

Compromises in Data Selection in a Meta-Analysis of Biodiversity in Managed and Unmanaged Forests: Response to Halme et al.

YOAN PAILLET,^{1,2} LAURENT BERGÈS,^{1,20} JOAKIM HJÄLTÉN,³ PÉTER ÓDOR,⁴ CATHERINE AVON,¹ MARKUS BERNHARDT-RÖMMERMANN,⁵ RIENK-JAN BIJLSMA,⁶ LUC DE BRUYN,^{7,8} MARC FUHR,² ULF GRANDIN,⁹ ROBERT KANKA,¹⁰ LARS LUNDIN,⁹ SANDRA LUQUE,² TIBOR MAGURA,¹¹ SILVIA MATESANZ,¹² ILONA MÉSZÁROS,¹³ M.-TERESA SEBASTIÀ,^{14,15} WOLFGANG SCHMIDT,⁵ TIBOR STANDOVÁR,⁴ BÉLA TÓTHMÉRÉSZ,¹⁶ ANNELI UOTILA,¹⁷ FERNANDO VALLADARES,¹² KAI VELLAK,¹⁸ AND RISTO VIRTANEN¹⁹

¹Cemagref, UR EFNO, Domaine des Barres, F-45290 Nogent-sur-Vernisson, France

²Cemagref, UR EMGR, 2 rue de la Papeterie BP 76, F-38402 Saint-Martin-d'Hères, France

³Department of Wildlife, Fish and Environmental Science, Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden

⁴Department of Plant Taxonomy and Ecology, Eötvös University, Pázmány P. stny. 1/C., H-1117 Budapest, Hungary

⁵Department Silviculture and Forest Ecology of the Temperate Zones, Georg-August-University Göttingen, Büsgenweg 1, D-37077 Göttingen, Germany

⁶Alterra Wageningen UR, Centre for Ecosystem Studies, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands

⁷Research Institute for Nature and Forest, Kliniekstraat 25, B-1070 Brussels, Belgium

⁸Evolutionary Ecology, Department of Biology, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

⁹Swedish Univ Agr Sci, Department of Aquatic Sciences and Assessment, Box 7050, SE-75007 Uppsala, Sweden

¹⁰Institute of Landscape Ecology, Slovak Academy of Sciences, Stefanikova Str. 3, SK-814 99 Bratislava, Slovakia

¹¹Hortobágy National Park Directorate, P.O. Box 216, H-4002 Debrecen, Hungary

¹²Instituto de Recursos Naturales, CSIC IRN-CCMA-CSIC, Serrano 115, E-28006 Madrid, Spain

¹³Department of Botany, University of Debrecen, P.O. Box 71, H-4010 Debrecen, Hungary

¹⁴Forest Technology Centre of Catalonia, Pujada del Seminari s/n, E-25280 Solsona, Spain

¹⁵Agronomical Engineering School, University of Lleida, Av. Rovira Roure 191, E-25198 Lleida, Spain

¹⁶Ecological Institute, Debrecen University, P.O. Box 71, H-4010 Debrecen, Hungary

¹⁷Faculty of Forestry, University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, Finland

¹⁸Institute of Ecology and Earth Sciences, University of Tartu, Lai Str., 40 Tartu EE-51005, Estonia

¹⁹Department of Biology, University of Oulu, P.O. Box 3000, FIN-90014 Oulu, Finland

Introduction

Meta-analysis (MA) is a powerful tool for revealing general trends and quantitatively synthesizing the results of independent studies. Nevertheless, the procedure has been criticized, particularly when it has been applied to ecological and conservation biology studies.

To provide an updated picture of the effect of forest management on biodiversity, we performed an MA with

data from recent publications to compare the species richness of managed and unmanaged forests in Europe (Paillet et al. 2010). Thus, the opening sentence of Halme et al. (2010 [this issue]) goes against the goal of our paper and the philosophy behind MA. Indeed, we provided a balanced view of the contrasting opinions on the effects of forest management on biodiversity.

Our MA provides basic ecological knowledge needed for conservation and ecologically sustainable forestry and

²⁰email laurent.berges@cemagref.fr

Paper submitted March 19, 2010; revised manuscript accepted March 26, 2010.

we found that forest management has a negative effect on the biodiversity of forest-dwelling species. We were aware of the limitations of our MA, so we used caution when discussing our results because the effect is strongly heterogeneous among taxa; there is only a *trend* for recovery of biodiversity once management has been abandoned; and low replication numbers mean no strong conclusion on the effect of different management types could be drawn from our data. Our obvious primary conclusion is that research on the subject in Europe was scarce and that more controlled studies may help answer the questions raised.

Halme et al. overlook the fact that we discussed confounding effects and MA limitations (pp. 109–110). Furthermore, they claim the data we selected for analysis have four major flaws that compromise our conclusions: independence of observations, distribution of the taxonomic groups regarding time since abandonment and management intensity, taxonomic generalizations, and criteria used for inclusion of papers in the MA.

Independence of Observations

We share Halme et al.'s concern about proper replication of observations in scientific studies. The pseudoreplication issue is, however, much more complex than Halme et al. indicate. Specifically, in the case of large-scale field experiments, the question of what comprises a replicate has been intensely debated (e.g., Underwood 1997; Oksanen 2001, 2004). In addition, it has been argued that the core ideas behind pseudoreplication are based on a misunderstanding of statistical independence, the nature of control groups in science, and contexts of statistical inference (Koehnle & Schank 2009; Schank & Koehnle 2009 but see also Hurlbert 2009). Unfortunately, this issue is too complicated to be explored here, but it underscores the complexity of pseudoreplication.

Although Hurlbert's (1984) paper identified a very important problem in ecological research, pseudoreplication cannot and should not be used as a universal criterion for accepting or rejecting experimental research. Any research work must be judged on its own merits. For example, spatiotemporal proximity does not automatically lead to statistical dependence and certainly not in a way that prohibits appropriate statistical inferences (Schank & Koehnle 2009). In addition, there are many other important methodological and statistical issues to consider when evaluating the quality of a research work. Contrary to manipulative experiments, one should accept that background variations cannot be fully controlled in mensurative experiments (Hurlbert 1984), and it is often impossible to spatially replicate an experiment on a large number of different sites. As we clearly state in our paper, the surface area of unmanaged forests is very lim-

ited in Europe. Comparatively, the number of managed forest stands is much higher. Thus, it is nearly impossible to control for important factors such as site conditions, patch size, landscape context, soil, stand age, tree species composition and land-use history (Paillet et al. 2010). Avoiding pseudoreplication is a desired prerequisite of many experiments, but it can be impossible when investigating unmanaged forests in Europe. Thus, we disagree with Halme et al. that some papers should have been excluded from our MA due to alleged problems with pseudoreplication. These papers were all published in peer-reviewed journals and thus were judged to have scientific merit. If we were to subjectively exclude from our analyses the papers we considered flawed (for methodological, statistical, or other reasons), our objectivity could be questioned. We therefore chose to include these studies in our analyses.

Halme et al. also question the independence of observations, but in the MA process, comparing a single control to several experimental groups is generally accepted (Gurevitch & Hedges 2001). These cases represent 22% of the total number of comparisons included in our MA (26 out of 120 comparisons in our data set). Sampling dependence in multiple-treatment studies can be solved by (1) using the unmanaged forest stands only once and randomly choosing one managed forest type and leaving out the other types, (2) combining all the managed forest types in one value, and (3) using an MA model with study as a random effect, which controls for this type of dependence (Gurevitch & Hedges 2001).

Distributions of Covariables

The added value of an MA relies on its ability to test the relationship between effect size and factors that were not testable in the individual studies (Gurevitch & Hedges 2001). The distribution of individual studies across different taxa and covariables is definitely unbalanced in our study, as is emphasized in Table 1 here and in Table 1 of Paillet et al. (2010). Nevertheless, we do not share Halme et al.'s opinion that the general trend observed between effect size and time-since-abandonment (TSA) is only an artefact of the unbalanced distribution between vascular plants studies on the one hand and fungi and carabid beetles on the other hand. Figure 2 in Paillet et al. (2010) clearly shows there are many negative effect sizes around 50 years and positive ones around 100 years; this partly counterbalances the distribution at the extreme end of the TSA gradient. Moreover, Halme et al. do not mention two important results: analyses separated by taxa almost always provided negative slopes, except for bryophytes and birds (Table 3 in Paillet et al. [2010]) and most of the negative slopes for taxa had much higher values than the slope for all groups, even though the effect of TSA

Table 1. Distribution of individual studies used in Paillet et al. (2010) relative to time since abandonment (TSA).

| | TSA classes (years) | | | | total |
|---------------------------------|---------------------|-------|--------|------|-------|
| | <50 | 50-75 | 75-100 | >100 | |
| All | 21 | 32 | 22 | 14 | 89 |
| Vascular plants ^a | 10 | 4 | 8 | 1 | 23 |
| Bryophytes | 2 | 3 | 3 | 0 | 8 |
| Lichens | 0 | 3 | 5 | 2 | 10 |
| Birds | 3 | 0 | 4 | 0 | 7 |
| Carabids | 2 | 4 | 0 | 0 | 6 |
| Saproxylic beetles ^b | 4 | 4 | 0 | 4 | 12 |
| Nonsaproxylic beetles | 0 | 4 | 0 | 2 | 6 |
| Fungi | 0 | 9 | 0 | 2 | 11 |

^aIncludes ferns.

^bIncludes bark beetles.

was significant only for carabids, saproxylic beetles, and fungi.

Halme et al. criticize our extrapolation of the regression equation because the TSA values for which effect size equalled zero were outside the range of observed TSA for carabids and fungi. Nevertheless, 43 years is very close to the minimum TSA for fungi (50 years). More generally, we trust readers will consider only the threshold values we provide as indicative of the time needed for biodiversity to recover because we do not claim these values constitute absolute references for forest management policy.

The example of management intensity, far from nullifying our results, actually confirms and strengthens the conclusion presented in the abstract (p. 102), results (p. 107), and discussion (p. 109) sections: low replication number and poor information on management methods are not sufficient to determine the effect of particular types of management. More generally, MA methods are still under development. The test of interaction between factors is not yet part of statistical software, and this is a challenging issue. Nevertheless, low replication number would prevent us from testing interactions in a robust way.

Taxonomic Generalizations

Our systematic research identified the studies on fungi and saproxylic beetles that were available to us. Certainly, the fact that the fungi kingdom is mainly represented by taxa dependent on deadwood should have been mentioned in the tables, but this is clearly stated in the discussion. In addition, we agree that some bark beetles are early-successional species favored by forestry (e.g., clearfelling), but the majority of them are not, and many are confined to old-growth forests. When we analyzed the two groups separately, we obtained the same trend and can consequently draw the same conclusions: the

mean effect size was negative and significant for bark beetles ($d+ = -0.76$, bootstrap 95% CI = -1.21 to -0.35 , $n = 6$) and negative but marginally significant for the other saproxylic beetles ($d+ = -0.65$, bootstrap 95% CI = -1.41 to -0.01 , $n = 11$). Contrary to Halme et al.'s statement, we did not exaggerate the interpretation of our results.

Criteria of Inclusion

The use of p values and other statistics to estimate an effect size is indeed possible although used relatively less often in MA procedures than mean, standard deviation, and sample size. We did not use such data in our MA for several reasons. First, the exact p , F , or t values need to be available, which was not always the case (e.g., threshold values for probability). Second, if those values were available two problems presented themselves: (1) the statistics could be extracted from a more-or-less elaborated model (i.e., with covariates), and it is not advised to mix different sources of effect sizes in an MA (Rosenberg et al. 2000), and (2) when several treatment classes are compared in a one-way analysis of variance (ANOVA), the statistic tests only whether the means significantly differ from each other. Therefore, it is impossible to transform the F or p value of the ANOVA into an effect size because the effect size has to be computed from control and treatment means. The use of summary statistics is thus the only way to incorporate such results in an MA. Consequently, contrary to Halme et al., we do not believe we overlooked "a great deal of relevant literature" in our MA.

Another point raised by Halme et al. concerns inclusion of the study by Sippola et al. (2002). These authors compared old-growth forests with stands 15 years after they have been clearcut, which we did not consider "young regeneration phase[s]" or clearfelling stands in our protocol. We assumed our selection protocol was restrictive enough regarding the number of studies finally included in our MA. Had we been more restrictive in our inclusion criteria (i.e., excluding young stands), we would have rejected this paper.

Conclusions

The goal of Paillet et al. was not to influence European forest and conservation policies; rather, we sought to provide decision-making tools derived from scientific facts. Both managed and unmanaged forests are needed to preserve European forest biodiversity, but because there are many managed forests and very few old-growth forests, a special effort should be made to create protected reserves, as we suggest.

Most of the comments of Halme et al. (except the suggestion to use p values and other statistics) would reduce the number of comparisons, decrease the power of our MA, and weaken our conclusions. The methodological choices we made have intrinsic limitations and cannot compensate for the weaknesses of the studies fed into our MA, but they are transparent. We chose a set of criteria to produce a standard protocol and followed it as a sound standard scientific practice. Then, we worked with the available data after following our protocol. Moreover, we highlighted that future studies comparing biodiversity of managed and unmanaged forests should better control for other sources of variation than management and should systematically provide summary statistics. Many open questions remain, and key ideas for future research lay ahead.

Literature Cited

- Gurevitch, J., and L. V. Hedges. 2001. Meta-analysis. Combining the results of independent experiments. Pages 347–369 in S. M. Scheiner, and J. Gurevitch, editors. *Design and analysis of ecological experiments*. Oxford University Press, New York.
- Halme, P., T. Toivanen, M. Honkanen, J. S. Kotiaho, M. Monkkonen, and J. Timonen. 2010. Flawed Meta-Analysis of Biodiversity Effects of Forest Management. *Conservation Biology* doi: 10.1111/j.1523-1739.2010.01542.x.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187–211.
- Hurlbert, S. H. 2009. The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology* **123**:434–443.
- Koehnle, T. J., and J. C. Schank. 2009. An ancient black art. *Journal of Comparative Psychology* **123**:452–458.
- Oksanen, L. 2001. Logic of experiments in ecology: is pseudoreplication a pseudoissue? *Oikos* **94**:27–38.
- Oksanen, L. 2004. The devil lies in details: reply to Stuart Hurlbert. *Oikos* **104**:598–605.
- Paillet, Y., et al. 2010. Biodiversity differences between managed and unmanaged forests: meta-analysis of species richness in Europe. *Conservation Biology* **24**:101–112.
- Rosenberg, M. S., D. C. Adams, and J. Gurevitch. 2000. *Metawin: statistical software for meta-analysis*. Sinauer Associates, Sunderland, Massachusetts.
- Schank, J. C., and T. J. Koehnle. 2009. Pseudoreplication is a Pseudoproblem. *Journal of Comparative Psychology* **123**:421–433.
- Sippola, A. L., J. Siitonen, and P. Punttila. 2002. Beetle diversity in timberline forests: a comparison between old-growth and regeneration areas in Finnish Lapland. *Annales Zoologici Fennici* **39**:69–86.
- Underwood, A. J. 1997. *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, Cambridge, United Kingdom.

