Species Observations Explorer

Biodiversity databases make temporally and spatially extensive primary biodiversity data (i.e. species observations) available to a wide range of users. However, because of the intrinsic nature of most of the observations (e.g. opportunistic or non-systematic and presence-only), biodiversity datasets have considerable limitations including: sampling bias in favour of recorder distribution, lack of assessment of sampling effort, and lack of coverage of the distribution of all organisms. This interactive application maps where, and for which taxonomic groups, there are enough species observations for ecological analyses in applied or theoretical contexts. It is a report of the quality of the data that goes beyond the data abundance, to inform you about the sampling effort and the relevance of Zeros (lack of data).

It uses the half-ignorance algorithm [[1]](https://greensway.shinyapps.io/ignapp-v08/_w_f7cc4ef3/#footnote-1) to map the ignorance (i.e. bias and lack of sampling effort) inherent to the species observations. This approach represents the data into a scale 0 to 1 (0 being a theoretical absolute credibility in the data and 1 being absolute ignorance). With the algorithm settings you decide when the lack of data for one species is equally likely due to absence of the species or to absence of observers (i.e. O0,5). Any further data in a cell builds up the trust in the data (i.e. real absences of species).

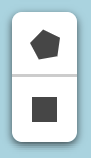
The rationale assumes that species groups share similar bias. Observations are reported by people with varied field skills and accuracy. However, observers are assumed to be fond of or specialist on one or more taxonomic groups (e.g. families, orders or even classes), rather than on individual species. Since it is likely that an entire group of species observed by similar methods (henceforth, a reference taxonomic group or **RTG**) will share similar bias, it is appropriate to use species' groups as a surrogate for sampling effort. Therefore, it is straightforward to assume that the lack of reports of any species from the RTG e.g. birds is likely due to a lack of ornithologists, rather than to the total absence of birds. The inverse logic also holds true. That is, the larger the number of observations of species from the RTG in a grid cell, the more likely it is that the lack of reports of a particular species reflects a true absence of a focal species from the grid cell. **So, even if you are interested in data for one particular species, you have to think in terms of its RTG**. This dynamic application allows you to explore any RTG at any spatial and temporal resolution. Currently it accesses data stored in the Global Biodiversity Information Facility [(GBIF)](http://www.gbif.org/), but it also allows filtering by institutions that provide data to GBIF.

You can read more about Ignorance Scores [here](https://devpost.com/software/mapping-ignorance-in-space-and-time-evtf4a) and here (https://devpost.com/software/mapping-ignorance-in-space-and-time-evtf4a)

HOW TO USE IT

On the first tab "**Map**" you can explore the spatial bias of the selected reference taxonomic group, in the selected time frame. On the map you will see in a scale of white-green the density of the data for the selected RTG. Beware that this is ALL the data that is available, but YOUR search will also depend on other search options like time window and the Basis of Record. This background data is only shown to give you some information about the presence and density of the data but is not a quality report. Your selected data will show only after you have finished your search, and the density map will be toggled off.

To perform a search, do like this:

1. Draw a polygon over the area you are interested in, using any of the shape-buttons  on the top-left. If you chose the rectangle-shape you can click in one corner and then drag it out to appropriate size, using the pentagram-shape you can determine the shape yourself by creating points/vertices on each click with the mouse.
2. On the panel to the left there are three tabs. The first one is called "**Grid Options**". There you can choose the width of the grid cells. If you want your grid to cover the full extent of your selected study area check the "Buffer" check-box. If you want squared grid cells, uncheck the "Hexagonal grid" check-box
3. Choose your Reference Taxonomic Group from the list. The groups shown in the lists are common taxas, but you can override this option in the tab “Search Options” (see below).
4. Click the "Grid" button  (on the "**Grid Options**" tab)
5. Do you like the grid? Click the "Search" button .
6. Don't you like it? Change the options and click "Grid" again, or start over by clicking on "Clear" .
7. Under " Ignorance Scores Assumptions" you can specify the parameters for the half-ignorance algorithm. The O0.5 parameter defines the number of observations required per grid cell to decrease the ignorance score (IS) to 0.5 (see the reference above). You can calculate IS over raw observations (using the O0.5RTG), over *Observation Indices* (i.e. the number of observations for the RTG divided by the observed number of species observed, per grid cell; using the O0.5 per species), or you can choose the combined Ignorance Score, that uses both parameters.
8. More search options are available if you look into the "**Search Options**" tab.
   1. Specifying a taxonomic ID will override the selected RTG. You will need to know the taxonomic ID to use this option, this can be found at <https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c> .
   2. You can determine the period (years) of observations to be included by dragging the two start-stop points, data from 1900 to 2017 is available (although the bar goes to 2020).
   3. The basis of records to be included can be determined, e.g. if you do not which to include fossil specimen. Select more classes using the Shift + Ctrl/Cmd keys.
   4. Also the country of observation, publishing country and publishing organization can be chosen.
9. When you have changed any search parameters, click the “Search” button again to see the new result.
10. Finally, after the search is performed, you can download the grid (.SHP file) and the data table (as .CSV file) by clicking "Download" button  on the "**Download**" tab. So long, there are only a few Coordinate Reference Systems to choose from, but more are coming on demand. Files will be compressed in a .TAR file (for compatibility across platforms) in your default Download folder. You can open the file with [7zip](http://www.7-zip.org/).

The search will take a while, depending on the number of cells, but it is rather quick.

On the second tab "**Data**" the data obtained for each grid cell is plotted and displayed as a table.

Author: Alejandro Ruete PhD, 2017  
Greensway AB. Uppsala, Sweden.

E-mail: [analys@greensway.se](mailto:analys@greensway.se)

[1] Ruete A. 2015. Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. Biodiversity Data Journal 3:e5361 [(article)](https://bdj.pensoft.net/articles.php?journal_name=bdj&id=5361" \t "_blank).

*Keywords*: citizen-science data, open-access biodiversity database, presence-only data, primary biodiversity data, sampling effort, spatial bias, species distribution model, taxonomic bias, temporal bias