were collated matching column tables, accuracy was performed computing summary statistics and manual checks. Taxonomic rank spelling was checked and corrected with NCBI as reference database using taxize package functions. All acronyms and field notes were checked and standardized to reduce redundancy or

ambiguity. Most of the original coordinates were in the Arc 1960 system (UTM zone 36 south). They were transformed using the ellipse 'clrk80' to decimal longitude-latitude projection with WGS84 datum; coordinates were verified by eye to lie in the study area.

B. Computer programs and data-processing algorithms

This procedure was performed in R (R Core Team 2017) using dplyr , tibble (Müller and Wickham 2017), ggplot2 (Wickham 2016), ggmap(Kahle and Wickham 2013), rgdal (Bivand et al. 2018) and rworldmap (South 2011) packages.

Acknowledgments:

We appreciate the guidance of Diane Srivastava and Matt Pennell for their suggestions with data management and comments with the manuscript. Also, we are grateful to two anonymous reviewers for their comments and suggestions. This data set was compiled under the Living Data Program, an initiative of the Canadian Institute of Ecology and Evolution (CIEE). We acknowledge support for publication costs from a Cluster Grant to the Biodiversity Research Centre from the University of British Columbia, VP Research.

Literature cited

Bibby, C.J., N. D. Burgess and D. A. Hill. 1992. Bird Census Techniques. Academic Press, New York.

Bivand, R., T. Keitt, B. Rowlingson, E. Pebesma, M. Sumner, R. Hijmans, E. Rouault, and M. R. Bivand. 2018. Package 'rgdal.'

Gill, F., and D. Donsker. 2016. IOC World Bird List v 6.4. https://www.worldbirdnames.org/. doi 10.14344/IOC.ML.6.4

Kahle, D., and H. Wickham. 2013. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144–161

Müller, K., and H. Wickham. 2017. tibble: Simple Data Frames. R package version 1.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sinclair, A. R. E., editor. 2008. Serengeti III: human impacts on ecosystem dynamics. University of Chicago Press, Chicago.

Sinclair, A. R. E., K. L. Metzge, S. A. R. Mduma, and J. M. Fryxell, editors. 2015. Serengeti IV: sustaining biodiversity in a coupled human-natural system. The University of Chicago Press, Chicago.

Sinclair, A. R. E., K. L. Metzger, J. M. Fryxell, C. Packer, A. E. Byrom, M. E. Craft, K. Hampson, T. Lembo, S. M. Durant, G. J. Forrester, J. Bukombe, J. Mchetto, J. Dempewolf, R. Hilborn, S. Cleaveland, A. Nkwabi, A. Mosser, and S. A. R. Mduma. 2013. Asynchronous food-web pathways could buffer the response of Serengeti predators to El Niño Southern Oscillation. Ecology 94:1123–1130.

South, A. 2011. rworldmap: A New R package for Mapping Global Data. R Journal 3.

Wickham, H., R. Francois, L. Henry, and K. Müller. 2015. dplyr: A grammar of data manipulation. R package version 0.4 3.

Wickham, H. 2016 ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.