

Toward more robust plant–soil feedback research: reply

MATTHEW J. RINELLA¹ AND KURT O. REINHART 

Citation: Rinella, M. J., and K. O. Reinhart. 2019. Toward more robust plant–soil feedback research: reply. *Ecology* 100(9):e02810. 10.1002/ecy.2810

Rinella and Reinhart (2018) discuss problems with methods of Teste et al. (2017), and Teste et al. (2019) assert their methods were unproblematic. Here we start out by showing that a methodological detail of Teste et al. (2017) and many other plant–soil feedback studies represents a strong departure from established scientific practice. To the extent the ideas behind this method spread within ecology and beyond, unreliable results will proliferate, and science will be damaged. We (Reinhart and Rinella 2016, Rinella and Reinhart 2017) and others (Smith-Ramesh and Reynolds 2017) have already warned about the method in question, and Rinella and Reinhart (2018) reiterate the warning and discuss additional concerns regarding Teste et al. (2017). Unfortunately, in addition to Teste et al. (2019), other groups have published confused rebuttals to our warning (Cahill et al. 2017, Gundale et al. 2017, 2019), thereby slowing the return to reliable scientific practice. We share blame for this, because our explanations of the problem have relied on unnecessarily complicated computer simulations and statistical examples. Therefore, we appreciate this chance to explain the problem more simply and clearly. After discussing the methodological problem that Teste et al. (2017) and many other studies hold in common, we discuss other aspects of the Teste et al. (2019) response to Rinella and Reinhart (2018).

THE COMBINED SOIL SAMPLE PROBLEM

The main problem with Teste et al. (2017) and many other plant–soil feedback studies is that they combined and mixed soil samples from multiple experimental units (e.g., sites, plots) that received the same treatment before measuring how organisms in the soils affected test plants. Teste et al. (2017) combined soil samples from

experimental units (i.e., plant neighborhoods) that received the same plant type treatment, and then measured survival and growth of test plants in subsamples drawn from the combined soil mixtures. To provide an example of problems caused by combining soil samples, Reinhart and Rinella (2016) explored the simple case where just one soil organism has a possibility of being present in the experimental units, and the response variable varies with that organism's density. In that case, combining soil samples by treatment approximates averaging data by treatment. Therefore, analyzing data from an improper experiment that combines soil samples approximates analyzing data from a proper experiment that maintains separate soil samples after replacing each raw data value with its corresponding treatment mean. In this sense, the logic behind combining soil samples is portable and could spread to other sciences where, instead of affecting soil handling decisions, it would transform standard statistical approaches. This is a concerning prospect because combining soil samples and replacing raw data with averages erases experimental-unit-to-experimental-unit variation from the data, thereby causing researchers to conclude their data contain stronger signals, and less noise, than they really do. Consequently, standard errors are incorrectly small and null hypotheses of no effect are rejected when they should not be. Elaborating this example from one soil organism to multiple interacting organisms only increases the ways combining soil samples encourage false conclusions.

Extensive confusion surrounds the combined soil problem. Teste et al. (2019) follow Cahill et al. (2017) in arguing that combining soil samples is logical when the goal is estimating treatment means, as opposed to quantifying variability. This is a distinction without a difference, as estimating treatment means is the near-universal goal of plant–soil feedback experiments, and experiments more generally. More importantly, the argument is untrue. The two components of a mean estimate are the most likely value, or point estimate, and the uncertainty estimate (confidence interval, standard error), and the previous paragraph shows combining soil samples leads to wrong uncertainty estimates.

Gundale et al. (2019) compound the confusion. These authors conducted experiments comparing combined and separate soil approaches and found “no consistent differences” between results from the two approaches, and Teste et al. (2019) construe this as empirical vindication of combined soil methods. But efforts to justify the combined soil method empirically are fundamentally misguided, because this method, like the method of substituting means for raw data, is not legitimate on analytic grounds, without reference to experiments. Had the two approaches given the same answer, that would be a coincidence, not an indication the combined soil method is

Manuscript received 20 February 2019; accepted 17 May 2019. Corresponding Editor: Joseph B. Yavitt.

Fort Keogh Livestock and Range Research Laboratory, U.S. Department of Agriculture Agricultural Research Service, 243 Fort Keogh Road, Miles City, Montana 59301 USA.

¹ E-mail: matt.rinella@ars.usda.gov

reliable. Moreover, the two approaches did not give the same answer. Instead, as expected, the combined soil method underestimated the standard errors (Table 4 of Gundale et al. 2019). Gundale et al. (2019) attempt to frame the incorrect standard errors as a virtue, indicating that “the (combined soil) approach... successfully reduce (s)... unexplained variation.” With this troubling statement, Gundale et al. (2019) in effect assert that replacing raw data with means is a logical recourse when data exhibit too much unexplained variation. Gundale et al. (2019) go on to provide concerning experimental design advice. For example, they advocate combining soil samples when determining if “an average soil community from a specific set of plots, stands or sites impact(s) an average subject.” Simple thought experiments reveal grave issues with this advice. Consider that 1 of 10 sites supports a soil mutualist. After soil samples from the sites are combined and mixed, 100% of subsamples drawn from the mixture will contain the mutualist. In this case, do subsamples drawn from the mixture really reflect the “average soil community?” Additionally, the mutualist will impact 100% of subjects grown in the mixture. Considering the mutualist occupied only 10% of sites, is it not misleading to report 100% of subjects being impacted? Similar thought experiments show results of Teste et al. (2017) may have been entirely driven by organisms present in one or a handful of their hundreds of soil samples (Rinella and Reinhart 2018). This seems all the more plausible given that Teste et al. (2017) combined soil samples from multiple sites. Unfortunately, Teste et al. (2017) indicate neither the number of sites in their study nor distances between sites.

STATISTICAL MODELS OF TESTE ET AL. (2017, 2019)

Because of combining soil samples, the Teste et al. (2017) data are not amenable to defensible statistical analysis. Rinella and Reinhart (2018) nevertheless reanalyzed the Teste et al. (2017) data to show that, because of missing model terms, their estimates would remain questionable had they followed a sound research protocol. Teste et al. (2019) perform additional analyses of their earlier data and arrive at different answers than Rinella and Reinhart (2018), but errors and omissions prevent determining why that is. Teste et al. (2019) indicate they “fitted... models... to [their] plant survival data using the model structure suggested by (Rinella and Reinhart [2018]).” Table 1 of Teste et al. (2019) indicates they modeled N-fixing and nonmycorrhizal survival data separately, while Rinella and Reinhart (2018) modeled these data jointly. Therefore, one of two things is incorrect—their Table 1, or their claim they used our model structure. Teste et al. (2019) used a Bayesian multilevel model, so their use of AIC for model selection was inappropriate (Chapter 6 of Gelman et al. 2014), and they indicate nothing about their prior distributions, so it is unclear what affect prior assumptions had on their estimates. In their fig. 1, panels a and b, Teste et al.

(2019) neglected to assign letters denoting significant differences to over half their treatments. Finally, Teste et al. (2019) reanalyze their survival data and claim their new estimates are “qualitatively similar” to their original estimates. This is confusing because their original and new estimates differ markedly both quantitatively and qualitatively (fig. 1 panel a vs. c of Teste et al. 2019), and rather than closely resembling their original estimates, their new estimates more closely resemble our estimates (fig. 1 panel c and d of Rinella and Reinhart 2018). The Teste et al. (2019) reanalysis raises more questions than it answers.

CONCLUDING REMARKS

Teste et al. (2019) state “Contrary to (the Rinella and Reinhart [2018]) belief, the overwhelming majority of soil microbes are not highly virulent pathogens that have large effects on plant growth and survival, even when present at low initial density.” We neither state nor imply this belief. Another false attribution is our “perception that all [plant–soil feedback] studies necessarily aim to estimate *variability* in the effects of soil biota on plant growth.” In fact, our perception is that these studies nearly always seek to estimate means, not variability. Teste et al. (2019) claim the argument that combining soil samples invalidates studies has been “thoroughly refuted by logical and empirical evidence (Cahill et al. 2017, Gundale et al. 2017, 2019).” We have shown here from first principles that combining soil samples invalidates studies. Also, according to the Web of Science, our three papers and Smith-Ramesh and Reynolds (2017) that explain insurmountable problems with combining soil samples have been cited 65 times, whereas the three papers that “thoroughly refute” concerns with combining soil samples have been cited 19 times. Finally, Teste et al. (2019) accuse Rinella and Reinhart (2018) of having “coerced” researchers into using “a single type of experimental design.” We agree with Teste et al. (2019) and Cahill et al. (2017) that there are no methodological “silver bullets.” Instead, there are widely adaptable, canonical methods set on solid theoretical footings, and there are ad hoc methods lacking theoretical support that encourage false conclusions.

LITERATURE CITED

- Cahill, J. F., J. A. Cale, J. Karst, T. Bao, G. J. Pec, and N. Erbilingin. 2017. No silver bullet: different soil handling techniques are useful for different research questions, exhibit differential type I and II error rates, and are sensitive to sampling intensity. *New Phytologist* 216:11–14.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
- Gundale, M. J., D. A. Wardle, P. Kardol, and M. Nilsson. 2019. Comparison of plant–soil feedback experimental approaches for testing soil biotic interactions among ecosystems. *New Phytologist* 221:577–587.

- Gundale, M. J., D. A. Wardle, P. Kardol, W. H. Van der Putten, and R. W. Lucas. 2017. Soil handling methods should be selected based on research questions and goals. *New Phytologist* 216:18–23.
- Reinhart, K. O., and M. J. Rinella. 2016. A common soil handling technique can generate incorrect estimates of soil biota effects on plants. *New Phytologist* 210:786–789.
- Rinella, M. J., and K. O. Reinhart. 2017. Mixing soil samples across experimental units ignores uncertainty and generates incorrect estimates of soil biota effects on plants. *New Phytologist* 216:15–17.
- Rinella, M. J., and K. O. Reinhart. 2018. Toward more robust plant–soil feedback research. *Ecology* 99:550–556.
- Smith-Ramesh, L. M., and H. L. Reynolds. 2017. The next frontier of plant–soil feedback research: unraveling context dependence across biotic and abiotic gradients. *Journal of Vegetation Science* 28:484–494.
- Teste, F. P., P. Kardol, B. L. Turner, D. A. Wardle, G. Zemunik, M. Renton, and E. Laliberté. 2017. Plant–soil feedback and the maintenance of diversity in Mediterranean-climate shrublands. *Science* 355:173–176.
- Teste, F. P., P. Kardol, B. L. Turner, D. A. Wardle, G. Zemunik, M. Renton, and E. Laliberté. 2019. Toward more robust plant–soil feedback research: a comment. *Ecology*. <https://doi.org/10.1002/ecy.2590>