ReadME - IDM for different taxonomic levels

2022-09-14

Introduction

This is a walkthrough of all the files in this GitHub repositiory for easy use of the code. There are three folders in this repository:

- idm_functions
- idm_simulations_new
- Data

The sections that follow will give a breakdown of each folder and the functions for easy understanding.

idm_functions

This is the first folder that should be looked through. It contains that set-up parameter choices and the functions that are used for estimating the indices and the metrics. It is a source to the `idm_simulations`.

1. parameters_for_simulation.R

This script is used to choose the parameter values for the simulation study. It has as inputs:

- constants : Number of sites, species, visits, replicates, etc.
- covariates : covariates simulated for the ecological and detection process.
- parameters : The values of parameters β and α chosen for the simulations.
- interraction: The interraction effect covariance matrix.

It outputs `parameters_50.RData`.

2. fnx_for_estimation.R

This script contains the functions for estimating hills indices, metrics for model evaluation, amongst others. It contains functions such as:

- mbias : Mean bias
- mse : mean square error
- hill_index : Hill indices
- richness : Richness
- shan_index : Shannon index
- subsetting_parameters : Retrieve values from MCMC output

3. function_for_simulation.R

This script is for simulating the data used in this paper. It has as inputs the results stored in `parameters_50.RData`, seed and `shared` which indicated whether all the datasets share the same parameters (*all*), share covariate and interractions (*covariate_inter*) or interraction effect (*interraction*). It outputs:

- mat.species : Species occupancy data
- mat.genus : Group count data
- pis : Proportion used for hills indices
- ecological_cov : covariate for ecological process
- detection_cov : Detection processscovariate
- p.tag : detection probability

- psi.s : occupancy probablity
- z : True presence absece
- richness : Richness estimates
- incidence_hills1 : Incidence hills for $q = 1$
- incidence_hills2 : Incidence hills for $q = 1$
- abundance_hills0 : Hills indices for $q = 0$
- abundance_hills1 : Hills indices for $q = 1$
- abundance_hills2 : Hills indices for $q = 2$

4. function_for_simulation_with_missing.R

This script has the same inputs and output as `function_for_simulation.R`. The difference between the two is that this one assumes that some sites have missing data and have *NA*s assigned to some sites that have missing data. We do not use this in the study.

idm_simulations

This folder contains the simulated data and the script for the analysis of the simulated data. This folder contains:

1. simulation_interraction.R

This script simulates the data needed for the analysis. It uses the `function_for_simulation.R` script and simulates 100 datasets for each `shared` variable (ie. all, interractions, covariate). It outputs the data stored in `sim_interractions_na.RData`.

2. nimble_simulations.R

This script analyses the data using `NIMBLE`. It has inputs:

- simulations_all: The simulated data stored in `sim_interractions_na.RData`.
- method : whether Integrated distribution model (hereafter, IDM), Insect group mode (hereafter, IG) or Species occupance model (hereafter, Spe).
- covariance_prior : whether `full` (use inverse wishart for the interraction covariace matrix) or `LV` (use multiplicative process shrinkage process).
- shared : whether all, interractions, covariate_inter (as discussed above).

It outputs the summary of the alpha's, beta's, z's, lambda's and correlation matrix as well as the values of alpha's and beta's that have converged.

3. idm_miss_na_50

This folder contains three folder for each of the `shared` configuration:

- interractions
- covariate_inter
- all

All these folder contains Rscripts for running the `nimble_simulations.R` script for each of the method (IDM, IG and Spe), and the results are stored in each folder.

4. Plot_simulations1.R

This script is used to plot the results from the simulation study presented in paper.

Data

This folder contains the PoMs data used to parameterize this model. The data used are `FIT_counts.csv` (FIT count data) and `pan_traps.csv` (occupancy data). The `data_format_new.R` was used to format the data to be used as input to the `nimble.R` script. The `nimble.R` script is run for each `shared` configuration and the results are stored in the folder with their respective names.

The `cross_validation` folder is used to run the (un)conditional predictive performance of the models for the 3 insect groups. Check the paper for the details of this approach. The `cross_validation.R` script performs the two-fold cross_validation using an alteration of the NIMBLE run_crossvalidation function (alteration done in `cross_valid_function.R`) script.