multifit *Multi-scale analysis for landscape ecology*

**Description**

Function *multifit* facilitates multi-scale data analysis in landscape ecology. The user provides a data.frame containing a column with the response variable and several columns depicting a particular landscape attribute at different spatial scales. Also, the user must provide the type of model to be applied for the analysis, along with a formula and any other relevant arguments, and the criterion to be used for the selection of the ‘best’ model. The function’s output includes the following elements: a plot depicting the strength of each model, an optional plot showing the estimates of the response variable for each model, and a list containing relevant information about the models (including the plot and the models themselves).

**Dependencies**

The function requires the correspondent package that contains the function that will be used to run the models for the multi-scale analysis.

**Usage**

multifit(mod, multief, formula = NULL, data = NULL, args = NULL,

criterion = "AIC", signif = TRUE, alpha = 0.05, print\_sum =

FALSE, plot\_est = FALSE, xlab = "Radio [m]", labels = NULL,

type = "b", ...)

**Arguments**

|  |  |
| --- | --- |
| mod | string depicting the type of model to be used (see details) |
| multief | character. A vector containing the column names of the data.frame holding the landscape attributes at different spatial scales |
| formula | formula to be applied to each model, labeling the landscape effect as multief (see details) |
| data | data.frame containing the response variable and the landscape attributes at different spatial scales |
| args | character vector with any additional argument/s for the models (see details) |
| criterion | character. Criterion of selection of the ‘best’ model. So far, one of three options (“AIC”, “BIC” or “R2”) or a user-defined function specifying the name of the function in a first element and the model-selection criterion (“max” or “min”) in a second one (see details). |
| signif | logical. Plot the significance codes of each model? |
| alpha | numeric value, between 0 and 1. Statistical significance level (only relevant for the plotting of significance stars) |
| print\_sum | logical. Print the summary of the best model? |
| plot\_est | logical. Plot the estimates of the models? |
| xlab | character. A title for the x axis |
| labels | character vector for the labels of the models at the x axis |
| type | character. What type of plot should be drawn (see details)? |
| ... | arguments to be passed to methods, such as graphical parameters (see ?plot or ?par). Only for aesthetic options of the plots |

**Details**

The aim of this function is to facilitate the analysis of ecological data in relation to any attribute of landscapes, via a multi-scale approach. In this way, *multifit* allows the user to run many statistical models at the same time (i.e. one model per spatial scale), and simplifies the analysis and selection of the appropriate spatial scale for the provided response variable.

First of all, the user must possess a data.frame with at least a column with the response variable to be analyzed, and several columns with information about the landscape attributes (e.g. habitat amount) at different spatial scales (i.e. one column per spatial scale). This data.frame must be specified in the argument data of the function.

The user must provide the statistical type of model to be applied in the analysis, specifying it as a character in the argument mod (e.g. “lm” for a classic linear model). Take into account that the user must have loaded the correspondent package containing the function before running the multi-scale analysis. In the argument multief, the user must provide a character vector depicting the names of the columns that contain the information of the landscape attribute in *data* (e.g. multief = c(“R\_500”, “R\_1000”, “R\_1500”), which refers to the landscape attribute at three different spatial scales: radius of 500, 1000 and 1500 m). Is important to provide the elements of the vector in an order that makes sense (which probably would be an order representing an increase in the spatial scale), as this order will be kept for the multi-scale analysis.

The argument formula must be fulfilled with the statistical formula to be applied to the models. This must include at least the main response variable and a predictor variable named ‘multief’ (e.g. formula = richness ~ multief). Function *multifit* will recognize this particular string in each model as the predictor variable containing the landscape attribute at a particular spatial scale. If the model definition of the function defined in mod does not contain an argument called formula, then the response and predictor variables can be defined in the argument args(see below).

The user may add as many arguments as needed to run the models at each spatial scale. These must be added in the argument args as a vector of characters, each element depicting a particular argument written as the user would in an individual analysis (e.g. assuming a classic linear model, args = c(“na.action = na.omit”, “singular.ok = FALSE”)). As explained above, argscan include the response and predictor variables for those functions that do not include an argument named formula by specifying them in the correspondent arguments.

The criterionargument must include the criteria to be used for the selection of the ‘best’ model among the different spatial scales (i.e. the one with the strongest relationship with the response variable). So far, *multifit* allows choosing between three options: “R2” (for R squared, i.e. coefficient of determination), “AIC” (for Akaike Information Criterion), and “BIC” (for Bayesian Information Criterion). The user must take into account if the type of model defined in mod allows the calculation of the specified criterion. If not, *multifit* would recommend the use of another one. The function allows as well the possibility of specify a user-defined function for the calculation of a different criterion. This user-defined function should have the possibility of calculate a particular criterion value from an object containing the statistical models at each spatial scale. If this is the case, the user must specify the name of the function in a first element of the vector, and the model-selection criteria (“max” or “min” of the value of the criterion) in a second element (e.g. criterion = c(“my\_function”, “max”)).

The argument signif asks if significance stars should be plotted in the cases of statistical significance, whereas the argument alphadefines the statistical level for the plotting of the stars. The argument print\_sumasks if the summary of the selected model (i.e. the ‘best’ model by the defined criterion) should be printed in the console. The argument plot\_est asks if a plot depicting the estimates of the response variables at each spatial scale should be drawn in a separated plot.

The user may change some aesthetic characteristics of the plot/s, such as xlab or type, or other graphical parameters to be specified in ... such as type, pch, lty, and col*.* Check ?plot and ?par for full description of these arguments and many others. Particularly, the argument labelsallows the user to specify a character vector with names for each model at the x axis.

**Value**

*multifit* returns a list with the following components:

|  |  |
| --- | --- |
| lands\_summary | a data.frame containing a summary of the defined landscape attributes at each spatial scale: n, min, max, range, mean and median |
| summary | a data.frame containing relevant information of the models, including the value of the criterion, the estimates and the p.values |
| plot | a plot that shows the strength of the models at each spatial scale by the defined criterion, along with an optional plot of the estimates for each model |
| models | a list containing the models of all spatial scales as individual R objects. These may be useful for a posteriorianalysis |
| warnings | a list containing the warnings, if they occurred, during the analysis of the models for each spatial scale |
| messages | a list containing the messages, if they occurred, during the analysis of the models for each spatial scale |

The function also returns a plot, which is the same that is included in the component plot of the returned list.

**Note**

So far, *multifit* has been tested in fitting models of the packages *stats* (e.g. “lm”), *lme4* (e.g. “glm”, “lmer”, “glmer”), *nlme* (e.g. “lme”) and *pscl* (e.g. “zeroinfl”, “hurdle”). The function should work with other models and packages, and it will be tested with more of them in the future.

**Author**

**Example**

# Read table: fake data simulating bird richness. You can find this

# data.frame in the following link:

# github.com/phuais/multifit/blob/master/fake\_data/fake\_data.txt

df <- read.table(“fake\_data.txt”, header = T, sep = " ")

# Create an object with multifit

# In this case, we fit the response variable S (richness) against the

# proportion of forest amount at ten different spatial scales,

# starting from 500 until 5000, by 500 m per step.

fits <- multifit(mod = "glmer", multief = rad, formula = S ~

multief + (1|site), args = c("family = poisson"), data = df,

criterion = "AIC", plot\_est = T)

# Once the fitting of the models finished, we can explore the object...

# Prints a summary table of the landscape attributes at each spatial

# scale

fits$lands\_summary

# Prints a summary table with relevant information for multi-scale analysis

fits$summary

# Prints the plot (that was already plotted when running multifit)

fits$plot

# Prints a particular model object, and its summary

fits$models$R\_2500

summary(fits$models$R\_2500)

# Check for possible warnings or messages of particular model

fits$warnings$R\_2500

fits$messages$R\_2500