Methods in Ecology and Evolution

Methods in Ecology and Evolution 2012, **3**, 257–259 doi: 10.1111/j.2041-210X.2011.00153.x

APPLICATION

SMATR 3 – an R package for estimation and inference about allometric lines

David I. Warton[1]*, Remko A. Duursma[2], Daniel S. Falster[3] and Sara Taskinen[4]

[1]School of Mathematics and Statistics and Evolution & Ecology Research Centre, The University of New South Wales, NSW 2052, Australia; [2]Hawkesbury Institute for the Environment, University of Western Sydney, NSW 2751, Australia; [3]Department of Biological Sciences, Macquarie University, NSW 2109, Australia; and [4]Department of Mathematics and Statistics, P.O.Box 35, FI-40014, University of Jyväskylä, Finland

Summary

1. The Standardised Major Axis Tests and Routines (SMATR) software provides tools for estimation and inference about allometric lines, currently widely used in ecology and evolution.

2. This paper describes some significant improvements to the functionality of the package, now available on R in smatr version 3.

3. New inclusions in the package include sma and ma functions that accept formula input and perform the key inference tasks; multiple comparisons; graphical methods for visualising data and checking (S)MA assumptions; robust (S)MA estimation and inference tools.

Key-words: common slope testing, model II regression, principal component analysis, robust estimation, standardised major axis

Biologists often wish to estimate how one variable scales against another and to test hypotheses about the nature of this relationship and how it varies across samples. The most common example of this is allometry (Reiss 1989); hence, we refer to this problem as one of estimation and testing about allometric lines. An example is given in Fig. 1a, where we wish to understand how leaf lifespan (longev) scales against leaf mass per area (lma) and how this relationship changes across sites with different rainfall (rain). longev and lma are log-transformed prior to analysis and are approximately linearly related on the transformed scale. This is common in allometry, and it means that their relationship approximately follows a power law, longev= *a*lma^b. The 'scaling exponent' *b* is the slope on log-transformed axes, and the magnitude of this parameter describes how steep the leaf lifespan–leaf mass per area relationship is. The 'proportionality coefficient' *a*, related to the elevation on log–log axes, is needed to understand how long-lived leaves of a given mass per area will be.

Estimating *a* and *b* is not a simple linear regression problem because we are not interested in predicting one variable from another – we are interested in estimating some underlying line of best fit (Warton *et al.*, 2006). Another way to understand this is to see that the problem is symmetric – the basic problem does not change if we plot lma on the *Y* axis instead of the *X* axis (Smith 2009). Hence, the appropriate methods for analysis have more in common with principal component analysis, a multivariate approach, than with linear regression (Warton *et al.*, 2006). Common approaches to estimating the line of best fit are standardised major axis (SMA) and major axis (MA) estimation, which will be collectively referred to as (S)MA, and which are widely used in ecology and evolution.

Warton *et al.* (2006) reviewed (S)MA techniques, proposed routines for comparing the parameters *a* and *b* amongst groups and developed software to implement the methods. The Standardised Major Axis Tests and Routines (SMATR) software, available in both R (R Development Core Team 2010) and C++, has since been used in over 200 publications. We have made significant improvements to the software in the recently released version 3 of the smatr R package, and this paper briefly describes this new functionality.

Formula input via the sma and ma functions

The new sma and ma functions are the key access point to the smatr package, performing all the available estimation and inference tasks. These functions behave similarly to the lm function used in R for linear regression, taking a formula as the primary input argument. The type of task to be performed is determined by the formula that is used in the function call. Some of the more common types of tasks that can be

*Correspondence author. E-mail: david.warton@unsw.edu.au
Correspondence site: http://www.respond2articles.com/MEE/

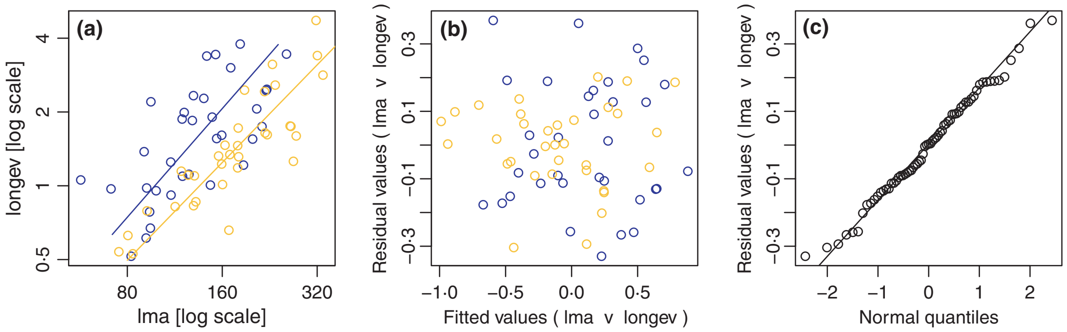

Fig. 1. Example graphs from the smatr package, available via applying `plot` to an `sma` object: (a) scatterplot of leaf longevity against leaf mass per area, with different labels for sites with high vs. low rainfall, and SMAs included; (b) residual vs. fits plot from SMAs fitted separately to high and low rainfall sites; (c) normal quantile plot of residuals.

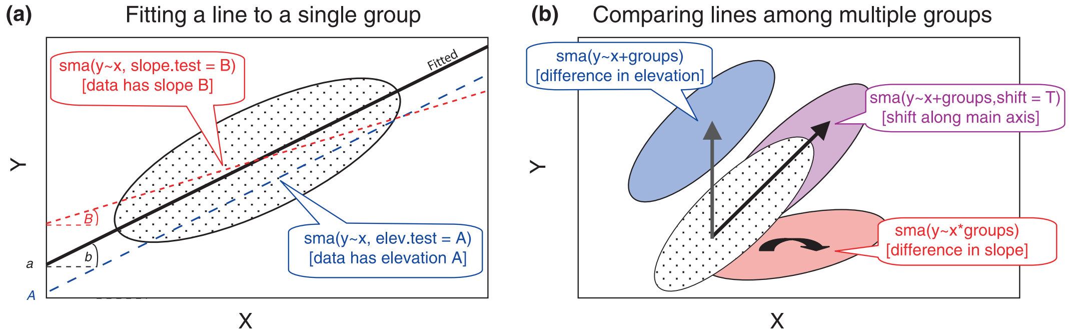

Fig. 2. Schematic diagram illustrating the main types of hypothesis tests that the `sma` and `ma` functions can be used for, by specifying different arguments to the function. These tests can involve (a) a single group of observations; (b) comparing lines from multiple groups of observations. For example, `sma(y~x,slope.test=B)` tests for evidence that a SMA constructed from a single group of observations has slope B.

performed using `sma` or `ma` are summarised schematically in Fig. 2 and listed below.

1 `sma(y~x)` will fit a SMA (for `y` against `x`) and return confidence intervals for the slope and elevation (Pitman 1939; Warton *et al.*, 2006).

2 `sma(y~x*groups)` will test for common slope (Warton & Weber 2002) amongst several SMAs (for `y` against `x`), fitted separately for each level of the factor `groups`.

3 `sma(y~x+groups)` will test for common elevation (Warton *et al.*, 2006) amongst several SMAs (for `y` against `x`), fitted with common slope but with separate elevations for each level of the factor `groups`.

The last two calls in the above perform a SMA equivalent of analysis of covariance (Sokal & Rohlf 1995), and the function calls are written in an analogous form to how this would be done for linear regression using the `lm` function. To use MA estimation instead of SMA estimation, the `ma` function is used instead of `sma`, which works in exactly the same way.

Additional arguments can be specified to perform some additional tasks:

1 `sma(y~x, slope.test=B)` will test the hypothesis that the SMA (for `y` against `x`) has slope `B` (Pitman 1939). The most common use of this command is to test for isometry, which usually implies a slope of one (Warton *et al.*, 2006). If `groups` is specified in the formula, results will be reported for a simultaneous test of whether the true slope is `B` for all groups, as well as separate results for each group.

2 `sma(y~x+groups, shift=T)` will test the hypothesis that several SMAs of common slope are centred on the same location along the SMA (Warton *et al.*, 2006).

3 The argument `multcomp=T`, when used in comparing multiple lines, will return pairwise comparisons of slopes (or elevations or locations along common-slope SMAs), and `multcompmethod="adjust"` will use adjusted *P*-values (via the 'Sidak adjust-ment', Westfall & Young, 1993) to control family-wise error rate in a conservative way.

4 The argument `intercept=F`, used in combination with most of the above functions, will force lines through the origin. This is necessary, for example, when analysing phylogenetically independent contrasts (Felsenstein 1985). The only situation in which this argument cannot be used is when testing for common elevation, because in that case, it is no longer applicable.

5 The argument `log="xy"` will $\log_{10}$-transform variables prior to analysis.

The output from any `sma` or `ma` call can be saved to an object (of type 'sma') for use in combination with generic `R` functions as below.

Graphing data

Applying the `plot` function to a `sma` object will by default produce a scatterplot with a (S)MA line added to the plot, or with multiple lines if appropriate. The argument `log="xy"` (in `sma` or `plot`) will ensure that the plot is on the log–log scale. A new function `defineAxis` can be used to customise tick spacing and axis labels. For example, Fig. 1a was produced using the following code:

```
ft < sma(longev~lma* rain, log="xy")
xax < defineAxis(major.ticks = c(80,160,320))
yax < defineAxis(major.ticks = c(0.5,1,2,4))
plot(ft, xaxis=xax, yaxis=yax)
```

Checking assumptions

The `plot` function can also be used to produce residual plots, to check the critical assumptions of linearity and equal variance at all fitted values (via a residuals vs. fitted values plot, Fig. 1b) and the assumption of normally distributed residuals (via a normal quantile plot, Fig. 1c). Normality can be important for inference when sample size is small. Figure 1b,c was generated using the following commands:

```
plot(ft, which="residual") #Fig 1b
plot(ft, which="qq") #Fig 1c
```

`summary, coef`, ... General-purpose functions such as `summary, coef, print` will now work with `ma` and `sma` objects, so that these functions can be used in just the same way as they can with `lm` objects. This will make the use of the `smatr` package more intuitive for those who are already familiar with linear regression modelling in `R` via `lm`.

Robust estimation

Standard methods of line fitting, including (S)MA methods, do not perform well in the presence of outliers (Taskinen & Warton 2011). A new robust option enables estimation of and inference about (S)MA in a manner that is insensitive to outliers, by adding `robust=T` to a `sma` or `ma` function call, e.g. `sma(y~x, robust=T)`. This method uses Huber's *M* estimation in place of least squares (Taskinen & Warton 2011). The method is currently only available when fitting a single line, and in future work, we plan to extend the robust approach to inference tasks involving several (S)MAs.

For more details, see the documentation associated with the package, which can be downloaded from the `CRAN` website http://cran.r-project.org/.

Acknowledgements

D.I.W. and D.S.F. are supported by the Australian Research Council, via the Discovery (project number DP0987729) and Laureate Fellow (FL100100080, awarded to Mark Westoby) schemes, and S.T. is supported by the Academy of Finland. Thanks to Ian Wright for data. For feedback that improved the manuscript and software, thanks to Brian Sidlauskas, an anonymous reviewer and editor.

References

Felsenstein, J. (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1–15.

Pitman, E.T.G. (1939) A note on normal correlation. *Biometrika*, **31**, 9–12.

R Development Core Team (2010) *R: A Language and Environment for Statistical Com-puting*. R Foundation for Statistical Computing, Vienna, Austria.

Reiss, M.J. (1989) *The Allometry of Growth and Reproduction*. Cambridge University Press, Cambridge.

Smith, R.J. (2009) Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, **140**, 476–486.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry – The Principles and Practice of Statistics in Biological Research*. W. H. Freeman, New York.

Taskinen, S. & Warton, D.I. (2011) Robust estimation and inference for bivariate line-fitting in allometry. *Biometrical Journal*, **53**, 652–672.

Warton, D.I. & Weber, N.C. (2002) Common slope tests for errors-in-variables models. *Biometrical Journal*, **44**, 161–174.

Warton, D.I., & Wright, I.J.Falster, D.S. & Westoby, M. (2002) Bivariate line-fitting methods for allometry. *Biological Reviews*, **81**, 259–291.

Received 28 April 2011; accepted 30 July 2011
Handling Editor: Robert Freckleton