**‘DiMBio’ Package: Tutorial**

March 2024

Version 1.0.1

Date 2024-03-16

Discrimination of Biological Models

Alexis Valerya, Roni Contrerasa, Rossana Timaureb, Marc Pansuc, Lesmes A. M. Jerezd

1. Grupo de Investigación en Biotecnología Agrícola y Ambiental (GIBAA). Universidad Nacional Experimental del Táchira UNET. San Cristóbal, Táchira, Venezuela. [avalery@unet.edu.ve](mailto:avalery@unet.edu.ve), [ronnyalberto.contreras@unet.edu.ve](mailto:ronnyalberto.contreras@unet.edu.ve)
2. Instituto de Estadística Aplicada y Computación (IEAC), Universidad de Los Andes ULA. Mérida Venezuela. [rttg@mail.com](mailto:rttg@mail.com)
3. UMR Eco & Sols, IRD, INRAe, CIRAD, Univ Montpellier, Supagro, Bât. 12, 2 Place Viala, 3406, Montpellier Cedex 2, France. [marc.pansu@gmail.com](mailto:marc.pansu@gmail.com)
4. Department of Earth and Environment. Florida International University. Miami, FL. USA. lmorajer@fiu.edu

DiMBio is a statistical tool that allows to discriminate results from simulated models of biological processes. This statistical tool package incorporates both univariate and multivariate statistics, avoiding the use of assumptions based on comparisons between two datasets for determining their potential differences.

**Univariate statistical tests** of this tool include:

* Coefficient Determination (*R2*)
* Pearson’s Correlation Coefficient (*ρ*)
* Student’s t-test for paired data (*t*)
* Nash-Sutcliffe efficiency index (*NSE*)
* Percentage of the root mean square error (*%RMSE*)
* Mean square error of the prediction (*MSEP*)
* Root of the mean square error of the prediction (*RMSEP*)
* Mean absolute error (*MAE*), (ix) mean absolute percentage error (*%MAE*)
* Theil’s coefficient of inequality (*U*), (xi) determination of errors for linear regression (*MC, SC,* and *RC*)
* Maximum likelihood (*Lj*)
* Akaike information criterion (*AIC*)

This tool also includes the following **multivariate statistical tests**:

* Mahalanobis distance (*D2*)
* Euclidean distance (*dE*)
* Manhattan distance (*dL1*)

**Install the package and run via R (DIMBIO\_0.2.0.tar.gz)**

Install.packages("DIMBIO”)

library(DIMBIO)

**#Load the data using the format that were saved**

data <- read.table("C:/……………/Data.xlsx")

**#Real data. Create a data.frame using only columns that correspond with either the experimental data set or the simulated data set from the first model.**

Reales <- data.frame(data$variable\_1, data$variable\_2, …. data$variable\_n)

**#Simulated Data. Create a data.frame using only columns that correspond with either the experimental dataset or the simulated dataset from the second model.**

Simulados <- data.frame(data$variable\_1s, data$variable\_2s, …. data$variable\_ns)

**#log transformation of real and simulated data**

logReales<-log10(reales)

logSimulados<-log10(simulados)

**#Determination of multivariate statistics**

Manhattan(logReales, logSimulados)

Euclidiana(logReales, logSimulados)

Mahalanobis(logReales, logSimulados)

**#Determination of univariate statistics, identified with compared variables**

nombres <-c(“variable\_1”, “variable\_2”, …. “variable\_n”)

Univariantes(reales, simulados, nombres)

Check out provided example

**Notes –** Meaning of the following terms:

Reales: Experimental/observed data

Simulados: Simulated data

Nombres: Name/label of the dataset