Recreate UK Biodiversitt Trends Paper

Charlotte L. Outhwaite

08/06/2019

Contents

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 you will need to register with the website first, then you can follow the intructions 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 this folder is 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.

```r
# Where are the posterior sample files?
datadir <- paste0(getwd(), "/POSTERIOR_SAMPLES")

# List the files
```

```
files <- list.files(datadir)

# How many files are there?
length(files) # 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 lemani.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 packa
library(devtools)

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

# Load the package
library(UKBiodiversity)
```

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 taxa 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 subfolder of the directory that you specify, within a group specific subfolder. Use the `group_level` argument to select either "major_groups" or "taxa" level datasets.

Note that these functions can take a long time to run (estimated run time below).

```
# Run the combine posteriors function providing a data directory and and output directory.

## First for Major groups.
# Runtime of this function: 76 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"
```

```r
## Then for taxa.
# Runtime for this function: ~ 53 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"
```

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

Recreating Figure 1 and associated trend estimates

Figure 1 within the paper presents indicators of average occupancy over time for the four major groups. Within the text are associated group level trends and a trend estimte for all species since 1970.

First, use the `generate_fig1` function to determine the group level average change,associated indicator values and the plot of Figure 1. Then use the `group_trends` function to estimate group level trends. The `generate_fig1` function will save csv files of the indicator posterior values, the average indicator and associated credible interval values and a pdf file of the plot within a subdirectory of your specified folder.

This function will only work using the output files from the "major_groups" specification of the `combine_posteriors` function. These will be in a subfolder called "MajorGroups" in the outdir that you specified within that function.

```r
# Use the generate_fig1 function to generate figure 1 nad associated outputs.
# Runtime of this function: 15 mins
generate_fig1(postdir = paste0(getwd(), "/MajorGroups"),
              status = FALSE,
              save_plot = TRUE)
```
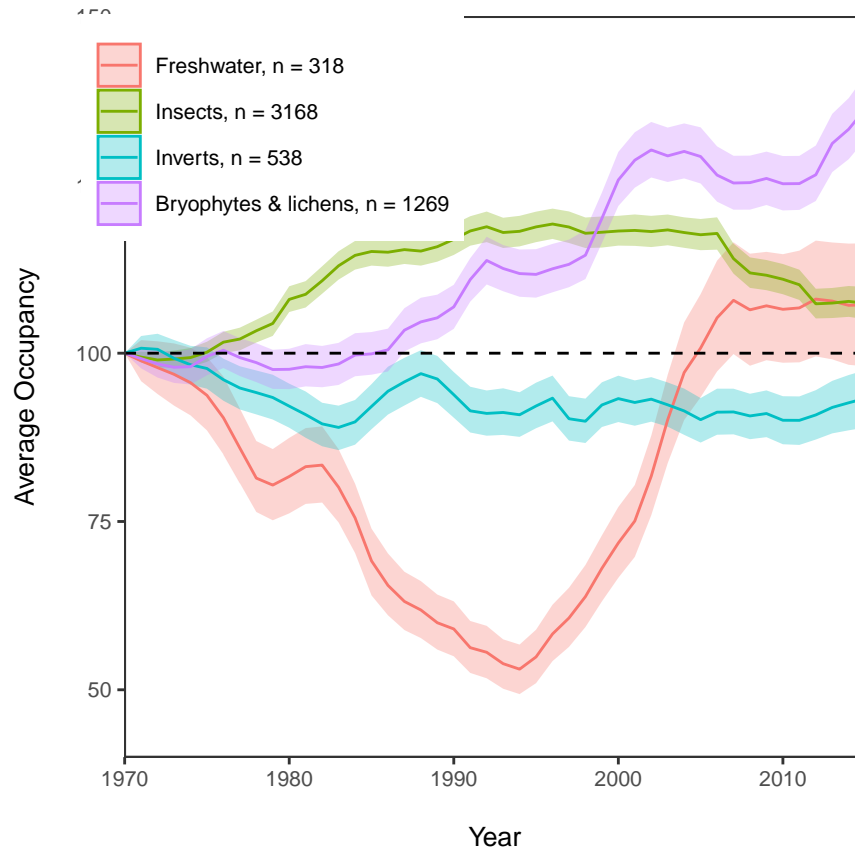
```r
# Use the group_trends function to estimate overall and major group level trends.
# Runtime of this function: 0.17 seconds
trends <- group_trends(datadir = paste0(getwd(), "/MajorGroups/geomeans"))

trends
```

```
##          Group                               Mean change LCI       UCI
## result "ALL"                                "12.233"    "10.035"  "14.384"
## result "FRESHWATER_SPECIES"                 "7.189"     "-3.268"  "18.985"
## result "LOWER_PLANTS"                       "36.269"    "30.791"  "42.298"
## result "TERRESTRIAL_INSECTS"               "7.406"     "4.711"   "10.242"
## result "TERRESTRIAL_NONINSECT_INVERTS"     "-6.713"    "-11.804" "-1.628"
```
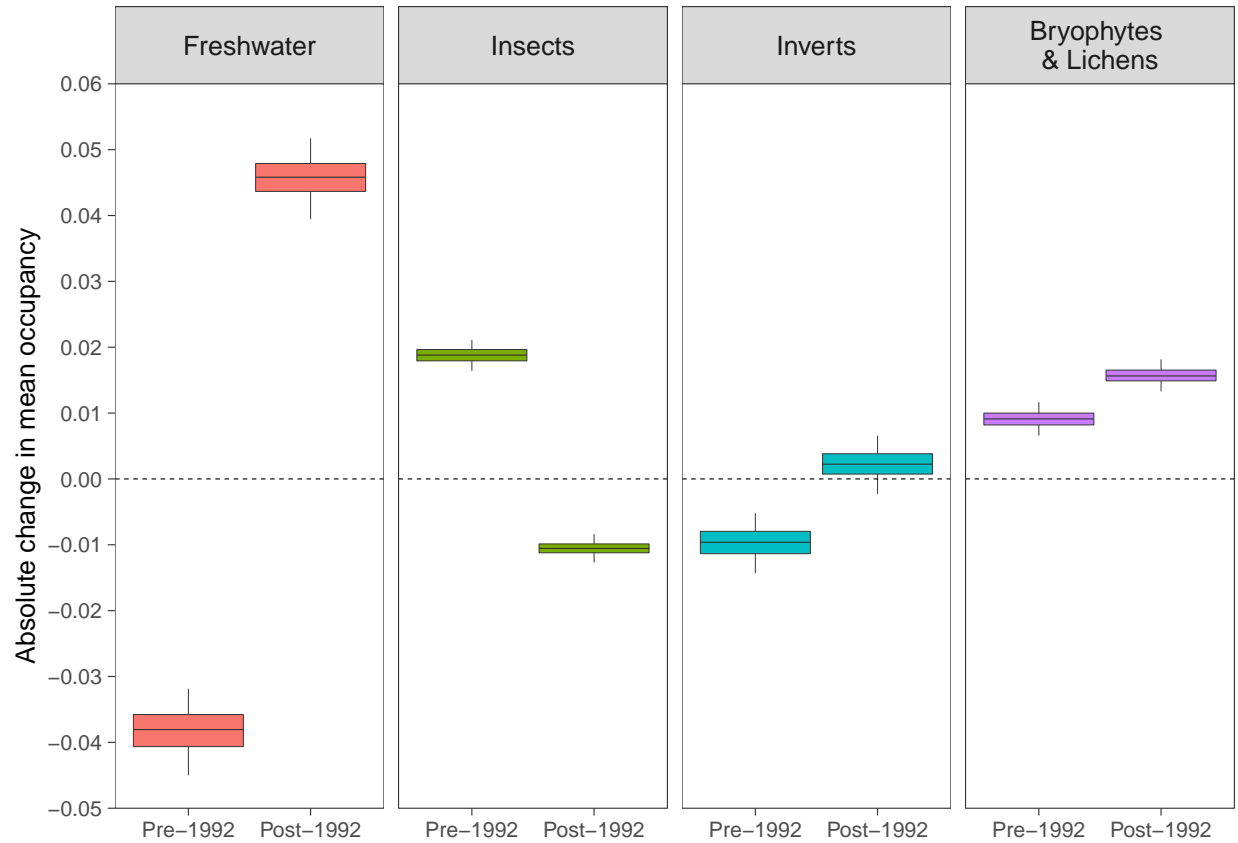
Recreating Figure 2

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 subfolder within the `/MajorGroups` folder called "geomeans". You will need to direct the function to this folder.

```r
## 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, saves this within a subfolder called "quantiles" as csv files and generates the figure.

```
## Use the generate_fig3 function to recreate figure 3.
# Runtime of this function: 7 mins
generate_fig3(postdir = paste0(getwd(), "/MajorGroups"),
              save_plot = TRUE)
```
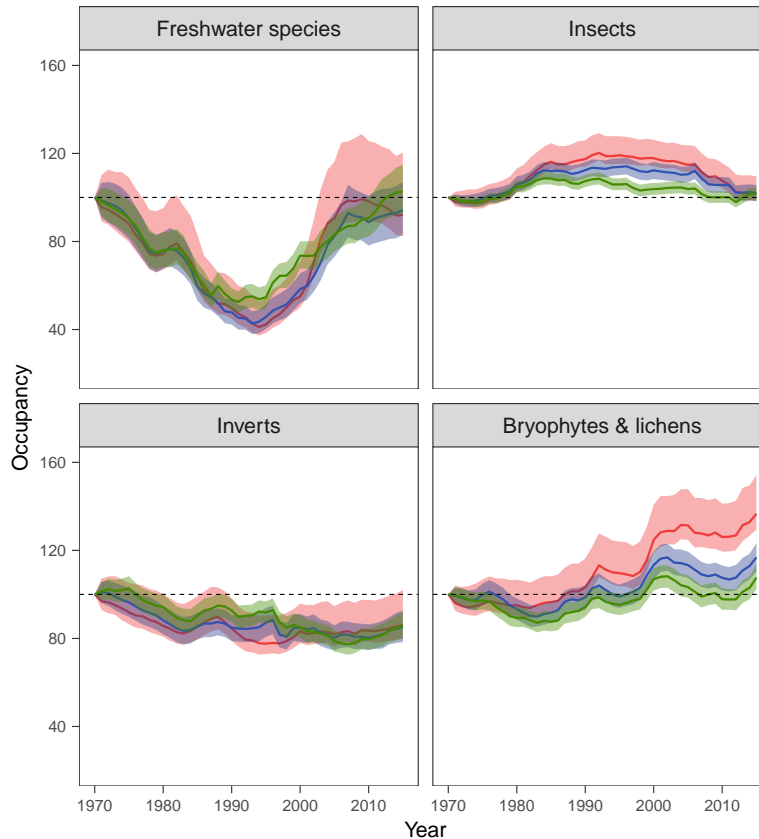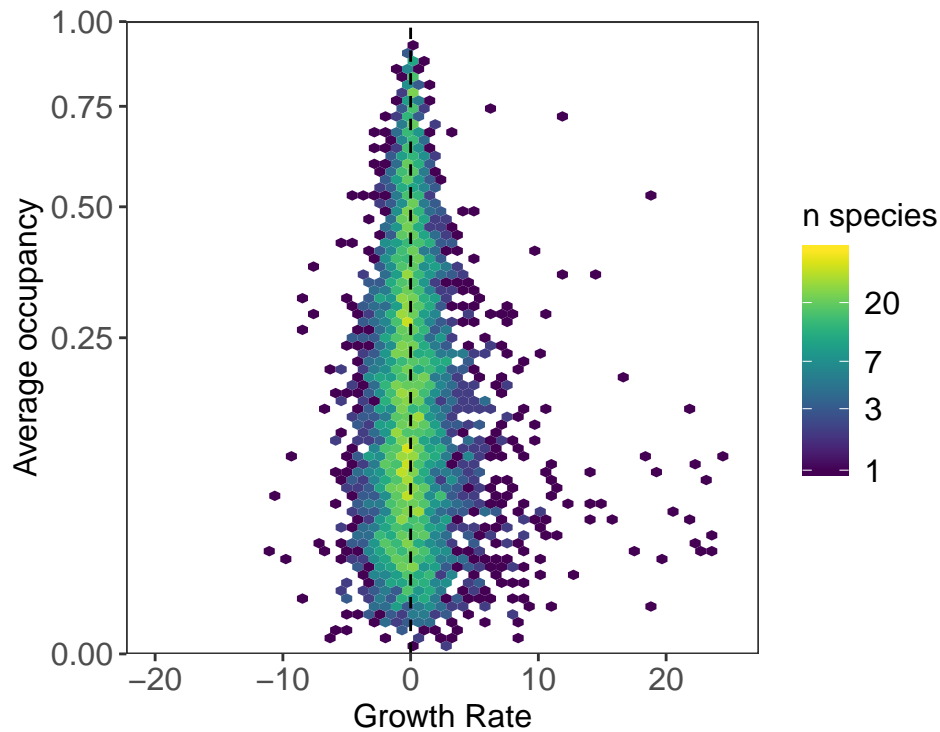
Recreating Figure 4

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. These are saved in a subfolder called "quantiles". These are then used to recreate Figure 4 which is saved as a PDF file.

```r
## Use the generate_fig4 function to recreate figure 4.
# Runtime of this function: 3.7 minutes
generate_fig4(postdir = paste0(getwd(), "/Taxa"),
              save_plot = TRUE)
```

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 within the "Species_Trends.csv" file downloaded from the repository.

```r
## 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)
```

Recreating Supplementary Figure 1

Supplementary Figure 1 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.

```
## Use the generate_fig1supp function to recreate the figure in the paper supplementary materials.
# Runtime for this function: 13 minutes
generate_fig1supp(postdir = paste0(getwd(), "/MajorGroups"),
                  sp_trends = read.csv(paste0(getwd(), "/Species_Trends.csv")),
                  status = FALSE,
                  save_plot = TRUE)
```
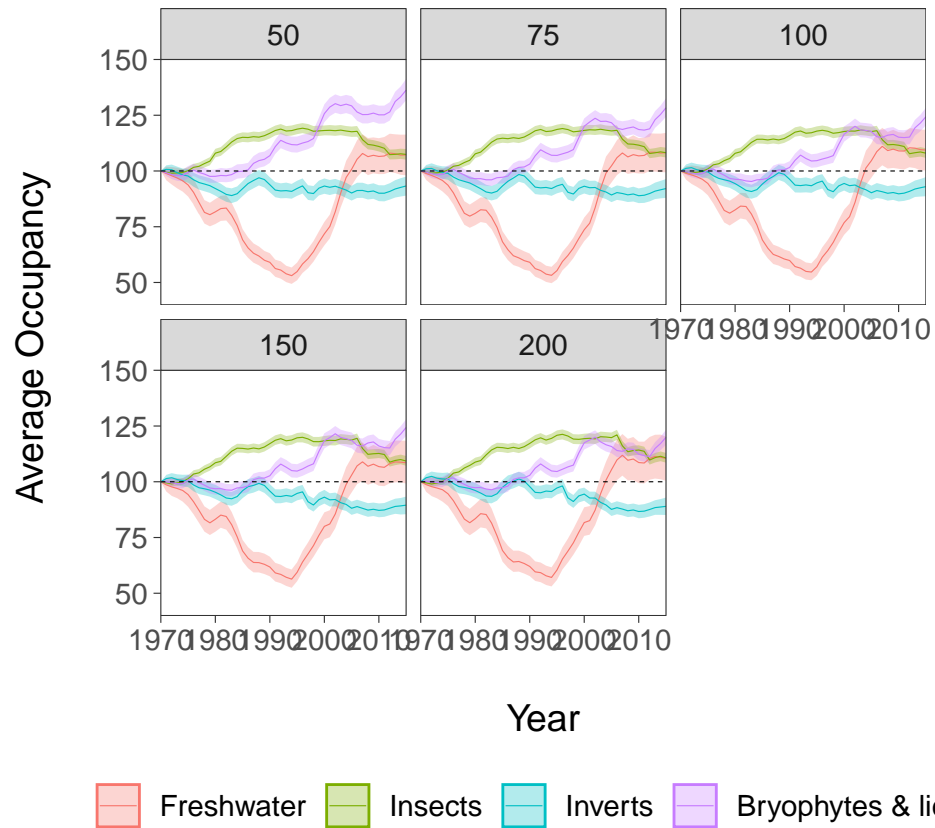
You have now recreated the analyses carried out in the paper by Outhwaite et al using the original data files!