

Recreate UK Biodiversity Trends Paper

Charlotte L. Outhwaite

08/06/2019

Contents

Set up R environment	1
Generation of group level posterior sample datasets	2
Recreating Figure 1	3
Recreating Figure 2	4
Recreating Figure 3	5
Recreating Figure 4	6
Recreating Figure 5	7
Calculating group and taxa level trends	8
Recreating Supplementary Figure 1	9
Recreating Supplementary Figure 2	10

This vignette uses the UKBiodiversity R package to recreate the figures generated in the paper “Complexity of biodiversity change revealed through long-term trends of invertebrates, bryophytes and lichens” by Outhwaite et al.

Set up R environment

First you will need to download the dataset used. These data are freely available from the EIDC [repository](#) and are described in full in the accompanying data paper: Outhwaite, C. L., Powney, G. D., August, T. A., Chandler, R. E., Rorke, S., Pescott, O. L., ... Isaac, N. J. B. (in review). Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK (1970 – 2015). *Scientific Data*.

To download the required datasets you will need to register with the website first, then you can follow the instructions to download the data. The complete dataset zip file should be downloaded.

In the repository you will find the following items:

- POSTERIOR_SAMPLES: zip file containing 1 file per species of the posterior samples of average occupancy.
- SUMMARY_TABLES: zip file containing 1 file per species of occupancy model output summary tables.
- Dataset_Information: csv file containing input dataset metadata.
- Species_Names: csv file containing information on species names.
- Species_Trends: csv file containing all estimates of species trends as annual percentage growth rates.

For this analysis you will need to unzip the POSTERIOR_SAMPLES folder.

Set your working directory so that the POSTERIOR_SAMPLES folder is a sub-folder within it, you can do this using the function `setwd()`.

You will also need to download the R package called UKBiodiversity from Github. How to do this is detailed below.

First, let's take a look at the POSTERIOR_SAMPLES data files.

```
# Where are the posterior sample files?
# These should be in the named subfolder of your working directory.
datadir <- paste0(getwd(), "/POSTERIOR_SAMPLES")

# List the files
files <- list.files(datadir)

# How many files are there?
length(files) # There should be 5293, one for each species.

## [1] 5293

# Take a look at the file list
head(files)

## [1] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_aquilonia.csv"
## [2] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_cunicularia.csv"
## [3] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_fusca.csv"
## [4] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_lemanii.csv"
## [5] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_lugubris.csv"
## [6] "AnnualSpeciesOccupancy_PosteriorSamples_Ants_Formica_picea.csv"

# Download the UKBiodiversity R package from GitHub,
# to do this you will first need the devtools package
library(devtools)

# Install the R package from GitHub
install_github("CharlieOuthwaite/UKBiodiversity", quiet = TRUE)

# Load the package
library(UKBiodiversity)
```

This process should have downloaded the required packages for this analysis.

Generation of group level posterior sample datasets

For some of the analyses we will need a matrix containing all the posterior samples of the species within each of the major groups and for each of the taxonomic level groups. In the repository, the posterior samples for each species are saved as individual csv files, one for each species. Use the `combine_posteriors` function to make the required combinations of the posteriors by major group and taxa. The function will save the outputs as .rdata files in a sub-folder of the directory that you specify, within a group specific sub-folder. Use the `group_level` argument to select either "major_group" or "taxa" level datasets.

Note that these functions can take a long time to run (estimated run time below) and may have high memory requirements.

```
# Run the combine_posteriors function providing a data directory and an output directory.
# select the group level you wish to generate datasets for, either "major_group" or "taxa".
# If status is TRUE, progress will be printed within the console.

## First for major groups.
# Runtime of this function: 67 mins (when status = FALSE)
```

```

combine_posteriors(datadir = paste0(getwd(), "/POSTERIOR_SAMPLES"),
                  outdir = getwd(),
                  group_level = "major_group",
                  status = FALSE)

# Take a look at the outputs in the function generated subdirectory
files <- list.files(paste0(getwd(), "/MajorGroups"), pattern = "posterior_samples")
head(files)

## [1] "ALL_posterior_samples_national.rdata"
## [2] "FRESHWATER_SPECIES_posterior_samples_national.rdata"
## [3] "LOWER_PLANTS_posterior_samples_national.rdata"
## [4] "TERRESTRIAL_INSECTS_posterior_samples_national.rdata"
## [5] "TERRESTRIAL_NONINSECT_INVERTS_posterior_samples_national.rdata"

# A file has been created for each of the major groups and for all species.

## Then for taxa.
# Runtime for this function: ~ 20 minutes
combine_posteriors(datadir = paste0(getwd(), "/POSTERIOR_SAMPLES"),
                  outdir = getwd(),
                  group_level = "taxa",
                  status = FALSE)

# Take a look at the outputs in the function generated subdirectory
files <- list.files(paste0(getwd(), "/Taxa"), pattern = "posterior_samples")
head(files)

## [1] "Ants_posterior_samples_national.rdata"
## [2] "AquaticBugs_posterior_samples_national.rdata"
## [3] "Bees_posterior_samples_national.rdata"
## [4] "Bryophytes_posterior_samples_national.rdata"
## [5] "Caddisflies_posterior_samples_national.rdata"
## [6] "Carabids_posterior_samples_national.rdata"

# A file has been created for each taxonomic group

```

These group level files are required for the following functions. Files must be maintained within the folder they are placed by the function.

For the analyses carried out here, rove beetles are excluded since they do not have data from 1970 to 1979. Therefore, there are 5,214 species included in the analyses carried out here. 79 Rove Beetle species are removed.

Recreating Figure 1

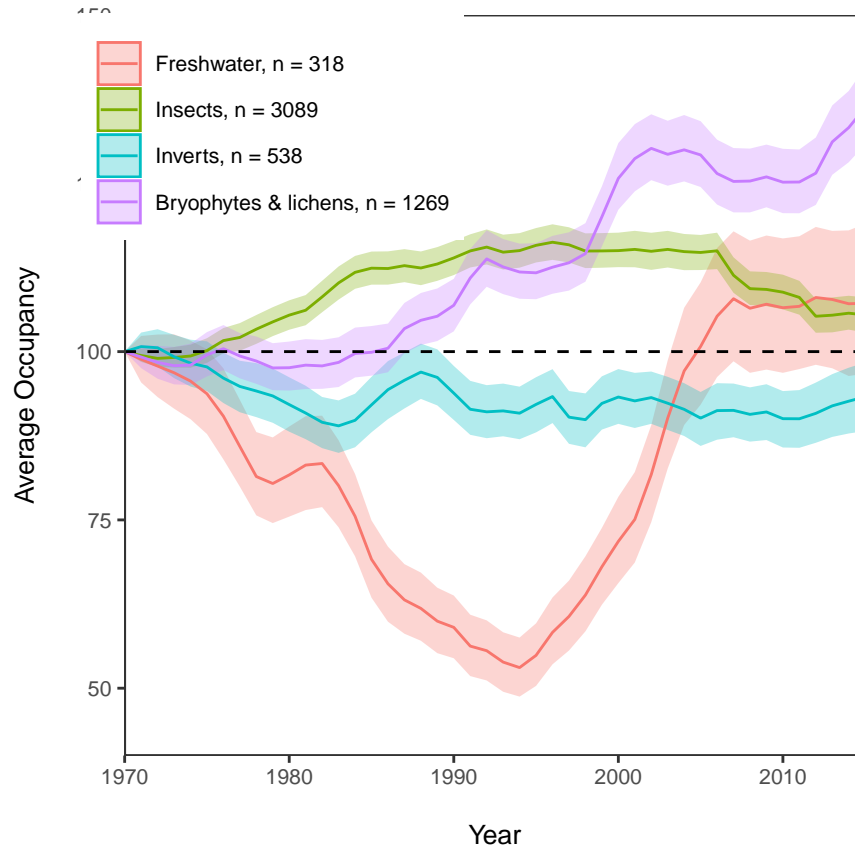
Figure 1 within the paper presents indicators of average occupancy over time for the four major groups.

Use the `generate_fig1` function to determine the group level average change associated indicator values and the plot of Figure 1. The `generate_fig1` function will save csv files of the indicator posterior values, the average annual indicator estimates and associated credible interval values, and a pdf file of the plot within a sub-directory of your specified folder called “geomeans”.

This function will only work if outputs have already been produced from the “major_group” specification of the `combine_posteriors` function. These should be in a sub-folder called “MajorGroups” in the `outdir` that you specified within that function.

```
# Use the generate_fig1 function to generate figure 1 nad associated outputs.
# Runtime of this function: 6 mins
generate_fig1(postdir = paste0(getwd(), "/MajorGroups"),
              status = FALSE,
              save_plot = TRUE,
              interval = 95)
```

```
## [1] "Warning: overwriting existing files"
```



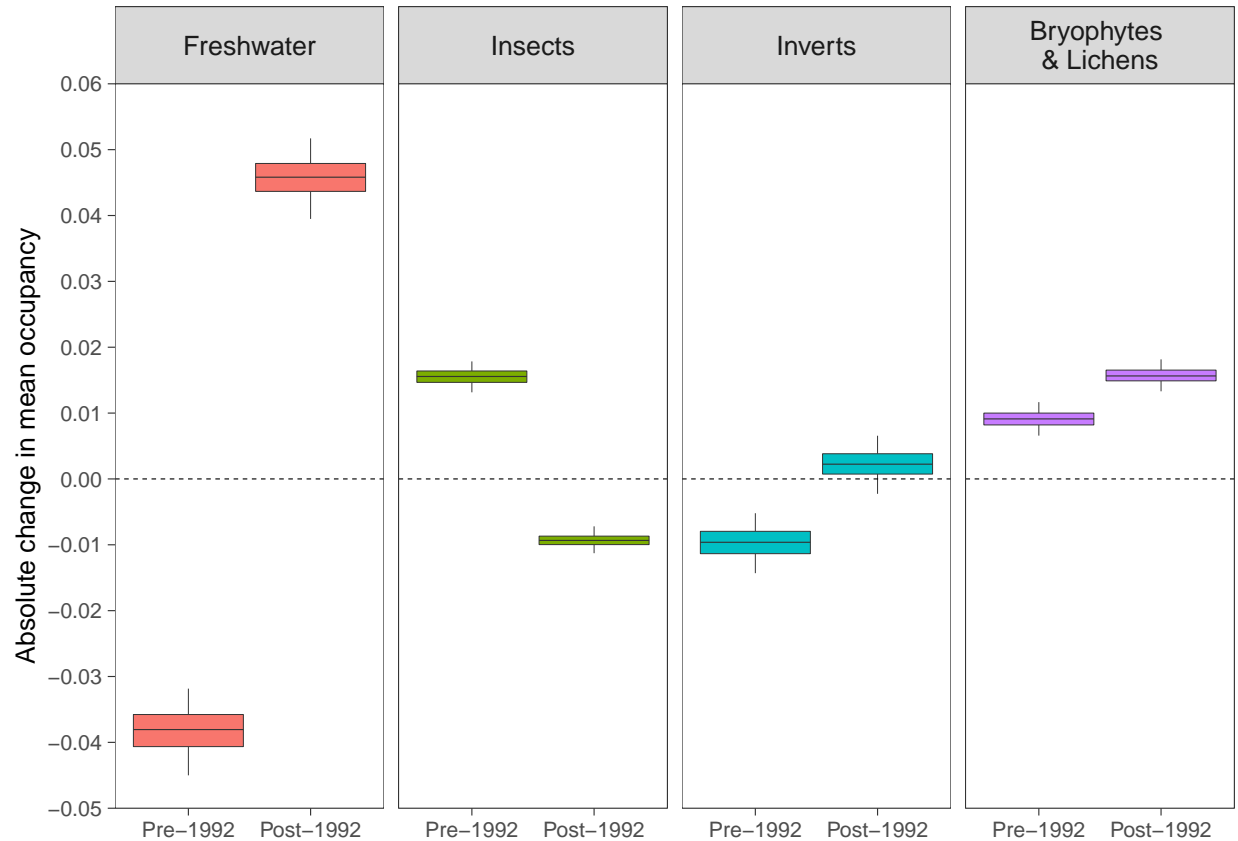
Recreating Figure 2

Figure 2 presents box plots of absolute change in average occupancy across 2 halves of the time period assessed for each major group. It shows estimates for 1970 to 1992 and for 1993 to 2015.

Again, this function uses the major group level posteriors of average occupancy across each group that were generated and saved within the `generate_fig1` function. These will have been saved within a sub-folder within the `/MajorGroups/geomeans`. You will need to direct the function to this folder.

```
## Use the generate_fig2 function to recreate the box plots.

# Runtime of this function: 0.55 seconds
generate_fig2(datadir = paste0(getwd(), "/MajorGroups/geomeans"),
              save_plot = TRUE)
```



Recreating Figure 3

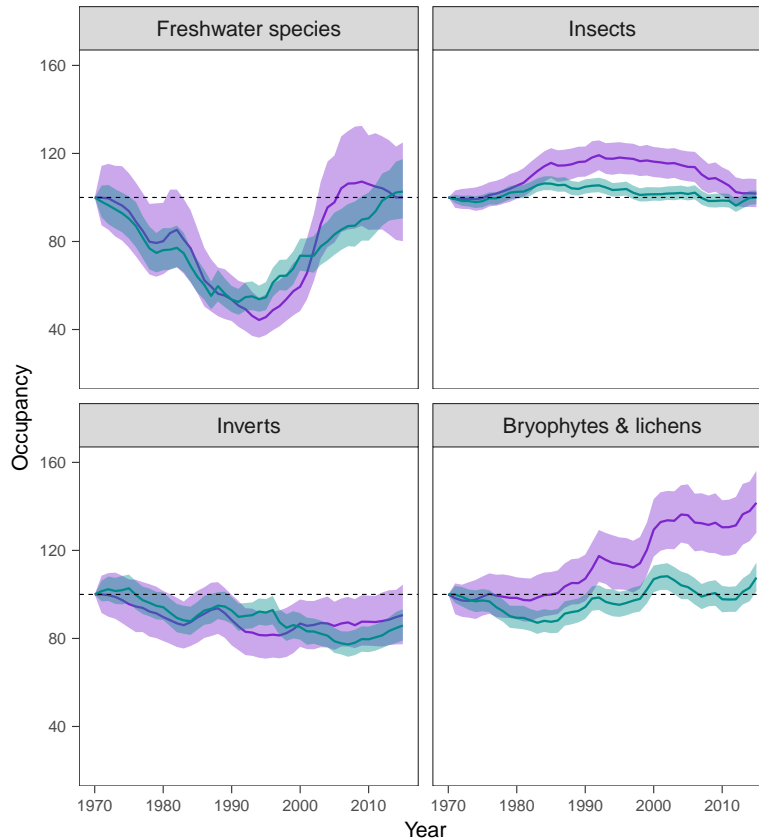
Figure 3 presents the changes in the quantiles of occupancy over time to assess how rarity and commonness has changed over time. This function estimates the quantile data from the combined posterior samples generated by the `combine_posteriors` function, saves this within a sub-folder called “quantiles” as csv files and generates the figure.

```
## Use the generate_fig3 function to recreate figure 3.
```

```
# Runtime of this function: 6 mins
```

```
generate_fig3(postdir = paste0(getwd(), "/MajorGroups"),
  save_plot = TRUE,
  interval = 95)
```

```
## [1] "Warning: overwriting existing files"
```



Recreating Figure 4

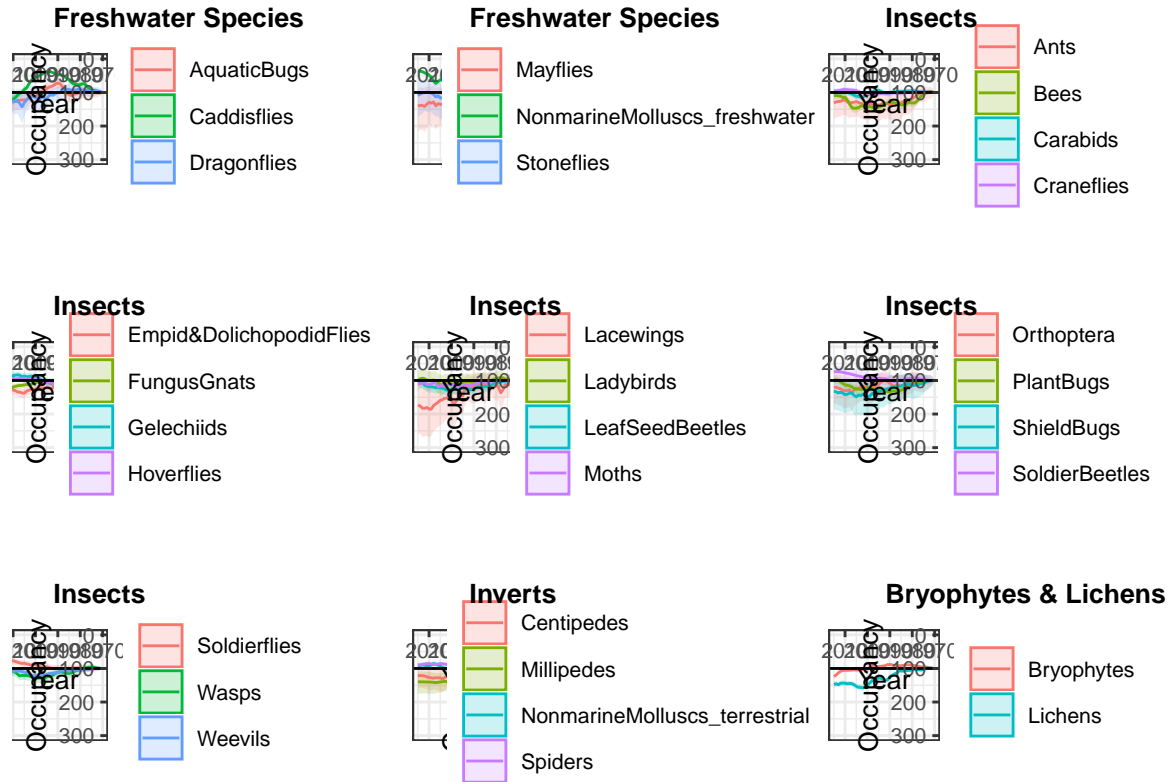
Figure 4 is a representation of Figure 1, except with a line for each taxonomic group. It uses the taxa level estimates generated from the `combine_posteriors` function to calculate the indicator posteriors and the average and 95% CI indicator values. The combined samples should be in a sub-folder called “Taxa” which was generated by the `combine_posteriors` function. The outputs are saved in a sub-folder called “Taxa/geomeans”. These are then used to recreate Figure 4 which is saved as a PDF file within that same sub-folder.

So that the trends are more visible, the Freshwater and Insect plots have been split into multiple panels.

Use the generate_fig4 function to recreate figure 4.

```
# Runtime of this function: ~ 2 minutes
generate_fig4(postdir = paste0(getwd(), "/Taxa"),
              save_plot = TRUE,
              interval = 95)
```

```
## [1] "Warning: overwriting existing files"
```

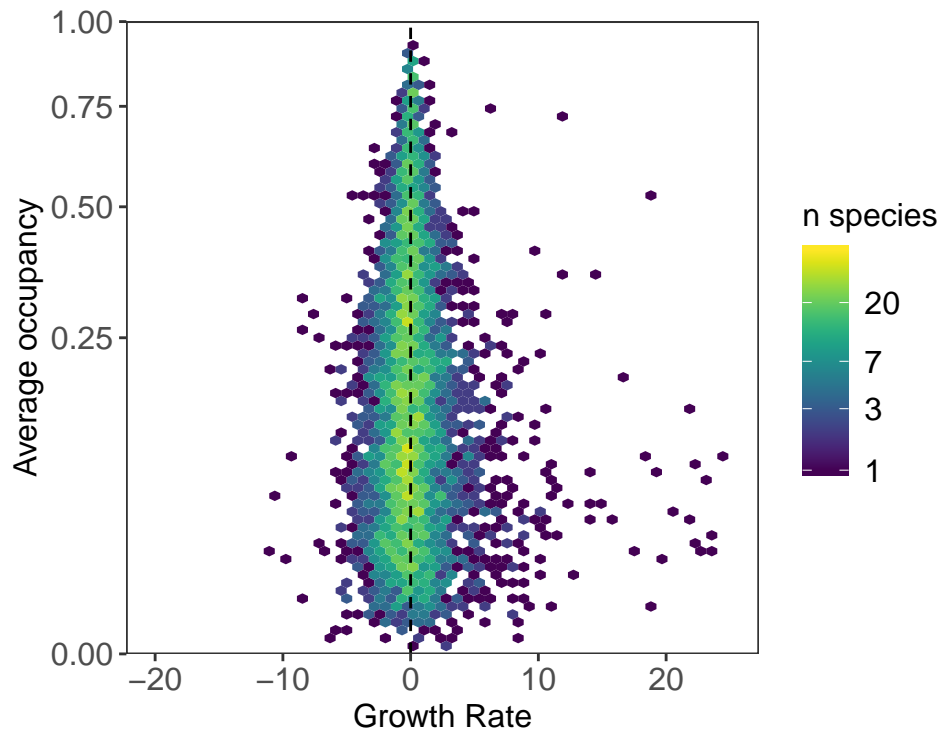


Recreating Figure 5

Figure 5 presents the average occupancy for each species plotted against its average annual growth rate. Average occupancy across the time period is estimated within the function, growth rates are supplied within the “Species_Trends.csv” file downloaded from the repository.

Use the generate_fig5 function to recreate figure 5.

```
# Runtime for this function: 9 minutes (when status = FALSE)
generate_fig5(postdir = paste0(getwd(), "/POSTERIOR_SAMPLES"),
  outdir = getwd(),
  sp_trends = read.csv(paste0(getwd(), "/Species_Trends.csv")),
  save_plot = TRUE,
  status = FALSE)
```



Calculating group and taxa level trends

Within the text are associated group level trends and an average trend estimate for all species since 1970.

Trends can be calculated once the associated posterior values have been generated via the `generate_fig1` and `generate_fig4` functions for major groups and taxa respectively. Use the `group_trends` function to calculate the trends relative to 1970 and the associated credible intervals. The credible interval value can be specified within the function, but defaults to 95%.

```
# Use the group_trends function to estimate overall and major group level trends
# and associated credible intervals.
# Runtime of this function: 0.17 seconds
```

```
# datadir should point to the main directory in which the "MajorGroups" and "Taxa" subdirectories
# can be found created within the generate_fig1 function.
```

```
trends <- group_trends(datadir = getwd(),
                       interval = 95)
```

```
trends
```

##	Group	Mean_change	LCI	UCI
## 1	ALL	10.924	8.785	12.948
## 2	FRESHWATER_SPECIES	7.189	-3.268	18.985
## 3	LOWER_PLANTS	36.269	30.791	42.298
## 4	TERRESTRIAL_INSECTS	5.456	2.849	8.061
## 5	TERRESTRIAL_NONINSECT_INVERTS	-6.713	-11.804	-1.628
## 6	Ants	31.329	-6.188	78.270

## 7	AquaticBugs	18.804	-6.582	50.517
## 8	Bees	9.402	-0.883	21.401
## 9	Bryophytes	23.626	16.939	30.408
## 10	Caddisflies	37.871	19.702	58.376
## 11	Carabids	0.861	-9.072	11.818
## 12	Centipedes	21.585	-5.976	58.529
## 13	Craneflies	-3.963	-11.578	4.319
## 14	Dragonflies	16.952	-15.204	51.311
## 15	Empid&DolichopodidFlies	19.772	7.611	33.650
## 16	FungusGnats	18.637	9.239	28.790
## 17	Gelechiids	-24.428	-35.265	-11.612
## 18	Hoverflies	-5.175	-14.521	4.182
## 19	Lacewings	73.504	10.549	160.431
## 20	Ladybirds	-0.474	-29.814	28.984
## 21	LeafSeedBeetles	2.506	-12.262	18.149
## 22	Lichens	47.548	38.768	57.119
## 23	Mayflies	40.707	-12.892	116.382
## 24	Millipedes	39.799	7.842	76.020
## 25	Moths	8.713	3.696	13.886
## 26	NonmarineMolluscs_freshwater	-64.156	-70.630	-57.504
## 27	NonmarineMolluscs_terrestrial	-8.121	-17.314	2.709
## 28	Orthoptera	20.742	-19.594	70.178
## 29	PlantBugs	-1.736	-10.830	8.932
## 30	ShieldBugs	31.504	-7.242	83.588
## 31	SoldierBeetles	-26.115	-43.157	-6.172
## 32	Soldierflies	-32.205	-40.620	-23.178
## 33	Spiders	-10.194	-15.984	-3.855
## 34	Stoneflies	9.051	-34.277	72.529
## 35	Wasps	4.512	-6.847	16.901
## 36	Weevils	19.885	11.747	28.702

Recreating Supplementary Figure 1

Supplementary Figure 1 presents the occupancy plots for each taxonomic grouping separately within a PDF document. These figures more clearly show the responses of individual groups which may otherwise be difficult to interpret from Figure 4 within the main text.

```
## Use the generate_fig1supp function to recreate the figure
## in the paper supplementary materials.

# Runtime for this function: 1 second
generate_supplefig1(postdir = "C:/Users/charl/Dropbox/PhD WORK/1. BIG PAPER/Package_testing2/Taxa",
                    interval = 95,
                    save_plot = TRUE)

## [1] "Warning: overwriting existing files"

## pdf
## 2
```

The PDF version of this plot will be saved in the postdir/geomeans.

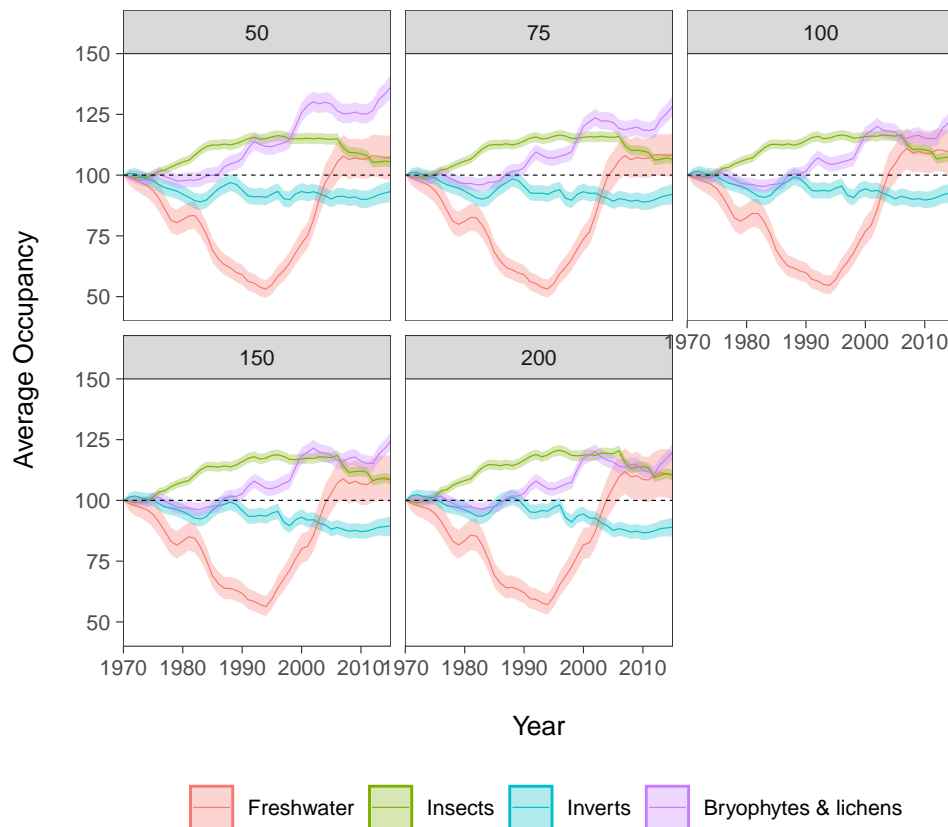
Recreating Supplementary Figure 2

Supplementary Figure 2 is a representation of Figure 1 in the main text except varying thresholds have been used to determine which species are included. These thresholds detail the number of records a species must have as a minimum before they are included within the generation of the indicators. The number of records within the raw data is specified within the “Species_Trends.csv” file that was downloaded from the repository. These values are used to subset the species within each group according to thresholds of number of records: 50, 75, 100, 150 or 200 records. In the main text, all species are used which have a minimum of 50 records.

```
## Use the generate_fig1supp function to recreate the figure  
## in the paper supplementary materials.
```

```
# Runtime for this function: 13 minutes
```

```
generate_supfig2(postdir = paste0(getwd(), "/MajorGroups"),  
  sp_trends = read.csv(paste0(getwd(), "/Species_Trends.csv")),  
  status = FALSE,  
  save_plot = TRUE)
```



You have now recreated the analyses carried out in the paper by Outhwaite et al using the original data files!