

multifit: an R function for multi-scale analysis in landscape ecology

Pablo Yair Huais 

Received: 8 January 2018 / Accepted: 24 May 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Context Multi-scale analyses are a common approach in landscape ecology. Their aim is to find the appropriate spatial scale for a particular landscape attribute in order to perform a correct interpretation of results and conclusions.

Objectives I present an R function that performs statistical analysis relating a biological response with a landscape attribute at a set of specified spatial scales and extracts the statistical strength of the models through a specified criterion index. Also, it draws a plot with the value of these indexes, allowing the user to choose the most appropriate spatial scale. This paper introduces the usage of *multifit* and demonstrates its functionality through a case study.

Conclusions The spatial scale at which ecologists conduct studies may change study outcomes and conclusions. Because of this, landscape ecologists commonly conduct multi-scale studies in order to establish an appropriate spatial scale for particular biological or ecological responses. The tool presented

here allows ecologists to simultaneously run several statistical models for a response variable and a specified set of spatial scales, automating the process of multi-scale analysis.

Keywords Landscape size · Spatial scale · Spatial extent · Buffer · Focal site design · Scale of effect · Scale of response

Introduction

One common task among landscape ecologists is to study how particular biological or ecological responses may relate with specific landscape attributes (Miguet et al. 2016). Landscape structure (e.g. habitat amount or number of patches) has important effects on species populations and communities (Fahrig 2003; Fahrig et al. 2011), and a key issue in every landscape ecology research is the spatial scale at which a particular landscape attribute is measured (Jackson and Fahrig 2015). In this way, the spatial scale at which a research is conducted can influence study outcomes (Holland et al. 2004; Smith et al. 2011; Flick et al. 2012), as the studied response may significantly relate with a landscape attribute at some spatial scales but non-significantly at others. For example, Flick et al. (2012) found positive relationships of butterfly species richness with the habitat amount at small

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10980-018-0657-5>) contains supplementary material, which is available to authorized users.

P. Y. Huais (✉)
Instituto Multidisciplinario de Biología Vegetal, FCEFYN
(CONICET, Universidad Nacional de Córdoba), Vélez
Sarsfield 1611, 495 Córdoba, Argentina
e-mail: phuais@imbiv.unc.edu.ar

spatial scales, but negligible correlations at higher extents. The main implication for ecologists is that important relationships may be missed if the study is not conducted at the scale in which the relationship is strongest (Jackson and Fahrig 2015).

However, ecologists do not usually know a priori the adequate spatial scale at which a new study should be examined, mainly due to a lack of a full understanding of the biological response (Jackson and Fahrig 2015). The selection of a spatial scale for a particular species or group of species is usually justified in literature under certain hypotheses, such as species mobility, dispersal movements or population density effects, although the majority of these have not been tested empirically (Miguet et al. 2016). An inappropriate selection of the spatial scale in landscape ecology studies may create misleading results and conclusions (Jackson and Fahrig 2015; Miguet et al. 2016), which can lead to improper landscape management recommendations for decision-makers.

A common tool used to select a spatial scale is to carry out a multi-scale analysis empirically (Brennan et al. 2002; an example of its application can be seen in Melo et al. (2017)). Through this approach, the landscape attribute of study (e.g. habitat amount, mean patch area, patchiness index, etc.) is measured at various spatial scales, starting from a central point and performing consecutive calculations of the landscape attribute at increasing radii. Each spatial scale is depicted as a circular buffer area of a particular radius from the focal point where the measurements of the response were performed (Fig. 1). This concept of scale is not the only one, for example, another one consists in selecting an entire patch rather than a central point from which radii are measured, e.g. Litteral and Wu (2012). Next, the biological response is related to the landscape attribute through a statistical model at each spatial scale, and the strength of relationship is measured at each scale. The strength of the models can be measured through any of three criteria that measures how strong the response variable relates with the predictor variable, such as the coefficient of determination (R-squared), the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), leading to the selection of the best model. Finally, the scale at which the relationship is strongest is selected, which is often called the ‘scale of effect’ (Jackson and Fahrig 2012). This scale

depicts the spatial scale at which the response variable relates more strongly with the landscape attribute, supported by the strongest statistical model (measured by a specific criterion).

I present an R function (R Core Development Team 2017) to carry out multi-scale analysis for landscape ecologists. The main goal of the function is to automate the process of statistical model selection between a set of spatial scales, running several models at the same time and displaying an output for the model selection process.

The R function

The function *multifit* automates multi-scale analysis for landscape ecology. Fully executable R code and a detailed manual are provided at <https://github.com/phuais/multifit/> and in Supporting Information.

Inputs and parameters

The key parameters of the function are described in Table 1, while the rest of them are described in the manual of Online Appendix 1. First of all, the user must provide a data frame with at least one column containing the biological or ecological response (i.e. the response variable) and multiple columns holding the values for a given landscape attribute, with one column of values for each spatial scale (the predictor variable at increasing radii, i.e. spatial scales). This data frame must be specified in the argument `data`. Additionally, this data frame can contain any other relevant variables to be included in the statistical analysis beyond the multi-scale landscape attribute (e.g. blocking factors, co-variables, random effects, etc.). The function to be applied in the models must be specified in the argument `mod` (e.g. “lm” for a classic linear model; Table 1). Before running the function, the user must have loaded the necessary package that contains the statistical function to be applied at each spatial scale. For instance, if GLMMs models will be performed through the function ‘glmer’ of package *lme4* (Bates and Sarkar 2007), this must be previously loaded in the R environment. The argument `multief` must contain a vector with the names of the columns that hold the values of the tested landscape attribute at each spatial scale. The argument `formula` must hold the predictor variable labeled as `multief`, which the

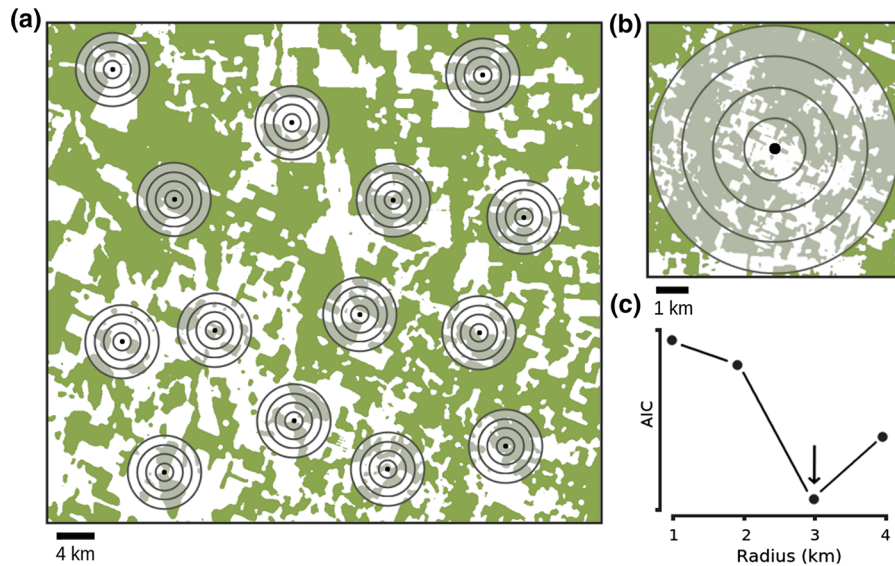


Fig. 1 **a** Typical experimental design of multi-scale analyses studies. A particular landscape attribute is measured at different sites separated in space (black dots) and at various spatial scales (in this case four). **b** Zoom of a particular site showing the set of spatial scales to be assessed: 1, 2, 3 and 4 km. **c** Common output

of a multi-scale analysis, showing the strength of the statistical models at each specified spatial scale. In this case, the model selection criterion was AIC (i.e. Akaike Information Criterion), and the strongest relationship (i.e. the scale of effect) occurred at 3 km. Adapted from Jackson and Fahrig (2015)

Table 1 Key parameters of the function. Detailed information about these parameters as well as the rest of the parameters can be found in the manual of Online Appendix 1

Parameter	Description
mod	String depicting the statistical function to be applied (e.g. 'lm' for a classic linear model)
multief	Character vector with the names of the columns holding the values of the landscape attribute in the data frame (specified in the argument 'data')
data	Data frame with at least one column with the response variable to be analyzed, and several columns holding the values of the landscape attribute at the various spatial scales (i.e. one column per spatial scale)
formula	String depicting 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. response_variable ~ multief)
criterion	String depicting the criterion to be used for the selection of the best model among the various spatial scales (i.e. the one with the strongest relationship with the response variable). This can be one of three options: 'R2' (for R-squared, i.e. coefficient of determination), 'AIC' (for Akaike Information Criterion), or 'BIC' (for Bayesian Information Criterion). Alternatively, the user may use an own function for the calculation of a different criterion. 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 criterion ('max' or 'min' of the value of the criterion, for R-squared-like and AIC-like criteria, respectively) in a second element (e.g. criterion = c('my_function', 'max'))

user must include as the main predictor variable (e.g. 'resp_variable ~ multief'). *multifit* will recognize this particular label as the landscape attribute at each spatial scale during the running of the models. Finally, in the argument *criterion* the user must specify the model selection criterion (Table 1), in order to select the model in which the strength of the

relationship between the response variable and the landscape attribute is maximum (i.e. the best models among the tested spatial scales). So far, *multifit* allows users to choose between three criteria: 'R-squared', 'AIC' or 'BIC'. Another useful alternative is the possibility of defining a particular criterion beyond these (see Table 1 and Online Appendix 1 for detailed

information), with a user-defined function in the R environment. This function must have the capability of calculating a particular criterion value from an R object that depicts the statistical model for each spatial scale (i.e. the output of the function specified in argument `mod`, if the models for each spatial scale are run individually).

With this inputted information, *multifit* runs one model for each specified spatial scale, relating the provided response variable with the landscape attribute in each case. Pertinent information of the models is extracted, including the strength of the relationship with the specified criterion, p-values and estimated model coefficients.

Outputs

The function returns a plot and an R list. Regarding the plot, the x axis holds the specified spatial scales of the multi-scale analysis, while the y axis holds the values of the models strength through the specified criterion (e.g. values of AIC). This allows the user to compare between various spatial scales, and eventually choose the most appropriate one (i.e. the scale of effect). An optional plot can be drawn aside, depicting the estimated model coefficients (i.e. the slopes of relationships between the response and the landscape attribute, on the y axis) along the tested spatial scales (on the x axis). Additionally, the function offers the possibility of differentiating between models with non-significant and significant estimated model coefficients at the plots, in order to graphically show the relationships that are statistically significant beyond its statistical strength (by plotting the dots with different styles, see Online Appendix 1).

The returned R list contains several important components as R objects:

- (a) a summary table (R `data.frame`) with descriptive statistics of the landscape attribute at each spatial scale: n, min, max, range, mean and median.
- (b) a table (R `data.frame`) with relevant information of the models: strength of the

models, estimated model coefficients and p values.

- (c) the plot described above as an R object (including the optional plot with the estimated model coefficients aside). This may be useful for the user in the future, because calling this object draws the plot again in the R graphic device.
- (d) an R list with the models as R objects, one per tested spatial scale.
- (e, f) two R lists containing possible warnings and messages that may arise during the running of the models. These may arise when *multifit* individually runs the models at each spatial scale (e.g. a warning describing problems of statistical convergence of the model).

Within this R list, the user may use the model's objects (component 'd' of the returned list) for any posterior analysis, such as model assumptions checking or analysis of estimated model coefficients and statistical significance of the variables (see the following Section and Online Appendix 1).

Case study

I present a simple case study where I applied the function *multifit*. Data was collected for a study assessing the effects on biodiversity of the spatial structure of agricultural landscapes. For details on the sampling design, see Fahrig et al. (2015). Data contains a column named 'Resp' which represents the abundance of *Pterostichus melanarius* (continuous variable), a very abundant carabid beetle, and eight columns that contain the amount of corn cover (i.e. the predictor variable) at eight consecutive spatial scales (ranging from 0.25 km radius until 2 km, 0.25 km per step). Original data contained more spatial scales, but here I selected a subset for example purposes. I applied classic linear models, specifically the function 'lm' from the package *stats* (R Core Development Team 2017). Also, I chose the R-squared criterion for the selection of the best model (i.e. the scale of effect). The function was specified as follows:

```
fits <- multifit(mod = "lm", multief = colnames(data)[3:10], formula =

  Resp ~ multief, data = data, criterion = "R2",

  signif = TRUE, alpha = 0.05, labels = seq(0.25, 2,

  0.25), xlab = "Radius [km]")
```

The output of the function can be seen in Fig. 2a, and shows the strength of the models at the various spatial scales. From a visual analysis of this plot, we may select the model including the amount of corn cover at 0.5 km radius as the best model, and 0.5 km as the scale of effect for this particular biological response. Additionally, the user may add the argument `plot_est = TRUE`, and a plot with the estimated model coefficients (i.e. the slopes of the relationships) at each spatial scale will be drawn (Fig. 2b). In order to see the summary of the results of the model with highest statistical strength (i.e. the result of applying the generic function ‘summary’ to any statistical model at R), the user may add the argument `plot_sum = TRUE` and the summary will automatically be printed at the console. Also, the user can run the generic R function ‘summary’ (e.g. `summary(fits$model$'label_of_the_spatial_scale')`) to see the summary of results of any

model. The user may perform the necessary tests and analysis to the returned R models by selecting them from the returned list (i.e. component ‘d’ of the returned list, see the Section Outputs above and Online Appendix 1).

Conclusion and future directions

Landscape ecologists commonly use a multi-scale approach in order to establish an appropriate spatial scale for particular biological or ecological responses. Selecting an improper spatial scale in ecological studies would lead to lose the strongest relationships of the response variable with a particular landscape attribute. This issue has important implications for the development of recommendations relevant for biodiversity conservation. The function presented here

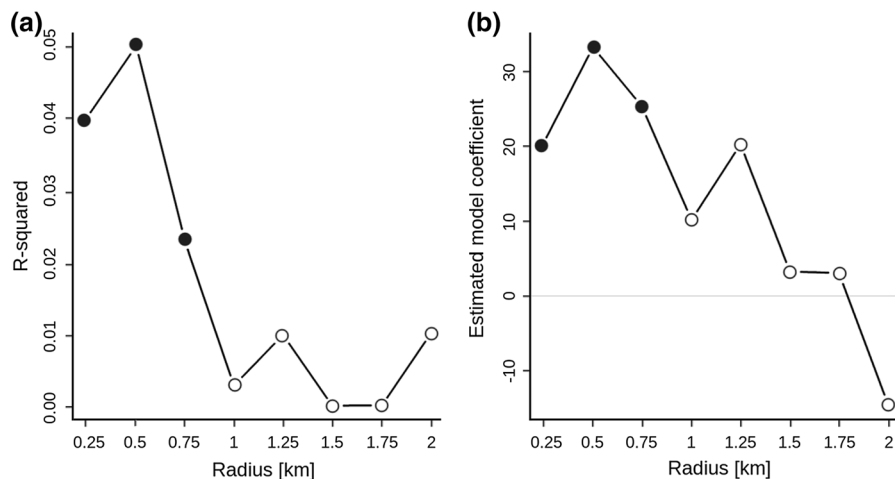


Fig. 2 **a** Plot output of the model selection between the set of specified spatial scales, using R-squared to index the strength of the relationship. In this example, the plot suggests that the spatial scale at the radius of 0.5 km was the best scale explaining the relationship between the abundance of beetles and the amount of corn cover. **b** Plot output of the estimated model coefficients (i.e. slope of relationship between the response and

the landscape attribute) at each spatial scale. Likewise Fig. 2a, the plot shows that the steepest slope was found at the spatial scale of 0.5 km. White-filled and black-filled dots represent models with non-significant and significant estimated model coefficients, respectively. Horizontal gray line of Fig. 2b marks the value of zero

automates the process of multi-scale selection, allowing ecologists to simultaneously run statistical models for a response variable and a landscape attribute at each of a specified set of spatial scales. The function has the potentiality of expanding its options in order to embrace more statistical models from other R packages. Additionally, this first version of *multifit* does not allow to simultaneously estimate scales of effect of more than one landscape attribute on the same response variable, a more complex approach carried out in landscape studies. I encourage R programmers and ecologists to modify the code to satisfy these needs and others, and expand the usage of *multifit*.

Acknowledgements I am grateful to two anonymous reviewers for suggestions and comments that improved previous versions of the manuscript and R code. I specially thank to María del Carmen Romero and Gustavo Giménez for introducing me to R programming. I thank to Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Universidad Nacional de Córdoba (UNC) for financial support, only possible with sustainable public policies for science.

Author contributions PYH conceived the idea, programmed the R code and wrote the manuscript.

Data availability The R function's code and the manual are available as online Supporting Information and at <https://github.com/phuais/multifit/>. Data for the case study was acquired from Miguet et al. (2017a), which is available at the Dryad Digital Repository (Miguet et al. 2017b).

References

- Bates D, Sarkar D (2007) lme4: linear mixed-effects models using S4 classes. R package version 0.9975-11
- Brennan JM, Bender DJ, Contreras TA, Fahrig L (2002) Focal patch landscape studies for wildlife management: optimizing sampling effort across scales. In: Liu J, Taylor WW (eds) Integrating landscape ecology into natural resource management. Cambridge University Press, Cambridge, pp 68–91
- Fahrig L (2003) Effects of habitat fragmentation on biodiversity. *Annu Rev Ecol Syst* 34(1):487–515
- Fahrig L, Baudry J, Brotons L, Burel FG, Crist TO, Fuller RJ, Martin JL (2011) Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. *Ecol Lett* 14(2):101–112
- Fahrig L, Girard J, Duro D, Pasher J, Smith A, Javorek S, Tischendorf L (2015) Farmlands with smaller crop fields have higher within-field biodiversity. *Agric Ecosyst Environ* 200:219–234
- Flick T, Feagan S, Fahrig L (2012) Effects of landscape structure on butterfly species richness and abundance in agricultural landscapes in eastern Ontario, Canada. *Agric Ecosyst Environ* 156:123–133
- Holland JD, Bert DG, Fahrig L (2004) Determining the spatial scale of species' response to habitat. *Bioscience* 54(3):227
- Jackson HB, Fahrig L (2012) What size is a biologically relevant landscape? *Landscape Ecol* 27(7):929–941
- Jackson HB, Fahrig L (2015) Are ecologists conducting research at the optimal scale? *Glob Ecol Biogeogr* 24(1):52–63
- Litteral J, Wu J (2012) Urban landscape matrix affects avian diversity in remnant vegetation fragments: evidence from the Phoenix metropolitan region, USA. *Urban Ecosyst* 15(4):939–959
- Melo GL, Sponchiado J, Cáceres NC, Fahrig L (2017) Testing the habitat amount hypothesis for South American small mammals. *Biol Conserv* 209:304–314
- Miguet P, Fahrig L, Lavigne C (2017a) How to quantify a distance-dependent landscape effect on a biological response. *Methods Ecol Evol* 8(12):1717–1724
- Miguet P, Fahrig L, Lavigne C (2017b) Data from: how to quantify a distance-dependent landscape effect on a biological response. Dryad Digital Repos. <https://doi.org/10.1111/2041-210X.12830>
- Miguet P, Jackson HB, Jackson ND, Martin AE, Fahrig L (2016) What determines the spatial extent of landscape effects on species? *Landscape Ecol* 31(6):1177–1194
- R Core Development Team (2017) R: a language and environment for statistical computing. R Core Development Team, Vienna
- Smith AC, Fahrig L, Francis CM (2011) Landscape size affects the relative importance of habitat amount, habitat fragmentation, and matrix quality on forest birds. *Ecography* 34(1):103–113