Biodiversity in managed and unmanaged forests: compromises in data selection for meta-analysis do not mean compromised results. Reply to Halme et al.

Journal:	*Conservation Biology*
Manuscript ID:	Draft
Wiley - Manuscript Category:	Comment
Date Submitted by the Author:	
Complete List of Authors:	Paillet, Yoan; Cemagref, UR EFNO
Abstract:	

Biodiversity in managed and unmanaged forests: compromises in data selection for meta-analysis do not mean compromised results. Reply to Halme et al.

Yoan Paillet[1,2], Laurent Bergès[1,*], Joakim Hjältén[3], Péter Ódor[4], Catherine Avon[1], Markus Bernhardt-Römermann[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], Risto Virtanen[19]

*corresponding author, mailto: laurent.berges@cemagref.fr

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Running head: Reply to Halme et al.

Word count: 2129 words

Introduction

Meta-analysis (hereafter MA) is a powerful tool to assess general trends and quantitatively synthesize the results of independent studies. However, this procedure has received criticisms, particularly when it has been applied to ecological and conservation biology studies.

In an attempt to provide a state of the art of the effect of forest management on biodiversity, we performed a MA comparing the species richness of managed and unmanaged forests in Europe (Paillet et al. 2010). We intended to review and analyze the recent publications regarding the biodiversity of management of forests. Thus, the opening sentence of Halme et al. (this issue) goes against the goal and the basic philosophy of our paper. Indeed, Paillet at al. (2010, p. 103) provided a balanced view of the contrasting opinions.

Our MA provides basic ecological knowledge needed for conservation and ecologically sustainable forestry. In this paper, we showed that forest management has a negative effect on the biodiversity of forest dwelling species. Because we were aware of the limitations of our MA, we used caution when discussing the results considering that: (i) the effect is strongly heterogeneous between different taxa; (ii) there is a *trend* for recovery of biodiversity once management has been abandoned; (iii) no strong conclusion on the effect of different management types could be drawn from our data due to low replication number. The obvious main conclusion of this paper was that research on the subject in Europe was scarce and that more controlled studies may help answer the questions raised.

However, Halme et al. surprisingly overlook the fact that confounding effects and MA limitations were largely discussed in our paper (p. 109-110). Further, they have claimed that our data selection has 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.

Independence of observations

We share Halme et al.'s concern on proper replication in scientific studies. However, the pseudoreplication issue is much more complex than Halme et al. indicate. Specifically in the case of large scale field experiments, the question on what comprises a replicate has been intensely debated (e.g. Oksanen 2001, 2004; Underwood 1997). 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 in such a short reply but this underscores the complexity of the pseudoreplication question.

Although Hulbert's (1984) paper pinpointed a very important problem in ecology, because of the complexity of this issue, it cannot and should not be used as a universal criterion for accepting or rejecting experimental research; all research 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 piece of research. Contrary to manipulative experiments, we should accept that background variations cannot be fully controlled in mensurative experiments (Hurlbert 1984), and it is often impossible to spatially replicate on a large number of different sites. As clearly stated in our article, the surface area of unmanaged forests is very limited in Europe. Comparatively, the number of managed forest stands is much higher. Thus, it is nearly impossible to control for important factors like site conditions, patch size, landscape context, soil, stand age, tree species composition and land use history. Avoiding pseudoreplication is a desired prerequisite of many experiments, but it is practically impossible to fulfil when investigating unmanaged forests in Europe. Thus, we disagree with Halme et al. that some papers should have been excluded from our meta-analysis due to supposed pseudoreplication problems. These papers have all been published in peer-reviewed journals and have thus been judged to have scientific merit. If we were to subjectively exclude from our analyses the papers we consider flawed (for methodical, statistical or other reasons), our objectivity could be questioned. We therefore chose to include these studies in our analyses.

Halme et al. also questioned 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 finally represent 22% of the total number of comparisons (26 out of 120 in our dataset). Sampling dependence in multiple-treatment studies can be solved in three ways: (i) by using the unmanaged forest stands just once and randomly choosing one managed forest types and leaving out the other types; (ii) by mixing all the managed plots in one; (iii) by using a meta-analysis model with study as a random effect, which controls for this type of dependence (Gurevitch & Hedges 2001).

Distributions of co-variables

The added-value of a 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 across different taxa and covariables is definitely unbalanced, as emphasized by Table 1 and Table 1 p. 104 in Paillet et al. (2010). However, we do not share the opinion that the general trend observed between effect size and Time Since Abandonment (TSA) effect was only an artefact of the unbalanced distribution between vascular plants studies on the one hand and fungi and carabid beetles on the other. Figure 2, p. 107 in Paillet et al. (2010) clearly shows that 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 that analyses separated by taxa almost always provided negative slopes, except for bryophytes and birds (see Table 3, p. 107). Finally, even if the effect of TSA was significant only for carabids, saproxylic beetles and fungi, most of the negative slopes for taxa have much higher value than the slope for all groups.

Halme et al. criticised the 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. However, 43 years is very close to the minimum TSA for fungi (50 years). More generally, we trust the readers of our article to only consider the threshold values we provided as indicative since we never claimed that these constitute absolute references for forest management policy.

The example of management intensity, far from nullifying our results, actually confirms and strengthens the conclusion clearly 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 conclude on the effect of management type. More generally, MA methods are still under development: the test of interaction between factors is not yet implemented in statistical software and this is a challenging issue. However, low replication number would prevent us from testing interactions in a robust way.

Taxonomic generalisations

Concerning fungi and saproxylic beetles, our systematic research identified the studies currently available. 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. Concerning bark beetles, although we agree that some are early-successional species favoured by forestry, e.g. clear-felling, the majority of them are not, and many are confined to old-growth forests. If we analyse the two groups separately, we obtain 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 CI= -1.21 to -0.35, n=6) and negative but marginally significant for the other saproxylic beetles (d+=-0.65, bootstrap 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 relatively less often used in MA procedures than mean, SD and sample size. Several reasons can explain why we did not use such data in our MA. First, the exact p, F or t-values need to be available, which is not always the case (e.g. threshold values for probability). Second, when those values are available, there could be two subsequent problems: (i) 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 a meta-analysis (see Rosenberg et al. 2000, p. 20); (ii) when several treatment classes are compared using a one-way analysis of variance, the statistic simply tests if 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. Coming back to the summary statistics is thus the only way to incorporate such results in a MA. Consequently, contrary to Halme et al., we do not believe that we have overlooked "a great deal of relevant literature" in our MA.

Another point raised by Halme et al. concerns the inclusion of the study by Sippola et al. (2002): this paper compares old growth with 15 years-old stands, which were not considered as "young regeneration phases" nor "clearfelling stands" in our protocol. We assume that our selection protocol was restrictive enough regarding the number of studies finally included in our MA; if we had been more restrictive in our inclusion criteria (i.e. excluding young stands), we would have rejected this paper.

Conclusions

The paper we published does not aim at influencing European forest and conservation policies in any way, but to provide decision-making tools based on scientific facts. Both managed and unmanaged forests are needed to preserve European forest biodiversity, but since there are many managed forests and very few old-growth ones, a special effort should be allocated to create protected reserves, as suggested by Paillet et al. (2010).

Most of the comments of Halme et al. (except the suggestion to use p-values and other statistics) would lead to reduce the number of comparisons, decrease the power of our meta-analysis and weaken our conclusions. The methodological choices we made have intrinsic limitations and cannot compensate for the weaknesses of the studies fed into it, but 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, USA.

Halme, P., T. Toivanen, M. Honkanen, J. S. Kotiaho, M. Monkkonen, and J. Timonen. this issue. Biodiversity effects of forest management: flawed data scrutiny compromises conclusions from meta-analysis. Conservation Biology.

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., L. Bergès, J. Hjältén, P. Ódor, C. Avon, M. Bernhardt-Römermann, R.-J. Bijlsma, L. De Bruyn, M. Fuhr, U. Grandin, R. Kanka, L. Lundin, S. Luque, T. Magura, S. Matesanz, I. Mészáros, M.-T. Sebastià, W. Schmidt, T. Standovár, B. Tóthmérész, A. Uotila, F. Valladares, K. Vellak, and R. Virtanen. 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.

Table 1: Distribution of individual studies used in Paillet et al. (2010) relative to Time Since Abandonment (TSA).

	TSA classes (years)				
	<50	50-75	75-100	>100	Total
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
Non-saproxylic beetles	0	4	0	2	6
Fungi	0	9	0	2	11

[a] including ferns
[b] including bark beetles