available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/funeco

Commentary

Comment on Fungal under-representation is (slowly) diminishing in the life sciences

SUMMARY

In this commentary I argue that the data presented by and conclusions drawn in the paper by Pautasso (2013) lack statistical support.

Bibliometric data are often used to assess literature impact. In reaction to a recent letter, I argue that the analysis and presentation of such data should take basic statistical principles into account, including uncertainty estimates in the outcome.

The paper by Pautasso (2013) presents 30 scatter plots on the incidence of fungal topics in papers indexed by Web of Science (WoS). The scatter plots also contain linear (or in one case parabolic) fits, with the resulting fit formulae and an error assessment. I focus here on the one plot with a parabolic fit, but my major criticism concerns all others as well: the presentation neglects the role of round-off errors, there are no confidence limits for the results, and no attention is paid to alternative interpretations of the trends observed.

The plot focussed on here, is the one for the epidemi*-keyword search. Fig 1 shows a replot of the formula provided by Pautasso (2013). Obviously, the trend is completely

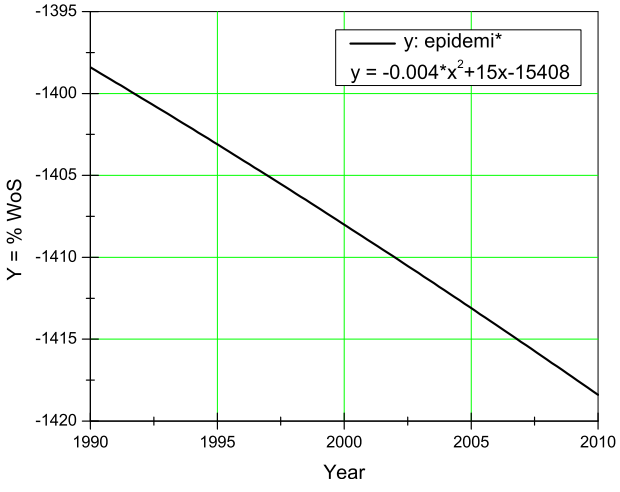

Fig 1 – Calculations on the basis of the epidemi*-plot in Pautasso (2013) (p. 134, top left). Replot of the fit function on the basis of the numbers given by Pautasso (2013).

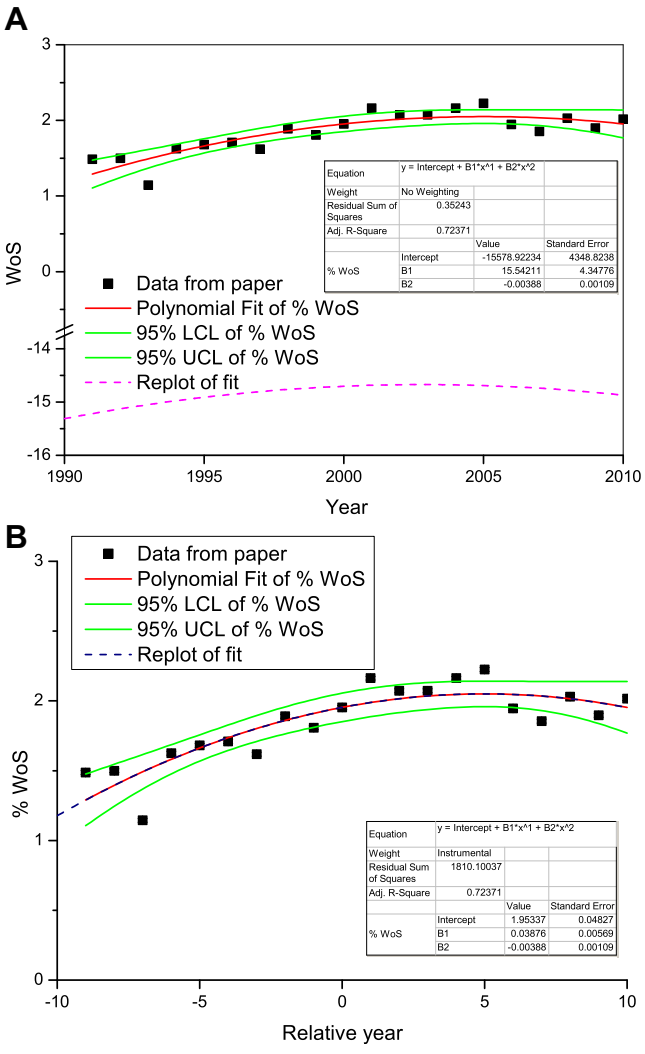

Fig 2 – Calculations on the basis of the epidemi*-plot in Pautasso's paper (p. 134, top left). A of the 'same' data (measured with a ruler from the paper), with a replot (pink) of the fit result. Note the scale break in the vertical axis. (B) As A, but the *x*-variable 'year' reduced by 2 000.

different from that of the curve in Pautasso (2013), and the resulting percentages are negative and a factor of about 1 000 off. The reason is largely the lack of precision in the numbers provided in Pautasso (2013). To illustrate this, I have reproduced the data (just by measuring the location of the data points with a ruler, from a print of the paper; I assume an uncertainty of ±0.15 mm), and refitted it, using Origin 8 software (OriginLab Corp., Northampton, MA, USA), default settings, no weighting. The fit curve (Fig 2A) is indistinguishable from the one in Pautasso (2013), but I give the fit parameters as returned by Origin. They are much more precise (but look at the error estimates for an indication of accuracy) than that in Pautasso (2013), but still a replot of the reported fit curve fails to reproduce the actual fit (pink curve; note the axis break). There is still too much round-off error! In this case, the large scale difference between the dependent and the independent variables is to blame: the x-values are about 3 orders of magnitude larger than the y-values. In a 2nd-order polynomial, $2\,000^2$ becomes 4 million, so up to the 6th decimal place in the associated fit parameter becomes important. This precision requirement can simply be avoided by reducing the 'year'-variable. This is illustrated in Fig 2B, which is almost the same as Fig 2A, but I use a 'relative year': year minus 2 000. Most fit parameters become much smaller, and the replot of the fit (pink curve) falls on top of the actual fit.

Pautasso (2013) has neglected round-off error in providing figures for his fits, and the fit formulae provided therefore miss information. This may be sloppy, but in itself does not invalidate the trends observed. Unfortunately, the author also does not provide uncertainty estimates for the fit parameters; as can be seen in the example provided in Fig 2A, the Origin 8 software indicates standard errors of about 25%. The significance of the results presented in Pautasso (2013) is, therefore, hard to judge. The sheer immensity of the number of papers involved – although this number is not given in Pautasso (2013), I found 9 249 papers between 1991 and 2004 on the keyword 'species richness' (including quotes; WoS accessed on January 5, 2013; Arts & Humanities and Social Sciences excluded), 2% of which were stated in Pautasso (2013) to be about fungi – also makes it effectively impossible to judge how many papers really dealt in depth with fungi and the chosen keywords.

Finally, there may be a more down to earth explanation for the observed trends: the more diverse keywords you use, the more likely it becomes that your paper will be found in a keyword-search, and the more likely it is that you will be cited in forthcoming papers; and unfortunately, we scientists tend to be judged by our H-index.

REFERENCE

Pautasso M, 2013. Fungal under-representation is (slowly) diminishing in the life sciences. *Fungal Ecology* **6**: 129–135.

ARTICLE INFO

Accepted 25 February 2013

Corresponding editor: Lynne Boddy

KEYWORDS

Bibliometrics
Fungi
Uncertainty estimates

*Nico DAM**

Hooischelf 13, 6581 SL Malden, The Netherlands

* Tel.: +31 402472117.
E-mail address: n.dam@science.ru.nl

1754-5048/$ – see front matter

http://dx.doi.org/10.1016/j.funeco.2013.03.003