# How to use the code appendix?

Author: Michael Kalyuzhny

Code last updated: 27/04/2023

This code appendix is aimed at reproducing all the analyses, figures and tables. The code is written in the Matlab™, with a single event of data transfer to an R script. The master script that calls all other scripts is ‘Full\_analysis.m’, which calls all other scripts and generates all results. Some graphics saving pieces of the code may be sensitive to the version of Matlab or screen resolution. Also, some parts may depend on the stochastic realization that is used (i.e., what sample of the simulation to visualize?).

**Code input**

The code requires two inputs: one is ‘Tree\_data\_reduced.xlsx’, a data file on tree characteristics that is provided. The second is a set of .csv files, each containing data from one BCI survey. The R data for that can be downloaded from <https://datadryad.org/stash/dataset/doi:10.15146/5xcp-0d46> and it needs to be separated into separate CSV files as explained in the code.

**Code output**

The code generates many intermediate data files. Because they may take long to generate (many days using a 16 core CPU), some intermediate result files (with extension ‘.mat’) are included. The final output is a set of figures and tables.

**Folder structure**

The provided folder contains six sub folders:

* ‘Results’, which contains several scripts and intermediate results
* ‘Clustered\_disp’, which performs the analysis of the clustered dispersal
* ‘CNDD\_HNDD’, which is used to perform the theoretical analysis of the magnitudes of CNDD and HNDD and contains intermediate results as well.
* ‘Dioecious’ – theoretical comparison of spatial patterns in Dioecious vs. Monoecious species
* ‘Edge’
* ‘Meta’ - theoretical comparison of a large closed metacommunity with speciation to the standard DL approach

**Structure of specific scripts**

The master script, ‘Full\_analysis.m’, calls many scripts and their children. The bullets below presents the direct children of the master script and their own children, in hierarchy. If a script is called by (“child of”) more than one script, it appears here only under the first parent in the linear order in which the results are generated.

* ‘generate\_sp\_dat\_table1.m’ – produces the basic data for each species required for their analysis
* ‘run\_sims\_null1\_2008.m’ – runs the standard null model with the HML2008 dispersal kernel. This script sets the parameters for the run and calls the function of the null. The output (as well as the output of other simulations) is generated in the ‘Results’ folder.
  + ‘Null1.m’ – The script of the Standard null model
* ‘run\_sims\_null1\_2001.m’ – runs the standard null model with the HML2001 dispersal kernel. This script sets the parameters for the run and calls the function of the null.
  + ‘Null1\_exp.m’ – the script of the standard null model with an exponential distance distribution.
* ‘run\_sims\_null2\_lag\_2008.m’ - runs the ‘Lagged’ model with the HML2008 dispersal kernel. This script sets the parameters for the run and calls the function of the null.
  + ‘Null2\_lag.m’ – the script of the ‘Lagged’ null model
* ‘run\_sims\_null3\_LDD\_2008.m’ - runs the Standard null model with the LDD dispersal kernel. This script sets the parameters for the run and calls the function of the null.
  + ‘Null3\_LDD.m’ – the script of the standard null model with LDD
* ‘run\_sims\_null4\_initial\_2008.m’ – runs the transient null model, initiating with the observed distribution and randomizing the observed number of demographic events. This script sets the parameters for the run and calls the function of the null.
  + ‘Null4\_init.m’ - the script of the transient null model with LDD.
* ‘run\_sims\_null6\_2008.m’ – runs the fixed null model, where the dynamics happens only on the observed distribution of trees of all species. This script sets the parameters for the run and calls the function of the null.
  + ‘Null6\_Fixed\_Trees.m’ - the script of the Fixed null model with LDD.
* ‘Analyze\_all\_stats3\_log.m’ – Performs the analysis of all the statistics of all the nulls. Saves them separately also updates the main data table.
  + ‘Analyze\_nulls4.m’ – performs the analysis for a single variant of the null model for all species, saves resulting stats.
    - ‘sum\_pop.m’ – generates a data structure of species composition by samples (time) from the samples in space and time
    - ‘dist\_pvals.m’ – performs a two sided test comparing observed statistic to reference distribution
    - ‘Export\_ERL\_env.R’ – an R script for computing the global envelopes, using the GET R package.
  + ‘Analyze\_nulls\_dem3.m’ – performs the analysis of the transient null model, where only new trees are analyzed.
* ‘Analyze\_all\_stats\_CSR1.m’ – performs the analysis with CSR as null, adding the resulting statistics
* ‘Analysis\_CNDD\_HNDD\_Delta7.m’ – runs all theoretical simulations considering the effects of CNDD and HNDD, assuming Delta (range of CNDD and HNDD) is 7 meters. Also generates Figure 3 of the main text and related supplementary figures.
  + ‘sim3.m’ – runs a simulation model with CNDD and HNDD. There are many unused functionalities in this script.
  + ‘sim2\_N.m’ – runs a null simulation, without CNDD or HNDD.
  + ‘Stats\_regime3.m’ – analyzes a simulation regime (null/with DD) computing many summary statistics.
  + ‘Compose\_pieces.m’ – service script to assemble simulation results if simulations were stopped/crashed.
  + ‘complete\_sims.m’ – service script to complete all simulations that finished prematurely
    - ‘sim3\_continue.m’ – continues the stopped simulations to completion (hopefully).
* ‘Analysis\_CNDD\_HNDD\_Delta20.m’ – runs supplementary simulations considering the effects of CNDD and HNDD, assuming Delta (range of CNDD and HNDD) is 20 meters. Also generates appropriate supplementary figure.
* ‘generate\_fig1\_figS1\_ver2.m’ – Generates Figure 1 and S1
* ‘Create\_tab1\_S1\_3.m’ – Generates result tables
* ‘Generate\_Fig2\_and\_SI\_GLOBAL.m’ – Generates Figure 2 and SI Figures with global envelopes.
* ‘generate\_fig\_EA\_dist1.m’ – generates the figure of the theoretical EA(r)
* ‘analyze\_sim\_clustered\_dispersal2.m’ – master script for all analyses of clustered dispersal
  + ‘Run\_sim\_clustered\_dispersal1.m’ – runs the simulations of clustered dispersal, their respective null models, and also runs their analysis
    - ‘Sim\_clustered.m’ – function of the actual simulation with cluster dispersal
    - ‘Stats\_regime\_clustered\_disp.m’ – analyzes the clustered simulation results
    - ‘sim2\_NC.m’ – function of the actual null model simulation incorporating two stages of dispersal
* ‘Analysis\_Dioecious.m’ – located within the ‘Dioecious’ folder, it performs the comparison of spatial patterns between Dioecious vs. Monoecious species. It also generates the plot.
  + ‘sim2\_Die.m’ – simulate a neutral model for a Dioecious species.
* ‘Analysis\_meta.m’ – Located within the ‘Meta’ folder. It performs the comparison of a large closed metacommunity with speciation to the standard DL approach of a smaller metacommunity with species input (‘immigration’) and torus boundary.
  + ‘sim2\_N\_spec\_nobc.m’ – the simulation model of a closed metacommunity with speciation and absorbing boundary.
  + ‘complete\_sims\_meta\_nobc.m’ – a help script used to complete crashed simulations, uses the following simulation version that starts from the last sample taken:
    - ‘sim2\_N\_spec\_nobc\_continue.m’ – a version of the simulation that starts from the last sample.
  + ‘Stats\_regime\_meta.m’ – performs the analysis of the spatial patterns on a data file without predetermined species (as they emerge by speciation). This version also truncates the observed data to a smaller L.
* ‘Analysis\_edge21.m’ – Located within the ‘Edge’ folder. It compares the statistics obtained with torus boundary conditions, as we use, to those obtained by considering neighbors OUTSIDE the observed plot (but that are simulated on a larger landscape), hence without edge effects.
  + ‘Stats\_regime\_unobs.m’ – computes distances while considering neighbors outside the truncated plot.
* ‘Analysis\_power21.m’ - Located within the ‘Edge’ folder. Performs power analysis using the same simulations generated by the ‘Analysis\_edge21.m’ script. Also generates the plot.
  + ‘Stats\_regime\_meta\_export.m’ – exports the raw values of the statistics, to be used for the power analysis