MS18: Selection of the monitoring and inventory projects: selection of species (groups), spatial and temporal extent

30/11/2023

Author(s): Ward Langeraert, Katelyn Faulkner, Tsungai Zengeya, Matilde Martini, Duccio Rocchini, Lissa Breugelmans, Rocìo Beatriz Cortès Lobos & Toon Van Daele

Funded by
the European Union

This project receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Table of contents

Summary

B-Cubed aims to investigate the factors that determine the reliability of models, trends, and status by comparing structured monitoring data with unstructured data. These factors are related to both data quality and species characteristics. The rise of citizen science and remote sensing technology has brought about a new era of "big unstructured data" in biodiversity conservation. The actual potential of these data in accurately describing species status and trends from aggregated occurrence cubes needs to be studied.

This report focuses on selecting structured monitoring and inventory projects that will serve as reference datasets for comparing with results from data cubes based on unstructured data from GBIF. To aid and streamline this selection, general guidelines are developed for selecting monitoring projects and their associated datasets. These guidelines are often related to a balance between sample size (enough data spread along taxonomic, spatial, and temporal dimensions) and applicability (practical use of the data).

For Flanders (the northern part of Belgium), the monitoring programs are selected by INBO (https://pureportal.inbo.be/nl/). For South Africa, the selection is conducted by SANBI (https://www.sanbi.org/).

This report is the outcome of the first milestone of task 4.5 of B-Cubed. Both INBO and SANBI provide descriptions of the selected dataset(s) for this task. Besides the selected datasets, other useful monitoring projects are listed that can be used in the B-Cubed project. The report provides the most important context and references, along with clear links to download the data.

List of abbreviations

EU	European Union
INBO	Instituut voor Natuur en Bosonderzoek (Research Institute for Nature and Forest)
SANBI	South African National Biodiversity Institute
GBIF	Global Biodiversity Information Facility
ABV	Algemene Broedvogelmonitoring Vlaanderen (Common Breeding Bird Survey Flanders)
UTM	Universal Transverse Mercator
VIS	Vis Informatie Systeem (Fish Information System)
SABAP2	Southern African Bird Atlas Programme 2
SAPIA	Southern African Plant Invaders Atlas
SANSA	South African National Survey of Arachnida

1 Introduction

The aim of this report is to examine the conditions that determine the reliability of models, trends and status by comparing aggregated cubes of structured monitoring data with cubes generated from unstructured data. These conditions relate to both data quality (Burgess et al., 2017; Isaac et al., 2014; Van Eupen et al., 2021) and species characteristics (such as abundance, detection probability, and spatial and temporal dynamics) (Callaghan et al., 2018; Kamp et al., 2016).

The emergence of technologies such as mobile phone apps, environmental DNA, automated detectors, computer vision and remote sensing technology has ushered in a new age of "big unstructured data" within biodiversity conservation. Unstructured data can be defined as *"data collected for loosely defined "observatory purposes""* (Bayraktarov et al., 2019). The true potential of these data in accurately describing national or regional species status and trends from aggregated occurrence cubes needs to be studied. In B-Cubed we will conduct a comparative analysis between (extensive) unstructured data and high-quality structured datasets based on consistent, scientifically-determined methods (Bayraktarov et al., 2019) in Belgium and South Africa.

This report outlines the selection of structured monitoring and inventory datasets which we will use as a reference to compare with results from data cubes based on unstructured data from GBIF. We first highlight common considerations and criteria with regards to the selection of the datasets so that teams in Belgium and South Africa can follow a common approach when evaluating datasets. As clear and unified reports are desired, it is key to document a common framework from the beginning. After the general considerations, a description is provided for the choice of the final structured datasets by each institute. For Flanders (northern part of Belgium), the monitoring programs are selected by INBO (https://pureportal.inbo.be/nl/). For South Africa, this will be done by SANBI (https://www.sanbi.org/). The proposed datasets in this report can also be used in other tasks within the B-Cubed project.

2 General considerations for data selection

In this section, we propose general considerations and criteria for the selection of monitoring projects and their associated datasets that will ultimately be used to evaluate the conditions for reliable national or regional species status and trends from heterogeneous data from GBIF. The following considerations are often related to a balance between sample size (enough data (spread) along taxonomic, spatial and temporal dimensions) and the applicability (practical use of the data). The aim is a selection that covers a large spectrum of cases, but leaves out the most extreme situations. The interpretation varies from dataset to dataset. Therefore, these considerations and criteria are rather guidelines than strict rules.

The criteria below are specifically made for datasets to be used in evaluating the conditions for reliability of calculations of species status and trends. Other datasets mentioned in this report might be used elsewhere in the project, and might not be subject to these guidelines. Nevertheless, potential shortcomings of these extra datasets will be discussed in further sections.

1. Species characteristics
    a. Preference for datasets with a sufficient number of different species (e.g., > 10-20 species)
    b. Preference for a spectrum of rare to common species, but no extreme cases
        i. Very rare species are likely to have only few or no unstructured observations and are often concentrated in few sites or cells (see point c)
        ii. Very common species might also have a relatively low number of observations or biased records as people tend not to record them (e.g., pigeons)
    c. Spatial distribution
        i. Preference for datasets that cover multiple habitat types
        ii. General vs. very specific habitats
    d. Preference for species that can be rather easily observed and identified species
        i. Species that are not easily observed will have too few records
        ii. Records for species that are not easily identified, or that are often confused with other species, will be unreliable
    e. Preferably representative for other regions and/or species groups
2. Structured monitoring
    a. Preference for long term monitoring (e.g., > 10 years)
        i. Long term monitoring allows capturing of increased observation effort over time in non-structured datasets
        ii. Long-term monitoring is likely to capture more accurate trends
    b. Preference for datasets with a large spatial extent
        i. Observations should cover a sufficient spatial extent to allow a comparison between different data cube grid resolutions (e.g.,1 km, 5 km, 10 km …)
    c. Preference for datasets with good knowledge of (sources) of spatial uncertainty
        i. Ideally are point coordinates with uncertainty
        ii. Data with the observation location blurred on a relatively small scale (e.g., 5 km based on grid) are less desirable.

   d. Preferably good knowledge of potential biases
      i. Geographical – data with knowledge about the over- or undersampling of some cells or sites are preferred.
      ii. Temporal – data with knowledge about the changes in sampling effort in time are preferred (more sampling over certain periods of the year/or in certain years)
   e. Good knowledge of sampling protocol, and information that could be used to estimate sampling effort
3. Unstructured monitoring (GBIF data)
   a. Spatial uncertainty
      i. Preferable not larger than the minimal size of the cubes
         → depends on spatial resolution of the cube to be analysed (see also ii
         → depends on spatial resolution/grid of the structured dataset
      ii. Suited for multiple resolutions of spatial grid of cube
         → test interaction between uncertainty of observations and different grid resolutions
   b. Knowledge of link with original data sources
      i. E.g., iNaturalist, waarnemingen.be (local website Belgium) …
      ii. Knowledge on possible erroneous records, validation procedures …
      iii. Awareness is necessary to filter out potential duplicate data published on GBIF (same data from structured monitoring and unstructured data sources)
   c. Needs to cover the same period of time as the structured dataset
   d. Ideally a proxy for sampling effort can be derived (for example: total number of observations per day per grid cell, number of observers per grid cell …)

3 Selection of monitoring projects Flanders (INBO)

3.1 Monitoring projects in Flanders

The current resources published by INBO through the Integrated Publishing Toolkit (IPT) on GBIF can be found on this website: https://ipt.inbo.be/. These datasets are useful for the B-Cubed workflow, because they are all available under permissive licences and are well described by metadata

A well-defined monitoring network in Flanders is called Meetnetten.be (https://meetnetten.be/). It consists of a series of monitoring schemes that were designed to report to Europe within the framework of the Habitat and Birds Directives, but also for other species that hold significance for Flemish nature policy. Not all datasets from Meetnetten.be are useful as some focus on uncommon species which tend to have restricted spatial distributions and/or lack unstructured data (see Section 2).

3.2 Common breeding birds as reference dataset in task 4.5

The "Common Breeding Bird Survey Flanders" (ABV) project monitors a set of ca. 100 general breeding bird species in Flanders since 2007. The monitoring protocol starts from a random sample of Universal Transverse Mercator (UTM) 1x1 km grid cells stratified according to the relative proportions of certain land uses (agriculture, urban, forest, suburban, heather and dunes, swamps and open water) (Vermeersch, 2007). The final sample consists of 1200 grid cells with a goal of 300 grids cells visited in a three-year rotation (1-300 visited in year 1, 4, 7 …, 301-600 visited in year 2, 5, 8 …, 601-900 visited in year 3, 6, 9 …). Each grid cell contains six monitoring locations from which birds are counted. Figure 1 shows a grid cell with the monitoring locations (green dots). Each grid cell is visited three times a year within fixed time periods (at least two weeks apart) (Onkelinx et al., 2023; Piesschaert et al., 2022; Vermeersch et al., 2018, 2021).

This dataset conforms well to the general considerations outlined in Section 2. First, the monitoring project includes relatively common species. Since it involves a lot of distinct species, there is still enough variation between species in terms of rarity and spatial extent to serve as an interesting test dataset. Secondly, it involves a long-term monitoring program. The sampling protocol is well-defined and the data has been analysed and maintained for an extensive period (2007 until now) (Onkelinx et al., 2023). Lastly, birds are typically a popular group for observation by the general public which means they are easily identified and reported by citizen scientists.

3.3 Description of the dataset

The dataset from the ABV project is published on GBIF as two separate datasets: from 2007-2016 (Vermeersch et al., 2021) and from 2017-2021 (when they joined Meetnetten.be) (Piesschaert et al., 2022). The data can be downloaded via the DOI of the references (Table 1). The occurrences are spatially generalised to the level of the UTM 1x1 km grid square. This means that the coordinates in the dataset of birds seen in grid cell x are set to the coordinates of the center of that grid cell. The coordinate uncertainty in meters is the radius of a circle encompassing the grid cell: for each occurrence $\sqrt{500^2 + 500^2} = 707$ meters (Figure 1). The data are not aggregated in time, exact dates are available for each observation.

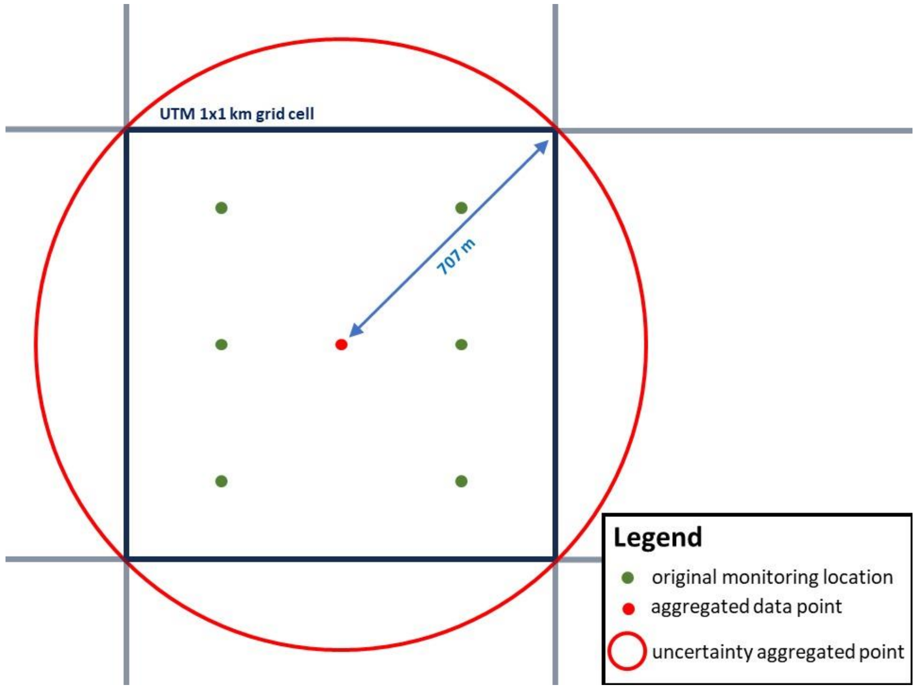

Figure 1. Visualisation of sampling protocol and data aggregation “Common Breeding Bird Survey Flanders”. Data were collected on the original monitoring locations (6 per 1 km grid square), but aggregated to a single data point in the ABV datasets. The coordinate uncertainty in meters for each datapoint in the dataset is then the radius of a circle encompassing the grid cell.

3.4 Networking with other tasks

The data presented above are essential to assess model accuracy over space and time. Commonly, unstructured data are biased and how they affect model prediction remains understudied. Using temporal moving windows and/or spatial resampling, we could assess how models are able to infer natural processes from species observations (occurrences). In particular, changes in model performance over time are often used as metrics to gain insight into a model's consistency over time under different environmental conditions. Task 4.1 deals with Habitat Suitability Modelling. The selected datasets, e.g. “Common Breeding Bird Survey Flanders” (ABV) project, might be used to evaluate error in simulate species distribution by measuring models’ performance within different time-series periods under different environmental predictors (e.g. microclimate data).

3.5 Other useful monitoring projects for B-Cubed

Some other datasets from monitoring projects in Flanders might be useful for other tasks in B-Cubed but are probably less suitable for the evaluation of reliability of species status and trends.

3.5.1 Butterfly monitoring scheme in Flanders

Data from the “Butterfly monitoring scheme in Flanders” are published on GBIF (Maes et al., 2022). These are part of a larger database on butterflies in Flanders and the Brussels capital region (Maes et al., 2016). The monitoring scheme contains transect data for butterfly species using a standardised protocol since 1991 (Maes et al., 2020, 2022).

The data are generalised to the level of the UTM 1x1 km and UTM 5x5 km depending on the species (related to rarity). This dataset might still be useful for task 4.5 after filtering out rare species with large coordinate uncertainties.

3.5.2 Fishes in inland waters in Flanders

The Fish Information System (VIS) database contains data collected to monitor the status of fishes and their habitats in Flanders. The occurrence data are published on the INBO IPT as two separate datasets (Brosens et al., 2015): “Fishes in inland waters in Flanders” (Van Thuyne, Breine, & Brosens, 2021; Van Thuyne, Breine, Verreycken, et al., 2021) and “Fishes in estuarine waters in Flanders” (Breine, Brosens, et al., 2021; Breine, Verreycken, et al., 2021). Together these datasets represent a complete overview of the distribution and abundance of fish species pertaining in Flanders from late 1992 to the end of 2018. The amount of unstructured data available is likely to be low for these species in Flanders, because there are very few other sources of fish data openly published to GBIF for Flanders.

3.6 Accessibility of the datasets

Table 1. Summary of proposed datasets from monitoring projects in Flanders.

Name	Temporal scope	Geographic scope	Taxonomic scope	Sources	Download link
Common Breeding Bird Survey Flanders	2007-2016	Flanders	Birds	(Onkelinx et al., 2023; Vermeersch et al., 2021)	https://doi.org/10.15468/xj0ikb
Common Breeding Bird Survey Flanders (post 2016)	2017-2021	Flanders	Birds	(Onkelinx et al., 2023; Piesschaert et al., 2022)	https://doi.org/10.15468/pj2v6h
Butterfly monitoring scheme in Flanders	1991-2020	Flanders	Butterflies	(Maes et al., 2016, 2020, 2022)	https://doi.org/10.15468/wwrgcd
Fishes in inland waters in Flanders	1992-2012	Flanders	Fish	(Brosens et al., 2015; Van Thuyne, Breine, Verreycken, et al., 2021)	https://doi.org/10.15468/gzyxyd
Fishes in inland waters in Flanders (post 2013)	2013-2018	Flanders	Fish	(Brosens et al., 2015; Van Thuyne, Breine, & Brosens, 2021)	https://doi.org/10.15468/klsy8u
Fishes in estuarine waters in Flanders	1992-2012	Flanders	Fish	(Breine, Verreycken, et al., 2021; Brosens et al., 2015)	https://doi.org/10.15468/estwpt
Fishes in estuarine waters in Flanders (post 2013)	2013-2018	Flanders	Fish	(Breine, Brosens, et al., 2021; Brosens et al., 2015)	https://doi.org/10.15468/jhv16z

See the data management plan of B-Cubed for more detailed information on these datasets.

4 Selection of monitoring projects South Africa (SANBI)

4.1 Monitoring projects in South Africa

In addition to various datasets that comprise of only ad hoc records that are often collected by citizen scientists (e.g., Southern African Frog Atlas Project), there are several datasets with records collected through structured surveys. These include projects (e.g., on ants or alien plants) where standardised, repeated sampling is done across transects that have often been set up to monitor biodiversity across altitudinal gradients (Bishop et al., 2014; Botes et al., 2006; Kalwij et al., 2015; Munyai & Foord, 2012). These datasets are unsuitable for task 4.5, as the amount of unstructured data available is likely to be low, but could be useful for other tasks. There are also a number of national or regional scale atlassing projects that include records from structured surveys. These include the Southern African Bird Atlas Programme 2 (SABAP2; https://sabap2.birdmap.africa/), the Southern African Plant Invaders Atlas (SAPIA), and the South African National Survey of Arachnida (SANSA). The data from some of these atlassing projects are appropriate for task 4.5 (based on the general considerations in Section 2), while the others might be useful for other tasks. These three atlassing projects are detailed below.

4.2 Southern African Bird Atlas Programme 2 as a reference dataset in task 4.5

Data for the whole of South Africa from the Southern African Bird Atlas Programme 2 (SABAP2), and higher quality data from SABAP2 for the Hessequa Atlas Area could be used for task 4.5.

4.2.1 SABAP2 – South Africa

SABAP2 is an ongoing bird monitoring project that began in 2007 (Brooks et al., 2022). The project aims to map the distribution and relative abundance of birds in South Africa, Lesotho, Botswana, Namibia, Mozambique, Eswatini, Zimbabwe, and Zambia (Brooks et al., 2022). The data are gridded and are available at a 5-minute^2 spatial resolution, which equates to a resolution of ~ 8.2 km^2 in southern Africa (Brooks et al., 2022). The data are collected by citizen scientists. Two types of records can be submitted: ad hoc records, and records following a standardised protocol (i.e., BirdMap protocol) (Brooks et al., 2022). The standardised protocol is as follows (for full details see Brooks et al., 2022):

At least two hours, but a maximum of five days, are spent recording as many different bird species in the grid cell as possible, with the end of each hour of the survey noted. Species are recorded in the order that they are seen or heard. The hour in which each species was first detected is also recorded, thus yielding approximate information on the time-to-detection for each species. The observer is expected to identify at least 90% of the species encountered, and an effort should be made to reach all or most of the representative habitats within the grid cell. These factors, and whether observations included night-time observations, are self-reported when data are submitted. Note that while some grid cells have been surveyed more than once (Figure 2), repeated sampling over time is not standardised, except for in the Hessequa Atlas Area (see Section 4.2.2).

Ad hoc records are any records that do not meet the requirements of the standardised protocol (Brooks et al., 2022). Ad hoc records can be easily filtered out of the dataset.

The dataset conforms well to the general considerations outlined in Section 2. The dataset includes distribution data for > 1000 bird species, including common and rare species, and native and common alien species. Alien species in South Africa are the focus of one of the case studies in WP6, but alien species can be filtered out of the dataset for task 4.5. Birds are observed and identified with relatively ease, and thus there should not be major issues in terms of the identifications. In addition, all records submitted to SABAP2 undergo a quality check. Records are vetted against existing records for the species in question, and if a record falls outside of the known geographical distribution of the species, it is flagged and vetted by regional atlas committees (Brooks et al., 2022). Further input from the observer may also be obtained to correct or validate the record (Brooks et al., 2022). SABAP2 includes data from 16 years of surveys, but from 2012 the ease and accuracy of data collection improved when related phone applications (Lynx BirdTicks Southern Africa (2012) and then BirdLasser (2014)) were launched (Brooks et al., 2022). There are > 10 years of this higher quality data available. The project has a large spatial extent (South Africa on its own is 1 221 037 km^2), the survey protocol is well-defined, there is good knowledge of the spatial uncertainty of the data, and information on sampling effort is readily available (see Figure 2). Biases in surveys have also been studied (Hugo & Altwegg, 2017).

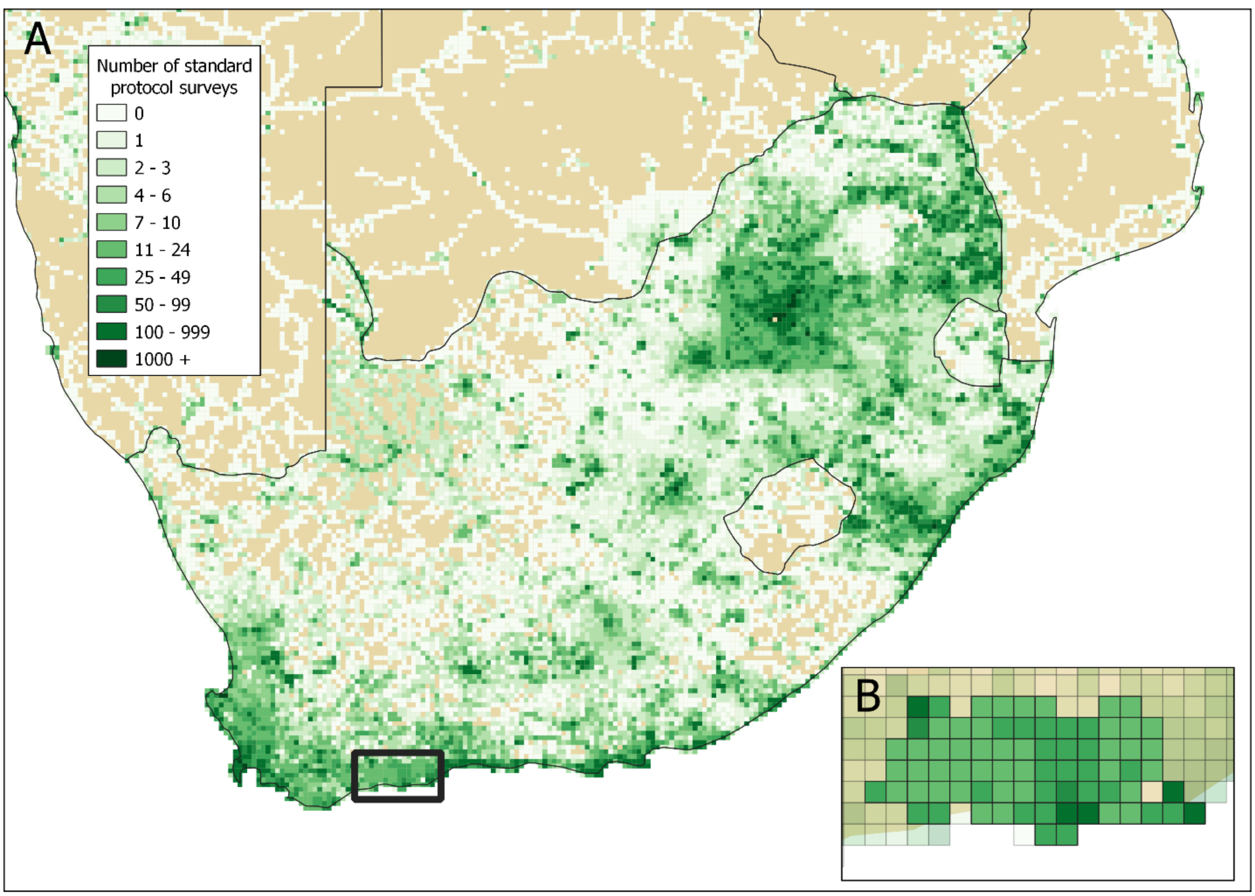

Figure 2. Number of SABAP2 standard protocol surveys per 5-minute^2 grid cell in South Africa (panel A), and in the 75 grid cells of the Hessequa Atlas Area (panel B). The inset in panel A shows the location of the Hessequa Atlas.

4.2.2 SABAP2 – Hessequa Atlas Area

The Hessequa Atlas Area is a ~110 x 55 km area near Stillbaai in the Western Cape of South Africa (Figure 2, van Rooyen, 2018). In this area of the country, ongoing directed, systematic bird surveys following the SABAP2 standard protocol have been conducted since October 2014 (van Rooyen, 2018; van Rooyen & Underhill, 2020). The area covers 75 grid cells, in which more than 300 bird species have been recorded. Although the frequency of the surveys has changed over time, every grid cell has been surveyed at least once every year by the same group of people (van Rooyen, 2018; van Rooyen & Underhill, 2020). The surveys are detailed in van Rooyen (2018), and van Rooyen and Underhill (2020), but briefly are as follows (Johan van Rooyen, pers. comm., 2023):

- 2015 and 2016: at least one standardised survey of each grid cell per year
- 2017: at least two standardised surveys of each grid cell
- 2018–2021: at least two standardised surveys of each grid cell per year, evenly spread over the seasons
- 2021–current: at least one standardised survey of each grid cell per year, distributed evenly over the seasons.

The SABAP2 data for the Hessequa Atlas Area conforms well to the general considerations outlined in Section 2 and, due to the directed, repeated sampling, represents higher quality data than the SABAP2 data for the rest of South Africa. Although the spatial extent of the data is restricted (~110 x 55 km), preliminary analyses indicate that the amount of unstructured data available in GBIF could be suffice — since 2015, 412 bird species have been recorded in the area, and for 53 of these bird species there are more than 100 records. The data from SABAP2 for the Hessequa Atlas Area could be used in task 4.5 as a high-quality data test case.

4.3 Description of the dataset

SABAP2 data are uploaded every two weeks to GBIF (Brooks & Ryan, 2023). As the data uploaded to GBIF follows GBIF standards, not all the information (e.g., sampling effort) can be obtained through GBIF (Table 2). The data are open access and can also be obtained through the website for the project https://sabap2.birdmap.africa/) and using the R package **rabm** (https://github.com/davidclarance/rabm). The records are spatially generalised to 5 km^2 grid cells. Thus, the coordinates of the records in the dataset are those at the center of the grid cell (cf. Figure 1). The data are not aggregated in time, with the start and end date of the survey available for each record. Information can be obtained on how many times a grid cell has been surveyed and when, which grid cells each species was recorded in, and how many times and when each species was found in each grid cell.

4.4 Other useful monitoring projects for B-Cubed

Other atlas projects in South Africa that are less suitable for task 4.5 might be useful for other tasks in B-Cubed

4.4.1 Southern African Plant Invaders Atlas

The Southern African Plant Invaders Atlas (SAPIA) comprises records of alien plants collected in an ad hoc manner and through standardised roadside surveys (Henderson, 2007; Henderson & Wilson, 2017). The roadside surveys were conducted between 1974 and 2018, but excluding 2017 (Lesley Henderson, pers. comm., 2023). Data from these surveys could be useful for other tasks, but are unsuitable for task 4.5. This is because the dataset focuses on alien plants in South Africa, and there is an overlap with the case study (WP6) on biological invasions in South Africa. All the data from the roadside surveys are available at a quarter-degree grid cell resolution, which is equal to 15 minute^2 or ~27.4 km^2, but many records are assigned a point locality, with notes on precision (Henderson, 2007; Henderson & Wilson, 2017). The survey protocol is well-defined (Henderson, 2007), there is some knowledge on biases (Henderson & Wilson, 2017), and survey effort could be calculated from the dataset (Lesley Henderson, pers. comm., 2023). Some of the data have been uploaded to GBIF as part of the Botanical Database of Southern Africa (BODATSA): Botanical Collections dataset (Ranwashe, 2023). The full dataset is not publicly accessible, but participants in B-Cubed can obtain it upon request.

4.4.2 South African National Survey of Arachnida

The South African National Survey of Arachnida (SANSA) comprises records of arachnids collected in an ad hoc manner and through standardised surveys (Dippenaar-Schoeman et al., 2015). The standardised surveys were largely conducted between 2008 and 2011 (Stefan Foord, pers. comm., 2023). The data are unsuitable for task 4.5, partly due to the relatively short period over which the surveys were conducted, but could be useful for other tasks. The data are available at a degree grid cell resolution (~110 km^2). The standardised surveys were conducted in 40 grid cells, mainly in the Savanna and Grassland Biomes, although isolated grid cells were also surveyed in the other biomes (Dippenaar-Schoeman et al., 2015). The survey protocol was well-defined (Haddad & Dippenaar-Schoeman, 2015), and there is relatively good knowledge of survey effort (Stefan Foord, pers. comm., 2023). As arachnids are relatively difficult to observe and identify, one constraint for SANSA is the lack of good taxonomic revisions for many of the large spider families in Africa (Dippenaar-Schoeman et al., 2015). The full dataset (including ad hoc records and those from standardised surveys) is available on GBIF, but ad hoc records would need to be filtered out.

4.5 Accessibility of the datasets

Table 2. Summary of proposed datasets from monitoring projects in South Africa.

Name	Temporal scope	Geographic scope	Taxonomic scope	Sources	Download link
Southern African Bird Atlas Programme 2	2007-2023	Southern Africa	Birds	(Brooks et al., 2022; Brooks & Ryan, 2023)	https://www.gbif.org/dataset/906e6978-e292-4a8b-9c39-adf6bb0f3323
Southern African Plant Invaders Atlas	1974-2018	Southern Africa	Alien plants	(Henderson, 2007; Henderson & Wilson, 2017; Ranwashe, 2023)	upon request
South African National Survey of Arachnida	2008-2011	South Africa	Arachnids	(Dippenaar-Schoeman et al., 2015; Haddad & Dippenaar-Schoeman, 2015)	https://doi.org/10.15468/6sqilm

See the data management plan of B-Cubed for more detailed information on these datasets.

5 Acknowledgements

We thank Thierry Onkelinx, Els Lommelen, Floris Vanderhaeghe, Hans Van Calster, Pieter Verschelde, Raïsa Carmen and Toon Westra (all INBO) for the feedback and help they provided on monitoring projects in Flanders. Lesley Henderson (SAPIA), Stefan Foord (SANSA), and Johan van Rooyen (SABAP2 - Hessequa Atlas Area) are thanked for information they provided on monitoring projects in South Africa. Finally, our gratitude goes to Quentin Groom and Shawn Dove for their helpful comments to improve the text.

6 References

Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E. L., Nguyen, H. A., McRae, L., Possingham, H. P., & Lindenmayer, D. B. (2019). Do Big Unstructured Biodiversity Data Mean More Knowledge? *Frontiers in Ecology and Evolution*, *6*, 239. https://doi.org/10.3389/fevo.2018.00239

Bishop, T. R., Robertson, M. P., van Rensburg, B. J., & Parr, C. L. (2014). Elevation–diversity patterns through space and time: Ant communities of the Maloti-Drakensberg Mountains of southern Africa. *Journal of Biogeography*, *41*(12), 2256–2268. https://doi.org/10.1111/jbi.12368

Botes, A., McGeoch, M. A., Robertson, H. G., van Niekerk, A., Davids, H. P., & Chown, S. L. (2006). Ants, altitude and change in the northern Cape Floristic Region. *Journal of Biogeography*, *33*(1), 71–90. https://doi.org/10.1111/j.1365-2699.2005.01336.x

Breine, J., Brosens, D., & Desmet, P. (2021). *VIS - Estuarine monitoring in Flanders, Belgium (post 2013)* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/JHV16Z

Breine, J., Verreycken, H., De Boeck, T., Brosens, D., & Desmet, P. (2021). *VIS - Fishes in estuarine waters in Flanders, Belgium* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/ESTWPT

Brooks, M., Rose, S., Altwegg, R., Lee, A. T., Nel, H., Ottosson, U., Retief, E., Reynolds, C., Ryan, P. G., Shema, S., Tende, T., Underhill, L. G., & Thomson, R. L. (2022). The African Bird Atlas Project: A description of the project and BirdMap data-collection protocol. *Ostrich*, *93*(4), 223–232. https://doi.org/10.2989/00306525.2022.2125097

Brooks, M., & Ryan, P. (2023). *Southern African Bird Atlas Project 2* [dataset]. FitzPatrick Institute of African Ornithology. https://www.gbif.org/dataset/906e6978-e292-4a8b-9c39-adf6bb0f3323

Brosens, D., Breine, J., Van Thuyne, G., Belpaire, C., Desmet, P., & Verreycken, H. (2015). VIS – A database on the distribution of fishes in inland and estuarine waters in Flanders, Belgium. *ZooKeys*, *475*, 119–145. https://doi.org/10.3897/zookeys.475.8556

Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., HilleRisLambers, J., Tewksbury, J., & Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, *208*, 113–120. https://doi.org/10.1016/j.biocon.2016.05.014

Callaghan, C. T., Martin, J. M., Major, R. E., Kingsford, R. T., Callaghan, C. T., Martin, J. M., Major, R. E., & Kingsford, R. T. (2018). Avian monitoring – comparing structured and unstructured citizen science. *Wildlife Research*, *45*(2), 176–184. https://doi.org/10.1071/WR17141

Dippenaar-Schoeman, A. S., Haddad, C. R., Foord, S. H., Lyle, R., Lotz, L. N., & Marais, P. (2015). South African National Survey of Arachnida (SANSA): Review of current knowledge, constraints and future needs for documenting spider diversity (Arachnida: Araneae). *Transactions of the Royal Society of South Africa*, *70*(3), 245–275. https://doi.org/10.1080/0035919X.2015.1088486

Haddad, C. R., & Dippenaar-Schoeman, A. S. (2015). Diversity of non-acarine arachnids of the Ophathe Game Reserve, South Africa: Testing a rapid sampling protocol. *Koedoe*, *57*(1), 1255. https://doi.org/10.4102/koedoe.v57i1.1255

Henderson, L. (2007). Invasive, naturalized and casual alien plants in southern Africa: A sum-mary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, *37*(2), 215–248. https://doi.org/10.4102/abc.v37i2.322

Henderson, L., & Wilson, J. R. U. (2017). Changes in the composition and distribution of alien plants in South Africa: An update from the Southern African Plant Invaders Atlas. *Bothalia*, *47*(2). https://doi.org/10.4102/abc.v47i2.2172

Hugo, S., & Altwegg, R. (2017). The second Southern African Bird Atlas Project: Causes and consequences of geographical sampling bias. *Ecology and Evolution*, *7*(17), 6839–6849. https://doi.org/10.1002/ece3.3228

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254

Kalwij, J. M., Robertson, M. P., & van Rensburg, B. J. (2015). Annual monitoring reveals rapid upward movement of exotic plants in a montane ecosystem. *Biological Invasions*, *17*(12), 3517–3529. https://doi.org/10.1007/s10530-015-0975-3

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, *22*(10), 1024–1035. https://doi.org/10.1111/ddi.12463

Maes, D., Piesschaert, F., Brosens, D., & Desmet, P. (2022). *Butterfly monitoring scheme in Flanders, Belgium* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/WWRGCD

Maes, D., Piesschaert, F., T'jollyn, F., & Van Dyck, H. (2020). *Dagvlindermonitoring in Vlaanderen: 1991-2019* (Rapporten van Het Instituut Voor Natuur- En Bosonderzoek 25). Instituut voor Natuur- en Bosonderzoek. https://doi.org/10.21436/inbor.18344476

Maes, D., Vanreusel, W., Herremans, M., Vantieghem, P., Brosens, D., Gielen, K., Beck, O., Van Dyck, H., Desmet, P., & Natuurpunt, V. (2016). A database on the distribution of butterflies (Lepidoptera) in northern Belgium (Flanders and the Brussels Capital Region). *ZooKeys*, *585*, 143–156. https://doi.org/10.3897/zookeys.585.8019

Munyai, T. C., & Foord, S. H. (2012). Ants on a mountain: Spatial, environmental and habitat associations along an altitudinal transect in a centre of endemism. *Journal of Insect Conservation*, *16*(5), 677–695. https://doi.org/10.1007/s10841-011-9449-9

Onkelinx, T., Vermeersch, G., & Devos, K. (2023). *Trends op basis van de Algemene Broedvogelmonitoring Vlaanderen (ABV): Technisch achtergrondrapport voor de periode 2007-2022* (Rapporten van Het Instituut Voor Natuur- En Bosonderzoek 2023). Instituut voor Natuur- en Bosonderzoek. https://doi.org/10.21436/inbor.89419879

Piesschaert, F., Vermeersch, G., Brosens, D., Westra, T., Desmet, P., Feys, S., Van De Poel, S., & Pollet, M. (2022). *ABV - Common breeding birds in Flanders, Belgium (post 2016)* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/PJ2V6H

Ranwashe, F. (2023). *Botanical Database of Southern Africa (BODATSA): Botanical Collections* [dataset]. South African National Biodiversity Institute. https://doi.org/10.15468/2AKI0Q

Van Eupen, C., Maes, D., Herremans, M., Swinnen, K. R. R., Somers, B., & Luca, S. (2021). The impact of data quality filtering of opportunistic citizen science data on species distribution model performance. *Ecological Modelling*, *444*, 109453. https://doi.org/10.1016/j.ecolmodel.2021.109453

van Rooyen, J. A. (2018). Systematic atlasing in Hessequa—Moving from mapping to monitoring. *Biodiversity Observations*, *9*(10), 1–13. https://doi.org/10.15641/bo.v9i0.508

van Rooyen, J. A., & Underhill, L. G. (2020). Systematic atlasing in Hessequa – Report on the first cycle of seasonal monitoring. *Biodiversity Observations*, *11*(3), 1–14. https://doi.org/10.15641/bo.933

Van Thuyne, G., Breine, J., & Brosens, D. (2021). *VIS - Reference freshwater monitoring in Flanders, Belgium (post 2013)* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/KLSY8U

Van Thuyne, G., Breine, J., Verreycken, H., De Boeck, T., Brosens, D., & Desmet, P. (2021). *VIS - Fishes in inland waters in Flanders, Belgium* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/GZYXYD

Vermeersch, G. (2007). *Sampling framework for the common breeding bird survey in Flanders, Belgium* [dataset]. Zenodo. https://doi.org/10.5281/zenodo.10103472

Vermeersch, G., Anselin, A., Herremans, M., & Brosens, D. (2021). *ABV - Common breeding birds in Flanders, Belgium* [dataset]. Research Institute for Nature and Forest (INBO). https://doi.org/10.15468/XJ0IKB

Vermeersch, G., Ledegen, H., & Feys, S. (2018). *Methodehandleiding bij het project Algemene Broedvogelmonitoring Vlaanderen (ABV)*. https://doi.org/10.21436/inbor.15674942