Repeatability and transparency in ecological research

Aaron M. Ellison[1]

Harvard University, Harvard Forest, 324 North Main Street, Petersham, MA 01366 USA

INTRODUCTION

A fundamental tenet of science is that results must be reproducible by other scientists before they are accepted as factual. However, because ecological phenomena are context-dependent, and because that context changes through time and space, it is virtually impossible to reproduce precisely or quantitatively any single experimental or observational field study in ecology. Yet many ecological studies can be repeated. In particular, *ecological synthesis* – the assembly of derived datasets and their subsequent analysis, re-analysis, and meta-analysis – should be easy to repeat and reproduce. Such syntheses also demonstrate qualitative and quantitative consistency among many ecological studies (Gurevitch *et al.* 1992, Warwick and Clarke 1993, Jonsen *et al.* 2003, Walker *et al.* 2006, Cardinale *et al.* 2006, Marczak *et al.* 2007, Vander Zanden and Fetzer 2007) and provide strong support for general ecological theories .

It should come as no surprise that meta-analysis by Mittelbach *et al.* (2001) of the effect of productivity on species richness has led to the development of a cottage industry focused on empirical testing of this relationship (post-2001 examples abound in Appendix A of Whittaker 2009). But it is much more surprising that continual re-analyses of the *same* datasets (Whittaker and Heegaard 2003, Gillman and Wright 2006, Pärtel *et al.* 2007) have yielded such disparate results that Whittaker (2009) has suggested abandoning the effort to obtain consistent results from the available data. He goes even further, suggesting that ecology may not yet be ready for

[1] E-mail: aellison@fas.harvard.edu

meta-analysis and data synthesis. For two reasons, I respectfully suggest that Whittaker's critique is misplaced. First, of all the studies critiqued by Whittaker (2009), only Mittelbach *et al.* (2001) actually conducted a formal meta-analysis. The others, as pointed out by Whittaker (2009: ms. p. 4, line 7) undertook extensive primary analyses but the authors did not conduct formal meta-analyses (Gurevitch and Hedges 1999). Second, and more importantly, if ecological synthesis is transparent – data, models, and analytical tools are available freely to the research community – then it should yield consistent, repeatable results. We may then disagree on the *interpretation* of the resulting synthesis, but at least we will be able to agree on the reproducibility of the results themselves.

Requirements for repeatable ecological synthesis

In a nutshell, ecological synthesis proceeds by assembling available datasets into a common, derived dataset and then applying one or more (statistical) models to this derived dataset to test the prediction of a hypothesis of interest (Ellison *et al.* 2006). Repeatability and reproducibility of ecological synthesis requires full disclosure not only of hypotheses and predictions, but also of the raw data, methods used to produce derived datasets, choices made as to which data or datasets were included in, and which were excluded from, the derived datasets, and tools and techniques used to analyze the derived datasets. Of all the papers under discussion by Whittaker (2009), Mittelbach *et al.*'s (2001) paper comes closest to achieving such transparency, although neither the raw data nor the derived dataset they analyzed are publicly available.

But achieving this level of disclosure and transparency is difficult. First and foremost, researchers must be committed to transparent production of ecological knowledge. We may be blissfully unaware of our own intellectual biases, but there are no excuses for not making data, methods, and tools freely available in a timely fashion. Yet despite mandates from funding agencies and research networks that data be made available publicly (Arzberger *et al.* 2004), raw data are not easily accessed. Research teams can spend many weeks searching data archives only to find summary statistical tables, lists of means, or concise graphs. Contacting individual investigators may yield raw data in digital form or in yellowing notebooks, or it may yield nothing at all. Fortunately, archives of ecological data are growing (examples include ESA's data registry,[2] *Ecological Archives*,[3] the data repository of the National Center for Ecological Analysis and Synthesis [NCEAS],[4] the data archive of the Long Term Ecological Research Network[5], and Oak Ridge's Distributed Active Archive Center[6] among many others), but archiving ecological data is not yet a requirement for publication in any journal. Ecologists also have developed standard methods for describing ecological datasets with *descriptive metadata* (Michener *et al.* 1997, Jones *et al.* 2006, Madin *et al.* 2008) that make it easier to interpret and hence re-use them. Software tools such as Morpho[7] that help investigators create descriptive metadata also are maturing.

But it is not enough simply to find a dataset and understand its origin and structure. Once datasets are obtained, it is usually necessary to transform the data into common units and scales (*e.g.,* species/ha or kg/ha). Interpolated values may need to be substituted for missing data, and methods of interpolation will vary among investigators (Ellison *et al.* 2006). Finally, and usually after still further manipulations and making decisions as to which data to include or exclude (*cf.*

[2] <http://data.esa.org/esa/style/skins/esa/index.jsp>
[3] <http://www.esapubs.org/archive/>
[4] <http://knb.ecoinformatics.org/knb/style/skins/nceas/>
[5] <http://metacat.lternet.edu/knb/>
[6] <http://daac.ornl.gov/>
[7] <http://knb.ecoinformatics.org/morphoportal.jsp>

Whittaker and Heegard 2003 and Appendix A of Whittaker 2009), a derived dataset is ready for analysis.

Each step – *e.g.,* digitization, rescaling, interpolation, inclusion or exclusion – requires individual judgment and provides an opportunity to introduce bias or error. If subsequent synthesis is to be repeatable, users must have confidence in the reliability of the derived dataset. Thus it is imperative that researchers document clearly each of the steps used to produce derived datasets. This *process metadata* – the documentation of the processes used to produce a dataset – provides one way to assess the reliability of a derived dataset (Osterweil *et al.* 2005, Ellison *et al.* 2006). Storage of the original datasets *and* the processes applied to create the derived dataset provides the mechanism to reproduce it.

Such audit trails that include archived datasets and tools allow can allow future users to determine effects of changing particular processes on the structure and subsequent analysis of the derived dataset (Ellison *et al.* 2006). For example, Mittelbach *et al.* (2001) classified the relationship between species richness and productivity in one of five categories (unimodal humped or U-shaped, monotonic positive or negative, or no relationship) whereas Laanisto *et al.* (2008) classified this same relationship simply as unimodal or not. Whittaker and Heegard (2003) and Whittaker (2009) excluded data that Mittelbach *et al.* (2001) included. Gillman and Wright (2006) used some of the regression results reported by Mittelbach *et al.* (2001) but also reanalyzed some of the original datasets using different software and without specifying which data were reanalyzed. Clearly results will differ if the same data are classified differently; if different subsets of data are analyzed, or if individual datasets are treated differently. Importantly, we can assess these differences by running new analyses on available datasets. The resulting differences in approach to and analysis of the data may reflect differences in questions on the part of the investigators, honest disagreements regarding the "best" available evidence (*sensu* Slavin 1995), or strongly held opinions regarding the most appropriate statistical analysis (*e.g.,* ordinary least-squares regression *versus* general linear models with a variety of error distributions and link functions). However, these differences and disagreements do not in and of themselves invalidate the activity of ecological synthesis.

It is equally important to document and whenever possible archive the statistical tools and models used for analysis and synthesis (Thornton *et al.* 2005); such an archival record should be a requirement for publication of any meta-analysis or data synthesis. The various authors critiqued by Whittaker (2009) all used different statistical tools (Table 1), and it would be impossible to repeat precisely any of the author's analyses.

Documentation and archiving of analytical processes, including those processes used to create derived datasets and the statistical tools and models applied to them, is difficult, and software tools for such documentation and archiving are rudimentary. It may seem wasteful to archive software, but numerical precision of arithmetic operations changes with new integrated circuit chips and different operating systems, functions work differently in different versions of software, and implementation of even "standard"" statistical routines differ among software packages (a widely unappreciated example of relevance to ecologists is the different sums-of-squares reported by SAS, S-Plus, and R for analysis of variance and other linear models (Venables 1998)). Finally, there are no standards for process metadata (Osterweil *et al.* 2005, Ellison *et al.* 2006) and no easy way to archive model code used by, or specific versions of, commercial software packages. While open-source software tools such as R (R Development Core Team 2007) is an attractive (and affordable) alternative, they evolve even more rapidly than their commercial counterparts, and regular changes in functionality of familiar routines are not uncommon (implementation of the cor function for calculation of Pearson's correlation coefficient in early versions of R is a notorious example). But without archiving software, tools, and associated process metadata, it is unlikely that we will be able to accurately reproduce any ecological synthesis.

Moving forward

More and more ecologists are following federal guidelines (Office of Management and Budget Circular A-110)[8] and making their data freely available within a short time of collection and publication. Cultural impediments to data sharing among ecologists are disappearing as more and more ecologists recognize not only that sharing of data benefits the entire scientific enterprise (Baldwin and Duke 2005) but also results in successful collaborations and subsequent publications such as those facilitated by NCEAS.[9] Rapid development of data archiving and sharing tools has been facilitated by funding initiatives focused on development of software for production of descriptive metadata and distributed access to permanently and stably archived data.[10] There is increasing recognition that similar efforts must be undertaken to document analytical tools and processes and to archive the software tools themselves (Thornton *et al.* 2005, Ellison *et al.* 2006). Software tools in development for creating process metadata, including documentation of dataset provenance and storage of analytical tools applied to derived datasets, include Kepler (Ludäscher *et al.* 2006) and the Analytic Web (Osterweil *et al.* 2009). Ecologists should work with these software development teams, and others like them, to learn how better

[8] <http://www.whitehouse.gov/omb/circulars/a110/a110.html>; for analysis and agency-specific implementation of this regulation, see <http://thecre.com/access/index.html>

[9] <http://nceas.ucsb.edu/products>

[10] <http://www.nsf.gov/dir/index.jsp?org=OCI>

documentation and archiving of scientific processes and work-flows can advance our science and to provide challenging tests of these evolving systems (Boose *et al.* 2007).

Rather than abandon data synthesis and meta-analysis as Whittaker (2009) suggests, ecologists should embrace these activities as the very essence of our science. With appropriate attention to documentation of data *and* analytical processes and a commitment to unbiased inquiry and full transparency of analytic activities, data synthesis and meta-analysis will become the most repeatable and reproducible activities that ecologists undertake. The results of such syntheses and meta-analyses will be the grist for the mill of ecological forecasting, perhaps the most important endeavor of 21st century ecology (Clark *et al.* 2001).

ACKNOWLEDGEMENTS

I thank Don Strong for inviting this commentary on Robert Whittaker's thought-provoking article. Gary Mittelbach discussed availability of the original species richness-productivity dataset and Tom Mitchell-Olds answered questions about the availability of his Pascal software written in 1987. Brad Cardinale provided helpful comments on early versions of the manuscript. Work on this manuscript was supported by the Analytic Web project (NSF grant CCR-0205575) and by the Harvard Forest Long Term Ecological Research Program (NSF grant DEB 06-20443).

LITERATURE CITED

Arzberger, P., P. Schroeder, A. Beaulieu, G. Bosker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir, and P. Wouters. 2004. An international framework to promote access to data. Science **303**:1777-1778.

Baldwin, J. D., and C. Duke. 2005. Society summit on data sharing and archiving policies. Bulletin of the Ecological Society of America **86**:61-66.

Boose, E., A. M. Ellison, L. J. Osterweil, R. Podorozhny, L. Clarke, A. Wise, J. L. Hadley, and D. R. Foster. 2007. Ensuring reliable datasets for environmental models and forecasts. Ecological Informatics **2**:237-247.

Cardinale, B. J., D. S. Srivastava, J. E. Duffy, J. P. Wright, A. L. Downing, M. Sankaran, and C. Jouseau. 2006. Effects of biodiversity on the functioning of trophic groups and ecosystems. Nature **443**:989-992.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Jr. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: an emerging imperative. Science **293**:657-660.

Ellison, A. M., L. J. Osterweil, J. L. Hadley, A. Wise, E. Boose, L. Clarke, D. R. Foster, A. Hanson, D. Jensen, P. Kuzeja, E. Riseman, and H. Schultz. 2006. Analytic webs support the synthesis of ecological datasets. Ecology **87**:1345-1358.

Gillman, L. N., and S. D. Wright. 2006. The influence of productivity on the species richness of plants: a critical assessment. Ecology **87**:1234-1243.

Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in ecological meta-analyses. Ecology **80**:1142-1149.

Gurevitch, J., L. Morrow, A. Wallace, and J. Walsh. 1992. The meta-analysis of competition in field experiments. American Naturalist **140**:539-572.

Jones, M. B., M. P. Schildhauer, O. J. Reichman, and S. Bowers. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Review of Ecology, Evolution, and Systematics **37**:519-544.

Jonsen, I. D., R. A. Myers, and J. M. Flemming. 2003. Meta-analysis of animal movement using state-space models. Ecology **84**:3055-3063.

Laanisto, L., P. Urbas, and M. Pärtel. 2008. Why does the unimodal species richness-productivity relationship not apply to a woody species: a lack of clonality or a legacy of tropical evolutionary history. Global Ecology and Biogeography **17**:320-326.

Ludäscher, B., I. Altintas, C. Berkeley, D. G. Higgins, E. Jaeger-Frank, M. Jones, E. Lee, J. Tao, and Y. Zhao. 2006. Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience **18**:1039-1065.

Madin, J. S., S. Bowers, M. P. Schildhauer, and M. B. Jones. 2008. Advancing ecological research with ontologies. Trends in Ecology and Evolution **23**:159-168.

Marczak, L. B., R. M. Thompson, and J. S. Richardson. 2007. Meta-analysis: trophic level, habitat, and productivity shape the food web effects of resource subsidies. Ecology **88**:140-148.

Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications **7**:330-342.

Mitchell-Olds, T., and R. G. Shaw. 1987. Regression analysis of natural selection: statistical influence and biological interpretation. Evolution **41**:1149-1161.

Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? Ecology **82**:2381-2396.

Osterweil, L. J., L. A. Clarke, A. M. Ellison, E. Boose, R. Podorozhny, and A. Wise. 2009. Clear and precise specification of ecological data management processes and dataset provenance. IEEE Transactions on Automation Science and Engineering **in press**.

Osterweil, L. J., A. Wise, L. Clarke, A. M. Ellison, J. L. Hadley, E. Boose, and D. R. Foster. 2005. Process technology to facilitate the conduct of science. Pages 403-415 *in* M. Li, B. Boehm, and L. J. Osterweil, editors. Lecture Notes in Computer Science - SPW 2005. Springer-Verlag, Berlin, Germany.

Pärtel, M., L. Laanisto, and M. Zobel. 2007. Contrasting plant productivity-diversity relationships across latitude: the role of evolutionary history. Ecology -1091.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Slavin, R. E. 1995. Best evidence synthesis: an intelligent alternative to meta-analysis. Journal of Clinical Epidemiology **48**:9-18.

Thornton, P. E., R. B. Cook, B. H. Braswell, B. E. Law, W. M. Post, H. H. Shugart, B. T. Rhyne, and L. A. Hook. 2005. Archiving numerical models of biogeochemical dynamics. Eos **86**:431.

Vander Zanden, M. J., and W. W. Fetzer. 2007. Global patterns of aquatic food chain length. Oikos **116**:1378-1388.

Venables, W. N. 1998. Exegeses on linear models. Pages 1-25 *in* S-Plus User's Conference, Washington, DC.

Waide, R. B., M. R. Willig, C. F. Steiner, G. Mittelbach, L. Gough, S. I. Dodson, G. P. Juday, and R. Parmenter. 1999. The relationship between productivity and species richness. Annual Review of Ecology and Systematics **30**:257-300.

Walker, M. D., C. H. Wahren, R. D. Hollister, G. H. R. Henry, L. E. Ahlquist, J. M. Alatalo, M. S. Bret-Harte, M. P. Calef, T. V. Callaghan, A. B. Carroll, H. E. Epstein, I. S. Jónsdóttir, J. A. Klein, B. Magnússon, U. Molau, S. F. Oberbauer, S. P. Rewa, C. H. Robinson, G. R. Shaver, K. N. Suding, C. C. Thompson, A. Tolvanen, Ø. Totland, P. L. Turner, C. E. Tweedie, P. J. Webber, and P. A. Wookey. 2006. Plant community responses to experimental warming across the tundra biome. Proceedings of the National Academy of Sciences, USA **103**:1342-1346.

Warwick, R. M., and K. R. Clarke. 1993. Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. Marine Ecology Progress Series **92**:221-231.

Whittaker, R. J. 2009. Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. Ecology **90**:000.

Whittaker, R. J., and E. Heegaard. 2003. What is the observed relationships between species richness and productivity? Comment. Ecology **84**:3384-3390.

Table 1. Analytical methods used in the syntheses of the species richness-productivity relationship.

Author	**Analytical method(s) used**	**Analytical tool(s) used**	**Comments**
Waide *et al.* (1999)	Linear and quadratic regressions	None specified	Not repeatable
Mittelbach *et al.* (2001)	Ordinary least-squares regression	SYSTAT 8.0	Possibly repeatable; current available version is 12.0
	Poisson regression	NAG Statistical Add-in for Excel	Not repeatable; software discontinued
	"Mitchell-Olds & Shaw test" (Mitchell-Olds and Shaw 1987)	None specified	Not repeatable; software unavailable (but algorithm available). Which of three tests proposed by Mitchell-Olds and Shaw) was also not specified.
	Chi-square Exact test	StatXact	Possibly repeatable; no version given.
	Meta-analysis using mixed-effects model	MetaWin 2.0	Repeatable; commercial software version still available

Whittaker and Heergard (2003)	Poisson regression	Not specified	Not repeatable
Gillman and Wright (2006)	Ordinary least-squares regression on "some" datasets of Mittelbach *et al.* (2001)	Software not specified; datasets re-analyzed not specified	Not repeatable
Pärtel *et al.* (2007)	Multinomial logit regression	Statistica 6.1	Possibly repeatable; current release is 8.0
Laanisto *et al.* (2008)	Fisher exact tests	Not specified	Possibly repeatable using available algorithms
	General linear model	Statistica 6.1	Possibly repeatable; current release is 8.0