Received: 30 November 2020 | Revised: 19 February 2021 | Accepted: 9 April 2021

DOI: 10.1111/ejss.13120

ORIGINAL ARTICLE

European Journal of Soil Science WILEY

# Importance of sources of variability, scales and experimental design: A case study about the effects of biochar and slurry application on soil properties in agricultural silty loam soils

**Bernard Ludwig**[1] | **Xiaona Song**[1,2] | **Anna Gunina**[1] | **Isabel Greenberg**[1] | **Michaela A. Dippold**[2] | **Hans-Peter Piepho**[3]

[1]Department of Environmental Chemistry, University of Kassel, Witzenhausen, Germany

[2]Department of Biogeochemistry of Agroecosystems, University of Göttingen, Göttingen, Germany

[3]Institute of Crop Science, Biostatistics Unit, University of Hohenheim, Stuttgart, Germany

**Correspondence**
Bernard Ludwig, Department of Environmental Chemistry, University of Kassel, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany.
Email: bludwig@uni-kassel.de

**Abstract**

In designed experiments, different sources of variability and an adequate scale of measurement need to be considered, but not all approaches in common usage are equally valid. In order to elucidate the importance of sources of variability and choice of scale, we conducted an experiment where the effects of biochar and slurry applications on soil properties related to soil fertility were studied for different designs: (a) for a field-scale sampling design with either a model soil (without natural variability) as an internal control or with composited soils, (b) for a design with a focus on amendment variabilities, and (c) for three individual field-scale designs with true field replication and a combined analysis representative of the population of loess-derived soils. Three silty loam sites in Germany were sampled and the soil macroaggregates were crushed. For each design, six treatments (0, 0.15 and 0.30 g slurry-N $kg^{-1}$ with and without 30 g biochar $kg^{-1}$) were applied before incubating the units under constant soil moisture conditions for 78 days. $CO_2$ fluxes were monitored and soils were analysed for macroaggregate yields and associated organic carbon (C). Mixed-effects models were used to describe the effects. For all soil properties, results for the loess sites differed with respect to significant contributions of fixed effects for at least one site, suggesting the need for a general inclusion of different sites. Analysis using a multilevel model allowed generalizations for loess soils to be made and showed that site:slurry:biochar and site:slurry interactions were not negligible for macroaggregate yields. The use of a model soil as an internal control enabled observation of variabilities other than those related to soils or amendments. Experiments incorporating natural variability in soils or amendments resulted in partially different outcomes, indicating the

Bernard Ludwig and Xiaona Song should be considered joint first authors.

need to include all important sources of variability.

**Highlights**

- Effects of biochar and slurry applications were studied for different designs and mixed-effects models were used to describe the effects.
- Including an internal control allowed observation of, e.g., methodological and analytical variabilities.
- The results suggested the need for a general inclusion of different sites.
- Analysis using a multilevel model allowed generalizations for loess soils.
- The results indicated the need to include all important sources of variability.

**KEYWORDS**

aggregates, biochar, experimental design, field scale, generalization, naturally-occurring variability, random intercept model, sampling design, slurry

## 1 | INTRODUCTION

Ideally, a soil science study should include careful consideration of sources of variability, choice of scale of measurement and experimental (or sampling) design as key components. Such studies would present a priori hypotheses based on the literature and choose the appropriate scale(s) of sampling. Moreover, they would have information on the required degrees of freedom obtained by preliminary experiments to detect differences in the response variables affected by the factors studied, which are statistically significant and of practical relevance (Crawley, 2012; Welham, Gezan, Clark, & Mead, 2014). Such an ideal study would thus focus on variabilities of soils at different scales, amendment variabilities, variabilities due to the specific experimental methodologies used, and analytical variabilities. Although such ideal studies are difficult to carry out with limited funding and project time, many real-world studies may benefit from a stronger emphasis on sources of variability, scales and experimental design.

For example, the choice of field scale versus larger scales, that is specific regions or soil types, is a key consideration for the inference space (McLean, Sanders, & Stroup, 1991) of a study. Studies may or may not involve compositing soils, and foci may involve amendments or soils for which the naturally occurring variability is considered (Welham et al., 2014). We have considered these topics in a case study based on our previous findings (Grunwald et al., 2018) and now further study them using an improved-design experiment on the effects of biochar and slurry treatments on soil biological and physical properties, namely cumulative $CO_2$ emission in an incubation, and macroaggregate formation and associated carbon (C) contents.

Both slurry and biochar have been studied in detail regarding their roles in soil biological, chemical and physical processes (Abiven, Menasseri, & Chenu, 2009; Sun & Lu, 2014). Input of fresh organic matter (OM) such as slurry may markedly stimulate microbial activity and C mineralization in soils, possibly leading to a priming effect (Kuzyakov & Bol, 2006). Additionally, it is well established that fresh OM may improve soil aggregation, mediated by the stimulated microbial activity, which generates binding agents in the aggregation formation process (Abiven et al., 2009; Six, Bossuyt, Degryze, & Denef, 2004). Quantitative relationships are also affected by soil texture, clay mineralogy and contents of cations, as well as aluminium and iron oxides. Moreover, studies utilizing isotopically labelled C and rare earth oxides have indicated that aggregate turnover rate is related to C dynamics and that aggregate formation, stabilization and breakdown follow a first-order kinetic model (Peng, Zhu, Zhang, & Hallett, 2017).

In contrast to fresh OM, the effects of biochar on aggregate stability may be more complex and both positive effects and no effects on macroaggregate dynamics have been reported depending on soil and biochar characteristics as well as biochar amounts (Liu et al., 2014; Sun & Lu, 2014; Zhang, Du, Lou, & He, 2015). As recalcitrant OM, biochar may increase aggregate stability by acting as a binding agent, adsorbing labile organic compounds, so they can be used by microorganisms colonizing the biochar particle surface (thereby promoting excretion of mucilage or development of fungal hyphae), cation bridge formation due to high cation exchange capacity, and adsorption of carboxylic and phenolic functional groups on the surface of biochar (Li, Rubæk, & Sørensen, 2017; Liu et al., 2014; Soinne, Hovi, Tammeorg, & Turtola, 2014; Zheng et al., 2018).

Combined applications of biochar and fresh OM may have different effects to individual applications. For instance, Fungo et al. (2017) reported significant promotion of aggregate formation in an agricultural field experiment with combined application of biochar and green manure on a tropical Ultisol. For German agricultural silty loam soils, results of incubation experiments suggested that slurry promotes the formation of biochar–mineral interactions (Grunwald et al., 2018). Overall, the effects of combined applications on soil properties are not yet sufficiently understood, especially regarding the effects of variabilities, scales and experimental design.

Based on these findings and with a focus on amendment and soil variabilities, we hypothesized that: (a) the use of a design without inclusion of any naturally occurring soil or amendment variability (i.e., use of soil from just one sampling location and amendments from one source) serves as an important internal control, which shows all combined additional sources of variabilities, such as methodological and analytical variabilities; (b) combined effects of biochar and slurry on $CO_2$ emissions and macroaggregate formation may be different for different field-scale sites, and more unexplained variance (residual variance) of the response variable studied is expected compared to the internal control design; and (c) a multilevel analysis using mixed effects models will allow assessment of the generalizability of the results for a population of loess soils. Additionally, we assumed that (d) an adequate sampling design using independent sampling locations, which allows generalization of the results at the respective scales, will produce a model with more unexplained variance for the response variables than that with a problematic design (i.e., compositing soils). Finally, we hypothesized that (e) the importance of the sources of variation for the response variables decreases in the following order: variation of soils (from different sampling locations) > slurries (from different farms) > biochars (from different batches). The objectives were to study the effects of biochar and slurry applications on soil properties for (a) a field-scale sampling design with either a model soil (without natural variability) as an internal control or with composited soils, (b) a design with a focus on amendment variabilities, and (c) three individual field-scale designs with true field replication and a combined analysis representative of the population of loess-derived soils.

## 2 | MATERIALS AND METHODS

### 2.1 | Designs

Seven designs were used, with foci on (I) providing an internal control, (II) the use of composited soils, (IIIa to IIIc) investigating field scales, and (IVa and IVb) the use of a model soil (without natural variability) with a focus on amendment variability. Additionally, a general analysis (IIId) of the combined field scales (IIIa to IIIc) was carried out. The designs are:

I. Internal control (design without any naturally-occurring soil and amendment variability, $n = 3$ for each of the six treatments described below, thus $N$ [total number of replicates over the treatments] $= 18$ for design I): Soil from one sampling location at the Friemar site (characterized subsequently) serving as a model soil was used. Model slurry s1 (without natural variability) and model biochar (without natural variability, mixed biochars b1 + b2 + b3) were applied in the treatments with $n = 3$ (amendments and treatments described below).

II. Design using composited soils as a different soil sampling strategy for the Friemar site ($N = 18$): The setup was the same as in I, except that three composited soils (each representative of one third of the site) were used as field replicates.

III. Designs using three different field scales (IIIa to IIIc) and general analysis of the combined field scales (IIId): Three designs consisted of three loess field-scale sites (IIIa, Friemar; IIIb, Lüttewitz; IIIc, Zschortau; sites characterized subsequently, $N = 18$ for each site). Again, model slurry s1 and model biochar were applied in the treatments with three replicates each, where the replicates are true field replicates for each design. The design IIId ($N = 54$), consisting of the combined data from designs IIIa to IIIc, refers to the combined data analysis representative of a population of loess soils (described below).

IV. Designs with a focus on amendment variabilities: One model soil from Friemar was used in the designs IVa and IVb, where the naturally occurring variabilities of the amendments were incorporated into the experiment by utilizing various biochars (IVa: b1, b2, b3) and slurries (IVb: s1, s2, s3) for the $n = 3$ replicates of each of the six experimental treatments described below (thus $N = 18$ for IVa and IVb).

Figure 1 outlines the different soil and amendment variabilities in the designs.

### 2.2 | Site descriptions and soil sampling

Soils were taken in Spring 2018 from the three long-term field experiments: Friemar (Thuringia), Lüttewitz (Saxony) and Zschortau (Saxony) in Germany. All samples were taken from the 0–30-cm depth of conventionally tilled treatments. The soils in Friemar, Lüttewitz and Zschortau are a Haplic Phaeozem, a Haplic Luvisol and a Gleyic Luvisol, respectively. All three soils have developed on loess and have a silty

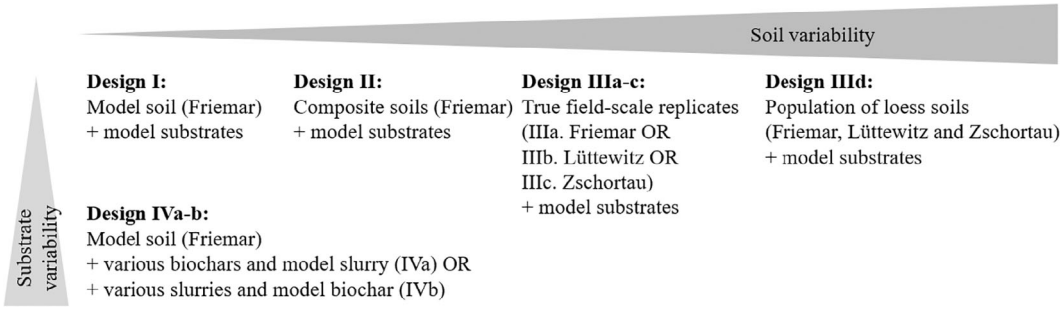


**FIGURE 1** Designs I to IV with respect to soil and amendment variability

clay loam (Friemar: 5% sand, 65% silt, 31% clay) or a silt loam texture (Lüttewitz: 4% sand, 80% silt, 16% clay; Zschortau: 28% sand, 56% silt, 16% clay). Soil organic C contents (in %, means, with standard deviation in parentheses, $n = 3$ true field replicates per site) were 1.5 (0.1), 1.1 (0.1) and 1.3 (0.4) for Friemar, Lüttewitz and Zschortau, respectively. The pH($H_2O$) values (means, with standard deviation in parentheses, $n = 3$ true field replicates per site) were 7.2 (0.2), 7.3 (0.1) and 7.45 (0.04) for Friemar, Lüttewitz and Zschortau, respectively. The sites from different parts of Germany were chosen because they are considered to be representative of a population of loess soils. Detailed site information is given in Koch, Dieckmann, Büchse, and Märländer (2009) and Andruschkewitsch, Geisseler, Koch, and Ludwig (2013).

For the designs IIIa to IIIc for the three sites, three replicate samples of soil (approx. 10 kg per sampling location) were collected from three randomly selected locations using an auger. For designs I, IVa and IVb, only soil from one of the three selected locations was used. For design II using composited soils of the Friemar site, three composited soils ($r = 3$) were obtained by putting a W-shaped path with nine sampling locations over each third of the entire site for each of the three composite soils.

## 2.3 | Soil amendments

The three biochar variants used were produced by Verora GmbH (Switzerland) in different batches but on the same day by pyrolizing wood chips (0–20 mm) from tree and shrub cuttings from settlements and agriculture, mainly consisting of hardwood with a diameter of up to approx. 12 cm, at an approximate mean temperature of 575°C. Variability between biochar batches is caused by slight but continuous changes in reactor conditions. In the designs I to III and IVb, a model biochar, created from mixing the three biochar batches, was used to amend the soil. For design IVa, biochar b1 (C and nitrogen (N) contents: 77.5 and 0.7%), biochar b2 (C and N contents: 78.9 and 0.6%) and b3 (C and N contents: 77.6 and 0.6%) were used separately to investigate variability between biochar batches. The biochar was dried at 40°C and ground to a diameter < 250 μm.

Additionally, in the designs I to III and IVa, a model cattle slurry was used as soil amendment. This model slurry s1 was obtained from the experimental farm at Neu-Eichenberg and contained 6.8% C and a C/N ratio of 14. For design IVb, in addition to the model slurry s1, slurries s2 (from a farm in Frankenhausen with 4.3% C and a C/N ratio of 14) and s3 (from a farm in Dransfeld with 6.5% C and a C/N ratio of 16) were used.

## 2.4 | Setup of the incubation experiment and treatments

To ensure complete disruption of the macroaggregate fraction, each air-dried soil was ground to pass a 250-μm sieve (Andruschkewitsch, Geisseler, Dultz, Joergensen, & Ludwig, 2014; Helfrich, Ludwig, Potthoff, & Flessa, 2008). For each experiment, 180 g dry soil was added to a 1-L incubation vessel. In the biochar treatments, biochar was added at a rate of 30 g $kg^{-1}$ soil ($B_{30}$), which is in the range chosen in previous laboratory studies (e.g., Grunwald et al., 2018; Jones et al., 2011; Zimmerman, Gao, & Ahn, 2011). Slurry was added at rates of 0.15 ($S_{0.15}$) and 0.3 g N $kg^{-1}$ soil ($S_{0.3}$), corresponding to common annual field application rates of 75 and 150 kg N $ha^{-1}$. All mixtures were adjusted to a bulk density of approx. 1.2 g $cm^{-3}$ and a water content of 60% of their water holding capacity (WHC).

In total, we analysed six treatments in a balanced $2 \times 3$ factorial design with two levels of biochar and three levels of slurry: (a) control (no additives); (b–d) biochar addition without slurry ($B_{30}$), with low-level slurry addition ($B_{30}S_{0.15}$), and with high-level slurry addition ($B_{30}S_{0.3}$); and (e–f) low-level slurry addition ($S_{0.15}$) and high-level slurry addition ($S_{0.30}$), both without biochar.

Incubation experiments for the different replications ($n = 3$), treatments ($t = 6$) and designs were carried out as described by Grunwald et al. (2018). Units were incubated in incubation vessels in climate chambers at 15°C for 78 days at 60% WHC. The units were not pre-incubated because the initial microbial activity after water addition might already be connected to aggregate formation.

## 2.5 | Determination of macroaggregate yields

The wet-sieving fractionation method described by Cambardella and Elliott (1993), with later modifications by Jacobs, Rauber, and Ludwig (2009), was used to analyse water-stable macroaggregates. Dried soil (30 g) in a 250-μm sieve was submerged in water for 10 min. Then, the sieve was elevated and resubmerged 50 times in order to separate the fraction >250 μm, which was vacuum-filtered (>0.45 μm), dried at 40°C for 48 h, and weighed.

## 2.6 | Analytics and soil characterization

$CO_2$ emission during the incubation was monitored by chromatographic measurement of the $CO_2$ concentration in the exhaust air, which was collected by constant flushing of each headspace of the vessels with fresh air (Loftfield, Flessa, Augustin, & Beese, 1997).

Macroaggregate fractions and bulk soils were ball-milled and analysed for total C concentrations using a CN elemental analyser (ElementarVario El, Heraeus, Hanau, Germany), which correspond to organic C owing to the absence of carbonates. Soil pH was measured at a soil: water ratio of 1:2.5 (weight/weight). Air-dried soil (10 g) and 25 mL of deionised water were shaken together for 1 min and left to settle for 30 min. This step was repeated once more before pH was determined with a pH electrode (Luo, Durenkamp, De Nobili, Lin, & Brookes, 2011).

## 2.7 | Statistical analyses

All statistical analyses were performed separately for the different designs with R version 4.0.2 (R Core Team, 2020). Combined analysis of variance (ANOVA) and regression analyses were carried out for the response variables cumulative $CO_2$ emission, macroaggregate yield and macroaggregate C contents using linear fixed-effects models for design I and mixed-effects models (random-intercept models) for designs II to IV. This approach allowed variance heterogeneity between designs I to IV. Mixed-effects modelling was performed using the packages lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017).

All initial models included the fixed treatment effects of the factor biochar (levels: $B_{30}$ and no biochar), the quantitative predictor slurry (addition of 0, 0.15 and 0.3 g slurry-N addition $kg^{-1}$ soil) as a regression term, a squared contribution of slurry, and the interaction of biochar and slurry. For the field-scale designs II and IIIa to IIIc, and for designs IVa and IVb with a focus on amendment variability, the effects of block (block levels refer to the sampling locations [II and IIIa to IIIc], biochar batches [IVa] or farms that provided the slurries [IVb]) were considered random effects in the mixed-effects models.

For the analysis that uses all three sites (IIId), the following random effects were additionally included: blocks nested within site (thus, the site main effect and block: site nested effect) and treatment:site interactions (biochar:site, slurry:site, and biochar:slurry:site).

Model simplifications were carried out, in which first a non-significant interaction between the main effects biochar and slurry and then non-significant main effects were eliminated (Crawley, 2012). Thus, fixed effects were only considered in the final models in the case of significant contributions. Non-significant effects of the main effects were only included in the case of a significant interaction. Coefficients of determination were calculated for the fixed-effects models in design I and are labelled as $R_f^2$. For the other designs, where mixed-effects models were used, marginal ($R_m^2$) and conditional ($R_c^2$) pseudo-coefficients of determination were calculated, which account for the variance explained by fixed effects ($R_m^2$) and by both fixed and random effects ($R_c^2$) (Nakagawa & Schielzeth, 2013). Calculations were carried out using the package MuMIn (Barton, 2020).

Residuals of the final model for each variable and design were checked for homoscedasticity graphically and for normal distribution by the Shapiro–Wilk test and graphically by inspecting QQ-plots. For design IIId, we additionally tested for variance homogeneity between sites graphically and using Levene's test. The data for macroaggregate yield in designs IIId and IVb and for -macroaggregate-C in designs IIId, IVa and IVb were log-transformed, because the residual plots indicated non-normality. For these response variables and designs, a comparison of marginal and conditional pseudo-coefficients of determination with those without transformation is problematic and a visual inspection of modelled and measured data becomes important.

In total, two extreme values were removed: an exceptionally small cumulative $CO_2$ emission (which was half of the mean of the other two replicates probably due to a leaky microcosm) in design I in the $B_{30}S_{0.15}$ treatment (replicate 1) and an exceptionally high macroaggregate

yield (which was 2.4 times higher than the mean of the other two replicates) in the same treatment (replicate 3).

## 3 | RESULTS

### 3.1 | Analysis of the internal control (no naturally-occurring soil and amendment variability)

The internal control (design I) had coefficients of determination ($R_f^2$ ) of 0.995, 0.66 and 0.79 for the response variables cumulative $CO_2$ emission, macroaggregate yield and macroaggregate C content, respectively (Figures 2 to 4). For $CO_2$, especially slurry addition explained the variation, whereas the biochar effect was much smaller. Reproducibility was slightly less for the medium addition of slurry compared to no or a high addition (Figure 2). For the macroaggregate yield of the internal control, reproducibility was worse for the treatments without biochar at medium and high levels of slurry addition (Figure 3). For the macroaggregate-C content, a significant interaction was found and reproducibility decreased with increasing slurry addition (Figure 4).

### 3.2 | Analysis using composited soils as a different soil sampling strategy

For all three response variables, the experiments using composited soils (design II) resulted in smaller conditional and marginal pseudo-coefficients of determination $R_c^2$ and $R_m^2$ relative to the $R_f^2$ of the internal control, as expected (Figures 2 to 4). The precisions relative to the $R_c^2$ for the Friemar site using field replicates without compositing (design IIIa) showed no consistent pattern: the use of composited soils resulted in a slightly smaller $R_c^2$ (0.95 vs. 0.99) and markedly smaller $R_c^2$ (0.57 vs. 0.79) for $CO_2$ and macroaggregate yield, respectively, compared to using field replicates, whereas $R_c^2$ of composited soils was higher (0.58 vs. 0.47) for macroaggregate-C compared to using field replicates.

Besides the precision, also the estimates of the effects differed between designs II and IIIa. For $CO_2$, biochar increased the cumulative $CO_2$ emission by 80 mg C $kg^{-1}$ soil in design IIIa using three non-composited field replicates, whereas the use of three composited soils in design II did not have a significant biochar effect (Table 1, Figure 2). The effect of biochar on macroaggregate yield was more pronounced when composited soils were used (−141 g $kg^{-1}$ soil) compared to the use of non-

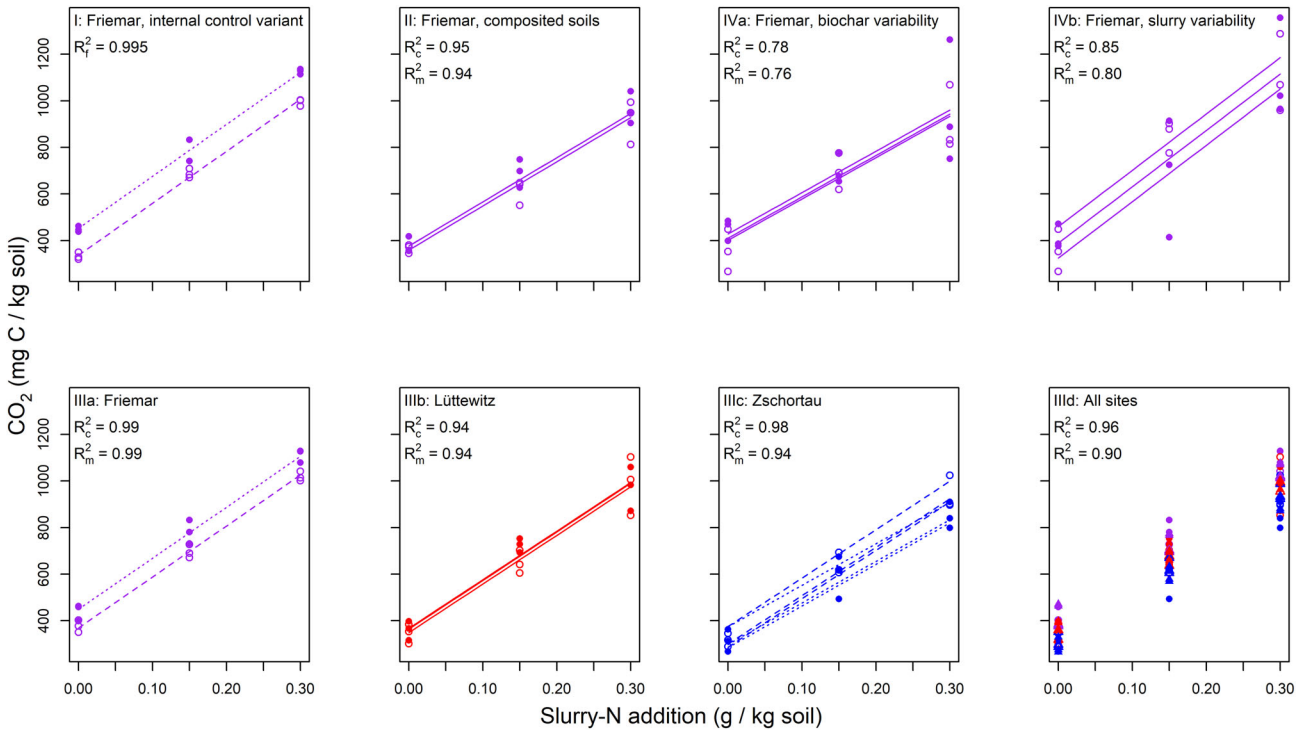


FIGURE 2 Cumulative $CO_2$ emission in response to slurry-N application. Closed (treatments with biochar addition) and open (treatments without biochar addition) circles refer to experimental data. Lines refer to the results of linear fixed and mixed-effects regression models. Solid lines show results for cases where the only significant fixed effect is the slurry application. Dotted lines refer to fixed effects of biochar addition and dashed to those without biochar. Colours refer to the sites Friemar (purple), Lüttewitz (red) and Zschortau (blue). Multiple lines of the same type refer to random effects given in Table 1. For design IIId, triangles refer to model results

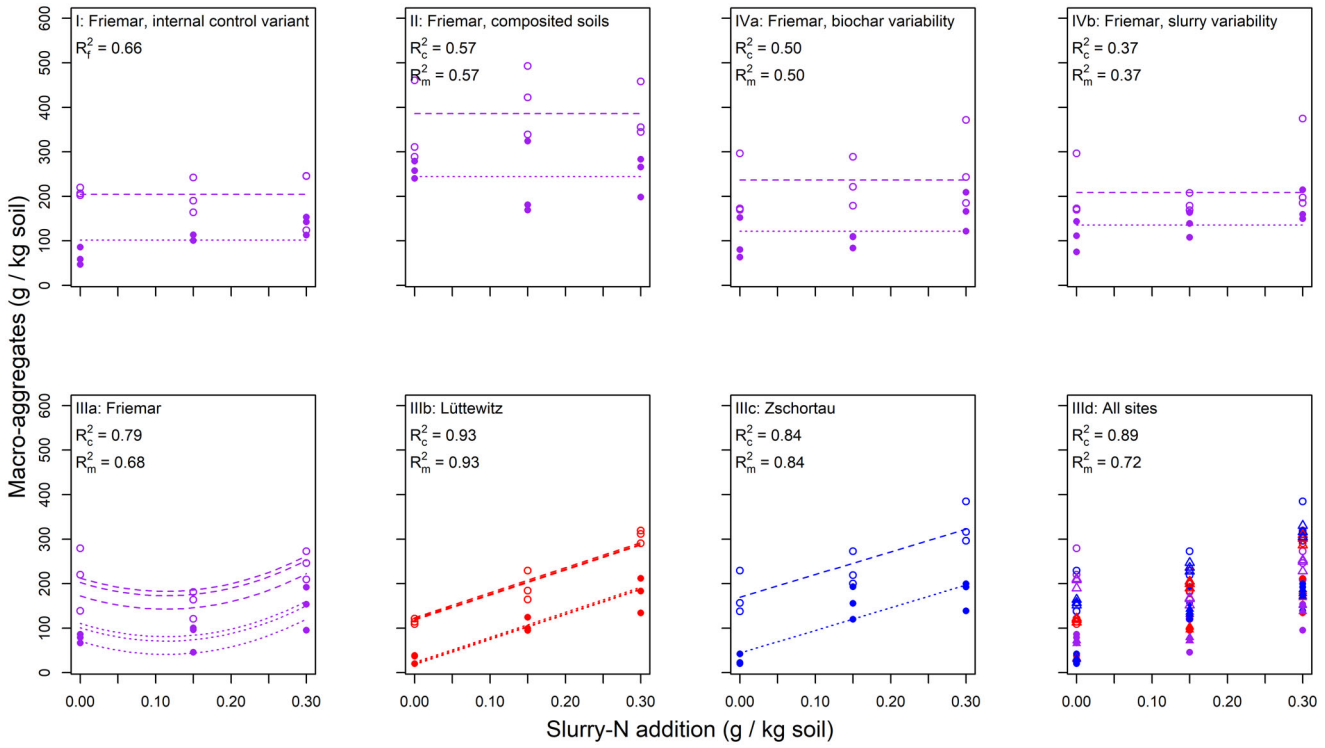


**FIGURE 3** Macroaggregate yield in response to slurry-N application. Closed (treatments with biochar addition) and open (treatments without biochar addition) circles refer to experimental data. Lines refer to the results of linear fixed and mixed-effects regression models. Dotted lines refer to fixed effects of biochar addition and dashed to those without biochar. Colours refer to the sites Friemar (purple), Lüttewitz (red) and Zschortau (blue). Multiple lines of the same type refer to random effects given in Table 2. For design IIId, triangles refer to model results

composited soils ($-102\ g\ kg^{-1}$ soil); however, estimated standard errors were considerable (30 and $16\ g\ kg^{-1}$ soil; Table 2). Macroaggregate-C was significantly positively affected by slurry application in design IIIa, whereas biochar increased macroaggregate-C only when composited soils (design II) were used (Table 3, Figure 4).

## 3.3 | General analysis at different scales

As expected, the inclusion of the naturally occurring soil variability at the Friemar site (design IIIa, three independent sampling locations) compared to the internal control (use of three pseudo-replicates) reduced the variance explained: $R_f^2$ (design I), $R_c^2$ (design IIIa) and $R_m^2$ (design IIIa) were 0.995, 0.988 and 0.988 for the cumulative $CO_2$ emission, and 0.79, 0.47 and 0.47 for macroaggregate C contents. However, for the macroaggregate yield, the opposite was true: the variance explained was less in the internal control ($R_f^2 = 0.66$) than in design IIIa ($R_c^2 = 0.79$, $R_m^2 = 0.68$), suggesting a problematic reproducibility in general and/or that the significant squared contribution of slurry on the macroaggregate yield for Friemar when true field replicates were used (which contributed to the increased $R_c^2$ and $R_m^2$ of 0.79 and 0.68, respectively) may not be reproducible.

Precisions with respect to $R_c^2$ 2 for the sites Lüttewitz (design IIIb) and Zschortau (design IIIc) were higher for macroaggregate yield and macroaggregate C contents, not only relative to the Friemar site with true field replicates, but also to the internal control (Figures 3 and 4). For all three response variables, results differed at least for one site. For $CO_2$, either a positive effect of biochar (Friemar: + $80\ mg\ C\ kg^{-1}$ soil), no significant effect (Lüttewitz) or a significant interaction between biochar and slurry addition (Zschortau) was found (Table 1). Macro-aggregation in the presence of biochar was reduced for all three sites relative to the treatment without biochar, with reductions in the range of 100 to $125\ g\ kg^{-1}$ soil (Figure 3, Table 2). However, the effect of slurry was either linear (Lüttewitz and Zschortau) or quadratic (Friemar). Overall, the reproducibility of the results for the macroaggregate yield may be of concern, because the internal control achieved only an $R_f^2$ of 0.66 (Figure 3). For the macroaggregate-C content (Figure 4), the significance of the effects was site dependent, with slurry having a significant positive effect at all sites, whereas the biochar effect was non-significant (Friemar)

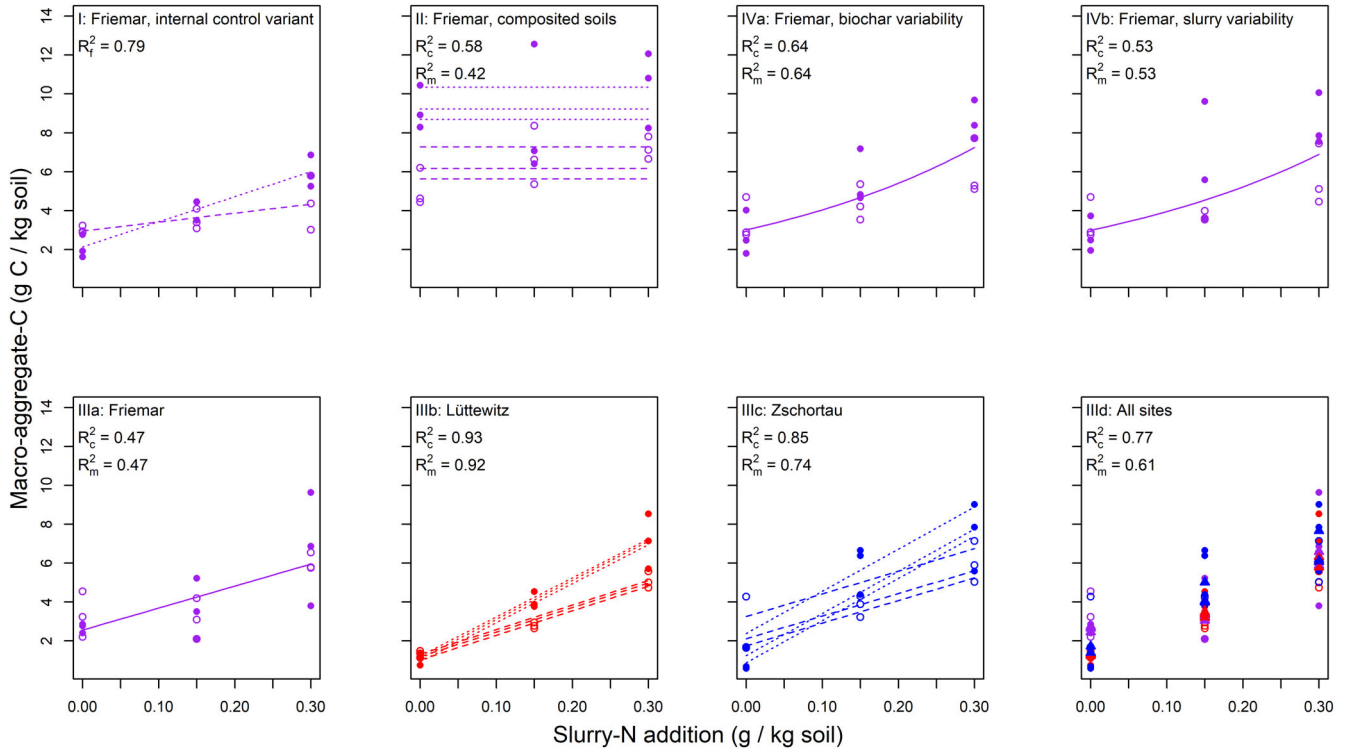


**FIGURE 4** Macroaggregate-C content in response to slurry-N application. Closed (treatments with biochar addition) and open (treatments without biochar addition) circles refer to experimental data. Lines refer to the results of linear fixed and mixed-effects regression models. The solid line shows results for the case where the only significant fixed effect is the slurry application. Dotted lines refer to fixed effects of biochar addition and dashed to those without biochar. Colours refer to the sites Friemar (purple), Lüttewitz (red) and Zschortau (blue). Multiple lines of the same type refer to random effects given in Table 3. For design IIId, triangles refer to model results

or part of a significant interaction with slurry (Lüttewitz, Zschortau, Table 3).

The analysis using all sites as a sample representative of loess soils (design IIId) with a multilevel model resulted in pseudo-coefficients of determination $R_c^2$ of 0.96, 0.89 and 0.77 for the cumulative $CO_2$ emission, macroaggregate yield and macroaggregate C content, respectively. For all response variables, fixed effects were of primary importance, as indicated by $R_m^2$ of 0.90, 0.72 and 0.61 (Figures 2 to 4); however, random effects were not negligible. With respect to the effect of slurry, the importance of a site effect was evident from the estimated variance of the site: slurry interaction for macroaggregate yields and associated organic carbon (C), whereas for $CO_2$, the estimated variance of site–slurry interaction was zero. Moreover, also the site:biochar:slurry interaction was important for macroaggregate yields (Tables 1 to 3).

## 3.4 | Analysis with a focus on amendment variabilities

For the biochar design IVa, the use of three different biochar production batches decreased $R_c^2$ and $R_m^2$ for $CO_2$ and macroaggregate yields relative to the $R_f^2$ of the internal control (Figures 2 and 3). The use of three different biochars affected $R_c^2$ and $R_m^2$ also more than the use of three different soils (IIIa: different sampling locations at the Friemar site) for these response variables.

For the macroaggregate-C in design IVa, a log transformation was required for modelling and the pattern of the data was similar to the one of design IVb (Figure 4).

Modelling indicated that the slurry variability affected the cumulative $CO_2$ emissions differently to the soil variability (Figure 2): in design IVb, the slurry effect was slightly more pronounced than in design IIIa, as indicated by the regression coefficients for the slurry application, and biochar did not have a significant effect, in contrast to design IIIa (Table 1). However, the variation in the response variables of several replicates from different slurry treatments ($CO_2$ and macroaggregate-C: variation especially at the addition of 0.15 g N $kg^{-1}$ soil, Figures 2 and 4; macroaggregate yield: variation especially at the addition of 0.3 g N $kg^{-1}$ soil, Figure 3) hampers the interpretation.

**TABLE 1** Parameters of the final models for the response variable cumulative $CO_2$ emission for all designs. Units for the intercept and biochar effect are given in the first column. The unit for the regression term is kg $g^{-1}$ multiplied by the unit given in the first column

| Site and variant | Intercept | Slurry regression term (mean and *SE*) | Biochar and interaction effect (mean and *SE*) | Random components (assumed mean of 0 and variance) |
|---|---|---|---|---|
| $CO_2$ (mg C $kg^{-1}$ soil) | | | | |
| I. Analysis of the internal control variant (no naturally occurring soil and substrate variability) | | | | |
| Friemar, internal control | 337 | 2,231 (45)<br>[$p < 2 \times 10^{-16}$] | 116 (11)<br>[$p = 7.2 \times 10^{-8}$] | Residual ~ N (0, 544) |
| II. Analysis using composited soils as a different soil sampling strategy | | | | |
| Friemar, composited soils | 371 | 1897 (111)<br>[$p = 8.5 \times 10^{-11}$] | n.s. | Block ~ N (0, 294)<br>Residual ~ N (0, 3,300) |
| III. General analysis at different scales | | | | |
| Friemar | 369 | 2,189 (58)<br>[$p < 3 \times 10^{-16}$] | 80 (14)<br>[$p = 4.8 \times 10^{-5}$] | Block ~ N (0, 0)<br>Residual ~ N (0,919) |
| Lüttewitz | 360 | 2087 (128)<br>[$p = 1.7 \times 10^{-10}$] | n.s. | Block ~ N (0, 319)<br>Residual ~ N (0, 4,409) |
| Zschortau | 320 | 2079 (91)<br>[$p = 2.9 \times 10^{-11}$] | −1.4 (25)<br>[$p = 0.95$]<br>−296 (128)[a]<br>[$p = 0.04$] | Block ~ N (0, 2,469)<br>Residual ~ N (0, 1,114) |
| All three sites | 363 | 2069 (63)<br>[$p < 2 \times 10^{-16}$] | n.s. | Site:Slurry ~N (0, 0)<br>Block:Site ~ N (0, 712)<br>Site ~ N (0, 3,681)<br>Residual ~ N (0, 3,225) |
| IV. Analysis with a focus on substrate variabilities | | | | |
| Friemar, biochar variability | 413 | 1775 (233)<br>[$p = 2.4 \times 10^{-6}$] | n.s. | Block ~ N (0, 783)<br>Residual ~ N (0, 14,665) |
| Friemar, slurry variability | 391 | 2,419 (252)<br>[$p = 1.5 \times 10^{-7}$] | n.s. | Block ~ N (0, 6,513)<br>Residual ~ N (0, 17,127) |

[a]Biochar:Slurry interaction term.
[a]n.s., not significant; *SE*, standard error.

## 4 | DISCUSSION

### 4.1 | Analysis of the internal control (no naturally-occurring soil and amendment variability)

The topic of pseudo-replication is of fundamental importance in many fields of environmental science (Davies & Gray, 2015; Hurlbert, 1984, 2004). Specifically for soil science, problems with the designs of experiments and analyses have been reported (Andren et al., 2008; de Fatima Tavares, de Carvalho, & Machado, 2016), and Webster (2007) pointed out how pseudo-replication affects the hypothesis that is tested. In summary, it is mandatory to "sample all replicates in the field; sampling in the laboratory (pseudo-replication) is no substitute" (Webster, 2007). Especially, for laboratory studies in soil science, it is important to note that a correct sampling design (designs IIIa to IIId) with respect to naturally-occurring soil variability (in contrast to design I with pseudo-replication) does not have any negative effect on the costs and duration of the experiment.

With the above in mind, we see many benefits of including both — i.e., design I as internal control and designs IIIa to IIId with naturally-occurring soil variability — in experimental studies. The benefits are: design I indicates a maximal $R_f^2$ to be achieved and the residual variance in this design is the combined result of analytical inaccuracies, operator variabilities (i.e., weighing errors and heterogeneity of the amendments, which was especially challenging for slurries compared to soils and biochars in the present experiment), and variations induced by the methodology of the incubation experiment (e.g., duration and intensity of the initial mixing of the soil and amendments, variations in temperature and

**TABLE 2** Parameters of the final models for the response variable macroaggregate yield for all designs. Units for the intercept and biochar effect are given in the first column. The units for the first order and second order regression terms are kg $g^{-1}$ and $kg^2$ $g^{-2}$ multiplied by the unit given in the first column

| Site and variant | Intercept | Slurry regression terms (mean and *SE*) | Biochar effect (mean and *SE*) | Random components (assumed mean of 0 and variance) |
|---|---|---|---|---|
| Macroaggregates (g $kg^{-1}$ soil) | | | | |
| I. Analysis of the internal control variant (no naturally occurring soil and substrate variability) | | | | |
| Friemar, internal control | 204 | n.s. | −103 (19)<br>[$p = 7.7 \times 10^{-5}$] | Residual ~ N (0, 1,544) |
| II. Analysis using composited soils as a different soil sampling strategy | | | | |
| Friemar, composited soils | 386 | n.s. | −141 (30)<br>[$p = 2.4 \times 10^{-4}$] | Block ~ N (0, 0)<br>Residual ~ N (0, 4,062) |
| III. General analysis at different scales | | | | |
| Friemar | 196 | −524 (239)<br>[$p = 0.05$]<br>2,300 (764)[a]<br>[$p = 0.01$ ] | −102 (16)<br>[$p = 4.0 \times 10^{-5}$] | Block ~ N (0, 581)<br>Residual ~ N (0, 1,181) |
| Lüttewitz | 121 | 562 (46)<br>[$p = 1.7 \times 10^{-8}$] | −100 (11)<br>[$p = 7.0 \times 10^{-7}$] | Block ~ N (0, 29)<br>Residual ~ N (0, 572) |
| Zschortau | 169 | 510 (77)<br>[$p = 7.7 \times 10^{-6}$] | −125 (19)<br>[$p = 7.4 \times 10^{-6}$] | Block ~ N (0, 0)<br>Residual ~ N (0, 1,587) |
| All three sites[b] | 5.0 | 2.0 (0.9)<br>[$p = 0.04$] | −1.2 (0.2)<br>[$p = 2.0 \times 10^{-4}$]<br>2.7 (0.9)[c]<br>[$p = 0.02$] | Site:slurry:biochar ~ N (0, $3.7 \times 10^{-2}$)<br>Site:slurry ~ N (0, $4.8 \times 10^{-2}$)<br>Block:Site ~ N (0, $5.6 \times 10^{-3}$)<br>Site:biochar ~ N (0, 0)<br>Site ~ N (0, 0)<br>Residual ~ N (0, $5.7 \times 10^{-2}$) |
| IV. Analysis with a focus on substrate variabilities | | | | |
| Friemar, biochar variability | 236 | n.s. | −115 (28)<br>[$p = 8.6 \times 10^{-4}$] | Block ~ N (0, 0)<br>Residual ~ N (0, 3,545) |
| Friemar, slurry variability[b] | 5.3 | n.s. | −0.43 (0.14)<br>[$p = 6.2 \times 10^{-3}$] | Block ~ N (0, 0)<br>Residual ~ N (0, 0.08) |

[a]Squared term for slurry.
[b]For the variant, the response variable was log-transformed.
[c]Biochar:slurry interaction term.
n.s., not significant; *SE*, standard error.

water contents during the incubation) and aggregate fractionation procedure (among others, the intensity of elevation and re-submerging in the fractionation). A general inclusion of such an internal control in soil science studies may even trigger methodological ring trials for different experimental and analytical methods to improve their precision.

The higher $R_f^2$ for design I (Friemar internal control) versus $R_c^2$ and $R_m^2$ for design IIIa (Friemar field replicates) for $CO_2$ and macroaggregate-C is in agreement with our first hypothesis that the use of a design without any naturally-occurring soil and amendment variability serves as an important internal control, which shows all combined additional sources of variabilities. For the cumulative $CO_2$ emission measured by gas chromatography, the internal control of the Friemar soil indicated a very good reproducibility with respect to $R_f^2$ (0.995) and only slight deviations in the three replicate measurements (Figure 2). This shows the importance of the fixed effects studied (i.e., slurry and biochar) for the cumulative $CO_2$ emission and suggests that operator variabilities in the slurry applications and variations in incubation temperature and water contents were not marked in this experiment. In contrast, $R_f^2$ for the macroaggregate yield and macroaggregate-C content was only 0.66 and 0.79, respectively, without any naturally-occurring soil and amendment variability in this internal control. This points to the need for more standardization in the experimental setup and/or the aggregate fractionation procedure and more replications for future aggregation studies, especially for the response variable macroaggregate yield.

**TABLE 3** Parameters of the final models for the response variable macroaggregate-C content for all designs. Units for the intercept and biochar effect are given in the first column. The unit for the first order regression term is kg $g^{-1}$ multiplied by the unit given in the first column

| Site and variant | Intercept | Slurry regression terms (mean and S.E.) | Biochar and interaction effect (mean and S.E.) | Random components (assumed mean of 0 and variance) |
|---|---|---|---|---|
| Macroaggregate-C (g C $kg^{-1}$ soil) | | | | |
| I. Analysis of the internal control variant (no naturally occurring soil and substrate variability) | | | | |
| Friemar, internal control | 3.0 | 4.6 (1.9)<br>[$p = 0.03$] | −0.8 (0.5)<br>[$p = 0.15$]<br>8.3 (2.8)[a]<br>[$p = 9.2 \times 10^{-3}$] | Residual ~ N (0, 0.51) |
| II. Analysis using composited soils as a different soil sampling strategy | | | | |
| Friemar, composited soils | 6.4 | n.s. | 3.1 (0.7)<br>[$p = 1.0 \times 10^{-3}$] | Block ~ N (0, 1.0)<br>Residual ~ N (0, 2.5) |
| III. General analysis at different scales | | | | |
| Friemar | 2.6 | 11.3 (2.9)<br>[$p = 1.3 \times 10^{-3}$] | n.s. | Block ~ N (0, 0)<br>Residual ~ N (0, 2.3) |
| Lüttewitz | 1.2 | 13 (2)<br>[$p = 4.3 \times 10^{-6}$] | −0.1 (0.4)<br>[$p = 0.83$]<br>7 (2)[a]<br>[$p = 6.3 \times 10^{-3}$] | Block ~ N (0, 0.05)<br>Residual ~ N (0, 0.35) |
| Zschortau | 2.4 | 12 (3)<br>[$p = 9.8 \times 10^{-4}$] | −0.9 (0.7)<br>[$p = 0.26$]<br>10 (4)[a]<br>[$p = 0.02$] | Block ~ N (0, 0.75)<br>Residual ~ N (0, 0.98) |
| All three sites[b] | 0.57 | 4.4 (0.8)<br>[$p = 9.8 \times 10^{-4}$] | n.s. | Site:slurry ~ N (0, $6.2 \times 10^{-2}$)<br>Block:Site ~ N (0, $1.5 \times 10^{-2}$)<br>Site ~ N (0, $9.0 \times 10^{-11}$)<br>Residual ~ N (0, 0.11 ) |
| IV. Analysis with a focus on substrate variabilities | | | | |
| Friemar, biochar variability[b] | 1.1 | 2.9 (0.5)<br>[$p = 4.7 \times 10^{-5}$] | n.s. | Block ~ N (0, 0)<br>Residual ~ N (0, 0.08) |
| Friemar, slurry variability[b] | 1.1 | 2.8 (0.6)<br>[$p = 4.4 \times 10^{-4}$] | n.s. | Block ~ N (0, 0)<br>Residual ~ N (0, 0.11) |

[a]Biochar:slurry interaction term.
[b]For the variant, the response variable was log-transformed.
n.s., not significant; *SE*, standard error.

## 4.2 | Analysis using composited soils as a different soil sampling strategy

The benefits and potential disadvantages of compositing soils are well established with respect to analytical determination of element contents. For instance, Alter (2012) summarizes that compositing of soils (or units) saves analytical costs, but is only recommended if the discrete units are expected to have similar chemical composition. With respect to contamination, compositing soils with differing degrees of contamination may result in missing important clues regarding site contamination (Alter, 2012). However, much less is known with respect to compositing soils for designed experiments. We hypothesized that an adequate sampling design using independent sampling locations (designs IIIa to IIId), which allows generalization at the respective scales, will have smaller $R_c^2$ values for the response variables than that with a problematic design (compositing soils, design II), simply because compositing may reduce the naturally-occurring variability and thus move the results closer to those of the internal control. Our result for the macroaggregate-C content is in agreement with this hypothesis: $R_f^2$ and $R_c^2$ values decrease in the order 0.79 (internal control) > 0.58 (design II using composited soils) > 0.47 (design IIIa using independent sampling

locations). For the other two response variables, however, $R_c^2$ for the composited soils was less than that for soils from independent sampling locations (Figures 2–4).

A comparison of the significant contributions shows the results of the experiment using composited soils partly contradicted the field-scale and larger-scale results. In summary, we suggest that the use of composited soils for designed soil science experiments is problematic, because the naturally-occurring variability is reduced, as was found for the macroaggregate-C content. However, our suggestion is not supported by the results for the other two response variables, presumably because of the small number of replicates (sampling locations) used. We argue that it is valid to have (A) a homogeneous (model) soil in experiments in order to isolate the naturally-occurring variability given by growing plant individuals and/or by using substrates or amendments from different batches or locations; and it is equally valid to focus on (B) naturally-occurring variability using soils from different sampling locations. Compositing soils for a designed soil science experiment does not belong to case B and is thus not well defined. However, it is important to have a sufficient number of sampling locations (preferably higher than $n = 3$ used in this study) in order to adequately consider the variation in the field.

### 4.3 | General analysis at different scales

For analysis using true field replicates at all three sites independently (designs IIIa-c), $CO_2$ emissions were greatly affected by the slurry addition (regression coefficients ranged from 2,100 to 2,200 kg $g^{-1}$ × mg C $kg^{-1}$ soil; Table 1), indicating C limitation for the microorganisms in the absence of slurry application. For the $CO_2$ emission from the Lüttewitz soil, biochar did not have a significant effect, in contrast to the significant positive and negative effects on emissions for Friemar and Zschortau soils, respectively (Figure 2). Published studies have shown decreases in $CO_2$ emissions in the presence of biochar and attributed this to OM adsorption by biochar, leaving part of it less accessible for microorganisms (Ameloot, Graber, Verheijen, & De Neve, 2013), or to reduced enzymatic activity (He et al., 2017). However, also a non-significant effect and an increase in $CO_2$ emissions have been reported, which may be related to positive priming effects of biochar (He et al., 2017). The combined application of biogas digestate (1 or 5% w/w) and biochar from orchard pruning residues of fruit trees (1% w/w) to an Italian arable Xerorthent gave significantly smaller cumulative $CO_2$ emissions after 100 days of incubation than the respective applications of biogas digestate without biochar (Cardelli, Giussani, Marchini, & Saviozzi, 2018). These results, which are similar to our results for the Zschortau site, were explained by a reduction of the soluble organic compounds due to biochar application.

For all three sites (designs IIIa-c), there was a marked negative effect of biochar on macroaggregate formation (−100 to −125 g $kg^{-1}$ soil; Table 2), whereas the effect of slurry was linear (Lüttewitz and Zschortau) or quadratic (Friemar; Table 2, Figure 3). A positive effect of fresh OM on macroaggregation (e.g., Andruschkewitsch et al., 2014) is well established and more replicates might have been required for more certain results with respect to a linear or quadratic effect of slurry, especially in light of the lack of any slurry effect for the internal control (Figure 3). The negative effect of biochar may partly be explained by a reduced aggregation efficiency of the added slurry in the presence of biochar and is in agreement with previous studies at loess sites (Grunwald et al., 2018; Grunwald, Kaiser, & Ludwig, 2016), but in contrast to other studies that showed positive or no effects (Wang, Fonte, Parikh, Six, & Scow, 2017; Zhang et al., 2015). Overall, interpretations of macro-aggregation results and comparisons with other studies are difficult, not only because of a suggested need for more standardization in the experimental setup and/or the aggregate fractionation procedure (see above), but also due to differences between the studies in soil environmental conditions and biochar feedstock, application rate, and particle size.

Macroaggregate-C contents increased with slurry application and the biochar addition was either non-significant (for Friemar, where the data were highly variable) or showed a significant interaction with slurry (Table 3, Figure 4). In agreement with the results for Lüttewitz and Zschortau at increasing slurry addition (Figure 4), Wang et al. (2017) reported an increase in C incorporated into macroaggregates due to biochar addition to a Californian agricultural soil.

Overall, for all three response variables, results for the three loess sites differed with respect to significant contributions for at least one site. These findings and the smaller $R_c^2$ and $R_m^2$ results of design IIIa in comparison to the $R_f^2$ results of the internal control (except for macroaggregate yield) are in agreement with our second hypothesis that combined effects of biochar and slurry on $CO_2$ emissions and macroaggregate formation may be different for different field-scale sites, with coefficients of determinations for true field-scale replicates being less than for an internal control at a given site. Thus, there is a need for a general inclusion of different sites and true field replicates in soil science studies.

Our third hypothesis was that a multilevel analysis using mixed-effects models will allow assessment of the generalizability of the results for the population of German

loess soils (design IIId). Overall, the final models were useful for the cumulative $CO_2$ emission ($R_c^2 = 0.96$, $R_m^2 = 0.90$), macroaggregate yield ($R_c^2 = 0.89$, $R_m^2 = 0.72$) and associated organic C ($R_c^2 = 0.77$, $R_m^2 = 0.61$). Site:slurry and site:biochar:slurry interactions were not negligible for macroaggregate yields.

Overall, with respect to the quantitative predictor slurry, the summarizing analysis (IIId) using a multilevel model outperformed the individual site analyses (IIIa to IIIc), because the increased residual degrees of freedom ensured a more accurate analysis (the total number of observations $N$ was 54 in IIId compared to 18 in IIIa to IIIc).

### 4.4 | Analysis with a focus on amendment variabilities

We hypothesized that the importance of the sources of variation for the response variables decrease in the order soils (from different sampling locations) > slurries (from different farms) > biochars (from different batches) for the Friemar site. The results for macroaggregate-C content were in agreement with the hypothesis, whereas for macroaggregate yield and the cumulative $CO_2$ emission, the trend contradicted our hypothesis. Overall, with respect to biochar (from different batches) and slurry (from different farms), some large dispersions of data were noted (e.g., for the $CO_2$ emission at high [design IVa] and moderate [design IVb] slurry additions, Figure 1). Causes are unclear, but may be related to different levels of homogenization of the slurry designs and thus different accuracies for the applied rates.

In summary, experimental designs incorporating naturally occurring variability in soils and amendments resulted in partially different outcomes regarding the effects of slurry and biochar on the response variables compared to the use of homogeneous model soils and amendments, indicating the need to include all important sources of variability for meaningful studies. However, higher accuracies would be required for more specific results with respect to amendment variabilities. This could have been achieved by having more laboratory replicates (and averaging the results of the laboratory replicates for the data analyses) or by including more biochars and slurries from different batches and farms to increase the residual degrees of freedom.

## 5 | CONCLUSIONS

The experimental (in field studies) and/or sampling (in laboratory-based studies) design is an important part of any soil science study. The inclusion of an internal control (without any naturally occurring variability), field scales and a larger scale consisting of several fields – as in this study – has several benefits: The internal control provides information on a maximal $R_f^2$ to be achieved and the obtained residual variance for such a variant is a combined estimate of analytical inaccuracies, operator variabilities and variations induced by the experimental methodologies. The use of several fields gives information on significance of the factors studied for different fields and shows ranges of effect sizes (e.g., differences in regression coefficients). Provided that the different fields studied can be approximately regarded as a random sample of a population of fields of a specific underlying substrate (as in this study), region, soil type or land use, a multilevel analysis using mixed-effects models will allow an assessment of the generalizability of the results for the population.

Future studies may focus more on the relationship between internal controls and the required number of replicates depending on the response variables (and thus the underlying laboratory methodologies). In this study, the number of replicates was sufficient for the response variable cumulative $CO_2$ emission, whereas for macroaggregate yield and macroaggregate-C, a larger number of replicates would have been beneficial, as suggested by the dispersion of the data in the different variants.

### ACKNOWLEDGEMENTS

We would like to thank Anja Sawallisch and her team for technical assistance. We thank the Südzucker AG and Institute of Sugar Beet Research at Georg-August-Universität Göttingen for allowing access to their field experiments. Open access funding enabled and organized by Projekt DEAL.

### AUTHOR CONTRIBUTIONS

**Bernard Ludwig:** Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; writing-original draft. **Xiaona Song:** Data curation; investigation; validation; writing-original draft. **Anna Gunina:** Investigation; writing-review & editing. **Isabel Greenberg:** Writing-review & editing. **Michaela Dippold:** Writing-review & editing. **Hans-Peter Piepho:** Conceptualization; validation; writing-review & editing.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Bernard Ludwig* https://orcid.org/0000-0001-8900-6190
*Isabel Greenberg* https://orcid.org/0000-0002-4762-8474

## REFERENCES

Abiven, S., Menasseri, S., & Chenu, C. (2009). The effects of organic inputs over time on soil aggregate stability – A literature analysis. *Soil Biology & Biochemistry*, *41*, 1–12.

Alter, B. (2012). *Environmental consulting fundamentals*. Boca Raton, FL: CRC Press.

Ameloot, N., Graber, E. R., Verheijen, F. G. A., & De Neve, S. (2013). Interactions between biochar stability and soil organisms: Review and research needs. *European Journal of Soil Science*, *64*, 379–390.

Andren, O., Kirchmann, H., Kätterer, T., Magid, J., Paul, E. A., & Coleman, D. C. (2008). Visions of a more precise soil biology. *European Journal of Soil Science*, *59*, 380–390. https://doi.org/10.1111/j.1365-2389.2008.01018.x

Andruschkewitsch, R., Geisseler, D., Dultz, S., Joergensen, R.-G., & Ludwig, B. (2014). Rate of soil-aggregate formation under different organic matter amendments - a short-term incubation experiment. *Journal of Plant Nutrition and Soil Science*, *177*, 297–306.

Andruschkewitsch, R., Geisseler, D., Koch, H.-J., & Ludwig, B. (2013). Effects of tillage on contents of organic carbon, nitrogen, water-stable aggregates and light fraction for four different long-term trials. *Geoderma*, *192*, 368–377.

Barton, K. (2020). MuMIn: Multi-model inference. R package version 1.43.17. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Cambardella, C. A., & Elliott, E. T. (1993). Carbon and nitrogen distribution in aggregates from cultivated and native grassland soils. *Soil Science Society of America Journal*, *57*, 1071–1076.

Cardelli, R., Giussani, G., Marchini, F., & Saviozzi, A. (2018). Short-term effects on soil of biogas digestate, biochar and their combinations. *Soil Research*, *56*, 623–631.

Crawley, M. J. (2012). *The R book* (2nd ed.). Chichester: Wiley.

Davies, G. M., & Gray, A. (2015). Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). *Ecology and Evolution*, *5*, 5295–5304. https://doi.org/10.1002/ece3.1782

de Fatima Tavares, L., de Carvalho, A. M. X., & Machado, L. G. (2016). An evaluation of the use of statistical procedures in soil science. *Revista Brasileira de Ciência do Solo*, *40*. e0150246. https://doi.org/10.1590/18069657rbcs20150246

Fungo, B., Lehmann, J., Kalbitz, K., Thiongo, M., Okeyo, I., Tenywa, M., & Neufeldt, H. (2017). Aggregate size distribution in a biochar-amended tropical Ultisol under conventional hand-hoe tillage. *Soil and Tillage Research*, *165*, 190–197. https://doi.org/10.1016/j.still.2016.08.012.

Grunwald, D., Kaiser, M., & Ludwig, B. (2016). Effect of biochar and organic fertilizers on C mineralization and macroaggregate dynamics under different incubation temperatures. *Soil and Tillage Research*, *164*, 11–17.

Grunwald, D., Kaiser, M., Piepho, H.-P., Koch, H.-J., Rauber, R., & Ludwig, B. (2018). Effects of biochar and slurry application as well as drying and rewetting on soil macro-aggregate formation in agricultural silty loam soils. *Soil Use and Management*, *34*, 575–583.

He, Y., Zhou, X., Jiang, L., Li, M., Du, Z., Zhou, G., et al. (2017). Effects of biochar application on soil greenhouse gas fluxes: A meta-analysis. *Global Change Biology Bioenergy*, *9*, 743–755. https://doi.org/10.1111/gcbb.12376

Helfrich, M., Ludwig, B., Potthoff, M., & Flessa, H. (2008). Effect of litter quality and soil fungi on macroaggregate dynamics and associated partitioning of litter carbon and nitrogen. *Soil Biology & Biochemistry*, *40*, 1823–1835.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, *54*, 187–211.

Hurlbert, S. H. H. (2004). On misinterpretations of pseudoreplication and related matters: A reply to Oksanen. *Oikos*, *104*, 591–597.

Jacobs, A., Rauber, R., & Ludwig, B. (2009). Impact of reduced tillage on carbon and nitrogen storage of two Haplic Luvisols after 40 years. *Soil and Tillage Research*, *102*, 158–164.

Jones, D. L., Murphy, D. V., Khalid, M., Ahmadd, W., Edwards-Jones, G., & DeLuca, T. H. (2011). Short-term biochar-induced increase in soil $CO_2$ release is both biotically and abiotically mediated. *Soil Biology and Biochemistry*, *43*, 1723–1731.

Koch, H., Dieckmann, J., Büchse, A., & Märländer, B. (2009). Yield decrease in sugar beet caused by reduced tillage and direct drilling. *European Journal of Agronomy*, *30*, 101–109.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Kuzyakov, Y., & Bol, R. (2006). Sources and mechanisms of priming effect induced in two grassland soils amended with slurry and sugar. *Soil Biology and Biochemistry*, *38*, 747–758.

Li, X., Rubæk, G. H., & Sørensen, P. (2017). Availability of potassium in biomass combustion ashes and gasification biochars after application to soils with variable pH and clay content. *Archives of Agronomy and Soil Science.*, *64*, 1119–1130. https://doi.org/10.1080/03650340.2017.1414198

Liu, Z., Chen, X., Jing, Y., Li, Q., Zhang, J., & Huang, Q. (2014). Effects of biochar amendment on rapeseed and sweet potato yields and water stable aggregate in upland red soil. *Catena*, *123*, 45–51.

Loftfield, N., Flessa, H., Augustin, J., & Beese, F. (1997). Automated gas chromatographic system for rapid analysis of the atmospheric trace gases methane, carbon dioxide, and nitrous oxide. *Journal of Environmental Quality*, *26*, 560–564.

Luo, Y., Durenkamp, M., De Nobili, M., Lin, Q., & Brookes, P. C. (2011). Short term soil priming effects and the mineralisation of biochar following its incorporation to soils of different pH. *Soil Biology and Biochemistry*, *43*, 2304–2314.

McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, *45*, 54–64. https://doi.org/10.1080/00031305.1991.10475767

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.

Peng, X., Zhu, Q., Zhang, Z., & Hallett, P. D. (2017). Combined turnover of carbon and soil aggregates using rare earth oxides

and isotopically labelled carbon as tracers. *Soil Biology & Biochemistry*, *109*, 81–94.

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Six, J., Bossuyt, H., Degryze, S., & Denef, K. (2004). A history of research on the link between (micro) aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research*, *79*, 7–31.

Soinne, H., Hovi, J., Tammeorg, P., & Turtola, E. (2014). Effect of biochar on phosphorus sorption and clay soil aggregate stability. *Geoderma*, *219–220*, 162–167.

Sun, F., & Lu, S. (2014). Biochars improve aggregate stability, water retention, and pore-space properties of clayey soil. *Journal of Plant Nutrition and Soil Science*, *177*, 26–33.

Wang, D., Fonte, S. J., Parikh, S. J., Six, J., & Scow, K. M. (2017). Biochar additions can enhance soil structure and the physical stabilization of C in aggregates. *Geoderma*, *303*, 110–117.

Webster, R. (2007). Analysis of variance, inference, multiple comparisons and sampling effects in soil research. *European Journal of Soil Science*, *58*, 74–82. https://doi.org/10.1111/j.1365-2389.2006.00801.x

Welham, S. J., Gezan, S. A., Clark, S. J., & Mead, A. (2014). *Statistical methods in biology. Design and analysis of experiments and regression*. Boca Raton: CRC Press.

Zhang, Q., Du, Z., Lou, Y., & He, X. (2015). A one-year short-term biochar application improved carbon accumulation in large macroaggregate fractions. *Catena*, *127*, 26–31.

Zheng, H., Wang, X., Chen, L., Wang, Z., Xia, Y., Zhang, Y., … Xing, B. (2018). Enhanced growth of halophyte plants in biochar-amended coastal soil: Roles of nutrient availability and rhizosphere microbial modulation. *Plant, Cell & Environment*, *41*, 517–532.

Zimmerman, A. R., Gao, B., & Ahn, M.-Y. (2011). Positive and negative carbon mineralization priming effects among a variety of biochar-amended soils. *Soil Biology & Biochemistry*, *43*, 1169–1179.

**How to cite this article:** Ludwig B, Song X, Gunina A, Greenberg I, Dippold MA, Piepho H-P. Importance of sources of variability, scales and experimental design: A case study about the effects of biochar and slurry application on soil properties in agricultural silty loam soils. *Eur J Soil Sci*. 2021; 72:1954–1968. https://doi.org/10.1111/ejss.13120