O P I N I O N
O P I N I O N

Opinion is intended to facilitate communication between reader and author and reader and reader. Comments, viewpoints or suggestions arising from published papers are welcome. Discussion and debate about important issues in ecology, e.g. theory or terminology, may also be included. Contributions should be as precise as possible and references should be kept to a minimum. A summary is not required.

Comment to Oksanen (2001): reconciling Oksanen (2001) and Hurlbert (1984)

Karl Cottenie and Luc De Meester, Lab. of Aquatic Ecology, Katholieke Universiteit, Leuven, De Beriotstraat 32, BE-3000 Leuven, Belgium (karl.cottenie@bio.kuleuven.ac.be).

17 years and more than 2000 citations later, the publication of the seminal paper of Hurlbert (1984) still manages to rile ecologists. Especially ecologists working with and experimenting on large systems struggle with the implications of pseudoreplication on how to conduct valid experiments. The mere publication of Oksanen's (2001) paper is a proof of this struggle. Oksanen's analysis adds some valuable ideas and suggestions to the debate, but, as we will show, in the end strengthens Hurlbert's suggestions for statisticians, for editors, and for experimentalists.

This paper will try to reconcile the viewpoint of both authors by dividing experimental ecologists into those who recognize the importance of replication and take whatever measures necessary to achieve a good dose of it, those who understand that appropriate scale in the end must always have priority over replication (and recognize the associated consequences of limits in scope and the need for goodwill of the reader), and those who, by mere lack of knowledge or insight, do not see any need for replication (modified from both Hurlbert 1984 and Oksanen 2001). We acknowledge that the first two types can and should actually be united in one and the same person, but we merely want to stress that there is very often an underlying conflict between the perfect experimental design and relevancy of the question addressed. The problem treated in extentio by Hurlbert (1984, Table 1), Sources of confusion) is basically very simple: without replication, a treatment effect is indistinguishable from chance effects. This holds irrespective of the system that is manipulated, and is the basic reason to replicate and disperse treatments. An important problem that led to Oksanen's reply is that creating the perfect experimental design often involves compromises on the scale of the research, which may in the end interfere with the core of the questions addressed. One gets a perfect answer, but ends by being disconnected from the original question. The third type of experimenter mentioned above will inevitably reach plainly incorrect conclusions. The reason for us to write this comment is because we fear that Oksanen's paper may be misinterpreted by many ecologists. We want to prevent researchers from falling into the last category, using Oksanen (2001) as an excuse for not putting their results in the right context.

The value in Oksanen's paper lies in stressing the second kind of experimental ecologist (only implicitly treated by Hurlbert 1984, p. 208, section 6), who experiments with ecosystems or large scale communities, involving very costly and/or intensive manipulations, with the credo of better to do solid science on relevant questions then superb science on 'irrelevant' questions. Although Hurlbert (1984) did not really overlook this category, his main issue was a different one. This leaves many ecologists with the uneasy feeling that the internal dynamics of peer reviewed research may promote well-designed small-scale studies against field studies carried out at a relevant scale, but more difficult to replicate (Carpenter 1990). Given that scale matters, Oksanen (2001) has a valuable point stressing that proper replication and interspersion is only one aspect of good ecological research.

Oksanen (2001) provides a theoretical foundation for unreplicated large scale experiments, the falsificationist perspective, which allows "the experimentalist to check 'yes' or 'no' boxes in a pre-existing test-protocol" (Oksanen 2001). Since predictions are made in the test-protocol, the researcher does not only test for differences (which would be superfluous since no two systems are the same), but also predicts the direction of change. The use of classic (inferential) statistics can, in this context, indeed be considered to be gentleman's behavior towards the reader, supplying objective information on which box to check. The results of the statistical tests are then used merely as an extension of descriptive statistics. This approach, however, has as a severe consequence (touched upon by Hurlbert 1984, p. 193), that it limits the scope of the information gained by the experiment to the specific system tested. To make statements on the population of systems (which is ultimately the goal of many research projects), the correct variation should be taken into account by replication of the experimental units, such that inferential statistics can be used and properly interpreted (Hurlbert 1984). One sample is enough for confirmation at the level of the individual system, but replication is necessary for population inference. To argue that variation will be small (i.e. zero) in biological systems (Oksanen 2001, p. 36), and thus making extrapolations possible to a population of ecosystems of the type studied, is stretching the goodwill of the reader to a degree that can not be justified anymore. As the reader is interested in patterns in the statistical population, the researcher should allow him to judge what the data tell about the population. If this is not possible, the author should explicitly delineate the scope of the results and the patterns found.

Of course Oksanen realizes the above implicitly by advocating the (very valuable) use of meta-analysis, which is nothing else than the replication of the studied systems, resulting in valid statements on the population of the type of systems involved. The possibility to use meta-analysis later on may warrant publication of such unreplicated studies. Of course, this line of reasoning is no excuse for publication of the results of unreplicated experiments that easily could have been replicated, but it is a valuable extra argument for publishing results of studies that report on large-scale experiments that, due to the scale and the complexity of the system, were intrinsically difficult to replicate.

Our position that inferential statistics can be used to assess the reliability of descriptive statistics and as a way to better inform the reader, can be viewed as being in between Oksanen (2001), who advocates inferential statistics in this context, and Hurlbert (1984), who completely condemns this and advocates the sole use of descriptive statistics. It should be stressed that, from a scientific-logical point of view, we completely agree with Hurlbert (1984). It can not be stressed enough that understanding the full implications of Hurlbert's paper should be considered essential for each ecologist. If all authors are fully aware of the potential pitfalls, however, there is no problem in using inferential statistics to inform the reader on the value of the descriptions given. It is imperative, however, that authors should always be very explicit in stating that these statistics can not be used to extrapolate the results to a wider population of systems in case there was no proper replication and interspersion.

Oksanen's rejection of Hurlbert's viewpoint on the need for properly used inferential statistics seems to be partly due to Oksanen's confusion of compound treatments with experimental manipulation, and Oksanen's misinterpretation of Hurlbert's supposed inductionism. What Oksanen calls compound treatments (called randomized but with inter-dependent replicates by Hurlbert 1984, design B-4, Fig. 1) is an (although hidden) example of pseudoreplication, since the basis of statistical tests, independence of experimental units, is violated (Hurlbert 1984, p. 196). This, however, does not imply that all experiments are pseudoreplicated, as Oksanen (2001) claims. In the exclosure example of Oksanen and Henttonen (cited in Oksanen 2001), there is no problem of pseudoreplication if different cages are used as experimental units, rather then subsampling in one cage. The major problem Oksanen alludes to is how experimental manipulation changes "the natural situation", and how conclusions from the experiment are applicable to nature. This is, however, inherent to experimentation, with good experiments trying to minimize the effects of experimental manipulation. Large scale experiments may for certain questions be essential in order to minimize these side effects and to maximize the applicability to the natural situation. To jump from that observation to the conclusion that well-replicated experiments cannot touch the right questions is, however, incorrect. Hurlbert himself (1984: p. 188) explicitly warns: "Clearly the hypothesis is of primary importance, for if it is not, by some criterion, "good," even a well-conducted experiment will be of little value". Pseudoreplication is thus not a pseudoissue, but a valid and important statistical problem that should be taken into account by referees when applicable.

The choice to the experimentalist in planning his experiment does not lie in the position of the work in the logical framework of scientific progress (sensu Oksanen 2001), but in the scope of the interpretation of the results obtained by the experiment. The choice is thus determined by a trade-off between staying as closely as possible to the natural situation and replication with experimental side effects. We believe experimental ecologists should try to make statements on well defined populations of systems, and thus should replicate experimental units and use inferential statistics. However, when practical considerations make replicated experiments difficult to obtain, the resulting experiment can still be used as a check of existing theory on the particular system. In this case, the limited scope of the results should be explicitely stated. For questions that require experiments at a scale at which replication is difficult, meta-analysis on a number of studies may in the end provide a powerful tool to check whether the results are applicable to the population of relevant systems. Since the studies used in such a meta-analysis are then to be considered replicates, however, the researcher carrying out this meta-analysis should also be fully aware of the pitfalls of pseudoreplication.

Acknowledgements – We wish to thank all the members of the Aquatic Ecology Workgroup, in which the basic ideas for this paper were discussed and prepared. We would also like to thank Joost Raeymakers for his contribution of valuabel ideas in the preparation of the paper. K. Cottenie was research assistant of the Fund for Scientific Research-Flanders (Belgium) (F.W.O) when preparing the paper. We acknowledge financial support by the National Fund of Scientific Research-Flanders (grant G.0358.01), from EU project BIOMAN (EVK2-CT-1999-00046), and of the K.U.Leuven Research Fund (OT/00/14).

References

Carpenter, S. R. 1990. Large-scale perturbations: opportunities for innovation. – Ecology 71: 2038–2043.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. – Ecol. Monogr. 54: 187–211.

Oksanen, L. 2001. Logic of experiments in ecology: is pseudoreplication a pseudoissue? – Oikos 94: 27–38.