**Brief Guide to the iNEXT.3D package and R code for graphics in the MEE paper by Chao et al. (2021)**

First, the following packages must be installed and imported:

library(chaoUtility)

library(readr)

library(ape)

library(dplyr)

library(ggplot2)

library(Rcpp)

library(reshape2)

library(ggpubr)

library(gg.gap)

library(devtools)

Next, install and import the package “**iNEXT.3D”** from Anne Chao’s Github.

install\_github("AnneChao/iNEXT.3D")  # Press 'Enter' to skip number selection.

library(iNEXT.3D)

In the following, we briefly introduce the three main functions in the “**iNEXT.3D”** package.

####################################################################

We use three functions in the **iNEXT.3D** package to compute TD, PD and FD and make plots: “estimate3D”, “asy3D” and “obs3D”.

(1) Function “estimate3D”⸺ computes interpolation and extrapolation of Hill number of order q with defaults:

estimate3D(data, diversity = 'TD', q = c(0,1,2), datatype = "abundance", base = "coverage",

level = NULL, nboot=50, conf = 0.95, nT, PDtree, PDreftime = NULL, PDtype = 'meanPD',

FDdistM, FDtype = "AUC", FDtau = NULL)

(2) Function “asy3D” ⸺ computes the estimated asymptotic diversity of order q with defaults:

asy3D(data, diversity = "TD", q = seq(0, 2, 0.2), datatype = "abundance", nboot = 50, conf = 0.95,

nT, PDtree, PDreftime = NULL, PDtype = "meanPD",

FDdistM, FDtype = "AUC", FDtau = NULL)

(3) Function “obs3D” ⸺ computes the empirical (observed) diversity of order q with defaults:

obs3D(data, diversity = "TD", q = seq(0, 2, 0.2), datatype = "abundance", nboot = 50, conf = 0.95,

nT, PDtree, PDreftime = NULL, PDtype = "meanPD",

FDdistM, FDtype = "AUC", FDtau = NULL)

The description for each argument in the above three functions is given in the following table.

|  |  |
| --- | --- |
| **Argument** | **Description** |
| data | a matrix, data.frame (species by sites), or list of species abundance/incidence frequencies/incidence raw. If datatype = "incidence\_freq", then the first entry of the input data must be total number of sampling units in each column or list. |
| diversity | selection of diversity type: 'TD' = Taxonomic diversity, 'PD' = Phylogenetic diversity, and 'FD' = Functional diversity. |
| q | a numerical vector specifying the diversity orders. |
| datatype | data type of input data: individual-based abundance data (datatype = "abundance"), sampling-unit-based incidence frequencies data (datatype = "incidence\_freq") or species by sampling-units incidence matrix (datatype = "incidence\_raw"). |
| base | Selection of sample-size-based (base="size") or coverage-based (base="coverage") rarefaction and extrapolation. |
| level | a sequence specifying the particular sample sizes or sample coverages (between 0 and 1). If base="size" and level=NULL, then this function computes the diversity estimates for the minimum sample size among all sites extrapolated to double reference sizes. If base="coverage" and level=NULL, then this function computes the diversity estimates for the minimum sample coverage among all sites extrapolated to double reference sizes. |
| nboot | an integer specifying the number of replications. |
| conf | a positive number < 1 specifying the level of confidence interval, default is 0.95. |
| nT | needed only when datatype = "incidence\_raw", a sequence of named nonnegative integers specifying the number of sampling units in each assemblage. If names(nT) = NULL, then assemblage are automatically named as "assemblage1", "assemblage2",..., etc. It is necessary when datatype = "incidence\_raw" and class of data is data.frame.. |
| PDtree | a phylo object describing the phylogenetic tree in Newick format for all observed species in the pooled assemblage. It is necessary when diversity = 'PD'. |
| PDreftime | Select a reference time point for diversity = 'PD'. Default is NULL. |
| PDtype | Select phylogenetic diversity type: PDtype = "PD" for Chao et al. (2010) phylogenetic diversity and PDtype = "meanPD" for mean phylogenetic diversity (phylogenetic Hill number). It will be use when diversity = 'PD'. Default is "meanPD". |
| FDdistM | a pairwise distance matrix for all pairs of observed species in the pooled assemblage. It will be use when diversity = 'FD'. |
| FDtype | a binary selection for FD. FDtype = "tau\_values" computes diversity under certain threshold values. FDtype = "AUC" computes an overall FD which integrates all threshold values between zero and one. Default is "AUC" |
| FDtau | a sequence between 0 and 1 specifying tau. If NULL, threshold = dmean. Default is NULL. It will be use when diversity = 'FD' and FDtype = “tau\_values”. |

Use '?iNEXT3D', '?estimate3D', '?asy3D', '?obs3D' for help.

#####################################################################

To plot Figures 1 to 4 in the MEE paper, read source code provided in the repository:

source("Source R code.txt")

1. **Figure 1 (yearly abundance data)**

Figure 1 (a), (b), (c) plot the yearly temporal diversity patterns for (a) TD, (b) PD, and (c) FD of orders q = 0, 1, and 2 based on four estimation methods with estimates from high to low: asymptotic estimates (column 1), standardized estimates under a coverage value of *Cmax* (column 2), observed diversity (column 3), and standardized estimates under a coverage of *Cmin* (column 4).

**1a. R code for Figure 1a (TD)**

Species yearly abundance data are included in the file “Fish Abundance data.csv” and are stored

as a species(row) by plot (column) data matrix; each element in the matrix represents the number of individuals.

First, copy the file “Fish Abundance data.csv” into a working directory, and load (or import) the data file in your R console:

Abun <- read.csv("Fish abundance data.csv", row.names = 1, header= TRUE)

Then run the following code obtain *Cmin* and *Cmax*, where *Cmin* is the smallest observed coverage among all assemblages, and *Cmax* is the smallest coverage after extrapolating double sample size among all assemblages.

Cmax <- apply(Abun, 2, function(x) iNEXT.3D:::Chat.Ind(x, 2\*sum(x))) %>% min %>% round(., 4)

Cmin <- apply(Abun, 2, function(x) iNEXT.3D:::Chat.Ind(x, sum(x))) %>% min %>% round(., 4)

Then separately use “estimate3D”, “obs3D” and “asy3D” to obtain the TD estimates.

1. Use function “estimate3D” to calculate the standardized taxonomic diversity with order q=0, 1, and 2 at the coverage levels of *Cmax* and *Cmin*.

TD\_est <- estimate3D(data = Abun, diversity = 'TD', q = c(0, 1, 2), datatype = 'abundance',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0)

1. Use function “obs3D” to calculate the observed taxonomic diversity with order q=0, 1,and 2.

TD\_obs <- obs3D(data = Abun, diversity = 'TD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0)

1. Use function “asy3D” to calculate the asymptotic taxonomic diversity with order q=0, 1,and 2.

TD\_asy <- asy3D(data = Abun, diversity = 'TD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0)

Finally, run the following code to obtain Figure 1a. The code for fitting of quartic and linear trends is included in the function fig\_1\_or\_3.

out\_TD <- rbind(TD\_est %>% select(Assemblage, Order.q, qD, goalSC),

TD\_obs %>% select(Assemblage, Order.q, qD) %>% mutate(goalSC = 'Observed') ,

TD\_asy %>% select(Assemblage, Order.q, qD) %>% mutate(goalSC = 'Asymptotic') )

fig\_1\_or\_3(out\_TD, y\_label = 'Taxonomic diversity')

**1b. R code for Figure 1b (PD)**

To compute PD, species abundance data (“Fish abundance data.csv”), and species phylogenetic tree data (“Fish phyloTree.txt”) in Newick format are required. Copy the files “Fish abundance data.csv” and "Fish phyloTree.txt" in your working directory and load (or import) the data sets in your R console:

Abun <- read.csv("Fish abundance data.csv", row.names = 1, header= TRUE)

tree <- read.tree("Fish phyloTree.txt")

Then, separately use “estimate3D”, “obs3D” and “asy3D” to obtain the PD estimates.

1. Use function “estimate3D” to calculate the standardized phylogenetic diversity with order q=0, 1, and 2 at the coverage levels of *Cmin* and *Cmax*.

PD\_est <- estimate3D(data = Abun, diversity = 'PD', q = c(0, 1, 2), datatype = 'abundance',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0, PDtree = tree, PDreftime = 1)

1. Use function “obs3D” to calculate the observed phylogenetic diversity with order q=0, 1, and 2.

PD\_obs <- obs3D(data = Abun, diversity = 'PD', q = c(0, 1, 2), datatype = 'abundance',

nboot = 0, PDtree = tree, PDreftime = 1)

1. Use function “asy3D” to calculate the asymptotic phylogenetic diversity with order q=0, 1, and 2.

PD\_asy <- asy3D(data = Abun, diversity = 'PD', q = c(0, 1, 2), datatype = 'abundance',

nboot = 0, PDtree = tree, PDreftime = 1)

Finally, execute the following code to obtain Figure 1b. The code for fitting of quartic and linear trends is included in the function fig\_1\_or\_3.

out\_PD <- rbind(PD\_est %>% select(Assemblage, Order.q, qPD, goalSC),

PD\_obs %>% select(Assemblage, Order.q, qPD) %>% mutate(goalSC = 'Observed'),

PD\_asy %>% select(Assemblage, Order.q, qPD) %>% mutate(goalSC = 'Asymptotic') )

fig\_1\_or\_3(out\_PD, y\_label = 'Phylogenetic diversity')

**1c. R code for Figure 1c (FD)**

To compute FD, species abundance data (“Fish abundance data.csv”), and species trait data (“Fish traits.csv”) are required. Copy the two files “Fish abundance data.csv” in a working directory and load (or import) the data sets in your R console:

Abun <- read.csv("Fish abundance data.csv", row.names = 1, header= TRUE)

traits <- read.csv("Fish traits.csv", row.names = 1, header= TRUE)

Run the following code to obtain the Gower distances between any two species:

for (i in 1:ncol(traits)) { if (class(traits[,i]) == "character") traits[,i] <- factor(traits[,i], levels =unique(traits[,i]))}

distM <- cluster::daisy(x = traits, metric = "gower") %>% as.matrix()

Then, separately use “estimate3D”, “obs3D” and “asy3D” to obtain the FD estimates.

a. Use function “estimate3D” to calculate the standardized functional diversity with order q=0, 1, and 2 at the two coverage levels of *Cmin* and *Cmax*.

FD\_est <- estimate3D(data = Abun, diversity = 'FD', q = c(0, 1, 2), datatype = 'abundance',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0, FDdistM = distM)

b. Use function “obs3D” to calculate the observed functional diversity with order q=0, 1, and 2.

FD\_obs <- obs3D(data = Abun, diversity = 'FD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0, FDdistM = distM)

c. Use function “asy3D” to calculate the asymptotic functional diversity with order q=0, 1, and 2.

FD\_asy <- asy3D(data = Abun, diversity = 'FD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0, FDdistM = distM)

Finally, run the following codes to obtain the Figure 1c. The code for fitting of quartic and linear trends is included in the function fig\_1\_or\_3.

out\_FD <- rbind(FD\_est %>% select(Assemblage, Order.q, qFD = qAUC, goalSC),

FD\_obs %>% select(Assemblage, Order.q, qFD = qAUC) %>% mutate(goalSC = 'Observed') ,

FD\_asy %>% select(Assemblage, Order.q, qFD = qAUC) %>% mutate(goalSC = 'Asymptotic'))

fig\_1\_or\_3(out\_FD, y\_label = 'Functional diversity')

1. **Figure 2**

Figure 2 (yearly abundance data with TD, PD and FD being plotted in the same figure for each order q = 0, 1 and 2) based on the same output as Figure 1. Construct Figure 2a, b and c by running the following code:

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 0)

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 1)

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 2)

1. Figure 3 (three-year incidence data)

Figure 3 (a), (b), (c) plot the three-year temporal diversity patterns for (a) TD, (b) PD, and (c) FD of orders q = 0, 1, and 2 based on monthly incidences for four estimation methods (same as those in Figure 1). For incidence-based data, there are two options for data input:

incidence-raw data: for each assemblage, input

data consist of a species-by-sampling-unit matrix; when there

are N assemblages, input data consist of N matrices via a list

object, with each matrix being a species-by-sampling-unit

matrix. In iNEXT, this type of data is speciﬁed by

datatype="inciden ce\_raw" . (ii) Incidence-frequen cy

data: input data for each assemblage consist of the number of

sampling units (T) followed by the observed incidence frequen-

cies (Y

1

, Y

2

, ..., Y

S

). When there are N assemblages, input

data consist of an S+1byN matrix or N lists of species

incidence frequencies. The ﬁrst entry of each column/list must

be the total number of sampling units, followed by the species

incidence frequencies. In iNEXT, this type of data is speciﬁed

by datatype="incidence\_freq"

incidence-raw data: for each assemblage, input

data consist of a species-by-sampling-unit matrix; when there

are N assemblages, input data consist of N matrices via a list

object, with each matrix being a species-by-sampling-unit

matrix. In iNEXT, this type of data is speciﬁed by

datatype="inciden ce\_raw" . (ii) Incidence-frequen cy

data: input data for each assemblage consist of the number of

sampling units (T) followed by the observed incidence frequen-

cies (Y

1

, Y

2

, ..., Y

S

). When there are N assemblages, input

data consist of an S+1byN matrix or N lists of species

incidence frequencies. The ﬁrst entry of each column/list must

be the total number of sampling units, followed by the species

incidence frequencies. In iNEXT, this type of data is speciﬁed

by datatype="incidence\_freq"

(1). Incidence-raw data: for each assemblage, input data consist of a species-by-sampling-unit matrix; when there are N assemblages, input data consist of N matrices via a list object, with each matrix being a species-by-sampling-unit matrix. This type of data is speciﬁed by datatype = "incidence\_raw". You could also input a column-binded incidence raw data matrix. For example, if there are 5 sampling units in Assemblage 1 and 7 sampling units in Assemblage 2, you can input an S (species) by 12 matrix. In this case, the number of sampling units (nT) for each assemblage must be specified in a separate file.

(ii) Incidence-frequency data: input data for each assemblage consist of the number of sampling units (nT) followed by the observed species incidence frequencies. When there are S species and N assemblages, input data consist of an S+1 by N matrix of species incidence frequencies. The ﬁrst entry of each column must be the total number of sampling units, followed by the species incidence frequencies. This type of data is speciﬁed by datatype="incidence\_freq".

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**3a. R code for Figure 3a (TD)**

One type of dataset is needed: three yearly species incidence data ("Fish incidence raw data.csv").

First, copy the file "Fish incidence raw data.csv" in your working directory.

and load(or import) the data set in your R console:

Inci\_raw <- read.csv("Fish incidence raw data.csv", row.names = 1, header= TRUE)

Second, the number of sampling units in each three yearly incidence data can be obtained by

nT <- read.csv('nT for incidence data.csv', row.names = 1)

and then execute following instructions to obtain *Cmin* and *Cmax*, where

*Cmin* is the smallest observed coverage among all assemblages, and

*Cmax* is the smallest coverage after extrapolating double sample size among all assemblages.(see Chao et al.(2019) for detail)

Cmax <- sapply(1:length(nT), function(i) rowSums( Inci\_raw[, (sum(nT[1:i]) - sum(nT[i]) + 1) : sum(nT[1:i])] )) %>% rbind(as.integer(nT),.) %>%

apply(., 2, function(x) iNEXT.3D:::Chat.Sam(x, 2\*x[1])) %>% min %>% round(., 4)

Cmin <- sapply(1:length(nT), function(i) rowSums( Inci\_raw[, (sum(nT[1:i]) - sum(nT[i]) + 1) : sum(nT[1:i])] )) %>% rbind(as.integer(nT),.) %>%

apply(., 2, function(x) iNEXT.3D:::Chat.Sam(x, x[1])) %>% min %>% round(., 4)

Then separately use “estimate3D”, “obs3D” and “asy3D” to obtain the estimates of taxonomic diversity at different levels.

1. Use function “estimate3D” to calculate the taxonomic diversity with order q=0, 1, and 2 at the sample coverage of *Cmax* and *Cmin*.

TD\_est <- estimate3D(data = Inci\_raw, diversity = 'TD', q = c(0, 1, 2), datatype = 'incidence\_raw',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0, nT = nT)

1. Use function “obs3D” to calculate the observed taxonomic diversity with order q=0, 1, and 2.

TD\_obs <- obs3D(data = Inci\_raw, diversity = 'TD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT)

1. Use function “asy3D” to calculate the asymptotic taxonomic diversity with order q=0, 1, and 2.

TD\_asy <- asy3D(data = Inci\_raw, diversity = 'TD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT)

Finally, execute the following two instructions to obtain the Figure 1a.

oout\_TD <- rbind(TD\_est %>% select(Assemblage, Order.q, qD, goalSC),

TD\_obs %>% select(Assemblage, Order.q, qD) %>% mutate(goalSC = 'Observed'),

TD\_asy %>% select(Assemblage, Order.q, qD) %>% mutate(goalSC = 'Asymptotic'))

fig\_1\_or\_3(out\_TD, y\_label = 'Taxonomic diversity')

**3b. Construct Figure 3b (theee yearly phylogenetic diversity pattern)**

Two types of dataset are needed: three yearly species incidence raw data ("Fish incidence raw data.csv"), and species phylogenetic tree data ("Fish phyloTree.txt").

First, copy the files "Fish incidence raw data.csv" and "Fish phyloTree.txt" in your working directory and load(or import) the data sets in your R console:

Inci\_raw <- read.csv("Fish incidence raw data.csv", row.names = 1, header= TRUE)

tree <- read.tree("Fish phyloTree.txt")

Second, the number of sampling units in each three yearly incidence data can be obtained by

nT <- read.csv('nT for incidence data.csv', row.names = 1)

Then, separately use “estimate3D”, “obs3D” and “asy3D” to obtain the estimates of phylogenetic diversity at different levels.

1. Use function “estimate3D” to calculate the phylogenetic diversity with order q=0, 1, and 2 at the sample coverage of *Cmin* and *Cmax*.

PD\_est <- estimate3D(data = Inci\_raw, diversity = 'PD', q = c(0, 1, 2), datatype = 'incidence\_raw',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0, nT = nT, PDtree = tree, PDreftime = 1)

1. Use function “obs3D” to calculate the observed phylogenetic diversity with order q=0, 1, and 2.

PD\_obs <- obs3D(data = Inci\_raw, diversity = 'PD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, PDtree = tree, PDreftime = 1)

1. Use function “asy3D” to calculate the asymptotic phylogenetic diversity with order q=0, 1,and 2.

D\_asy <- asy3D(data = Inci\_raw, diversity = 'PD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, PDtree = tree, PDreftime = 1)

Finally, execute the following two instructions to obtain the Figure 3b.

out\_PD <- rbind(PD\_est %>% select(Assemblage, Order.q, qPD, goalSC),

PD\_obs %>% select(Assemblage, Order.q, qPD ) %>% mutate(goalSC = 'Observed'),

PD\_asy %>% select(Assemblage, Order.q, qPD) %>% mutate(goalSC = 'Asymptotic'))

fig\_1\_or\_3(out\_PD, y\_label = 'Phylogenetic diversity')

**3c. Construct Figure 3c (three yearly functional diversity pattern)**

Two types of dataset are needed: three yearly species incidence data("Fish incidence frequency data.csv"), and species functional trait data ("Fish traits.csv").

First, copy the files "Fish incidence frequency data.csv" and "Fish traits.csv" in your working directory and load(or import) the data sets in your R console:

Inci\_raw <- read.csv("Fish incidence raw data.csv", row.names = 1, header= TRUE)

traits <- read.csv("Fish traits.csv", row.names = 1, header= TRUE)

Second, the number of sampling units in each three yearly incidence data can be obtained by

nT <- read.csv('nT for incidence data.csv', row.names = 1)

and execute following instructions to obtain functional distances among species by using Gower metrics.

for (i in 1:ncol(traits)) {if (class(traits[,i]) == "character") traits[, i] <- factor(traits[,i], levels = unique(traits[, i]))}

distM <- cluster::daisy(x = traits, metric = "gower") %>% as.matrix()

Then, separately use “estimate3D”, “obs3D” and “asy3D” to obtain the estimates of functional diversity at different levels.

1. Use function “estimate3D” to calculate the functional diversity with order q=0, 1, and 2 at the sample coverage of *Cmin* and *Cmax*, separately.

FD\_est <- estimate3D(data = Inci\_raw, diversity = 'FD', q = c(0, 1, 2), datatype = 'incidence\_raw',

base = 'coverage', level = c(Cmin, Cmax), nboot = 0, nT = nT, FDdistM = distM)

1. Use function “obs3D” to calculate the observed functional diversity with order q=0, 1, and 2.

FD\_obs <- obs3D(data = Inci\_raw, diversity = 'FD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, FDdistM = distM)

1. Use function “asy3D” to calculate the asymptotic functional diversity with order q=0, 1, and 2.

FD\_asy <- asy3D(data = Inci\_raw, diversity = 'FD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, FDdistM = distM)

Finally, execute the following two instructions to obtain the Figure 3c.

out\_FD <- rbind(FD\_est %>% select(Assemblage, Order.q, qFD = qAUC, goalSC),

FD\_obs %>% select(Assemblage, Order.q, qFD = qAUC) %>% mutate(goalSC = 'Observed'),

FD\_asy %>% select(Assemblage, Order.q, qFD = qAUC) %>% mutate(goalSC = 'Asymptotic'))

fig\_1\_or\_3(out\_FD, y\_label = 'Functional diversity')

1. Figure 4

Since Figure3 and Figure4 use identical running outputs, and show different types of presentation.

Figure3 presents the three yearly diversity pattern with order q=0,1,2 in the same figure for TD, PD, and FD.

Figure4 presents the three yearly patterns of TD, PD, and TD in the same figure for each order q.

Construct Figure 4a by executing the instruction.

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 0)

Construct Figure 4b by executing the instruction.

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 1)

Construct Figure 4c by executing the instruction.

fig\_2\_or\_4(TD.output = out\_TD, PD.output = out\_PD, FD.output = out\_FD, q = 2)