

Reproducibility

This document is a guide to reproducing the data and results in this package.

Running on the Peregrine High Performance Cluster Computer (HPCC):

Clone the DAISIEutils R package and the MadIsland R package in the home working directory. Here the home working directory refers to home/p-number/. Then change directory into MadIsland (`cd MadIsland`). Run the install bash script to ensure MadIsland is installed on the cluster (`sbatch inst/bash/install/install_MadIsland.sh`).

Run the extraction scripts with MadIsland as the working directory due to the file paths for saving, for compatibility with running the scripts locally (e.g. within Rstudio). The other option would be to have a platform identifier (e.g. `.Platform$OS.type`) to determine the file path as locally I use windows and the Peregrine HPCC uses Linux (i.e. unix). However, this platform specific option prevents reproducibility for those running the scripts locally on a Mac or Linux machine (because `.Platform$OS.type == "unix"` in those cases). The logs folder is already present at this point as it is made in the install of MadIsland, but good to check that the logs folder is in MadIsland.

Run each of the extraction scripts to produce the extracted data:

- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_complete_asr.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_complete_ds_asr.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_complete_ds.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_complete.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_dna_asr.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_dna_ds_asr.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_dna_ds.sh`
- `sbatch inst/bash/extract_data/amphibians/extract_amphibians_dna.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_complete_asr.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_complete_ds_asr.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_complete_ds.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_complete.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_dna_asr.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_dna_ds_asr.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_dna_ds.sh`
- `sbatch inst/bash/extract_data/birds/extract_birds_dna.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_complete_asr.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_complete_ds_asr.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_complete_ds.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_complete.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_dna_asr.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_dna_ds_asr.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_dna_ds.sh`
- `sbatch inst/bash/extract_data/mammals/extract_mammals_dna.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_complete_asr.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_complete_ds_asr.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_complete_ds.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_complete.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_dna_asr.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_dna_ds_asr.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_dna_ds.sh`
- `sbatch inst/bash/extract_data/squamates/extract_squamates_dna.sh`

Before running the analyses, make sure a logs folder is in the home directory. If this is not present create a logs folder.

Now that the island community data has been extracted and can be used for fitting the DAISIE inference models there is one more preparation step. The DAISIEutils package loads data using the `data()` function and thus it cannot be loaded from within the `inst/` folder where the extracted data is currently stored. Therefore, the data-raw scripts need to be executed to move the data sets we are fitting DAISIE to across to data. All data-raw scripts read the extracted data from `inst/extdata/extracted_data/...` and save it as data to be read by `data()`.

Run the analysis scripts the home working directory on the Peregrine HPC. This is because MadIsland interacts with the DAISIEutils package to run the DAISIE maximum likelihood models.

Run each of the analysis scripts to produce the DAISIE model output (parameter estimates and model likelihood and Bayesian Information Criterion):

- `sbatch MadIsland/inst/bash/analyse_data/analyse_amphibians/submit_amphibian_complete_cr_dd.sh`
-