

# Code to calculate protected area (PA) and other effective area-based conservation measures (OECMs) coverage of Key Biodiversity Areas (KBAs)

Version 3.0 – Instructions and recommendations

May 2024

Tom Scott ([tom.scott@birdlife.org](mailto:tom.scott@birdlife.org))

## Background

- The code detailed here is the most recent version of scripts originally developed in 2016 as project P90001: Indicator-related work to a) code to calculate national Red List Indices (RLIs), and b) calculate and provide updated data on the coverage by protected areas of marine, terrestrial and freshwater Key Biodiversity Areas.
- This analysis:
  - a. Overlaps KBAs with PAs and OECMs with the option to generate separate coverage statistics for these.
  - b. Assigns random years to those PAs/OECMs with no reported year of designation and accounts for these in the reported confidence intervals.
  - c. Generates initial outputs containing the coverage of KBAs by PAs/OECMs for every year in which a new overlapping PA/OECM was established.
  - d. Produces specific tabular and graphical outputs for both the IBAT ([Integrated Biodiversity Assessment Tool \(IBAT\) \(ibat-alliance.org\)](https://ibat-alliance.org/)) country profiles and Sustainable Development Goals indicator reporting in relation to the coverage of marine ([Metadata-14-05-01](#)), terrestrial/freshwater ([Metadata-15-01-02](#)) and mountain ([Metadata-15-04-01](#)) KBAs.
- These codes are written in R language and are stored at: [BirdLifeInternational/kba-overlap: spatial overlap codes for the SDG analyses \(github.com\)](https://github.com/BirdLifeInternational/kba-overlap)

## Code Structure

Unlike previous versions, this workflow is now consolidated into one “master” script which calls on functions coded in a second script. This allows for a more succinct and user-friendly process. This master script has 6 parts:

- 1) Initial set up and data loading/cleaning
  - a) Set working directory and other basic parameters
  - b) Call functions from second script
  - c) Load required packages
  - d) Build all necessary file directories
  - e) Read in all raw tabular data
- 2) Overlap analysis
  - a) Read in all raw spatial data
  - b) Clean spatial datasets
  - c) Set whether analysis should disaggregate PAs and OECMs to give separate coverage values

- d) Complete overlap
- 3) Randomising national data for IBAT
  - a) Rescale results
  - b) Perform randomisation for PAs with missing years
- 4) Summary tables for IBAT and KBA internal reporting
  - a) Summarise coverage per KBA
  - b) Summarise coverage of global KBA network
- 5) Graphical outputs for IBAT country profiles
  - a) Creation of IBAT figure 6 plots and legends
  - b) Creation of IBAT figure 5 plots and legends
  - c) Creation of IBAT table 7
  - d) Summarise coverage per country
- 6) Producing outputs for SDG reporting
  - a) Repeat part 3 for regional disaggregations
  - b) Format output into required SDG reporting layout

Whilst these parts do have to be run in sequence, they are not dependent on each other to be run in the same R session i.e. parts 2 and 3 could be run one day and then part 4 could be run the next day without having to reload any of part 2 or 3. **Is it essential to run part 1 every time the script is opened.**

## Required Input Files

- Spatial:
  - KBA shapefile containing boundaries for all confirmed KBAs (available at: [Key Biodiversity Areas GIS Data Request](#)). Ensuring that SitRecID column is named as such.
  - WDPA/WDOECM layer containing a cleaned spatial output provided by WCMC, for detail of this cleaning see metadata files linked above (4.c. Method of computation).
- Tabular (.csv format – all column headings much match exactly):
  - Classification of KBAs into marine, terrestrial, freshwater and mountain.

SitRecID	marine	terrestrial	Freshwater	mountain
11	Y	Y	N	N
12	Y	Y	N	N
17	N	Y	N	Y
18	Y	Y	N	N
19	N	Y	N	Y
21	N	Y	N	Y
24	N	Y	N	N
25	N	Y	N	Y

- ISO table of regional groupings. NOTE – this is subject to annual change as has to be updated to reflect any changes made by UNSD in their data request to the composition of various regional groupings if part 6 is to be run. Screenshot for visualisation but more columns for each regional grouping.

countryname	ordem	country_SIS	region	ISO3	ISO_BL	ISO_SDG	uname_BL	uname_SDG	Region_name	etc
Afghanistan	66	Afghanistan	West and	AFG	AFG	AFG	Afghanistan	Afghanistan	Region 1	...

			Central Asia							
Akrotiri and Dhekelia	256	Akrotiri and Dhekelia	Europe	XAD	CYP	CYP	Cyprus	Cyprus	Region 1	...
Aland	248	Aland Islands	Europe	ALA	ALA	FIN	Finland	Finland	Region 2	...

- Details of KBAs with no boundaries (only needed if producing IBAT outputs).

SitRecID	COUNTRY	ISO	percPA	terrestrial	Freshwater	marine	mountain	PA Coverage	WDPA_type
217	Lithuania	LTU	0	1	0	0	0	not assessed	OECMs
217	Lithuania	LTU	0	1	0	0	0	not assessed	Protected Areas
217	Lithuania	LTU	0	1	0	0	0	not assessed	Protected Areas and OECMs
485	Iceland	ISL	0	1	0	1	0	not assessed	OECMs

- Site status as KBA/IBA/AZE (only needed if producing IBAT outputs).

SitRecID	IbaStatus	KbaStatus	AzeStatus
49058	confirmed	confirmed	
49056		confirmed	
49054		confirmed	
49053		confirmed	confirmed

- UNSD regional groupings table lifted from annual data request (only needed if producing SDG outputs).

M49 Code	ISO Code	Country (or Area) Name	M49 Code(region)	Region Name	Reference Area Type [i.e. global, SDG groupings, MDG groupings, etc.]	2024 Submission [if data submission is required]
4	AFG	Afghanistan	1	World	1.0-Global	Required
8	ALB	Albania	1	World	1.0-Global	Required
12	DZA	Algeria	1	World	1.0-Global	Required

## Output Files

This code produces many output files which broadly fall into four classes:

1. General reporting
  - a. pa\_coverage\_per\_kba.csv – total protected area and OECM coverage for each KBA.
  - b. percentage\_coverage\_by\_country\_data\_global\_summary.csv – global summary of coverage, used by the KBA secretariat for annual reporting.
2. IBAT (to populate country profiles)

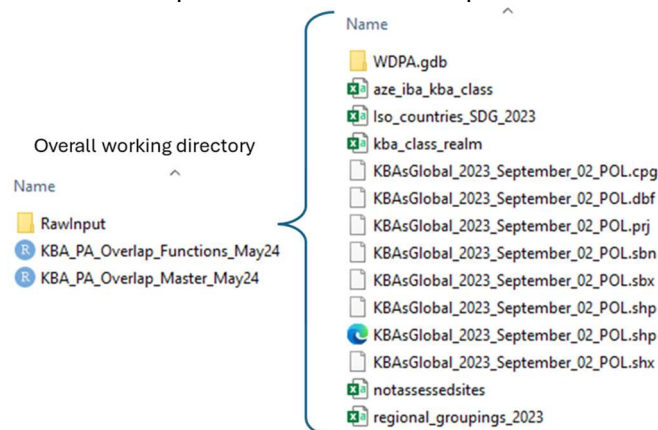
- a. IBAT figure 5 per country – for AZEs, IBAs and KBAs the reported coverage of sites classified into “complete” and “none” coverage with thresholds of  $\geq 98\%$  and  $\leq 2\%$  respectively with any remaining sites classed as having “partial” coverage. This is saved into IBATFig5 \*year\* folder.
  - b. IBAT\_fig5\_captions.csv – saved in top level of directory.
  - c. IBAT figure 6 per country - for all, marine, terrestrial, mountain and freshwater KBAs, change in mean PA coverage over time. This is saved into final graphs \*year\* folder.
  - d. IBAT\_fig6\_captions.csv – saved in top level of directory.
  - e. table7\_national\_kba\_totals.csv – count of AZEs, IBAs and KBAs for each country, saved in top level of directory.
  - f. percentage\_coverage\_by\_country\_data.csv – a reformatted version of the data seen in figure 6, saved in top level of directory.
3. SDG Reporting
- a. Output files named and formatted for SDG reporting. These are saved into the output tables \*year\* folder.
    - i. 119-14.5.1-2999-ER\_MRN\_MPA-4532.csv
    - ii. 119-15.1.2-2837-ER\_PTD\_FRHWRT-5654.csv
    - iii. 119-15.1.2-2839-ER\_PTD\_TERR-5654.csv
    - iv. 119-15.4.1-2838-ER\_PTD\_MTN-5654.csv
4. Intermediate files – used to produce outputs above
- a. \*countryname\*.csv – output from spatial overlap saved country by country, this allows for saved into files\_country\_\*year\* folder. SitRecID = KBA siteID, kba = kba area, ovl = area of overlap, year = year of protected area designation (0 if no protected areas overlap the kba), random = whether the designation year of the protected area was randomly assigned, nPAs = number of protected areas, percPA = percentage overlap of the kba with protected areas that year, ISO = ISO3 code, COUNTRY = country KBA is within, WDPA\_type = whether protected areas, OECMS or both. Note that when a kba has more than one year the overlap (ovl) and percentage overlap (percPA) refer to the increased overlap compared to the previous year (it is not total for that year); you would have to sum all overlaps across all years to determine the protected area overlap of that kba now.
  - b. Finaltab\_year.csv – raw output from spatial overlap combined into one file.
  - c. Input data for R \*regional grouping\*.csv – input data used for randomisation. csv files for each regional grouping, including each kba that falls within that grouping and its change in protected area coverage and the year its coverage changed. Saved into input tables \*year\* folder.
  - d. In\_out\_files\_\*year\*.csv - list of input and output file names for regional groupings which will be used in the randomisation step, saved in top level of directory.
  - e. Output data for R \*regional grouping\_system\*.csv – for all, terrestrial, marine, mountain and freshwater KBAs, with a kba-pa overlap value for each year from 1900 to the present year with confidence intervals, and how many kbas per regional grouping are fully protected. These are saved into the output tables \*year\* folder. year = year, 95Cllow/mid/high = 95% lower/middle/upper confidence interval, Count = the number of kbas completely covered by protected areas, Percentage = percentage of protected areas full covered, Percentage\_Area = mean percentage coverage of kbas by protected areas – all three of these are per regional grouping, region = regional grouping, code =

regional grouping category applied (e.g. country, ISO\_BL, etc), sset = system, WDPA\_type = whether protected areas, OECMS or both.

- f. SDG PA coverage \*system\* KBAs by country region \*year\*.csv – a combination and reformatting of the relevant “output data for R” files for SDG reporting, saved into the output files folder.

## Code Instructions

1. Initial work folder – your starting directory should be set up as the image below, with the master and functions R scripts in a folder with a sub folder called RawInput with the tabular and spatial input files above stored in. These input files do not need specific names and those of spatial nature do not need to be specific of the formats highlighted below, these can be inputted into the code in part 1.



- a.
2. Open master script – follow TODO markers in the script, anything labelled with one of these will need some action before running the respective lines. The active lines can be run as they are, nothing needs to be added within the parentheses for each one.
3. Run script from part 1. As mentioned above each part is dependent on the previous having been run within the working directory at some point however not necessarily in the same R session.

## Acknowledgements

This code was originally written by Maria Dias in 2016. Lizzie Pearmain and Ashley Simkins revised this in 2020 which was used as the basis for this current version (3.0).