RESEARCH ARTICLE

European Journal of Soil Science WILEY

The dilemma of analytical method changes for soil organic carbon in long-term experiments

Kathrin Grahmann[1] | Mariam Zwink[2] | Dietmar Barkusky[1] |
Gernot Verch[3] | Michael Sommer[1,4]

[1]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

[2]Environmental Informatics, University of Applied Sciences for Engineering and Economics (HTW), Berlin, Germany

[3]Research Station Dedelow, Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

[4]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Correspondence
Kathrin Grahmann, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany.
Email: kathrin.grahmann@zalf.de

Abstract

Long-term experiments (LTEs) have provided data to modellers and agronomists to investigate changes and dynamics of soil organic carbon (SOC) under different cropping systems. As treatment changes have occurred due to agricultural advancements, so too have analytical soil methods. This may lead to method bias over time, which could affect the robust interpretation of data and conclusions drawn. This study aims to quantify differences in SOC due to changes in dry combustion methods over time, using soil samples of a LTE established in 1963 that focuses on mineral and organic fertilizer management in the temperate zone of Northeast Germany. For this purpose, 1059 soil samples, collected between 1976 and 2008, have been analysed twice, once with their historical laboratory method right after sampling, and a second time in 2016 when all samples were analysed using the same elementary analyser. In 9 of 11 soil sampling campaigns, a paired t-test provided evidence for significant differences in the historical SOC values when compared with the re-analysed concentrations of the same LTE sample. In the sampling years 1988 and 2004, the historical analysis obtained about 0.9 g kg^{-1} lower SOC compared with the re-analysed one. For 1990 and 1998, this difference was about 0.4 g kg^{-1}. Correction factors, an approach often used to correct for different analytical techniques, could only be applied for 5 of 11 sampling campaigns to account for constant and proportional systematic method error. For this particular LTE, the interpretation of SOC changes due to agronomic management (here fertilization) deviates depending on the analytical method used, which may weaken the explanatory power of the historical data. We demonstrate that analytical method changes over time present one of many challenges in the interpretation of time series data of SOC dynamics. Therefore, LTE site managers need to ensure providing all necessary protocols and data in order to retrace method changes and if necessary recalculate SOC.

KEYWORDS
Bland–Altman, carbon stocks, data trueness, Deming regression, method bias, soil archive, soil survey

1 | INTRODUCTION

Soil organic carbon (SOC) is among the most important indicators of soil quality and sustainable agricultural practices and a key component in soil-based climate change mitigation strategies (Bünemann et al., 2018; Minasny et al., 2017; Stella et al., 2019). A wide range of methods exist to determine SOC concentrations per unit mass under laboratory conditions, in addition to indirect approaches such as mid- and near-infrared (MIR, NIR) spectroscopies (Biney et al., 2021; Leue et al., 2019), which have become increasingly common in assessing SOC at different spatial scales (O'Rourke et al., 2015). However, the most widely established SOC laboratory methods used presently are dry combustion techniques using automatic elemental analysers, which are recognized as gold standard methods (Chatterjee et al., 2009; Davis et al., 2018; FAO, 2020).

Agricultural practices, such as mineral and organic fertilization, have shown significant impacts on soil carbon dynamics and sequestration (Beillouin et al., 2022; Körschens et al., 2013). As SOC changes occur slowly in response to agricultural management (Grahmann et al., 2020; Smith, 2004), years or decades are required to reliably detect and quantify rates of change (Johnston & Poulton, 2018; Smith, 2004). For that reason, long-term experiments (LTEs) have provided the experimental platform to identify optimal combinations of fertilization management with respect to source and application rates and how they influence SOC stocks over time. Thus, LTEs have been fundamental in helping to design sustainable intensification pathways and to help mitigate global warming (Olson et al., 2014; Richter et al., 2007). In this sense, if measured, LTEs produce comprehensive and multidisciplinary SOC data sets to assess long-term sustainability, adaptation to climate change and productivity of cropping systems (Berti et al., 2016; Rasmussen et al., 1998).

When evaluating long-term soil data sets, researchers are confronted with several questions about interpreting and managing data uncertainties generated by analytics over time. Commonly, laboratory methods for a specific soil property may change over the lifespan of a LTE, and there is often a lack of method documentation and verification (Davis et al., 2018; Grahmann et al., 2022). However, to guarantee the quality of long-term soil data and ensuring its correct interpretation, issues associated with the accuracy and bias of analytical methods need to be considered, especially when different analytical equipment is used over time. Trueness in analytical studies refers to systematic error and bias and is defined as the systematic difference between the 'true' value, X, and the mean value of a large number of measurements (van Leeuwen et al., 2021). Systematic bias is expressed

Highlights

- A total of 1059 LTE soil samples taken between 1976 and 2008 were re-analysed for SOC in 2016
- Several methodological changes for SOC determination led to significant different SOC concentration in the same sample
- Interpretation and time series of LTE soil data suffer from consideration of analytical method changes and poor documentation of the same
- Soil archive establishment, thorough method protocols and diligent proficiency testing after soil method changes ameliorate the dilemma

by constant and proportional errors (Magari, 2004). Method bias is especially important for soil survey data of LTEs. Method bias exclusion and robust SOC laboratory results require the unification of methods through re-analysis of archived samples, which is seldom possible due to cost. Often, the application of conversion factors to compare SOC values between different methods is used (Petrokofsky et al., 2012). This is not only important for the assessment of SOC changes within the same experimental unit but also when comparing results across LTEs to address more universal questions. National and international assessment studies of SOC stocks, for example, require the application of conversion factors between different SOC methods (Lettens et al., 2007; Petrokofsky et al., 2012). However, this correction factor or mean recovery rate is not universal and depends on the divergence between both methods, but also on the soil type, sample SOC concentration range and its physical composition (Chan et al., 2011; Díaz-Zorita, 1999; Grahmann et al., 2022). The same applies for the impact of organic fertilizer application and its physical existence in the form of straw and decomposing particulate organic matter in the analyte, which may affect recovery rates between different methods (Carter, 2002; Meersmans et al., 2009).

Several studies have evaluated the effect of soil nutrient status on SOC extraction rates found between different laboratory methods. In particular, the discrepancy found in SOC values produced when using wet and dry combustion techniques, and with increasing SOC concentration (Davis et al., 2018; GLOSOLAN, 2019). Tropical soil samples with high organic carbon concentrations often show an incomplete oxidation using the Walkley–Black method (wet combustion) when compared with dry combustion (Tivet et al., 2012). The influence of laboratory method, mainly caused by different reaction temperatures, accounted for

29% of the SOC concentration variation in a forest soil (De Vos et al., 2007). Average extraction rates of SOC have been found to be about 20% higher with dry combustion methods compared with wet combustion methods (Meersmans et al., 2009; Mikhailova et al., 2003; Tivet et al., 2012). During wet combustion, less SOC is oxidized because the heating obtained by the sulphuric acid dilution (H_2SO_4) is less than that externally supplied (Mikhailova et al., 2003). This incomplete reaction resulted in only 59%–88% of all SOC being measured using the Walkley–Black method (Chatterjee et al., 2009).

This study focuses on the analytical error of SOC in LTE samples caused by different instruments being used over time. Prior to or during analysis, other sources of error may occur during sample collection and preparation (Menditto et al., 2007). For example, errors in the true SOC determination may occur during the sieving of a sample due to different applied forces and durations resulting in more or less contamination with particulate organic material such as straw and roots <2 mm (Díaz-Zorita et al., 2007; Kowalenko, 2001; Xu et al., 2017). Furthermore, a lack of soil sample homogenization by limited milling facilities (FAO, 2018), or bias due to changing laboratory operators over time (Grahmann et al., 2022), may affect the laboratory result. Changes in field sampling design (Olson et al., 2014), natural heterogeneous soil conditions (Brus, 2021; Poeplau et al., 2016) and landscape heterogeneity and topography (Goidts et al., 2009) may also affect the final SOC determination. Additional challenges of SOC determination in LTEs, which are not addressed in this study, are possible technogenic soil shifts in the sampling area, especially due to tillage, which may cause blurring in marginal areas of adjacent treatments (Sibbesen, 1986; Sibbesen et al., 2000). Storage time and soil archive conditions present another uncertain component for SOC determination in LTE studies that deserve more attention in future research (Bergh et al., 2022; Kühnel et al., 2019). However, the re-analysis of historical archived samples, with new state-of-the-art methods, might overcome the bias caused by different instrumentation being used over time (Jensen et al., 2018).

The objectives of this research were (1) to compare different historic SOC analytical methods against a new reference method using archived samples taken between 1976 and 2008 from a LTE; (2) to quantify how changing laboratory methods affect the variation in SOC extraction under different long-term mineral and organic fertilizer management practices; and (3) to assess the challenges and possible solutions that researchers face and seek when analysing and interpreting SOC survey data from LTE. We emphasize that our study cannot quantify other sources of error due to the availability of historical data and the circumstances of archived soil samples.

2 | MATERIALS AND METHODS

2.1 | Study site

The studied LTE 'V140' is located in Müncheberg, Brandenburg, Germany (lat: 52°31′01″ N, long: 14°07′19″ E, 62 m asl.) at the experimental station of the Leibniz Centre for Agricultural Research (ZALF) and was established in 1963. The site has a moderately continental climate and is characterized by a mean annual precipitation of 538 mm and a mean annual temperature of 9.3°C (30-year reference period from 1991 2020 of ZALF weather station). The dominant soil at the site is classified as Albic Luvisol (Arenic, Neocambic) (FAO et al., 2014) with predominant sandy texture of 740 g kg^{-1} (50 g kg^{-1} clay, 210 g kg^{-1} silt), low total carbon concentrations (4.3–5.2 g kg^{-1}), a CEC of 31.5–35.6 mmolc kg^{-1} and a pH (KCl) of 5.4–5.9 in the plough layer (0–25 cm) (Ellerbrock et al., 1999). The LTE area has an elevation difference of about 1.20 m from the north to south diagonal line of about 125 m.

The LTE 'V140' evaluates the effects of different mineral and organic fertilization strategies regarding fertilizer type and dose on crop yield and soil fertility. The crop rotation was cereals alternating with root crops (sugar beets, potatoes, maize) until 1994 and changed in 1995 to winter wheat–silage maize–winter rye–oil flax–winter rye–potatoes–spring barley–peas. The experimental design consists of 21 treatments, varying in mineral and organic N fertilizer source and N fertilizer application rates plus non-fertilized control plots with eight replications per treatment, which are sown with the same crop in 1 year. It has a full randomized block design with 168 individual plots. Fertilizer was either applied solely as mineral NPK (Nitrogen (N) as calcium ammonium nitrate, phosphorus (P) and potassium (K) as mixed mineral fertilizer or in form of pure P and K fertilizers, respectively), or NPK combined with 1.2 or 3.2 t ha^{-1} a^{-1} dry matter (DM) cow farmyard manure applied before the crops potato, silage maize or sugar beet (except for the years 1977–1981, when solid slurry was used) or NPK plus 4.0 t DM ha^{-1} straw residue retention after cereal harvest. The non-fertilized control (Zero NPK) only received a P and K fertilization between 1977 and 1981. Each plot has a size of 30 m^2 (6.0 × 5.0 m). The plots of V140 were regularly limed to maintain optimal pH levels. Long-term crop performance and grain production were evaluated for barley (Thai, Bellingrath-Kimura, et al., 2020) and wheat (Thai, Omari, et al., 2020).

2.2 | Soil sampling and SOC analysis

Since 1963, soil samples were taken regularly every 2 (until 1998) to 4 years (since 2004) by the staff of the

TABLE 1 Features of applied dry combustion laboratory methods for determining SOC (C_t = total carbon, C_t = SOC as soil samples were carbonate-free).

Period (years)	Lab standard	Method description	Method group	Sampling campaign
1970–1993	TGL 25418/04	C_t: Dry combustion at 1050°C in oxygen stream (CO_2–conductivity change) at ZALF	A	1976, 1982, 1984, 1986, 1988, 1990, 1992
1994–2006	DIN ISO 10694:1995	C_t: Dry combustion at 1350°C with CNS-2000 (LECO Corporation, St. Joseph, MI, USA) in Braunschweig, FAL	B	1994, 1998, 2004
2007–2018	DIN ISO 10694:1995	C_t: Dry combustion at 950°C with TruSpec CHN (LECO Corporation, St. Joseph, MI, USA) at ZALF	C	2008, 2016

Experimental Station of ZALF and analysed for several chemical soil properties in the institutional laboratory at ZALF. However, between 1994 and 2004, samples were analysed at FAL (Forschungsanstalt für Landwirtschaft, engl. Agricultural Research Institute) Braunschweig-Völkenrode (now Thünen Institute). Soil samples were always collected after grain harvest and before fertilization applications to a depth of 0–25 cm and 25–50 cm using a soil auger. At each sampling, one composite sample was taken from minimum five locations within each plot. Soil samples were air-dried, sieved <2 mm and milled to <0.5 mm particle size before analysis of SOC according to their respective dry combustion method (Table 1). The individual amount of each analysed soil sample was not recorded, but was assumed similar across treatments and years for the respective elementary analysers and varied between 200 and 500 mg of air-dried and milled soil. Soil moisture of air-dried samples was not determined and was assumed minimal in sandy soils, and hence negligible. This assumption is based on the fact that available unpublished data show an average soil moisture of 0.45 Vol.% (min. = 0.28 Vol.%, max. = 0.69 Vol.%) in air-dried, archived soil samples from 1982 determined by the gravimetric method (105°C) in 18 Ap horizons (top soil) of the experimental area where the LTE is located (Data S1). The amount of water in sandy soils, which is not plant available and below the permanent wilting point of pF 4.2, ranges between 2 Vol.% (coarse sand) and 5 Vol.% (fine sand) (Salter & Williams, 1965). The moisture content of the archived soil samples was less than 1 Vol.%, which is saved and retained in pores of sand particles, making it unavailable for microorganism and excluding microbial activity due to storage (Tecon & Or, 2017). After analysis, samples were packed in paper bags until 1994 and thereafter in sealed polyethylene bags. Bags were stored under unconditioned ambient air and unstable temperature and humidity conditions in buildings used for LTE sample archiving.

In 2015, a set of subsamples was selected for 11 years and 13 treatments (0–25 cm depth) from the sample archive for re-analysis (Table 2). Of five N fertilizer dose treatments, three of them were chosen: low, middle and high N dose (Table 3). For samples from some years, no re-analysis was possible as the amount or replicate number of required samples was not sufficient for re-analysis. A total of 1059 samples collected between 1976 and 2008 with eight soil sample replicates per treatment were re-analysed in 2016 with the reference method of dry combustion using the LECO TruSpec elementary analyser (Table 1). For the year 1976, only half of the re-analysed samples were considered in this data set as irrigation was applied in one single occasion for the other half of the plots. The detection range of total carbon for TruSpec provided by the manufacturer was between 0.02 and 200 mg with a RSD of 0.01 mg or 0.4%. Gazulla et al. (2012) reported a detection limit of 0.03% and a quantification limit of 0.08% determined from the measurements of a soil sample with a low concentration of the analyte of interest, SOC. Due to minimal levels of carbonates, soil samples were analysed for total C, which was assumed to be equal to SOC (C_t = SOC).

2.3 | Statistical analysis

Outliers were determined by the Z-score method for each sampling year within the range of −2.5 to 2.5 (Visser, 2006). The Z-score does not account for different treatments, as the SOC discrepancy between treatments for one sampling year might be quite large due to the contrasting agronomic practices and long running time of this LTE. This excluded from further analysis a total of 26 data pairs (Table 2). Thus, the data set consisted of 2066 data points, and 1033 pairs for method comparison (Table 4). Bland–Altman plots (Bland & Altman, 1999) were created to visualize and to examine the data for

TABLE 2 Number of re-analysed samples per year and summary of outlier removal.

Sampling campaign	Re-analysed samples	Outlier detection	
		Historic	Reference
1976	45	0	1
1982	104	2	3
1984	91	1	1
1986	100	1	1
1988	96	0	1
1990	104	2	1
1992	104	0	2
1994	103	0	0
1998	104	1	3
2004	104	1	4
2008	104	2	2

TABLE 3 Selected long-term experiment treatments and average mineral N fertilizer dose.

Abbreviation	N fertilizer type	N dose (kg/ha) in mineral NPK averaged per crop/crop rotation		
		Low	Medium	High
Zero NPK	No fertilization (except 1977 and 1981, see above)	0	-	-
NPK	Mineral NPK fertilizer	30	98	166
NPK + fym1	Mineral NPK fertilizer	30	98	166
	1.2 t ha^{-1} a^{-1} DM farm yard manure			
NPK + fym2	Mineral NPK fertilizer	0	64	132
	3.2 t ha^{-1}a^{-1} DM farm yard manure			
NPK + Straw	Mineral NPK fertilizer	30	98	166
	4.0 t ha^{-1} DM straw			

Abbreviation: DM, dry matter.

potential trends or patterns. They constitute a simple way to evaluate bias between the mean differences of two methods, and to estimate an agreement interval within which lies the 95% limit of agreement of the second method. In these plots, the x-axis is the mean value of the paired historic and reference SOC method, and the y-axis is the difference in g kg^{-1} of SOC expressed in percent (%) of the paired measurements (Zhu et al., 2021). For this purpose, we subtract the historic value from the reference value. Normal distribution was tested with Shapiro–Wilk for both the historic and reference data sets in each sampling year. Because data were normally distributed in almost all years (except reference data set in 2004), the Pearson coefficient r was used to test for correlation between methods, the coefficient of determination (R^2) was calculated to evaluate the quality of the model fit and the adjusted R^2 considered the impact of additional independent variables for each sampling year. The

concordance correlation was included to quantify the deviation from the 1:1 line. According to the scale developed by McBride (2005), Lin's concordance correlation coefficient (Lin, 1989) is considered poor if it is <0.90, moderate if 0.90–0.95, substantial if 0.95–0.99 and almost perfect if >0.99. The one-way ANOVA fixed effects model was used to calculate the 95% confidence interval (CI) of the intraclass correlation coefficient (ICC). The CI for a difference in proportions is calculated as follows:

$$\text{Confidence interval} = (p_1 - p_2) \pm z \times \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} \tag{1}$$

with p_1 and p_2 being sample 1 proportion and sample 2 proportion, z being the z-critical value based on the confidence level and n_1 and n_2 being sample 1 size and sample 2 size, respectively. Due to the small variances for

TABLE 4 Summary statistics for soil organic carbon (SOC, in g kg^{-1}) measurements in 11 years with the historic and reference laboratory method.

Year / n	1976 / 45	1982 / 99	1984 / 89	1986 / 98	1988 / 95	1990 / 102	1992 / 102	1994 / 103	1998 / 100	2004 / 100	2008 / 100
Range Reference SOC	5.04–7.58	3.70–7.04	4.08–7.68	3.92–7.01	3.88–7.92	3.65–7.39	3.71–7.88	3.52–6.80	3.63–7.24	3.65–7.13	3.66–7.02
Range Historic SOC	5.37–7.77	3.75–6.70	3.95–7.25	3.95–6.70	3.35–6.40	3.15–7.00	3.80–8.20	3.29–6.82	3.15–6.78	3.10–6.37	3.85–7.49
Mean Reference SOC	6.37	5.35	5.72	5.48	5.78	5.62	5.87	5.16	5.56	5.65	5.34
Mean Historic SOC	6.40	5.10	5.50	5.33	4.89	5.21	5.89	4.99	5.08	4.77	5.51
Mean difference (Reference-Historic)	−0.03	0.25	0.22	0.15	0.89	0.41	−0.02	0.17	0.48	0.88	−0.17
Shapiro–Wilk Test Reference	0.84	0.52	0.24	0.83	0.46	0.45	0.87	0.51	0.34	0.02	0.17
Shapiro–Wilk Test Historic	0.69	0.30	0.39	0.85	0.13	0.34	0.73	0.68	0.56	0.40	0.37
Pearson's r	0.70	0.75	0.83	0.85	0.89	0.88	0.82	0.95	0.92	0.76	0.92
R^2	0.49	0.56	0.70	0.73	0.80	0.78	0.67	0.90	0.85	0.58	0.85
Adjusted R^2	0.48	0.56	0.69	0.72	0.80	0.77	0.66	0.90	0.85	0.57	0.85
Concordance correlation	0.70	0.69	0.80	0.82	0.548	0.79	0.81	0.93	0.76	0.45	0.90
ICC1 (95% CI)	0.56–0.81	0.59–0.76	0.72–0.85	0.76–0.87	0.30–0.57	0.70–0.83	NA	0.90–0.95	0.65–0.81	0.16–0.45	0.86–0.92
Paired t-test (p-value)	0.55	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.78	<0.0001	<0.0001	<0.0001	<0.0001
White test	0.66	0.08	0.01	0.48	0.03	0.70	0.49	0.19	0.00	0.31	0.04
(Weighted) Deming intercept	0.08	0.83	−0.42	0.47	0.35	−0.34	−0.74	−0.29	−0.32	−1.08	−0.27
(Weighted) Deming slope	0.99	0.80	1.04	0.89	0.79	0.99	1.13	1.02	0.97	1.04	1.08
Intercept 95% confidence interval	−2.11 to 2.26	0.04 to 1.62	−1.33 to 0.48	−0.07 to 1.01	−0.05 to 0.75	−0.99 to 0.31	−1.59 to 0.11	−0.55 to −0.02	−0.67 to 0.03	−1.99 to −0.18	−0.74 to 0.19
Slope 95% confidence interval	0.65–1.34	0.64–0.95	0.88–1.19	0.79–0.99	0.71–0.86	0.88–1.1	0.98–1.27	0.97–1.08	0.9–1.04	0.88–1.19	0.99–1.17

the values in 1992, the ICC was not calculated for that year (mean difference = −0.02). A paired *t*-test was used to check for significant differences at alpha = <0.05 between the reference and historic method.

Method comparison studies mostly handle common linear regression analysis and coefficients of determination, which have been considered less effective to measure for constant and proportional systematic bias (Ellsäßer et al., 2021; Ludbrook, 2010). Improved methods for regression analysis were suggested to compare different analytical methods (Martin, 2000). One of these, the Deming regression accounts for both the reference and the historic method to contain errors and was applied in this study to compare the methods used for SOC measurements. Deming is different from simple linear regression where only the response variable, *Y*, is measured with error (Ludbrook, 2010) because there will always be some error as well in the *X* values, even if that method is very precise. Simple and weighted Deming regression methods exist (Martin, 2000). If data are heteroscedastic, the weighted Deming should be applied (Bahar et al., 2017). The White test demonstrated that some of the data were heteroscedastic (years 1984, 1988 1998, 2008), the remaining data sets were homoscedastic and analysed with simple Deming. The weighted Deming regression model used the jack-knife method to compute the 95% CIs (Linnets method, Linnet, 1993). Due to the error ratio of the data being unknown, the default value of 1 was used (Ellsäßer et al., 2021). The Deming regression was selected as the most suitable statistical tool for analytical method comparison (Linnet, 1999) when handling large samples of data pairs with *n* > 40, and was applied to confirm if the historic and reference values of the same sample may be used statistically interchangeably. This is only true when the confidence intervals of the Deming slope and intercept include one and zero, respectively (Ellsäßer et al., 2021), and hence include both constant and proportional systematic errors. Finally, the correlation coefficient for each compared year should be close to 1.0 (Ludbrook, 2010). Only if those conditions apply, a conversion factor can be used to correct the historic soil analysis results for standardized SOC determination based on one single method in long-term soil surveys. This conversion, or correction factor, if applicable, was calculated as the quotient of the reference and the historic mean SOC.

The scatter plot visualization was customized to group data points by 'N fertilizer type' as no visual evidence of N dose effect was observed. The Bland–Altman plots and outlier identification were carried out in Excel 2013. Further tests, calculations and plotting were computed in RStudio 1.4.1106, including the weighted Deming regression models. The packages readxl 1.3.1, dplyr 1.0.7, mcr 1.2.2, lmtest 0.9–39, lme4 1.1–27.1 and psych 2.1.9 were used. For concordance correlation, the package cccrm, version 2.1.0 and the method 'concordance correlation coefficient estimated by variance components' was used as data were not repeated.

3 | RESULTS

The mean SOC concentration decreased slightly, but not monotonically, for both comparative data sets over time. Highest average SOC was measured in 1976 with both methods, and the lowest average SOC was measured in 2004 using the historic method and in 1994 when samples were re-analysed. Except for the sampling campaigns in 1976 and 1992, there was high statistical evidence ($p < 0.0001$) that the reference and the historic SOC method resulted in significantly different values (Table 4). The mean difference in SOC between the historic and the reference method was highest with 0.89 g kg^{-1} in 1988 and 2004 and minimal in 1992 (Table 4). Only for the sampling campaign in 2008, did the original analysis show a higher mean SOC of 0.17 g kg^{-1} compared with the re-analysis in 2016, although the same elementary analyser was used; slightly higher historic values were also obtained in 1976 and 1992. We found highly correlated historic and re-analysed values with Pearson's over 0.80 in most sampling campaigns, and moderately lower correlations in 1976, 1982 and 2004 (Table 4).

3.1 | Bland–Altman plot

We selected Bland–Altman plots for a single year with no statistical evidence for differences between methods in 1992 ($p = 0.78$, Figure 1) and a single year with very strong evidence for method differences in 2004 ($p < 0.0001$, Figure 2). In 1992, the Bland–Altman plot showed evenly distributed differences within the CI (except for three samples). However, some patterns emerged in the data. The NPK + Straw treatment mostly resulted in higher SOC concentrations when the historic method was used, whereas the NPK + fym1 and NPK + fym2 treatments tended to have higher SOC using the reference method. In 2004, the Bland–Altman plot showed that more than half of the re-analysed samples resulted in SOC concentrations 20% higher than those produced using the historic method. No clear patterns were detected regarding the differences for N fertilizer type, which appeared to be randomly distributed across the range from 4.5 to 7.0 g kg^{-1}, although the NPK + fym1 treatment resulted in mostly lower method differences. In the same way, samples with higher average SOC showed less differences between methods than
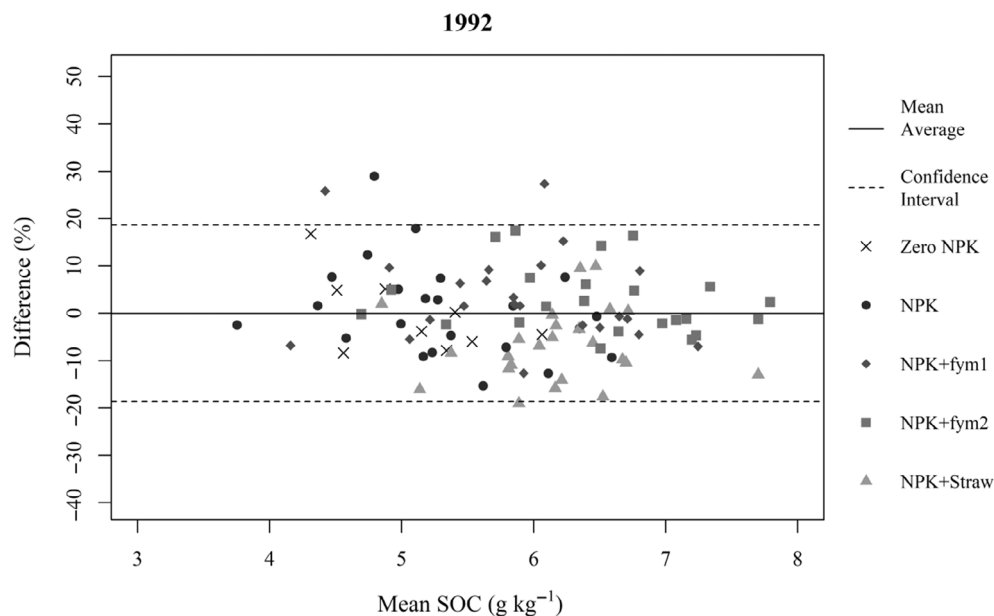
1992

FIGURE 1 Bland–Altman plot of the differences in soil organic carbon (SOC) concentration with the historic and reference method in 1992. Each point represents one replicate soil sample in the long-term experiment for five different treatments. The solid line indicates the mean difference between methods, and the dashed lines indicate 95% confidence intervals.
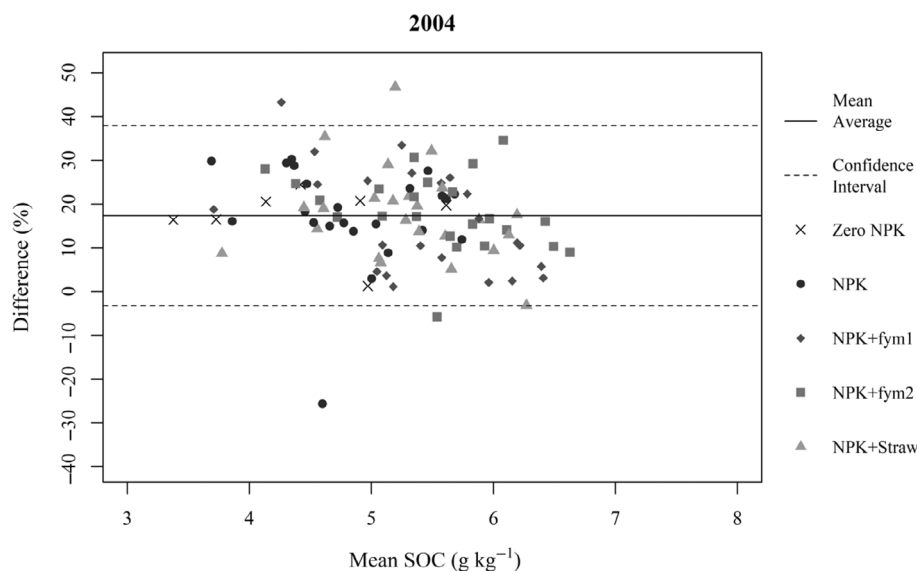
2004

FIGURE 2 Bland–Altman plot of the differences in soil organic carbon (SOC) concentration with the historic and reference method in 2004. Each point represents one replicate soil sample in the long-term experiment for five different treatments. The solid line indicates the mean difference between methods, and the dashed lines indicate 95% confidence intervals.

samples where SOC concentrations were at medium or low levels. The Bland–Altman plots for the remaining years can be found in the Data S2.
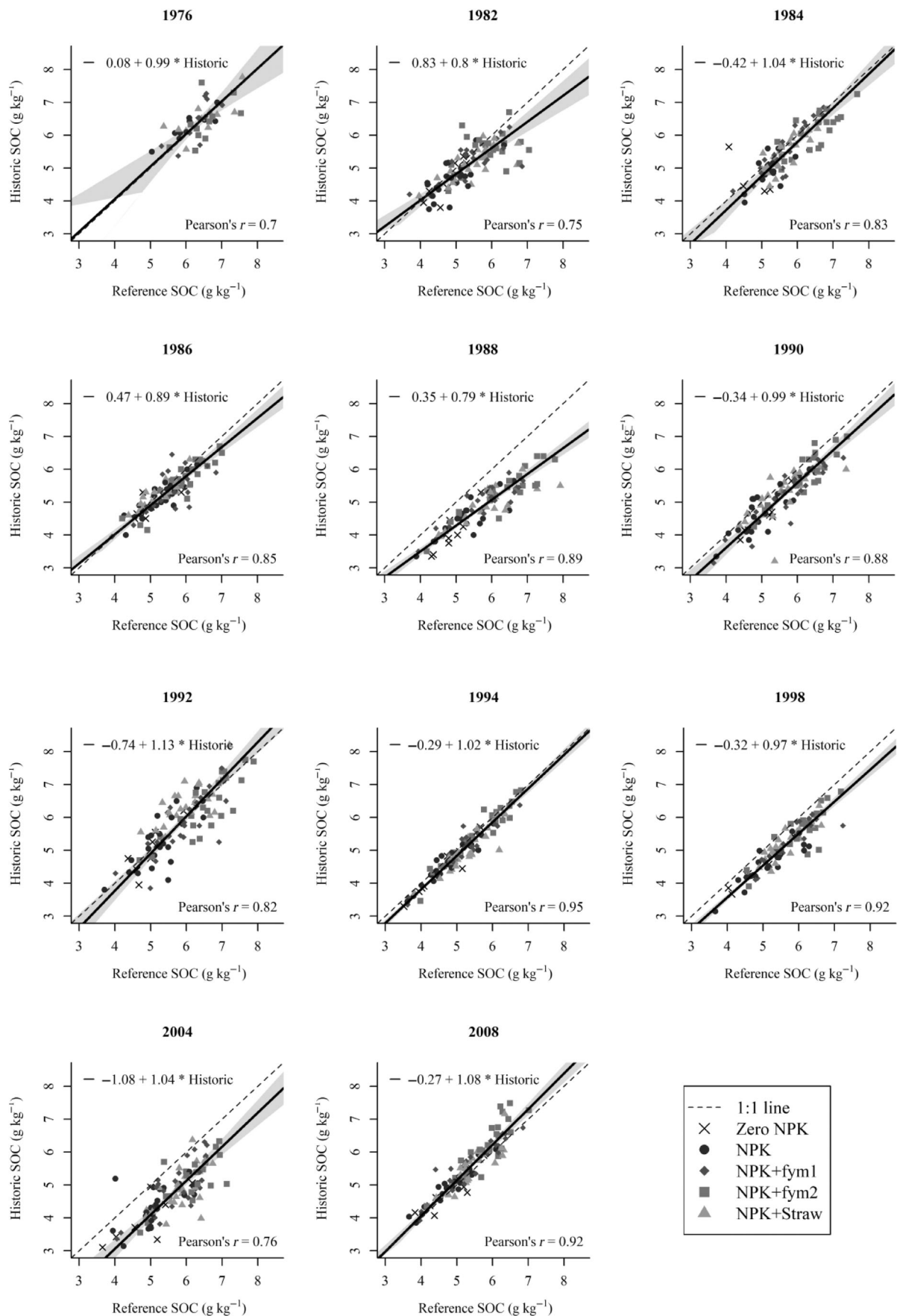
3.2 | Deming regression

The slope of the Deming regression line was slightly greater than the 1:1 line for 5 of 11 sampling years (Table 4, Figure 3) and consistent with the share of samples obtaining slightly minimal SOC differences between methods (except for 2004). The slope of the Deming regression model was far less than 1.0 in 1982, 1986 and 1988 with most samples obtaining higher SOC values

with the reference method. The Deming intercept was slightly negative in 7 of 11 years but showed an overall low value except in 2004. The concordance correlation performed moderately in 1994 and 2008, and was poor in the remaining years (Table 4).

The CIs of the slope of the Deming regression indicate whether there is a proportional systematic error between the methods, which increases proportionally with the magnitude of the predicted value. The 95% CI of the slope did not include 1 for the years 1982, 1988 and in 1986 the upper value was only slightly lower than 1. The remaining 8 years were not identified as having proportional errors between the respective SOC methods. If the CI for the Deming intercept does not include 0, both

FIGURE 3   Deming regression (solid line) for every sampling year. The dashed line of equality (1:1 line) is shown for reference. Data pairs were group according to N fertilizer type. The grey shaded area indicates the confidence region. SOC, soil organic carbon.

Year 1976	Zero NPK NA (4)	NPK 0.80 (10)	NPK + fym1 0.59 (12)	NPK + fym2 0.34 (11)	NPK + straw 0.55 (12)
1982	0.74 (7)	0.53 (24)	0.59 (24)	0.09 (21)	0.55 (23)
1984	0.71 (4)	0.46 (16)	0.80 (22)	0.63 (23)	0.68 (24)
1986	0.47 (7)	0.77 (23)	0.51 (21)	0.80 (23)	0.76 (24)
1988	0.85 (8)	0.71 (21)	0.84 (23)	0.80 (22)	0.65 (21)
1990	0.90 (7)	0.73 (24)	0.85 (23)	0.78 (23)	0.48 (24)
1992	0.72 (7)	0.61 (23)	0.60 (24)	0.71 (24)	0.52 (24)
1994	0.91 (8)	0.90 (24)	0.89 (24)	0.93 (23)	0.81 (24)
1998	0.96 (7)	0.79 (21)	0.87 (24)	0.75 (24)	0.81 (24)
2004	0.60 (8)	0.39 (24)	0.60 (24)	0.56 (21)	0.36 (23)
2008	0.79 (7)	0.93 (23)	0.86 (23)	0.77 (23)	0.74 (24)

TABLE 5 Coefficient of determination (R^2) between historic and reference soil organic carbon method grouped according to N fertilizer type.

Note: Numbers in brackets indicate the number of pairs used for comparison and correlation. Outlier values and incomplete data pairs were deleted from the data set.

methods differ by a constant amount and a constant systematic error occurs in one of the methods compared with the counterpart: this was true for the years 1982 and 2004 and slightly for 1994. When the CI for the intercept of the Deming regression includes 0, there is no constant systematic error between the two methods, which applied for the remaining seven sampling years. In 1982, the Deming regression depicted an increasing method deviation with increasing SOC sample concentration. Altogether, the analysis with Deming and Pearson´s *r* revealed six sampling years (1976, 1982, 1986, 1988, 1994 and 2004) with poor correlation, and a constant or proportional error either in the historic or the reference method, which prevents the application of correction factors for method alignment over time for this particular sampling year. For example, the results for 1982 neither show a one or zero in the slope and intercept CI, respectively, meaning that method correction is not appropriate. Under those circumstances, the identified data set should be excluded from the LTE time series analysis and cannot be aligned analytically as it most likely inherits a larger fraction of random errors that can hardly be corrected. The same applies for the year 1976 where a much smaller statistical association between historic and reference methods was found. The sampling years 1984, 1990, 1992, 1998 and 2008 were most suitable for method correction and allow an alignment through coefficients. In 1992 and 2008, the SOC values obtained with the historic and reference method were very similar and do not demand a correction factor. A correction factor of 1.04 in 1984, 1.08 in 1990 and 1.09 in 1998 can be applied to convert SOC values from the historic to the reference dry combustion method, which adjusts bias due to both, constant and proportional errors.

3.3 | Differences in SOC extraction due to N fertilizer type and analytical method

The alignment between different methods depends on the physical and chemical composition of soil samples and appears to be highly affected by fertilizer management in the present study. The correlation between historic and reference SOC method was relatively low in 1976 for all organic fertilizer type treatments (Table 5). For the samples taken between 1982 and 1992 and in 2004, consistently poor correlations were found in the NPK + Straw samples, and low correlation coefficients were found for other N fertilizer types in other years. For example, in 1982 for NPK + fym2 and in 2004 for NPK (Table 5). In 1986 and 1992, a higher SOC concentration was obtained using the historic method for most NPK + Straw samples, whereas most sample subsets of NPK + fym2 showed substantially higher SOC concentrations using the reference method in 1984, 1986 and 1992 (Figure 3). In 1982, 1984, 1992 and 2004, low correlations between historic and reference SOC method were obtained for NPK samples. The sampling campaigns 1994, 1998 and 2008 showered minimum differences in method alignment according to N fertilizer treatment.

4 | DISCUSSION

4.1 | Data trueness

This study had several possible limitations that should be considered in the interpretation of results.

Firstly, whilst we focus on the error obtained through different SOC analytical techniques, within this error will be any changes caused during the storage of soil samples.

Storage changes are most likely to be associated with microbial activity, but we consider these changes to be small, and they would also be difficult to quantify without setting up an experiment to specifically examine these effects. The degree of microbial activity in archived samples has been investigated in previous studies with varying results. Blake et al. (2000) did not find any impact on total C during storage for air-dried samples analysed first in 1959–1964, stored in a daylit room in lead-sealed or waxed cork-sealed glass bottles and re-analysed in 1991. However, Włodarczyk et al. (2014) found a storage effect on organic C concentration. Soil samples of different soil types were stored over 25 years in the Polish Bank of Soil under air-dry conditions and compared with fresh soil samples taken from the same site in the year of analysis. SOC of stored and fresh samples was determined with a TOC analyser and was found to be lower in stored soils compared with fresh soils, which authors attributed to continuous CO_2 release from stored soils. De Nobili et al. (2006) found similar accelerated CO_2 evolution in 103-year-old samples from the Rothamsted archive in contrast to fresh soil samples taken in an area of the same experiment and stored for only 2 years. Stored and fresh soil samples were air-dried, rewetted and subsequently analysed for SOC and microbial biomass activity and resulted in unexpected fractions of microbial biomass that survived storage (about 20% with 47 years and 10% with 80 years of storage; De Nobili et al., 2006). We remark that this increased SOC concentration in freshly taken soil samples for both studies (De Nobili et al., 2006; Włodarczyk et al., 2014) is likely to be related to land use changes and natural C turnover that took place over decades in the resampled soils, but not necessarily due to storage conditions. However, in the current study, soil moisture content was ~1 Vol. % in the archived air-dried sandy soil samples, which is considered low enough to minimize microbial activity (West et al., 1992). Also, if archiving and storage time would play a role for this set of samples, a deviation from the 1:1 line in a regular form, getting larger over time in one direction, would have been found, which is not the case (Figure 3).

Secondly, the information about discrepancies between SOC methods (Table 1) was collected to the best of our knowledge using the institutional laboratory's method documentation archive, but detailed information was lacking. For example, no detailed records were available on year-by-year historic sample archive conditions (e.g., temperature, humidity), as well as missing in-depth information for occasional soil sample survey years regarding field sampling design or particular laboratory analysis parameters (e.g., operator, soil moisture content of air-dried samples, sample weight). These are commonly reported weaknesses of LTE data collection inventories (Berti et al., 2016; Richter et al., 2007). It is strongly recommended that LTE data holders accurately collect and provide information on analytical methods (including lab protocols) as the experiments last for long periods during which laboratory techniques will most probably change (Paustian et al., 1995). This is very likely with the fast progress in automation, digitalization and technological development (Bergh et al., 2022; Demattê et al., 2019; Smith et al., 2020). We encourage LTE managers to ensure that the collected analytical method information corresponds to a unified and standardized catalogue such as those suggested by the GLOSOLAN initiative (FAO, 2019) or the BonaRes LTE fact sheets (Grosse et al., 2019).

Although we identified groups of soil sampling years where identical SOC methods were used, individual years often showed different statistical outcomes. This was particularly evident for Method Group A (Table 1), which included 7 of 11 re-analysed sampling years. Considerable differences between the historic and reference method accordance values were found between years, with no clear trends being detectable (Table 4). As historic soil samples were sieved at <2 mm and re-analysed without further sieving prior to analysis, pre-treatment effects were estimated to be minimal. In addition, all samples were milled prior to analysis to guarantee a homogenous analyte, thus reducing further sources of error.

As described in Section 3.2, only 5 soil sampling surveys of 11 campaigns allowed the application of a correction factor, and those sampling years were represented by years in different method groups and varied in their correction coefficient. This outcome underpins the challenge researchers' face to use and evaluate LTE soil data for time series analysis when methods have changed. Universal correction factors exist but have been mainly applied to SOC method changes to convert from wet oxidation to dry combustion (Shamrikova et al., 2022; Tivet et al., 2012). However, no SOC correction factors were needed in Blake et al. (2000) as the authors found very good alignment between the acidified dichromate method (Walkley–Black) and the Roboprep CN analyser (Europa Scientific Ltd, Crewe, England), although they included only 27 LTE samples. Within different dry combustion methods, correction coefficients are rarely used. As highlighted by Lettens et al. (2007), it is not recommendable to use universal correction factors. This is in accordance with our findings: in some years corrections factors can be applied, in others not. In the sampling campaigns where correction factors can be applied, they differ from year to year. Site- and year-specific correction factors would be the best strategy to improve data quality, but this is only possible for re-analysed LTE samples or

frequently provided proficiency results by the soil laboratory when SOC methods change over time, which is rarely done (Rayment et al., 2000). It is also a reasonable question to ask whether the use of correction factors is wise if we barely understand the changes to SOC that may occur during archival storage as previously discussed, again demonstrating the importance of re-analysis of samples if methodologies change over time. Data trueness is required by modellers who strive to obtain the 'truest value' possible for SOC dynamic modelling (Berti et al., 2016). As soil LTE data are often used for model calibration and up-scaling purposes, an accurate assessment is indispensable for the determination of SOC dynamics (Post et al., 2008). Errors in the input data of soil parameters derived from different standard soil methods will lead to higher model parameter uncertainty. Billings et al. (2021) draw attention to an advanced understanding of SOC dynamics across environmental gradients if laboratory methods would be standardized.

If correction factors were to be used, we have demonstrated that the SOC concentration differences between dry combustion methods appear to result in lower correction factors than those correction factors applied for conversion from wet to dry combustion methods. Dry combustion in oxygen at temperatures above 1000°C will typically oxidize all soil organic compounds (Shamrikova et al., 2022). Chatterjee et al. (2009) also summarized the importance of temperature for dry combustion for common elementary (auto-) analysers between 950°C and 1200°C. Dry combustion methods in the current study were carried out at different temperatures. In 1988, according to lab standard TGL 25418/04 (Table 1), the combustion temperature was 1050°C, but the re-analysed combustion of archived samples of the same year reached only 950°C (Table 1). All samples had higher SOC concentration with the lower combustion temperature of the reference method being the year with major SOC differences between methods (Figure 3). From 1994 to 2004, analyses at FAL were carried out at 1350°C with a CNS-2000 analyser, the combustion temperature for the re-analysed samples at ZALF during this time was again 950°C (Table 1). However, for this Method Group B, the clear increase in SOC concentrations due to lower combustion temperature of the reference method was only detected in 2004 and to a lesser extent in 1998, but not in 1994 with very similar SOC concentrations for both methods. Hence, we did not find a clear temperature effect across different methods and cannot provide any evidence that lower combustion temperatures were less effective. High temperatures improve and facilitate the combustion of carbonates. Therefore, higher SOC values may be determined at higher temperatures than at lower temperatures for soils containing carbonates (that does not apply for the LTE samples that are carbonate-free) if pre-treatment to remove carbonates is not undertaken.

Straw and root material leftovers in poorly sieved or highly enriched crop residue soils (e.g., LTE soil samples under conservation agriculture) provide disproportionate C input that may overestimate the actual SOC (Ge et al., 2021). Straw amendment in form of organic fertilizer source, as present in the studied LTE, decomposes and turns into particulate organic matter pools, which remain in the <2 mm sieved fraction, being sensitive to sample processing and subsequent analysis of C (Yan et al., 2022). This study has been unable to demonstrate a clear trend towards increased or decreased SOC concentrations with one or the other method when looking at increased portions of organic material through straw or manure inputs because LTE treatments showed varying inconsistent correlations (Section 3.3, Table 5). The gradient of soil aggregation caused by stable chemical organic compounds was found to reduce SOC recovery with wet oxidation methods (Díaz-Zorita, 1999). Forest soils, for example, have a larger proportion of macroaggregates in the top soil layer compared with agricultural soils that may protect the SOC contained against Walkley–Black oxidation (De Vos et al., 2007; Lettens et al., 2007). However, our data do not support any specific trend between fertilizer type (and hence organic compound alteration) and method bias. Apart from the soil sample composition caused by organic material like straw and root residues, soil type was found to affect method-dependent SOC extraction. De Vos et al. (2007) reported a strong correlation between SOC recovery percentage (ratio of Walkey–Black wet digestion and total organic carbon analyser SOC result) and soil textural classes. Recovery was higher by 3%–8% in sandy soils than in loam and silt-loam soils. Also, Lettens et al. (2007) found higher recovery of SOC in sandy soils compared with silt loam. Hence, the high recovery and small correction values in the present study capture a favourable situation for sandy soils, but might result in lower method alignment or even SOC overestimation in silty or clayey soils, requesting to account for clay content when SOC is determined with the loss-on-ignition approach (Jensen et al., 2018).

One further potential source of error that was not considered in this study because of poor record keeping is how dependent sample size was on the accuracy of the result. Within the different methodologies used (Table 1), soil weights between 200 and 500 mg were used. The weight of sample used is important as Shamrikova et al. (2022) summarized: the accuracy of dry combustion analytics increases with increasing SOC concentration of the analyte. Hence, low soil C levels in the present study are more proximal to the minimum detection limit (Gazulla et al., 2012) and therefore may result in higher analytical

variability, which conforms with Horwitz's trumpet that shows an increasing coefficient of variation with decreasing analyte concentration (Workman & Mark, 2006). This is important to consider when small SOC changes (e.g., after a short period of time between two soil sampling events) and variable carbon pools and fractions are investigated (Jandl et al., 2014). Again, the potential error associated with using different sample weights with different methods over time would be minimized by re-analysis using the same method, justifying the financial cost, as it would minimize laboratory bias due to method change. Therefore, these are reasons to encourage the implementation, extension and continuation of well-documented, standardized and easily accessible soil archives. Soil sample archives provide an invaluable tool for understanding long-term soil changes (Bergh et al., 2022), if established properly, for example, being fire-resistant and lockable (Ayres, 2019).

4.2 | Relevance for LTE treatment trends

Regular soil sampling campaigns for SOC are relatively common in LTE and face the challenge to be reliably interpretable if method changes occurred over time without reporting them in detail. This becomes even more critical when soils were only sampled once during the course of the LTE, which was regrettably found to happen quite frequently (Davis et al., 2018). The present study demonstrates the extent that methodologies may impact results over more than 30 years and 11 individual soil sampling campaigns. In addition, Körschens et al. (2013) highlight the need of SOC comparisons over many years to obtain detailed SOC behaviour, particularly as a result of agronomic management changes, which may impact soil organic matter pools differently (Roper et al., 2019).

The difficulties for reliable interpretation of SOC concentrations over time, due to SOC method changes, become obvious in Figure 4. Three contrasting treatments were selected to show how the SOC time series developed over 32 years for the studied LTE. It is apparent that due to methodological changes we may draw unrealistic conclusions from the data, if examined at various points in time. Firstly, we look at the Zero NPK treatment. It must be noted that the SOC of Zero NPK in 1976 was only measured with the historic method as no sample mass was left for re-analysis (Figure 4), so we are not sure of the error the re-analysis would have shown. However, the other two treatments analysed this difference of the first sampling year, which was small. Solely looking at 2008, one might conclude that there are no differences at all between methods. Considering only 2 years for a SOC trend determination, for example, 1994 and 2008 in the

NPK_{high} treatment, would have led to the assumption that no SOC alteration occurred over 15 years. However, considerable differences were noted in 1988 when the historic method indicated a decrease in SOC for all three treatments, but the re-analysed SOC resulted in an increase (fertilized treatments) or maintenance of SOC. Similarly, in 2004, the historic method obtained lower SOC values than in the previous sampling year 1998, whereas re-analysed samples increased (NPK + $Straw_{low}$, Zero NPK) or did not change (NPK_{high}) in their SOC concentration. Figure 4 clearly proves our point to study long periods and to conduct frequent samplings for SOC trends, which is of outstanding importance to monitor SOC changes expected through agronomic management practices. In most sampling years, fertilized treatments resulted in higher SOC when they were analysed with the reference method. In samples with overall lower SOC concentration (Zero NPK), differences between analytical methods were less conclusive and even absent in six sampling years. The variability of SOC between treatment replicates ($n = 8$ in fertilized treatments) appeared lower with high mineral NPK, whereas low organic straw fertilization showed higher variation between samples for both, the historic and the reference method. Generally, method differences were quite considerable in 1998 and 2004 when soil samples were analysed in a different soil laboratory in Braunschweig, which were than analysed again in ZALFs central lab since 2008. Although we do not have current ring proficiency tests for both laboratories, we advise that laboratory shifts should be avoided in LTE studies.

The current study does not provide an agronomic evaluation and statistical data analysis of the LTE and its respective fertilizer treatments (ongoing work by Barkusky et al., unpublished). However, we wanted to highlight that small annual differences in SOC concentrations may not result in practical significance and agronomic relevance for short time windows. The bias due to method changes is important and is highly relevant for SOC stock calculations, especially at the global scale. The methodological differences in sampling, sample preparation, uncertainties in calculating carbon stocks with regard to different tilling depths and bulk densities, and the precision of the laboratory analysis may have a considerable effect on results and trends (Goidts et al., 2009; Körschens et al., 2013) and make global comparisons of land use change almost impossible (Beillouin et al., 2022). Guo and Gifford (2002) conducted a meta-analysis of 537 observations of SOC in 74 studies exploring land use change and found that the diverse (mostly analytical) measurement methodologies limited the authors' conclusions to 'working hypotheses'. This was also criticized by Lee et al. (2009) who claimed that the quantification of soil C stocks can be biased,
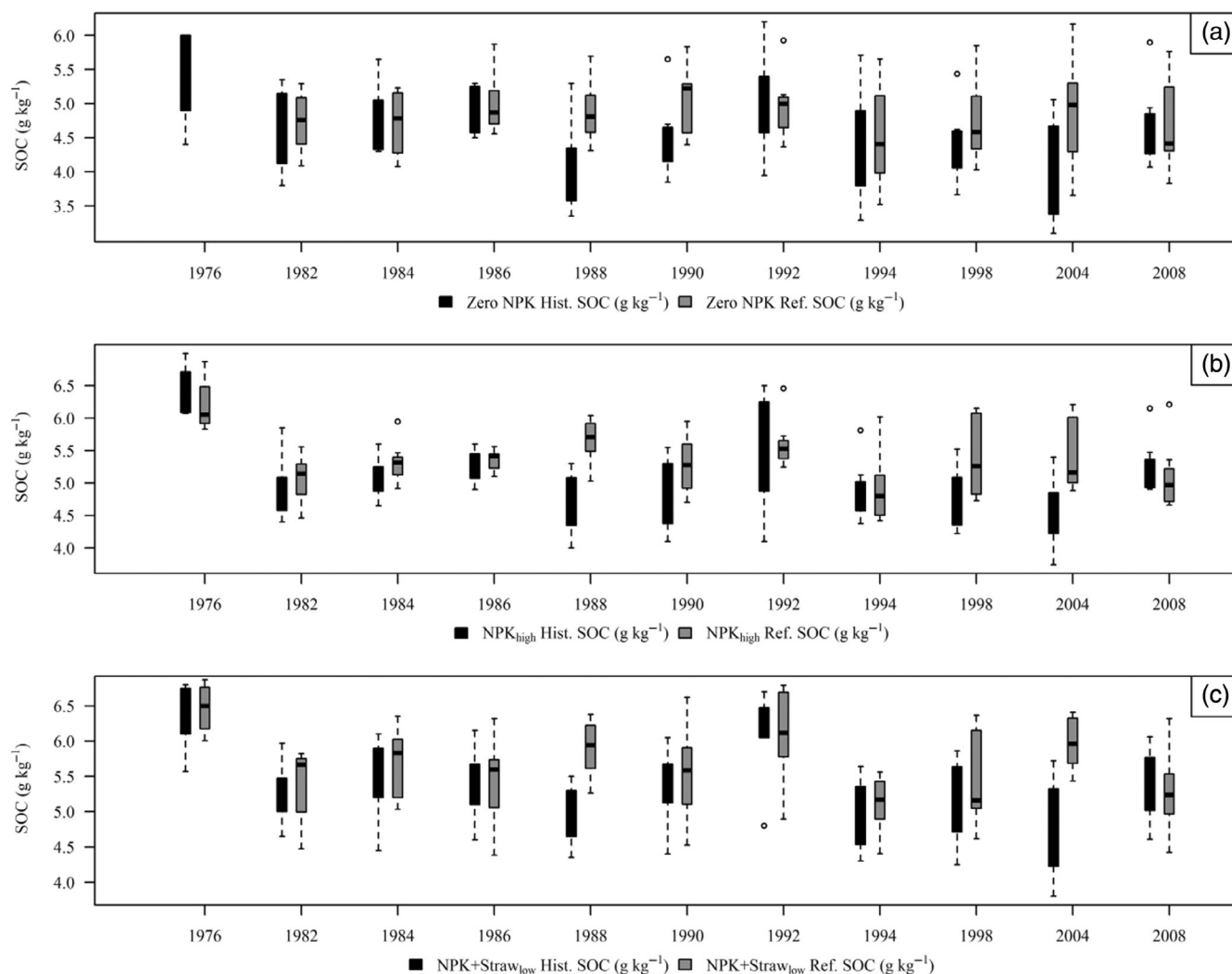
FIGURE 4 Comparison of soil organic carbon (SOC) concentration (g kg^{-1}) measured with different laboratory methods in (a) not fertilized (Zero NPK), (b) high level of mineral fertilized (NPK$_{high}$) and (c) low level of organically fertilized (NPK + Straw$_{low}$) treatments in 0–25 cm soil depth. Black box plots show historic (Hist.) SOC median and range, grey boxplots show re-analysed (Ref.) SOC median and range.

leading to over- or underestimate SOC changes in response to land use or management. However, most of the meta-analyses that summarize the impacts of agricultural management on SOC stocks and sequestration rates do not consider discrepancies in analytical methodologies for the original SOC values (Bai et al., 2018; Beillouin et al., 2022; Sun et al., 2020). Additionally, determination of bulk density on a regular basis is rarely done in LTE and further skews scientific conclusions on long-term SOC stock changes. Although BD measurements involve additional sampling errors and/or method inaccuracies (Kulmatiski & Beard, 2004), regular measurements of BD in this LTE were lacking, which will also increase the uncertainties in SOC stock changes (Walter et al., 2016).

To improve the handling and management of samples and data coming from LTE, several key recommendations are provided, which could, if implemented,

solve analytical method change bias in the future and would allow more robust SOC data interpretation from LTE.

A checklist for LTE site managers and LTE soil data users to improve SOC monitoring:

- Develop (internationally) standardized and meticulous protocols for sampling strategy (design, depth and timing), storage and analytical method
- Regular request of results from soil laboratory proficiency testing ('ring-tests') and avoidance of laboratory change
- Provision of year- and method-specific correction factor after each method change if applicable
- Soil archive with controllable air conditioning and/or stable room temperature and humidity as essential LTE infrastructure

- Commitment for periodic bulk density measurements to calculate carbon stocks

It is likely that many other studies and experiments measuring SOC change over long timescales might not have the same documentation that would be expected from a LTE. The dilemma of analytical method changes goes beyond LTEs and should be addressed in any study during planning, implementation and execution of soil sampling campaigns as a later evaluation of bias will be very limited.

5 | CONCLUSIONS

This study provides a guideline of statistical tools and recommendations to lower the hurdle of changes in analytical soil methods in LTE. We have shown that the Deming regression and Bland–Altman plots can be applied for reliable detection of method differences, but that the Pearson's correlation alone is not a robust statistical tool to decide on the suitability of correction factors to convert from one method to another. This study highlights the need to focus on the accuracy of laboratory analyses for chemical soil properties. Although concentration differences in SOC with different methods may appear small for a single soil sample, the discrepancies may increase at larger spatial and temporal scales (land use trends and global stocks) along with the potential for misinterpretation. Finally, the results of this study reinforce the fact that laboratory error is just one out of many error sources and uncertainties of SOC are most likely accumulative over time.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in BonaRes repository at https://doi.org/10.4228/zalf-acge-b683.

ORCID
Kathrin Grahmann https://orcid.org/0000-0002-9589-7441

REFERENCES
Ayres, E. (2019). Quantitative guidelines for establishing and operating soil archives. *Soil Science Society of America Journal*, *83*, 973–981. https://doi.org/10.2136/sssaj2019.02.0050

Bahar, B., Tuncel, A. F., Holmes, E. W., & Holmes, D. T. (2017). An interactive website for analytical method comparison and bias estimation. *Clinical Biochemistry*, *50*, 1025–1029. https://doi.org/10.1016/j.clinbiochem.2017.08.008

Bai, Z., Caspari, T., Gonzalez, M. R., Batjes, N. H., Mäder, P., Bünemann, E. K., de Goede, R., Brussaard, L., Xu, M., Ferreira, C. S. S., Reintam, E., Fan, H., Mihelič, R., Glavan, M., & Tóth, Z. (2018). Effects of agricultural management practices on soil quality: A review of long-term experiments for Europe and China. *Agriculture, Ecosystems and Environment*, *265*, 1–7. https://doi.org/10.1016/j.agee.2018.05.028

Beillouin, D., Cardinael, R., Berre, D., Boyer, A., Corbeels, M., Fallot, A., Feder, F., & Demenois, J. (2022). A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon. *Global Change Biology*, *28*, 1690–1702. https://doi.org/10.1111/gcb.15998

Bergh, E. L., Calderon, F. J., Clemensen, A. K., Durso, L., Eberly, J. O., Halvorson, J. J., Jin, V. L., Margenot, A. J., Stewart, C. E., Pelt, S. V., & Liebig, M. A. (2022). Time in a bottle: Use of soil archives for understanding long-term soil change. *Soil Science Society of America Journal*, *1–8*, 520–527. https://doi.org/10.1002/saj2.20372

Berti, A., Marta, A. D., Mazzoncini, M., & Tei, F. (2016). An overview on long-term agro-ecosystem experiments: Present situation and future potential. *European Journal of Agronomy*, *77*, 236–241. https://doi.org/10.1016/j.eja.2016.01.004

Billings, S. A., Lajtha, K., Malhotra, A., Berhe, A. A., de Graaff, M. A., Earl, S., et al. (2021). Soil organic carbon is not just for soil scientists: Measurement recommendations for

diverse practitioners. *Ecological Applications*, 31, 1–19. https://doi.org/10.1002/eap.2290

Biney, J. K. M., Saberioon, M., Borůvka, L., Houška, J., Vašát, R., Agyeman, P. C., et al. (2021). Exploring the suitability of UAS-based multispectral images for estimating soil organic carbon: Comparison with proximal soil sensing and spaceborne imagery. *Remote Sensing*, 13, 1–19. https://doi.org/10.3390/rs13020308

Blake, L., Goulding, K. W. T., Mott, C. J. B., & Poulton, P. R. (2000). Temporal changes in chemical properties of air-dried stored soils and their interpretation for long-term experiments. *European Journal of Soil Science*, 51, 345–353. https://doi.org/10.1046/j.1365-2389.2000.00307.x

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135–160. https://doi.org/10.1177/096228029900800204

Brus, D. J. (2021). Statistical approaches for spatial sample survey: Persistent misconceptions and new developments. *European Journal of Soil Science*, 72, 686–703. https://doi.org/10.1111/ejss.12988

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., et al. (2018). Soil quality – A critical review. *Soil Biology and Biochemistry*, 120, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030

Carter, M. R. (2002). Soil quality for sustainable land management: Organic matter and aggregation interactions that maintain soil functions. *Agronomy Journal*, 94, 38–47. https://doi.org/10.2134/agronj2002.3800

Chan, K. Y., Conyers, M. K., Li, G. D., Helyar, K. R. C., Poile, G., & Oates, A. (2011). Soil carbon dynamics under different cropping and pasture management in temperate Australia : Results of three long - term experiments. *Soil Research*, 49, 320–328.

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M. Z., & Ebinger, M. H. (2009). Evaluation of different soil carbon determination methods. *CRC Critical Reviews in Plant Sciences*, 28, 164–178. https://doi.org/10.1080/07352680902776556

Davis, M. R., Alves, B. J. R., Karlen, D. L., Kline, K. L., Galdos, M., & Abulebdeh, D. (2018). Review of soil organic carbon measurement protocols: A US and Brazil comparison and recommendation. *Sustainability*, 10, 4–8. https://doi.org/10.3390/su10010053

De Nobili, M., Contin, M., & Brookes, P. C. (2006). Microbial biomass dynamics in recently air-dried and rewetted soils compared to others stored air-dry for up to 103 years. *Soil Biology and Biochemistry*, 38, 2871–2881. https://doi.org/10.1016/j.soilbio.2006.04.044

De Vos, B., Lettens, S., Muys, B., & Deckers, J. A. (2007). Walkley-Black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. *Soil Use and Management*, 23, 221–229. https://doi.org/10.1111/j.1475-2743.2007.00084.x

Demattê, J. A. M., Dotto, A. C., Bedin, L. G., Sayão, V. M., & Souza, A. B. (2019). Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma*, 337, 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010

Díaz-Zorita, M. (1999). Soil organic carbon rcovery by the walkley-black method in a typic hapludoll. *Communications in Soil Science and Plant Analysis*, 30, 739–745. https://doi.org/10.1080/00103629909370242

Díaz-Zorita, M., Grove, J. H., & Perfect, E. (2007). Sieving duration and sieve loading impacts on dry soil fragment size distributions. *Soil and Tillage Research*, 94, 15–20. https://doi.org/10.1016/j.still.2006.06.006

Ellerbrock, R. H., Höhn, A., & Rogasik, J. (1999). Functional analysis of soil organic matter as affected by long-term manurial treatment. *European Journal of Soil Science*, 50, 65–71. https://doi.org/10.1046/j.1365-2389.1999.00206.x

Ellsäßer, F., Stiegler, C., Röll, A., June, T., Knohl, A., & Hölscher, D. (2021). Predicting evapotranspiration from drone-based thermography-a method comparison in a tropical oil palm plantation. *Biogeosciences*, 18, 861–872. https://doi.org/10.5194/bg-18-861-2021

FAO. (2018). Measuring and modelling soil carbon stocks and stock changes in livestock production systems. http://www.fao.org/3/I9693EN/i9693en.pdf

FAO. (2020). A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes-GSOC-MRV protocol. Rome, Italy. https://doi.org/10.4060/cb0509en

Food and Agriculture Organization of the United Nations. (2019). The Global Soil Laboratory Network. http://www.fao.org/global-soil-partnership/glosolan/en/

Food and Agriculture Organization of the United Nations, IUSS Working Group WRB and GSP. (2014). World Reference Base for soil resources 2014: International soil classification system for naming soils and creating legends for soil maps. Rome, Italy. https://doi.org/10.1038/nnano.2009.216

Gazulla, M. F., Rodrigo, M., Orduña, M., & Gómez, C. M. (2012). Determination of carbon, hydrogen, nitrogen and sulfur in geological materials using elemental analysers. *Geostandards and Geoanalytical Research*, 36, 201–217. https://doi.org/10.1111/j.1751-908X.2011.00140.x

Ge, Z., An, T., Bol, R., Li, S., Zhu, P., Peng, C., Xu, Y., Cheng, N., Li, T., Wu, Y., Xie, N., & Wang, J. (2021). Distributions of straw-derived carbon in Mollisol's aggregates under different fertilization practices. *Scientific Reports*, 11, 1–9. https://doi.org/10.1038/s41598-021-97546-3

Glosolan, Global Soil Laboratory Network. (2019). Standard operating procedure for soil organic carbon: Walkley-Black method.

Goidts, E., Van Wesemael, B., & Crucifix, M. (2009). Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. *European Journal of Soil Science*, 60, 723–739. https://doi.org/10.1111/j.1365-2389.2009.01157.x

Grahmann, K., Rubio Dellepiane, V., Terra, J. A., & Quincke, J. A. (2020). Long-term observations in contrasting crop-pasture rotations over half a century: Statistical analysis of chemical soil properties and implications for soil sampling frequency. *Agriculture, Ecosystems and Environment*, 287, 106710. https://doi.org/10.1016/j.agee.2019.106710

Grahmann, K., Terra, J. A., Ellerbrock, R., Rubio, V., Barro, R., Caamaño, A., & Quincke, A. (2022). Data accuracy and method validation of chemical soil properties in long-term experiments: Standard operating procedures for a non-certified soil laboratory in Latin America. *Geoderma Regional*, 28, e00487. https://doi.org/10.1016/j.geodrs.2022.e00487

Grosse, M., Heinrich, U., & Hierold, W. (2019). Fact sheet for the description of long-term field experiments/Steckbrief zur

Beschreibung von Dauerfeldversuchen. BonaRes Series. https://doi.org/10.20387/BonaRes-R56G-FGRW

Guo, L. B., & Gifford, R. M. R. (2002). Soil carbon stocks and land use change: A meta analysis. *Global Change Biology*, *8*, 345–360. https://doi.org/10.1046/j.1354-1013.2002.00486.x

Jandl, R., Rodeghiero, M., Martinez, C., Cotrufo, M. F., Bampa, F., van Wesemael, B., Harrison, R. B., Guerrini, I. A., Richter, D. B., Jr., Rustad, L., Lorenz, K., Chabbi, A., & Miglietta, F. (2014). Current status, uncertainty and future needs in soil organic carbon monitoring. *Science of the Total Environment*, *468–469*, 376–383. https://doi.org/10.1016/j.scitotenv.2013.08.026

Jensen, J. L., Christensen, B. T., Schjønning, P., Watts, C. W., & Munkholm, L. J. (2018). Converting loss-on-ignition to organic carbon content in arable topsoil: Pitfalls and proposed procedure. *European Journal of Soil Science*, *69*, 604–612. https://doi.org/10.1111/ejss.12558

Johnston, A. E., & Poulton, P. R. (2018). The importance of long-term experiments in agriculture: Their management to ensure continued crop production and soil fertility; the Rothamsted experience. *European Journal of Soil Science*, *69*, 113–125. https://doi.org/10.1111/ejss.12521

Körschens, M., Albert, E., Armbruster, M., Barkusky, D., Baumecker, M., Behle-Schalk, L., Bischoff, R., Čergan, Z., Ellmer, F., Herbst, F., Hoffmann, S., Hofmann, B., Kismanyoky, T., Kubat, J., Kunzova, E., Lopez-Fando, C., Merbach, I., Merbach, W., ... Zorn, W. (2013). Effect of mineral and organic fertilization on crop yield, nitrogen uptake, carbon and nitrogen balances, as well as soil organic carbon content and dynamics: Results from 20 European long-term field experiments of the twenty-first century. *Archives of Agronomy and Soil Science*, *59*, 1017–1040. https://doi.org/10.1080/03650340.2012.704548

Kowalenko, C. G. (2001). Assessment of Leco CNS-2000 analyzer for simultaneously measuring total carbon, nitrogen, and sulphur in soil. *Communications in Soil Science and Plant Analysis*, *32*, 2065–2078. https://doi.org/10.1081/CSS-120000269

Kühnel, A., Wiesmeier, M., Spörlein, P., Schilling, B., & Kögel-Knabner, I. (2019). Influence of drying vs. freezing of archived soil samples on soil organic matter fractions. *Journal of Plant Nutrition and Soil Science*, *182*, 772–781. https://doi.org/10.1002/jpln.201800529

Kulmatiski, A., & Beard, K. H. (2004). Reducing sampler error in soil research. *Soil Biology and Biochemistry*, *36*, 383–385. https://doi.org/10.1016/j.soilbio.2003.10.004

Lee, J., Hopmans, J. W., Rolston, D. E., Baer, S. G., & Six, J. (2009). Determining soil carbon stock changes: Simple bulk density corrections fail. *Agriculture, Ecosystems and Environment*, *134*, 251–256. https://doi.org/10.1016/j.agee.2009.07.006

Lettens, S., De Vos, B., Quataert, P., Van Wesemael, B., Muys, B., & Van Orshoven, J. (2007). Variable carbon recovery of Walkley-Black analysis and implications for national soil organic carbon accounting. *European Journal of Soil Science*, *58*, 1244–1253. https://doi.org/10.1111/j.1365-2389.2007.00916.x

Leue, M., Hoffmann, C., Hierold, W., & Sommer, M. (2019). In-situ multi-sensor characterization of soil cores along an erosion-deposition gradient. *Catena*, *182*, 104140. https://doi.org/10.1016/j.catena.2019.104140

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*, 255–268.

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry*, *39*, 424–432. https://doi.org/10.1093/clinchem/39.3.424

Linnet, K. (1999). Necessary sample size for method comparison studies based on regression analysis. *Clinical Chemistry*, *894*, 882–894.

Ludbrook, J. (2010). Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clinical and Experimental Pharmacology & Physiology*, *37*, 692–699. https://doi.org/10.1111/j.1440-1681.2010.05376.x

Magari, R. T. (2004). Bias estimation in method comparison studies. *Journal of Biopharmaceutical Statistics*, *14*, 881–892. https://doi.org/10.1081/BIP-200035450

Martin, R. F. (2000). General deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clinical Chemistry*, *46*, 100–104. https://doi.org/10.1093/clinchem/46.1.100

McBride, G. (2005). A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. http://www.sciencedirect.com/science/article/pii/S0140673686908378%5Cn%3CGotoISI%3E://A1989U124500022

Meersmans, J., Van Wesemael, B., & Van Molle, M. (2009). Determining soil organic carbon for agricultural soils: A comparison between the Walkley & Black and the dry combustion methods (North Belgium). *Soil Use and Management*, *25*, 346–353. https://doi.org/10.1111/j.1475-2743.2009.00242.x

Menditto, A., Patriarca, M., & Magnusson, B. (2007). Understanding the meaning of accuracy, trueness and precision. *Accreditation and Quality Assurance*, *12*, 45–47. https://doi.org/10.1007/s00769-006-0191-z

Mikhailova, E. A., Noble, R. R. P., & Post, C. J. (2003). Comparison of soil organic carbon recovery by Walkley-Black and dry combustion methods in the Russian Chernozem. *Communications in Soil Science and Plant Analysis*, *34*, 1853–1860. https://doi.org/10.1081/CSS-120023220

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z. S., Cheng, K., Das B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., ... Winowiecki, L. (2017). Soil carbon 4 per mille. *Geoderma*, *292*, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

O'Rourke, S. M., Angers, D. A., Holden, N. M., & McBratney, A. B. (2015). Soil organic carbon across scales. *Global Change Biology*, *21*, 3561–3574. https://doi.org/10.1111/gcb.12959

Olson, K. R., Al-Kaisi, M. M., Lal, R., & Lowery, B. (2014). Experimental consideration, treatments, and methods in determining soil organic carbon sequestration rates. *Soil Science Society of America Journal*, *78*, 348–360. https://doi.org/10.2136/sssaj2013.09.0412

Paustian, K., Elliott, E. T., Collins, H. P., Cole, C. V., & Paul, E. A. (1995). Use of a network of long-term experiments for analysis of soil carbon dynamics and global change: The north american model. *Australian Journal of Experimental Agriculture*, *35*, 929–939. https://doi.org/10.1071/EA9950929

Petrokofsky, G., Kanamaru, H., Achard, F., Goetz, S. J., Joosten, H., Holmgren, P., Lehtonen, A., Menton, M. C. S., Pullin, A. S., & Wattenbach, M. (2012). Comparison of methods for measuring

and assessing carbon stocks and carbon stock changes in terrestrial carbon pools. How do the accuracy and precision of current methods compare? A systematic review protocol. *Environmental Evidence*, 1, 1–21. https://doi.org/10.1186/2047-2382-1-6

Poeplau, C., Bolinder, M. A., & Kätterer, T. (2016). Towards an unbiased method for quantifying treatment effects on soil carbon in long-term experiments considering initial within-field variation. *Geoderma*, 267, 41–47. https://doi.org/10.1016/j.geoderma.2015.12.026

Post, J., Hattermann, F. F., Krysanova, V., & Suckow, F. (2008). Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling & Software*, 23, 125–138. https://doi.org/10.1016/j.envsoft.2007.05.010

Rasmussen, P. E., Goulding, K. W. T., Brown, J. R., Grace, P. R., Janzen, H. H., & Korschens, M. (1998). Long-term agroecosystem experiments: Assessing agricultural sustainability and global change. *Science (80-)*, 282, 893–896. https://doi.org/10.1126/science.282.5390.893

Rayment, G. E., Miller, R. O., & Sulaeman, E. (2000). Proficiency testing and other interactive measures to enhance analytical quality in soil and plant laboratories. *Communications in Soil Science and Plant Analysis*, 31, 1513–1530. https://doi.org/10.1080/00103620009370523

Richter, D. d B. B., Hofmockel, M., Callaham, M. A., Powlson, D. S., & Smith, P. (2007). Long-term soil experiments: Keys to managing Earth's rapidly chancing ecosystems. *Soil Science Society of America Journal*, 71, 266–279. https://doi.org/10.2136/sssaj2006.0181

Roper, W. R., Robarge, W. P., Osmond, D. L., & Heitman, J. L. (2019). Comparing four methods of measuring soil organic matter in North Carolina soils. *Soil Science Society of America Journal*, 83, 466–474. https://doi.org/10.2136/sssaj2018.03.0105

Salter, P. J., & Williams, J. B. (1965). The influence of texture on the moisture characteristics of soils: II. Available-water capacity and moisture release characteristics. *Journal of Soil Science*, 16, 310–317. https://doi.org/10.1111/j.1365-2389.1965.tb01442.x

Shamrikova, E. V., Kondratenok, B. M., Tumanova, E. A., Vanchikova, E. V., Lapteva, E. M., Zonova, T. V., Lu-Lyan-Min, E. I., Davydova, A. P., Libohova, Z., & Suvannang, N. (2022). Transferability between soil organic matter measurement methods for database harmonization. *Geoderma*, 412, 115547. https://doi.org/10.1016/j.geoderma.2021.115547

Sibbesen, E. (1986). Soil movement in long-term field experiments. *Plant and Soil*, 91, 73–85. https://doi.org/10.1007/BF02181820

Sibbesen, E., Skjøth, F., & Rubæk, G. H. (2000). Tillage caused dispersion of phosphorus and soil in four 16-year old field experiments. *Soil and Tillage Research*, 54, 91–100. https://doi.org/10.1016/S0167-1987(00)00088-X

Smith, P. (2004). How long before a change in soil organic carbon can be detected? *Global Change Biology*, 10, 1878–1883. https://doi.org/10.1111/j.1365-2486.2004.00854.x

Smith, P., Soussana, J. F., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., Batjes, N. H., Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J. E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., & Klumpp, K. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology*, 26, 219–241. https://doi.org/10.1111/gcb.14815

Stella, T., Mouratiadou, I., Gaiser, T., Berg-Mohnicke, M., Wallor, E., Ewert, F., & Nendel, C. (2019). Estimating the contribution of crop residues to soil organic carbon conservation. *Environmental Research Letters*, 14, 094008. https://doi.org/10.1088/1748-9326/ab395c

Sun, W., Canadell, J. G., Yu, L., Yu, L., Zhang, W., Smith, P., Fischer, T., & Huang, Y. (2020). Climate drives global soil carbon sequestration and crop yield changes under conservation agriculture. *Global Change Biology*, 26, 3325–3335. https://doi.org/10.1111/gcb.15001

Tecon, R., & Or, D. (2017). Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiology Reviews*, 41, 599–623. https://doi.org/10.1093/femsre/fux039

Thai, T. H., Bellingrath-Kimura, S. D., Hoffmann, C., & Barkusky, D. (2020). Effect of long-term fertiliser regimes and weather on spring barley yields in sandy soil in North-East Germany. *Archives of Agronomy and Soil Science*, 66, 1812–1826. https://doi.org/10.1080/03650340.2019.1697436

Thai, T. H., Omari, R. A., Barkusky, D., & Bellingrath-Kimura, S. D. (2020). Statistical analysis versus the m5p machine learning algorithm to analyze the yield of winter wheat in a long-term fertilizer experiment. *Agronomy*, 10, 1–19. https://doi.org/10.3390/agronomy10111779

Tivet, F., Carlos de Moraes Sá, J., Borszowskei, P. R., Letourmy, P., Briedis, C., Ferreira, A. O., & Burkner dos Santos Thiago Massao Inagaki, J. (2012). Soil carbon inventory by wet oxidation and dry combustion methods: Effects of land use, soil texture gradients, and sampling depth on the linear model of C-equivalent correction factor. *Soil Science Society of America Journal*, 76, 1048–1059. https://doi.org/10.2136/sssaj2011.0328

van Leeuwen, C. C. E., Mulder, V. L., Batjes, N. H., & Heuvelink, G. B. M. (2021). Statistical modelling of measurement error in wet chemistry soil data. *European Journal of Soil Science*, 73, 1–17. https://doi.org/10.1111/ejss.13137

Visser, R. G. (2006). Interpretation of interlaboratory comparison results to evaluate laboratory proficiency. *Accreditation and Quality Assurance*, 10, 521–526. https://doi.org/10.1007/s00769-005-0051-2

Walter, K., Don, A., Tiemeyer, B., & Freibauer, A. (2016). Determining soil bulk density for carbon stock calculations: A systematic method comparison. *Soil Science Society of America Journal*, 80, 579–591. https://doi.org/10.2136/sssaj2015.11.0407

West, A., Sparling, G., Feltham, C., & Reynolds, J. (1992). Microbial activity and survival in soils dried at different rates. *Soil Research*, 30, 209. https://doi.org/10.1071/SR9920209

Włodarczyk, T., Szarlip, P., Kozieł, W., Nosalewicz, M., Brzeziñska, M., Pazur, M., & Urbanek, E. (2014). Effect of long storage and soil type on the actual denitrification and denitrification capacity to N_2O formation. *International Agrophysics*, 28, 371–381. https://doi.org/10.2478/intag-2014-0027

Workman, J., & Mark, H. (2006). Limitations in analytical accuracy, part I: Horwitz's trumpet. Two technical papers recognized as significant early contributions in the discussion of spectroscopy 21, 19–24.

Xu, S., Silveira, M. L., Ngatia, L. W., Normand, A. E., Sollenberger, L. E., & Ramesh Reddy, K. (2017). Carbon and nitrogen pools in aggregate size fractions as affected by sieving method and land use intensification. *Geoderma*, 305, 70–79. https://doi.org/10.1016/j.geoderma.2017.05.044

Yan, Z., Zhou, J., Yang, L., Gunina, A., Yang, Y., Peixoto, L., Zeng, Z., Zang, H., & Kuzyakov, Y. (2022). Diversified cropping systems benefit soil carbon and nitrogen stocks by increasing aggregate stability: Results of three fractionation methods. *Science of the Total Environment*, *824*, 153878. https://doi.org/10.1016/j.scitotenv.2022.153878

Zhu, Y., Chen, Y., Ali, M. A., Dong, L., Wang, X., Archontoulis, S. V., Schnable, J. C., & Castellano, M. J. (2021). Continuous in situ soil nitrate sensors: The importance of high-resolution measurements across time and a comparison with salt extraction-based methods. *Soil Science Society of America Journal*, *1–14*, 677–690. https://doi.org/10.1002/saj2.20226

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.