

Reply to Veresoglou: Overdependence on “significance” testing in biology

In PNAS, we explore the effects of interacting global change factors on the functioning of decomposer communities and show how biotic interactions influence the strength of soil carbon feedbacks to climate change (1). Veresoglou (2) highlights that the highly interactive nature of our multifactor experiment can increase the likelihood of type I errors (i.e., “false positives”), an effect that he refers to as “*P* hacking.” We appreciate this perspective because it provides a platform to discuss what we believe is a critical topic in biology: an overdependence on significant *P* values.

P hacking refers to “data-dredging, snooping, fishing, significance-chasing and double-dipping” (3), essentially any instance of selectively searching for only significant *P* values. However, *P* hacking is not simply the reporting of a large number of *P* values. The multifactor experimental design necessitates the testing of multiple comparisons, but we focus on a specific set of preplanned hypotheses throughout. We discuss “significant” and “nonsignificant” relationships and, following convention, report only the relevant *P* values in the text. The significance values of all preplanned comparisons are presented in the accompanying figures.

Of course, Veresoglou (2) is correct that reporting a large number of *P* values inflates the likelihood of unveiling significant ($P < 0.05$) effects. However, statistics is far from black and white. Adjusting *P* values to protect against type I errors (false positives) comes at a considerable cost: it inflates the probability of type II errors (false negatives) (4). When false negatives are considered to be as

damaging as false positives, then adjusting for multiple comparisons is strongly discouraged (4, 5). That is, “scientists should not be so reluctant to explore leads that may turn out to be wrong that they penalize themselves by missing possibly important findings” (5). This is particularly important in complex ecological systems where mechanisms can be obscured by huge variability. Irrespective of the *P* values we report, many of the trends that we focus on will require further investigation, so a strict “accept/reject” framework would be counterproductive. Instead, we took great care to address a question from multiple angles using all available data and previous research. We then provide raw *P* values so that the reader can interpret relationships as they see fit [following Gelman et al. (4)].

We support Fisher’s original intention for the *P* value; it should be “just one part of a fluid, non-numerical process that blended data and background knowledge to lead to scientific conclusions” (3). The merit of our work does not hinge on the chance identification of some statistically significant effects that we selectively report (1). As highlighted by Veresoglou (2), the hypotheses we test are based on previous work and established theory. Accounting for multiple comparisons in our study would not change our findings. We encourage the reader to assess our work based on all of the evidence provided in the preceding literature, figures, and other statistical values, in addition to the reported *P* values. To conclude, we agree with Leek and Peng (6) that, “arguing about the *P* value is like focusing on a single misspelling, rather than on the faulty logic of a sentence.”

Thomas W. Crowther^{a,b,1}, Daniel S. Maynard^a, Stephen M. Thomas^c, Petr Baldrian^d, Kristofer Covey^a, Serita D. Frey^e, Linda T. A. van Diepen^e, and Mark A. Bradford^a

^aYale School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511; ^bDepartment of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6700AM, Wageningen, The Netherlands; ^cDepartment of Environmental Sciences, University of Helsinki, FI-00014 Helsinki, Finland; ^dLaboratory of Environmental Microbiology, Institute of Microbiology of the Academy of Sciences of the Czech Republic, v.v.i., 14220 Prague 4, Czech Republic; and ^eDepartment of Natural Resources and the Environment, University of New Hampshire, Durham, NH 03824

- 1 Crowther TW, et al. (2015) Biotic interactions mediate soil microbial feedbacks to climate change. *Proc Natl Acad Sci USA* 112(22):7033–7038.
- 2 Veresoglou SD (2015) *P* hacking in biology: An open secret. *Proc Natl Acad Sci USA* 112:E5112–E5113.
- 3 Nuzzo R (2014) Statistical errors. *Nature* 506(7487):150–152.
- 4 Gelman A, Hill J, Yajima M (2012) Why we (usually) don’t have to worry about multiple comparisons. *J Res Educ Eff* 5(2): 189–211.
- 5 Rothman KJ (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1(1):43–46.
- 6 Leek JT, Peng RD (2015) Statistics: *P* values are just the tip of the iceberg. *Nature* 520(7549):612.

Author contributions: T.W.C., D.S.M., S.M.T., P.B., K.C., S.D.F., L.T.A.v.D., and M.A.B. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: thomas.crowther@yale.edu.