**A compendium of earthworm data sources and associated information from the UK and Ireland 1891-2021 – *Supporting Information***

Kelly E. Mason, Frank Ashwood, Olaf Schmidt, Jack Cosby, Aidan Keith

Brief Description

This dataset presents a compendium of field-based earthworm data sources and associated meta-data from across the United Kingdom and Ireland (‘WormSource’). These data were compiled up to 2021 and include sources dating back to 1891. Source meta-data covers the type of quantitative earthworm data (i.e. incidence, abundance, biomass, taxa), methodological details (e.g. sampling method/s, location/s, whether sampled plots were natural or experimental, sampling year/s), and environmental information (e.g. habitat/land-use, inclusion of climate data and basic soil properties). Data sources were collected through literature searches on Web of Science and Google Scholar, as well as directly from original authors/data holders where possible. The data sources were compiled with the aim of gathering quantitative data on earthworm species and populations to develop earthworm abundance and niche models, and toward a modelling framework for earthworm impacts on soil processes. This work is part of the Soil Organic Carbon Dynamics (SOC-D) project funded by the NERC UK-SCAPE programme (Grant reference NE/R016429/1).

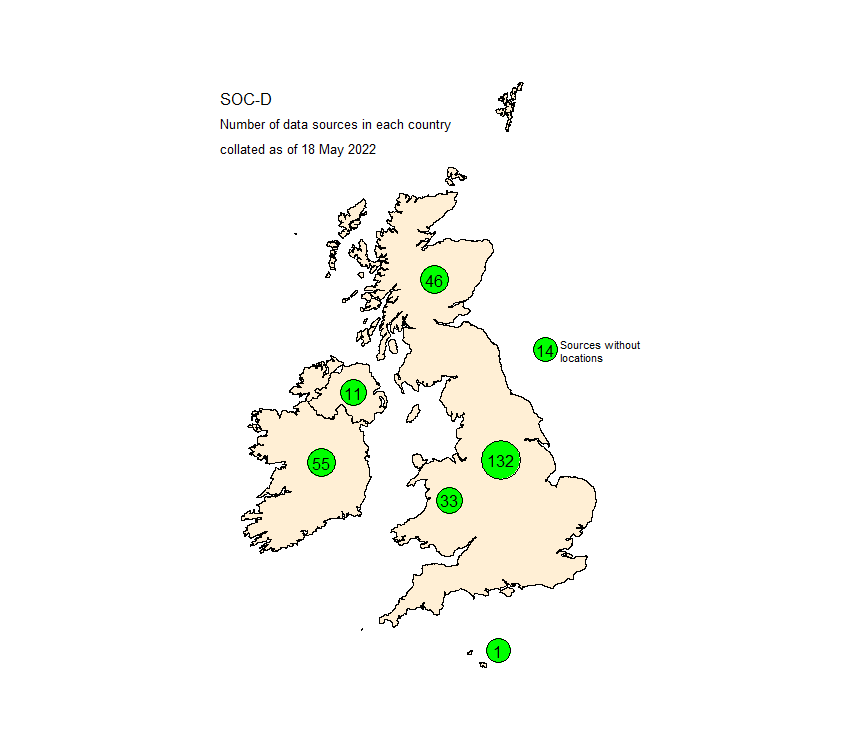
Literature Search and Methodology

A systematic literature search was conducted to find data sources using Web of Science and Google Scholar (in June/July 2020). This used the following search term:

(Earthworm\* OR Oligochaeta OR Annelid\* OR Lumbric\* OR Clitellat\*) AND (Diversity OR “Species richness” OR “OTU” OR Abundance OR individual\* OR Density OR “tax\* richness” OR “Number” OR Richness OR Biomass) AND (UK OR “United Kingdom” OR England OR Wales OR Scotland OR “NI” OR “Northern Ireland” OR “Ireland” OR “ROI”)

Google Scholar was searched using separate searches for individual countries (i.e. “UK” or “United Kingdom”; then “England”; “Wales”; “Scotland”; “Ireland”) due to a maximum search term character length. The WoS and Google Scholar searches yielded 364 and 38,000+ results, respectively. Of the Google Scholar results, approximately 15 pages were screened for each country searched, with 10 results per page, for a total of approximately 750 articles screened, most of which were repeats from the WoS search. Publications were selected and meta-data recorded if they contained location information and the presence of at least one species of earthworm or abundance/biomass data obtained from the field. Further publications were found through screening of review papers and contact with selected authors for additional information and resources, including published and unpublished datasets.

In total, 257 data sources were compiled. Source meta-data covers the type of quantitative earthworm data (i.e. incidence, abundance, biomass, taxa), methodological details (e.g. sampling method/s, location/s, whether sampled plots were natural or experimental, sampling year/s), and environmental information (e.g. habitat/land-use, inclusion of climate data and basic soil properties) (see table below). The most common sampling methods across data sources were hand-sorting (55%) and chemical extractant (42%), with 16% using these in combination. A large proportion of sources measured earthworm incidence (82%), abundance (68%), and taxa (68%), whilst only 49% contained information on biomass. Soil data of some kind (e.g., soil texture, bulk density, pH, carbon or organic matter content, nutrient content) was included in 43% of sources. The most common habitat category across data sources was grasslands (62%), followed by arable (32%) and woodland (22%). A map with the number of data sources in each country is provided in Figure 1. Data sources were dominated by sampling in England (51% of sources), followed by Republic of Ireland (21%), Scotland (18%), Wales (13%), and Northern Ireland (4%); though many data sources include sampling from multiple locations, regions, or countries. Original documents (i.e. published papers, book chapters) could not be obtained for 17 potential data sources at the time of compiling meta-data and are thus lacking any additional information.

*Figure 1. Number of data sources in each country in ‘WormSource.UKIreland.csv’.  A data source may include sampling in multiple countries and multiple regions within one country.*

Details of Data Structure

The dataset ‘WormSource.UKIreland.csv’ contains the following information:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Source\_ID | Unique identifier for data source |
| Study\_ID | Numbered identifier in chronological order |
| Source\_Summary | Summary of data source including author, year published, and title of source and publication/dataset |
| FirstAuthor\_LastName | Last name of first-named author |
| Source | Name of data source |
| Source\_Type | Type of output or publication providing data (i.e. Paper, Book chapter, Thesis, Report, Dataset) |
| DOI | Digital Object Identifier where available |
| Publication\_Year | Year published |
| Data\_Specific\_Location | Indication of availability of specific location data/coordinates (Yes = available at farm/forest/field-level and below, either given in source or obtained from authors/data managers; No = not readily available, though general location may be given at town or county level); newer datasets may have location data available upon request |
| Sampling \_Methods | List of sampling methods used; given as X + Y where combinations used for same sample (e.g. hand-sorting + chemical extractant) |
| Method\_Hand\_Sorting | Was hand-sorting used to sample earthworms? (1 = Yes; 0 = No) |
| Method\_Chemical\_Extractant | Was a chemical extractant used to sample earthworms (1 = Yes; 0 = No) |
| Method\_Under\_Feature\_Search | Was searching under natural or man-made features used to sample earthworms (e.g. rocks, logs, straw bales) (1 = Yes; 0 = No) |
| Method\_Electrical\_Octet | Was an electrical octet used to sample earthworms (1 = Yes; 0 = No) |
| Method\_Pitfall\_Traps | Was pitfall trapping used to sample earthworms (1 = Yes; 0 = No) |
| Method\_Surface\_Or\_Cast\_Counting | Was surface or cast counting used to sample earthworms (1 = Yes; 0 = No) |
| Method\_Other | Was another method used to sample earthworms (1 = Yes; 0 = No) |
| Data\_Incidence | Was incidence data collected? (1 = Yes; 0 = No) |
| Data\_Abundance | Was abundance/count data collected? (1 = Yes; 0 = No) |
| Data\_Biomass | Was biomass data collected? (1 = Yes; 0 = No) |
| Data\_Taxa | Was individual taxa data collected? (1 = Yes; 0 = No) |
| Data\_Species\_Specific | List of species where data for the entire earthworm community is not present in the data source |
| Data\_Climate | Are any climate data available? (1 = Yes, either given in source or available from authors/data managers; 0 = No, though may be possible to obtain data from nearest Met Office) |
| Data\_Soil | Is soil data available, specifically texture, pH, moisture, bulk density, soil organic matter, soil organic C, total C, total N? (1 = Yes, either given in source or available from authors/data managers; 0 = No) |
| Habitat\_Arable | Were samples collected from arable habitat (1 = Yes; 0 = No) |
| Habitat\_Grassland | Were samples collected from grassland habitat (1 = Yes, includes any type of grassland, e.g. extensive/managed/unmanaged/improved/unimproved grasslands, pasture, grassy field margins; 0 = No) |
| Habitat\_Wooded | Were samples collected from woodland or wood plantation habitat (1 = Yes; 0 = No) |
| Habitat\_Wetland\_Upland | Were samples collected from moorland or bog habitat (1 = Yes; 0 = No) |
| Habitat\_Brownfield | Were samples collected from brownfield habitat regardless of current overlying vegetation type (e.g. reclaimed waste site, mine site)? (1 = Yes; 0 = No) |
| Habitat\_Coastal | Were samples collected from coastal habitat (e.g. dune slack)? (1 = Yes; 0 = No) |
| Habitat\_Freshwater\_Shoreline | Were samples collected from riparian areas or adjacent to freshwater bodies? (1 = Yes; 0 = No) |
| Habitat\_Urban | Were samples collected from urban habitat? (1 = Yes, including parks, lawns/gardens, golf courses, football/sporting pitches, roadside verges, etc.; 0 = No) |
| Habitat\_Horticulture | Were samples collected from enriched soil environments such as flower/vegetable beds, compost/manure heaps, or botanical gardens? (1 = Yes; 0 = No) |
| Habitat\_Other | Were samples collected from any other habitat type e.g. greenhouses, botanical gardens, 'shearwater-impacted soils'? (1 = Yes; 0 = No) |
| Habitat\_Notes | Further details on habitat |
| Natural\_Or\_Experimental | Was the data from natural or experimental systems (Natural; Experimental; Both) |
| Spatial\_Extent | Spatial extent of sampling (Single Site = One farm, woodland, or similarly limited area; Multiple site-single region = Multiple sites/farms/etc. within a small region; Multiple site-multiple region = Multiple sites over a wider spatial scale). Note that this classification is somewhat subjective, so there may be overlap depending on the specific study, but is useful to broadly distinguish different sampling extents. |
| General\_Location | Approximate location of sampling sites (nearest village, research station, etc.) |
| Country | Country/Countries of sampling site(s) |
| Number\_Of\_Sites | Number of sites (e.g. farm-level, woodland, area-level, etc.) |
| Plot\_Types\_Sampled | Total number of plot types sampled at site-level regardless of replication within the site. EXAMPLES: (1) three farms with one field each = 3; (2) one farm with 3 fields sampled = 3; (3) one block experiment with 3 treatments = 3; (4) one farm with 3 fields, each with 3 treatments = 9); (5) one area such as a park or reclamation site with multiple sampling locations across 3 different habitat types = 3. Samples taken along a transect into a field are grouped together, except those taken from the grassy margin or hedgerow. |
| First\_Sampling\_Year | First year of earthworm sampling |
| Sampling\_Years | All earthworm sampling years |
| Notes | Additional notes on data source |