**Latest version (August 2022)**

**A Brief Guide to the iNEXT.3D Package and R code for Graphics in the MEE Paper by Chao et al. (2021)**

To make the graphics shown in Figures 1 to 4 in Chao et al.’s (2021) paper, all data and R code are included in a zipped file and also available in Anne Chao’s Github repository <https://github.com/AnneChao/MEE_iNEXT.3D>.

Before using the data and code, the following packages in CRAN must be installed and imported:

library(ape)

library(dplyr)

library(ggplot2)

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' key to skip update options

library(iNEXT.3D)

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

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

We use two functions (“estimate3D” and “ObsAsy3D”) in the **iNEXT.3D** package to compute TD, PD and FD and make the plots in Chao et al.’s (2021) paper.

(1) Function “estimate3D”⸺ computes interpolated and extrapolated diversity estimates of orders q = 0, 1 and 2 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)

Under the default diversity = "TD", all the arguments for PD (PDtree, PDreftime, PDtype) and for FD (FDdistM, FDtype, FDtau) can be omitted; input for these arguments are ignored.

(2) Function “ObsAsy3D” ⸺ computes asymptotic diversity of a sequence of diversity orders q with defaults:

ObsAsy3D (data, diversity = "TD", q = seq(0, 2, 0.2), datatype = "abundance", nboot = 50, conf = 0.95,

nT, method = c("Asymptotic", "Observed"), PDtree, PDreftime = NULL, PDtype = "meanPD",

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

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

|  |  |
| --- | --- |
| **Argument** | **Description** |
| data | (a) For datatype = "abundance", data can be input as a vector of species abundances (for a single assemblage), matrix/data.frame (species by assemblages), or a list of species abundance vectors.  (b) For datatype = "incidence\_freq", data can be input as a vector of incidence frequencies (for a single assemblage), matrix/data.frame (species by assemblages), or a list of incidence frequencies; the first entry for all types of input must be the number of sampling units in each assemblage.  (c) For datatype = "incidence\_raw", data can be input as a list of matrix/data.frame (species by sampling units); data can also be input as a matrix/data.frame by merging all sampling units across assemblages based on species identity; in this case, the number of sampling units (nT, see below) must be input. |
| diversity | selection of diversity type: "TD" = Taxonomic diversity, "PD" = Phylogenetic diversity, and "FD" = Functional diversity. |
| q | a numerical vector specifying diversity orders. |
| datatype | data type of input data: individual-based abundance data (datatype = "abundance"), sampling-unit-based incidence-frequency data (datatype = "incidence\_freq"), or species by sampling-units incidence matrix (datatype = "incidence\_raw") with all entries being 0 (non-detection) or 1 (detection) |
| base | selection of sample-size-based (base = "size") or coverage-based (base = "coverage") rarefaction and extrapolation. |
| level | A numerical vector specifying the particular sample sizes or sample coverages (between 0 and 1). If base = "coverage" (default) and level = NULL, then this function computes the diversity estimates for the minimum sample coverage among all samples extrapolated to double reference sizes. If base = "size" and level = NULL, then this function computes the diversity estimates for the minimum sample size among all samples extrapolated to double reference sizes. |
| nboot | an integer specifying the number of bootstrap replications. |
| conf | a positive number < 1 specifying the level of confidence interval. Default is 0.95. |
| nT | (required only when datatype = "incidence\_raw" and input data is matrix/data.frame) a vector of nonnegative integers specifying the number of sampling units in each assemblage. If assemblage names are not specified, then assemblages are automatically named as "assemblage1", "assemblage2",..., etc. |
| PDtree | (required only when diversity = "PD"), a phylogenetic tree in Newick format for all observed species in the pooled assemblage. |
| PDreftime | (required only when diversity = "PD"), a vector of numerical values specifying reference times for PD. Default is NULL (i.e., the age of the root of the PDtree). |
| PDtype | (required only when diversity = "PD"), select PD type: PDtype = "PD" (effective total branch length) or PDtype = "meanPD" (effective number of equally divergent lineages). Default is "meanPD", where meanPD = PD/tree depth. |
| FDdistM | (required only when diversity = "FD"), a species pairwise distance matrix for all species in the pooled assemblage. |
| FDtype | (required only when diversity = "FD"), select FD type: FDtype = "tau\_values" for FD under specified threshold values, or FDtype = "AUC" (area under the curve of tau-profile) for an overall FD which integrates all threshold values between zero and one. Default is "AUC". |
| FDtau | (required only when diversity = "FD" and FDtype = "tau\_values"), a numerical vector between 0 and 1 specifying tau values (threshold levels). If NULL (default), then threshold is set to be the mean distance between any two individuals randomly selected from the pooled assemblage (i.e., quadratic entropy). |

Use '?iNEXT3D', '?estimate3D', '? ObsAsy3D ' for help.

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

The source code for graphics (Figures 1 to 4) in the MEE paper is provided in the zipped file and the Github repository. First, import/load the source code into the R environment:

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 in each sub-figure), 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 import/load the data into the R console:

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

Run the following code to obtain the values of *Cmin* and *Cmax*, where *Cmin* is the smallest coverage value among all samples, and *Cmax* is the smallest coverage among all samples extrapolated to double reference sample sizes.

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

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

Then separately use “estimate3D” and “ObsAsy3D” to obtain the TD estimates.

Use function “estimate3D” to calculate the standardized TD of diversity orders 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)

Use function “ObsAsy3D” to calculate the observed TD of diversity orders q = 0, 1, and 2.

TD\_obs <- ObsAsy3D(data = Abun, diversity = 'TD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0,

method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic TD of diversity orders q = 0, 1, and 2.

TD\_asy <- ObsAsy3D(data = Abun, diversity = 'TD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0,

method = 'Asymptotic')

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

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

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

TD\_asy %>% select(Assemblage, Order.q, qTD) %>% mutate(SC = '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" into your working directory and import/load the two data sets into the R console:

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

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

Then, separately use “estimate3D” “ObsAsy3D” to obtain the PD estimates.

Use function “estimate3D” to calculate the standardized PD of orders 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)

Use function “ObsAsy3D” to calculate the observed PD of orders q = 0, 1, and 2.

PD\_obs <- ObsAsy3D(data = Abun, diversity = 'PD', q = c(0, 1, 2), datatype = 'abundance',

nboot = 0, PDtree = tree, PDreftime = 1 , method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic PD of orders q = 0, 1, and 2.

PD\_asy <- ObsAsy3D(data = Abun, diversity = 'PD', q = c(0, 1, 2), datatype = 'abundance',

nboot = 0, PDtree = tree, PDreftime = 1 , method = 'Asymptotic')

Finally, run 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, SC),

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

PD\_asy %>% select(Assemblage, Order.q, qPD) %>% mutate(SC = '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” into your working directory and import/load the two data sets into the 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”, “ObsAsy3D” to obtain the FD estimates.

Use function “estimate3D” to calculate the standardized FD of orders 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)

Use function “ObsAsy3D” to calculate the observed FD of orders q = 0, 1, and 2.

FD\_obs <- ObsAsy3D(data = Abun, diversity = 'FD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0,

FDdistM = distM , method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic FD of orders q = 0, 1, and 2.

FD\_asy <- ObsAsy3D(data = Abun, diversity = 'FD', q = c(0, 1, 2), datatype = 'abundance', nboot = 0,

FDdistM = distM , method = 'Asymptotic')

Finally, run the following code to obtain 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, SC),

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

FD\_asy %>% select(Assemblage, Order.q, qFD) %>% mutate(SC = '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 based on detection/non-detection in each month)**

Figure 3 (a), (b), (c) plots 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"

(i) Incidence-raw data (datatype = "incidence\_raw"): When there are N assemblages, input data consist of a list of N matrices/data.frames (species-by-sampling-unit matrices); each entry is either 0 (non-detection) or 1 (detection). Data can also be input as a large matrix/data.frame by merging all sampling units across assemblages based on species identity; in this case, the number of sampling units (nT, see below for an example) must be input. For example, if there are 5 sampling units in Assemblage I and 7 sampling units in Assemblage II, the data consist of a matrix/data.frame with S species (rows) and 12 sampling units (columns). Here, the number of sampling units for each assemblage must be specified in a separate data file (nT).

(ii) Incidence-frequency data (datatype = "incidence\_freq"): input data for each assemblage consist of the number of sampling units followed by the observed species incidence frequencies. When there are S species and N assemblages, input data consist of 1+S rows (sampling unit number plus S species incidence frequencies) and N columns (number of assemblages). Here the ﬁrst entry of each column must be the total number of sampling units, followed by species incidence frequencies.

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

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

Based on monthly incidence data, three-year incidence raw data are included in the file ("Fish incidence raw data.csv"). First, copy the file "Fish incidence raw data.csv" into your working directory and import/load the data into the R console:

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

Second, import/load the data specifying the number of sampling units (i.e., the number of months with data) in each three-year period. Note that data for some months were not collected.

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

Then run the following code to obtain the values of *Cmin* and *Cmax*. Note the two values based on incidence data are different from those based on abundance data.

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:::Coverage(x, 'incidence\_freq', 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:::Coverage(x, 'incidence\_freq', x[1])) %>% min %>% round(., 4)

Then, separately use “estimate3D” and “ObsAsy3D” to obtain the estimates of TD.

Use function “estimate3D” to calculate the standardized TD of orders q = 0, 1, and 2 at the coverage levels 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)

Use function “ObsAsy3D” to calculate the observed TD of orders q = 0, 1, and 2.

TD\_obs <- ObsAsy3D(data = Inci\_raw, diversity = 'TD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT , method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic TD of orders q = 0, 1, and 2.

TD\_asy <- ObsAsy3D(data = Inci\_raw, diversity = 'TD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT , method = 'Asymptotic')

Finally, run the following code to obtain Figure 3a.

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

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

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

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

**3b. Construct Figure 3b (PD)**

To compute PD, species incidence raw data ("Fish incidence raw data.csv") and species phylogenetic tree data ("Fish phyloTree.txt") are required. First, copy the files "Fish incidence raw data.csv" and "Fish phyloTree.txt" into your working directory and import/load the two data sets into your R console:

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

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

Second, import/load the data specifying number of sampling units in each three-year period:

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

Then, separately use “estimate3D” and “ObsAsy3D” to obtain the estimates of PD.

Use function “estimate3D” to calculate the standardized PD with orders q = 0, 1, and 2 at the two coverage levels 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)

Use function “ObsAsy3D” to calculate the observed PD of orders q = 0, 1, and 2.

PD\_obs <- ObsAsy3D(data = Inci\_raw, diversity = 'PD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, PDtree = tree, PDreftime = 1 , method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic PD of orders q = 0, 1, and 2.

PD\_asy <- ObsAsy3D(data = Inci\_raw, diversity = 'PD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, PDtree = tree, PDreftime = 1 , method = 'Asymptotic')

Finally, run the following code to obtain Figure 3b.

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

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

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

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

**3c. Construct Figure 3c (FD)**

To compute FD, species incidence data ("Fish incidence raw data.csv") and species functional trait data ("Fish traits.csv") are required. First, copy the files "Fish incidence raw data.csv") and "Fish traits.csv" into your working directory and import/load the two data sets into 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, import/load the data specifying number of sampling units in each three-year period:

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

Then run the following code to obtain species pairwise distances based on the Gower metric.

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” and “ObsAsy3D” to obtain the estimates of FD.

Use function “estimate3D” to calculate the standardized FD of orders 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)

Use function “ObsAsy3D” to calculate the observed FD of orders q = 0, 1, and 2.

FD\_obs <- ObsAsy3D(data = Inci\_raw, diversity = 'FD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, FDdistM = distM , method = 'Observed')

Use function “ObsAsy3D” to calculate the asymptotic FD of orders q = 0, 1, and 2.

FD\_asy <- ObsAsy3D(data = Inci\_raw, diversity = 'FD', q = c(0, 1, 2), datatype = 'incidence\_raw',

nboot = 0, nT = nT, FDdistM = distM , method = 'Asymptotic')

Finally, run the following code to obtain Figure 3c.

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

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

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

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

1. **Figure 4**

Figure 4 (three-year incidence data with TD, PD and FD being plotted in the same figure for each order q = 0, 1 and 2) is based on the same output as Figure 3. Construct Figure 4a, 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)