Letter to the editor


Brock et al. (2011) conclude: “Experiments with four replications can be sufficient for the detection of small STC (soil total carbon content)- and STN (soil total nitrogen content)-level changes […] if the assessment is based on repeated measurement in small plots […]”. This conclusion is misleading for several reasons:

1. The authors chose a field, which had received its last green manure three years ago and where the previous crop was corn that returns only little dead material due to the low radiation interception and the high harvest index (Brown et al., 1992; Riley, 2002). Hence fresh organic matter causing the main variation should contribute less in this experiment than in many other cropping systems.

2. The authors applied a repeated measurement at fixed locations, which greatly reduced the spatial variation from the differences. This is apparent from the larger standard deviation (SD) in their measurements (e.g., SD for STC in sampling scheme PT: 0.78 g kg⁻¹) than in the differences between two sampling occasions (SD for STC in sampling scheme PT: 0.03 g kg⁻¹) although Gaussian error propagation predicts that a difference must be subject to larger error than the error of the individual terms of the difference. The error expected from Gaussian error propagation is given by the square root of the sum of the squared individual errors (Schafer, 1978). Assuming that a SD of 0.78 g kg⁻¹ occurred at both sampling dates leads to an expected SD for the STC difference without sampling at fixed locations of 1.10 g kg⁻¹. Neglecting the precondition of excluding the spatial variation by resampling at fixed locations in the conclusion by Brock et al. (2011) is thus misleading by more than a factor of 30 (given by 1.10/0.03).

3. Resampling at fixed locations cannot be perfectly met in typical cases where tillage is applied because tillage will cause soil movement and dispersion of varying extent (Van Oost et al., 2000; Marques Da Silva et al., 2004). Resampling at the same location after tillage thus samples soil originating from unknown positions. The authors excluded this variation by sampling directly before seeding and shortly after harvest without reported tillage between both occasions. Thus, they also excluded soil movement and other tillage-induced sources of variation like unavoidable differences in tillage depth and heterogeneities caused by stratification that characterizes many tillage practices (Franzluebbe, 2002). Excluding tillage will not be appropriate in many cases and has to be at least mentioned as a precondition in the conclusions.

4. The authors compared two sampling schemes. One had 24 samples that were all measured PT while the other had 20 samples RT, but 5 were blocked and combined before measurement leading to 4 measurements in this case. Hence, both treatments almost did not differ in the number of samples (20 vs. 24). SD was smaller for sampling scheme RT (e.g., SD for STC change: 0.03 g kg⁻¹) than for PT (SD: 0.14 g kg⁻¹). Errorwise, the authors assumed that this reduction in SD was caused by the fewer measurements of RT. Applying the t-test function to calculate an unbiased SD (Cybranski et al., 2010) shows that SD should only decrease from 0.14 to 0.13 when only 4 instead of 24 samples would be drawn from the PT population. The main reason for the different SD of both sampling schemes thus is that they differed by two orders of magnitude in the sample volume over which the measurements integrated (scales: ~ 10² cm³ for PT vs. ~ 10⁴ cm³ for RT). A larger support causes regularization (Clark, 1977). Hence, the difference in SD is a striking example of the so-called “modifiable areal unit problem” (Openshaw, 1984) but it does not imply that the number of samples can be reduced from 24 to 4.

5. The low SD of the sampling scheme RT indicates low variation but it does not say anything about the reason. Likely this low SD is due to a high autocorrelation at this scale, which means that the four blocks are not independent measurements but pseudoreplicates (Hurlbert, 1984). Only one block would then be sufficient but many plots with one block would be needed.

6. The authors used only one cropping treatment but got significant but different results for their two sampling schemes. Sampling scheme PT indicated that STC had not significantly (α > 0.05) changed over the growing season, while STN had changed highly significantly (α < 0.01). In contrast, sampling scheme PT indicated exactly the opposite. It showed a significant change of STC (α < 0.05) but no significant change in STN (α > 0.05). No indications exist, which allow to judge, which of both results is in error. We may speculate that sampling scheme PT, which used 24 replicates, is closer to the truth than sampling scheme RT but a third sampling scheme with a much higher number of replicates would have been necessary to prove this. However, it could also well be that both apparently contradicting results indicate the so-called “ecological fallacy” that often is associated with the modifiable areal unit problem (Gowdy and Young, 2002).

Hence, the conclusion should read: “For experiments where homogeneous carbon distribution can be expected and where no tillage happens between samplings, repeated sampling at 20 fixed locations on small plots may be sufficient for the detection of small STC- and STN-level changes even if only four composite samples resulting from five blocked samples are measured, but even if significant these changes may be in error or apply only to the specific areal unit chosen for the sampling scheme and to the specific small plot.”
This long sentence may be shortened to: "Little can be said about the cropping-system impact on soil organic matter levels from short-term field experiments on small plots".

References


Karl Auerswald

Lehrstuhl für Grünlandlehre, Technische Universität München
Alte Akademie 12, 85350 Freising-Weihenstephan, Germany
e-mail: auerswald@wzw.tum.de

© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.plant-soil.com