README

This R project folder contains all necessary data and R scripts needed to reproduce the analysis in the paper: “Moth declines are most severe in broadleaf woodlands despite a net gain in habitat availability” – Blumgart et al 2022.

Scripts 1, 3 and 4 convert raw data into processed data and into species trend estimates. The scripts 2 and 5 do the analysis and produce the figures that are shown in the manuscript.

To get the scripts to run, you need change the directory to the place you have saved the data on your computer. Or, make an R project and make a folder called “Data” into which you put the CSV files and zipped folders. Make sure to unzip the folders first.

**Scripts**

Script\_1\_Making\_diversity\_stats\_data\_frame

This script takes the raw moth data in the CSV file “All\_years\_all\_species\_all\_sites.csv” and converts it into estimates of species richness and diversity per site-year. The outputs are saved as a CSV file “Diversity\_stats\_complete.csv” which is provided, so this script can be skipped if you want to go straight to the main analysis.

Script\_2\_General\_analysis

This script uses the Diversity\_stats\_complete.csv created above plus the raw abundance data and models annual abundance, biomass, richness and diversity across the seven habitat types and two regions. This script produces all the main estimates of percentage change as well as Figures 1 – 3 and many of the Figures in the Supporting Information. This script does the reanalysis of the Macgregor et al. 2019 paper that is mentioned in the discussion of the manuscript.

Script\_3\_Turning\_raw\_data\_to\_site\_indices

This script takes the raw abundance data for each species (saved in Raw\_moth\_species\_data folder) and uses Generalised Additive Models to estimate the completeness of sampling in each site-year for each species. It then produces ‘site indices’ which are estimates for how many moths the trap would have caught, had the trap been running continuously through the species flight period. If the trap was running continuously, then the site index is equal to the raw count. This script takes a very long time to run as it has to model 427 species. Running this script is not necessary to produce the main results and figures and is provided only for reference. The output files that this script produces are already provided in the Sindex\_files folder.

Script\_4\_Species\_models\_each\_habitat\_region

This script takes the files in the Sindex\_files folder which were produced in Script 3. This script models percentage change of each species in each habitat/region and saves the percentage changes to CSV files named Hab\_trends, Region\_trends and Hab\_region\_trends. It is not necessary to run this script to reproduce the analysis as these three data frames are already supplied in the Data folder.

Script\_5\_Species\_specific\_analysis

This script takes the percentage changes for each species that were estimated in Script 4. It models percentage change as a function a species traits and habitat/region. This script produces Figure 4 as well as many of the Figures in the Supporting Information. This script also produces the map in Fig. S1.

**Datasets**

The datasets provided and the meaning of their variables are provided here.

All\_years\_all\_species\_all\_sites.csv

Total number of moths caught of each species at each site in each year across the Rothamsted Insect Survey light-trap network.

|  |  |
| --- | --- |
| CalYear | Calendar year |
| RIS.TrapCode | Site number |
| Code.RISgeneric | Species number in the RIS database |
| Common.Name | Common name of species |
| binomial | Scientific name of species |
| Annual\_count | Total number of moths caught in that site-year |
| Site\_name | Name of site |
| Site\_name\_2 | Names of site with underscores replacing spaces (reduces errors) |
| Country | Country location of trap site |

Biomass\_df.csv

The estimated biomass (dry weight) for a single moth of each species as calculated in (Kinsella *et al.*, 2020)

|  |  |
| --- | --- |
| RIS\_code | Species number in the RIS database |
| Common\_name | Common name of species |
| Binomial | Scientific name of species |
| Biomass\_Kinsella\_direct | Dry weight of a single moth (mg) |
| Fam\_for\_biomass | Family that the species is in |

Diversity\_stats\_complete.csv

The species richness and diversity of moths in each site year calculated using the All\_years\_all\_species\_all\_sites.csv in Script 1

|  |  |
| --- | --- |
| SITE | Site number |
| Year | Calendar year |
| Actual\_sp\_richness | The actual number of species recorded in that site-year |
| Est\_sp\_richness | The estimated total number of species using rarefaction in iNEXT |
| Est\_sp\_richness\_se | The standard error of the total number of species using rarefaction |
| Est\_sp\_richness\_low | The lower 95% confidence interval of the total number of species using rarefaction |
| Est\_sp\_richness\_upp | The upper 95% confidence interval of the total number of species using rarefaction |
| Actual\_Shannon | The actual Shannon diversity of species recorded in that site-year |
| Est\_Shannon | The estimated Shannon diversity of species using rarefaction in iNEXT |
| Est\_Shannon\_se | The standard error of the Shannon diversity of species using rarefaction |
| Est\_Shannon\_low | The lower 95% confidence interval of the Shannon diversity of species using rarefaction |
| Est\_Shannon\_upp | The upper 95% confidence interval of the Shannon diversity of species using rarefaction |

Moth\_traits.csv

The traits of each moth species.

|  |  |
| --- | --- |
| RIS\_code | Species number in the RIS database |
| Bradley\_code | Species number in the Bradley system |
| Common\_name | Common name of species |
| Binomial | Scientific name of species |
| Family | Family of species |
| Family\_Kinsella\_biomass | Family as used in the Kinsella biomass equation |
| Feeding\_guild | Broad host plant group of larva – grass, forbs, shrubs etc. |
| Use\_Ellenbergs | If the species has three or fewer hostplants then this is ‘Y’ |
| Ellenberg\_mean\_Light | The mean Ellenberg value for light of the hostplants |
| Biomass\_Kinsella\_direct | Weight of dry biomass of the species from Kinsella et al. 2020 (mg) |
| Foodplant\_1 | Host plant 1 (for species with three or fewer host plants) |
| Foodplant\_2 | Host plant 2 (for species with three or fewer host plants) |
| Foodplant\_3 | Host plant 3 (for species with three or fewer host plants) |
| Ellenberg\_1\_Light | Ellenberg light value for host plant 1 |
| Ellenberg\_2\_Light | Ellenberg light value for host plant 2 |
| Ellenberg\_3\_Light | Ellenberg light value for host plant 3 |

Site\_info.csv

Information about the trap sites in the RIS network

|  |  |
| --- | --- |
| SITE | Site number |
| Site\_name | Site name |
| Years\_running | Number of years running (excluding gaps) |
| Region | Country within UK |
| N\_S | In the North or the South defined by the 4500 N gridline on British National Grid |
| Improved\_latitude | Latitude of site |
| Improved\_longitude | Longitude of site |
| Improved\_altitude | Altitude of site |
| Improved\_grid\_ref | British grid reference |
| Improved\_eastings | Easting on British National Grid |
| Improved\_northings | Northings on British National Grid |
| Dom\_land\_use\_500m | The dominant land use within a 500 m radius of site according to the CEH 2015 Land Cover map (LCM2015) |
| Percent\_dom\_land\_use\_500m | Percentage of radius filled by the dominant land use at 500 m |
| Deciduous\_woodland\_2015\_500m | The percentage coverage of each of the land use categories according LCM2015 within a 500 m radius |
| Arable\_2015\_500m |
| Improved\_grassland\_2015\_500m |
| Urban\_2015\_500m |
| Suburban\_2015\_500m |
| Fresh\_water\_2015\_500m |
| Conifer\_plantation\_2015\_500m |
| Ocean\_2015\_500m |
| Saltwater\_2015\_500m |
| Supra\_littoral\_rock\_2015\_500m |
| Littoral\_rock\_2015\_500m |
| Neutral\_grassland\_2015\_500m |
| Wetland\_2015\_500m |
| Saltmarsh\_2015\_500m |
| Acid\_grassland\_2015\_500m |
| Supra\_littoral\_sediment\_2015\_500m |
| Calcareous\_grassland\_2015\_500m |
| Littoral\_sediment\_2015\_500m |
| Heather\_2015\_500m |
| Heather\_grassland\_2015\_500m |
| Bog\_2015\_500m |
| Inland\_rock\_2015\_500m |
| Dom\_land\_use\_2km | The dominant land use within a 2000 m radius of site according to the CEH 2015 Land Cover map (LCM2015) |
| Percent\_dom\_land\_use\_2km | Percentage of radius filled by the dominant land use at 2000 m |
| Deciduous\_woodland\_2015\_2km | The percentage coverage of each of the land use categories according LCM2015 within a 2000 m radius |
| Conifer\_plantation\_2015\_2km |
| Arable\_2015\_2km |
| Improved\_grassland\_2015\_2km |
| Urban\_2015\_2km |
| Suburban\_2015\_2km |
| Inland\_rock\_2015\_2km |
| Fresh\_water\_2015\_2km |
| Neutral\_grassland\_2015\_2km |
| Heather\_grassland\_2015\_2km |
| Heather\_2015\_2km |
| Ocean\_2015\_2km |
| Saltwater\_2015\_2km |
| Supra\_littoral\_rock\_2015\_2km |
| Littoral\_rock\_2015\_2km |
| Wetland\_2015\_2km |
| Saltmarsh\_2015\_2km |
| Acid\_grassland\_2015\_2km |
| Supra\_littoral\_sediment\_2015\_2km |
| Littoral\_sediment\_2015\_2km |
| Calcareous\_grassland\_2015\_2km |
| Bog\_2015\_2km |
| Up\_lowland | Whether the site is upland (>300 m altitude) or lowland (<300 m altitude) |
| Habitat\_500m | Habitat type at 500 m radius (see Methods) |
| Habitat\_2km | Habitat type at 2000 m radius (see Methods) |
| Site\_name\_db | Site name as it appears in RIS database |
| Site\_name\_db\_nogaps | Site name as it appears in RIS database |
| Forest\_type | Site name as it appears in RIS database with underscores replacing spaces (for fewer errors) |
| Landscape\_forest\_cover | Percentage of landscape covered in all woodland types (including plantation) at 500 m – used in calculating deer damage estimate |
| Perennial\_cover | Percenatge of landscape that is grassland and heathland at 2 km radius – used in calculating deer damage estimate |
| GB\_tile | Grid location – used in calculating deer damage estimate |
| Road\_density | Length of road within 250 m of site – used in calculating deer damage estimate |
| Deer\_damage\_estimate | Deer damage estimate expressed as a probability of damage (0 – 1) |

Site\_year\_info.csv

Information about the completeness of sampling in each site-year.

|  |  |
| --- | --- |
| SITE | Site number |
| Year | Calendar year |
| Site\_name | Site name |
| Complete\_type\_Paul | The completeness score of the site-year. See Methods |
| Site\_name\_db\_nogaps | Site name with underscores replacing spaces |

UK\_coastline\_R.csv

The coordinates of the coastline for Great Britain and Ireland – used in making figures.

Hab\_trends.csv

The percentage change across the time series for each species in each habitat

|  |  |
| --- | --- |
| RIS\_code | Species number in the RIS database |
| Common\_name | Common name of the species |
| Binomial | Scientific names of the species |
| Habitat | Habitat type |
| Poptrend\_perc\_change | Percentage change in abundance across time series |
| Poptrend\_perc\_change\_upp | 95% CI of percentage change – upper bound |
| Poptrend\_perc\_change\_low | 95% CI of percentage change – lower bound |
| Sample\_size | Total number of moths used in the time series |
| Total\_sites | Total number of sites at which the species was caught |
| Total\_site\_years | Total number of site-years |
| Min\_year | First year in time series |
| Max\_year | Final year in time series |

Region\_trends.csv

Hab\_region\_trends.csv

These two datasets are essentially the same as Hab\_trends.csv but they also contain variables “N\_S” which is whether the trend is for the North or the South and “Hab\_region” which is whether the trend is for the North/South and in Broadleaf Woodland or Improved Grassland.

Hab\_trends\_ready.csv

Region\_trends\_ready.csv

Hab\_region\_trends\_ready.csv

These three datasets are modifications of the three datasets above but with outliers removed – the trends with outliers were re-modelled starting at the half-way point of the time series (see Methods). They also contain species traits which are explained above, and two additional variables: “Log\_sample\_size” which is just the natural log of the total number of moths in the subset, and “Trend\_trans” which is the natural log of the percentage change +100. The percentage change for each subset of moths within each subset of sites is used in Script 5 to model the interaction between species traits, region and habitat.

KINSELLA, R. S., THOMAS, C. D., CRAWFORD, T. J., HILL, J. K., MAYHEW, P. J. & MACGREGOR, C. J. 2020. Unlocking the potential of historical abundance datasets to study biomass change in flying insects. *Ecology and evolution,* 10**,** 8394-8404.