Geoderma Regional 28 (2022) e00487

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs

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

Kathrin Grahmann[a,*], José A. Terra[d], Ruth Ellerbrock[a], Valentina Rubio[b], Raquel Barro[c], Alejandro Caamaño[e], Andrés Quincke[b]

[a] Leibniz Centre for Agricultural Landscape Research (ZALF); Eberswalder Str.84, 15374 Müncheberg, Germany
[b] National Agricultural Research Institute, La Estanzuela, Ruta 50 km 11, Colonia, CP 70000, Uruguay
[c] Postdoctoral Scholarship ANII (National agency of innovation and research, Uruguay)
[d] National Agricultural Research Institute, Ruta 8 km 281, Treinta y Tres, CP 33000, Uruguay
[e] Service Unit for the Evaluation of Quality and Environmental Control, Department of Environmental Development, Intendancy of Montevideo, Uruguay

ARTICLE INFO

Keywords:
Internal reference material
Data precision
Trueness
Repeatability
Reproducibility
Soil organic carbon
Uruguay

ABSTRACT

Long-term agricultural experiments (LTE) are essential to detect cumulative treatment effects on soil properties and to design sustainable production systems. However, to ensure high quality of long-term soil data and their correct interpretation, several analytical issues regarding the accuracy and analytical laboratory bias need to be considered. This paper aims to (1) evaluate laboratory precision and trueness of analytical soil data for the evaluation of long-term trends in LTE and to (2) assess the limitations and challenges for non-certified soil laboratories that might compromise the quality of analytical soil data. A data set of internal reference soil materials (IRM) collected over 16 years and interlaboratory data from eleven years were analyzed to verify method precision, trueness, and the subsequent long-term dataset reliability for several soil quality parameters: organic carbon (SOC, determined either by wet or dry combustion), pH (water), extractable phosphorous (either Bray I or citric acid; Bray-P or citric acid-P), and exchangeable potassium (K_{exch}). Results showed that IRM used by the laboratory were homogenous in terms of physical and chemical composition and appropriate to confirm the precision of long-term soil survey data. The relative standard deviation for repeatability and reproducibility (RSD_R) ranged from 1.5% for SOC by wet combustion to 9.5% for citric acid-P. HorRat values (the ratio of the estimated standard deviations of reproducibility and the repeatability found for individual analytical procedures) for all chemical soil properties were within the acceptable ranges of <2.0. Interlaboratory trials for soil pH and SOC showed tolerable standard *Z*-Scores under 2.0 (Z-Score, calculated from the difference between laboratory results and the assigned value divided by the standard deviation), verifying the trueness of data. The results of this study reinforce the validity of analytical soil data originating from the non-certified laboratory in Uruguay obtaining both precise and true soil quality data over a long period of time for most soil analytes. Nevertheless, the analytical flaws in LTE soil monitoring were revealed and can only be minimized in ongoing and future studies through the inclusion of certified reference material. These recommendations should guide future research activities in LTE studies on analytical data quality management as a requirement for long-term soil monitoring. Finally, the paper proposes a proficiency testing procedure for soil laboratories to achieve and maintain high analytical quality for LTE soil research.

1. Introduction

Long-term agricultural experiments (LTE) are needed to evaluate the impacts of production systems and management practices on soil quality and contribute to assess their sustainability (Berti et al., 2016). Routine measurements of soil properties are useful to develop and calibrate models describing soil dynamics in the long run (Johnston and Poulton, 2018). The definition of "long-term" is still under debate in soil science,

---

* Corresponding author.
E-mail address: kathrin.grahmann@zalf.de (K. Grahmann).

https://doi.org/10.1016/j.geodrs.2022.e00487
Received 9 May 2021; Received in revised form 20 December 2021; Accepted 2 February 2022

but for tillage experiments, for example, a minimum of 15 years is recognized to provide consistent treatment effects (Kay and VandenBygaart, 2002). There are several challenges to manage LTE and their accumulation of large data sets correctly. Ensuring long-term funding, high operational costs, avoiding adverse effects of soil oversampling, or data access and management are some of the issues that must be faced (Berti et al., 2016). The main challenge in LTE research is that treatments remain relatively constant, but many other environmental and cropping technology factors change around them, jeopardizing the *ceteris paribus* assumption (Dyke, 1974). The same is true for operational aspects like laboratory analysis of soil samples which is a critical point that needs to be considered for their interpretation (Schut and Giller, 2020). Other sources of error, including sampling error or spatial soil heterogeneities at the plot scale (Poeplau et al., 2016), are not the focus of this paper.

When evaluating long-term soil datasets, researchers are confronted with several questions on interpreting and managing data uncertainties generated by analytics over time. For example, changes in laboratory methods for a specific soil property (Blake et al., 2000; Johnston and Poulton, 2018), higher precision for the laboratory analysis through improved purity of reagents or new technology, or different operators for particular equipment may affect the data obtained over a long period (Boone et al., 1999; Rayment et al., 2000). Thus, there is a crucial need to safely assert that possible effects observed in various decades of experimentation are attributed to the impact of treatments and not to laboratory bias.

Many studies assessed soil quality changes with a wide range of objectives and treatments. However, there is a lack of information about the reliability of soil data quality coming from these experiments that often run for more than 20 years. A large group of experiments includes point measurements of specific soil quality parameters for particular years (for example, SOC changes measured ten and 20 years after baseline; Verhulst et al., 2011; Congreves et al., 2015; Martínez et al., 2016; De Oliveira et al., 2018). A smaller group of studies conducts a regular soil sample collection over many years (Chan et al., 2011; Pravia et al., 2019). However, information on changes in the applied soil laboratory methods, laboratory fluctuation during the study period, or disclosure about the laboratories′quality control by presenting precision and trueness values of the respective method are limited and subsequently impede comparisons, linkages and meta-analysis of LTE soil studies around the globe. Differences between technical laboratory procedures must be considered when comparing, analyzing and evaluating SOC concentrations determined for recent soil sample sets with those of old databases (Grahmann et al., 2020; Skjemstad et al., 2000).

Analytical method validation is conducted to achieve quality assurance and quality control of chemical analysis with the ultimate goal of eliminating analytical bias (González et al., 2010). Bias is a systematic error defined as the difference between the measurement result and its unknown ‘true value’ (Hibbert, 2007). Data accuracy covers the concepts of data precision and data trueness (Menditto et al., 2007). Different tests can be conducted to quantify the random error by evaluating data precision which is stratified into repeatability and reproducibility. Repeatability represents within-laboratory changes over a short period of time, (within-laboratory) reproducibility covers the variation of one influencing factor (e.g. operator; Menditto et al., 2007). Trueness can be quantified as bias, the systematic error. Interlaboratory trials help to quantify data trueness, a performance parameter that is necessary to ensure that the generated data is comparable among time and from one laboratory to another (Quevauviller, 1995).

To distinguish between possible treatment effects and laboratory bias, a random and systematic method validation should be conducted in LTE soil surveys over the short and long-run. Necessary information for the validation can be provided from data of stable and non-certified reference materials analyzed during long-term trend studies together with field samples. A reference material is defined as “a material of whose property values are sufficiently homogeneous and well established to be used for the calibration of a device, the assessment of a measurement method, or for assessing values to materials” and are available as certified or internal reference material (Quevauviller, 1995). Reference materials are used for calibration and quality control and can detect the variation of a laboratory method between years or operators (Jochum and Brueckner, 2008). Commercial soil laboratories conduct regular method validation tests by interlaboratory trials or the use of certified reference materials (Van Zoonen et al., 1999). For quality control purposes and reliable analytical data, soil reference materials need to be stable, homogenous, and well-characterized. Therefore, it is usually necessary to test them over a more extended time period (minimum of 12 months) to verify the sample’s integrity (Jochum and Brueckner, 2008; Kennedy et al., 1994). However, due to financial constraints and limited availability, research institutes in developing countries often use internal reference soil materials (IRM; Kennedy et al., 1994; IAEA, 2003). This was reported to cause inequality in soil science and local research capacities, in this case particularly facing the publication impediment, that laboratory facilities do not fulfill international standards using certified reference materials (Berhe and Ghezzehei, 2020; Minasny et al., 2020).

The objective of this study was to evaluate the accuracy of soil data analysis from annual surveys conducted in the oldest agricultural LTE in Latin America using IRM in a non-certified soil laboratory to exclude laboratory bias. The source of error for different soil laboratory methods (SOC, Bray-P, citric acid-P, K_{exch}, and pH) was analyzed for up to six continuous years. Selected laboratory quality and risk analyses of potential error sources that may occur in long-term soil surveys were performed using the information of the laboratory’s IRM and interlaboratory trials. We suggest approaches for IRM handling and laboratory analytics in LTE studies, including the necessary statistical tests for data quality assurance. We hypothesized that (1) the soil laboratory provides accurate and adequate analytical data that did not obliterate or interfere with differences caused by the LTE treatments such that (2) long-term analytical data of the analyzed IRM can be used to control the quality of the LTE soil data.

2. Materials and methods

2.1. Experimental site

The study was carried out at INIA’s (National Research Institute of Agriculture) institutional soil laboratory located in the experimental station “La Estanzuela” in Uruguay (34°20′S, 57°41′W, 80 m a.s.l.). The site has a warm temperate climate with an average annual rainfall of 1100 mm and evapotranspiration of 1200 mm; monthly mean temperatures ranged from 10 °C in July to 23 °C in January. The typical soils are Mollisols (smectitic Vertic Argiudoll according to the USDA Soil Taxonomy; Soil Survey Staff, 2014). Internal reference soil materials were used by INIA’s soil laboratory during the study period and originated from the oldest cropping-system experiment in Latin America. This LTE was established in 1963 and consists of seven different rotations contrasting continuous cropping (CC) with crop-pastures rotations (CPR) of different proportions, lengths and species composition (supplementary Table A.1). Each experimental unit (plot) was 25 m wide and 200 m long. Detailed results on treatment differences, their impacts on several soil properties, and the importance of soil sampling frequency were presented in Grahmann et al. (2020). We emphasize that the present study will not discuss the changes in soil properties generated over time, but presents information on the quality of the analyzed data set.

2.2. Internal reference soil materials (IRM)

Four IRM bulk samples were taken between 2000 and 2016, from the northern end in two plots of the described LTE (supplementary Table A.2). In each IRM sampling, a large composite topsoil sample (50–100 kg) was taken in 0–15 cm sampling depth. Bulk samples were divided into smaller subsamples and then dried at 40 °C for 48 h, followed by milling and sieving to 2 mm and stored in 500 mL lockable plastic containers. Dry soil was used as IRM for various, routine soil laboratory analysis year-round at INIA's soil laboratory.

A dataset was compiled with measured analytical results of the laboratory methods SOC, Bray-P, citric acid-P, pH, and K_{exch}, for IRM obtained from daily routine analyses of the INIA soil laboratory since 2001. Supplementary Table A.2 depicts the utilization period and the available number of analytical repetitions for each tested IRM and its respective laboratory method. The data were used to assess the accuracy of the INIA laboratory having additional information about the year of analysis and the performing operator. Hence, an "unintentional" but worthwhile quality control of the LTE's annual soil survey was possible as IRM originated from the LTE. Note that the INIA laboratory (i) was established in 1963 associated and motivated by the implementation of the LTE, (ii) conducts yearly controlled interlaboratory soil and plant tests since 2002, and (iii) is only operating under ISO/IEC 17025 accreditation for SOC in soil and plant material (ISO 10694) since May 2018. For the regular interlaboratory trials, three reference soil samples were provided five times per year by the Brazilian Agronomy Institute of Campinas to more than 100 participating laboratories to evaluate data trueness (da Silva Dias et al., 2015). In this study, interlaboratory data of INIAs soil laboratory were available for the methods SOC, K_{exch}, and pH yearly between 2006 and 2019.

2.3. Soil analytical methods

SOC was determined using $K_2Cr_2O_7$ and heat as described by Tinsley (1967) between 1964 and 2011. Since 2011, SOC was measured by dry combustion at 900 °C followed by infrared detection with a LECO equipment (LECO TrueSpec CN-2000, St. Joseph, USA). Soil pH was determined potentiometrically with distilled water (1:2.5 soil/solution ratio; Beretta-Blanco et al., 2014). Phosphorus was measured colorimetrically after using two different extraction methods; first by the Bray-I method with a 1:7 soil/solution ratio which was applied until 2015 and increased to a 1:10 soil/solution ratio since 2016 and an extraction time of 5 min (Bray and Kurtz, 1945), and second with 0.5% citric acid (1:10 soil/solution ratio and 30 min extraction time; Hernández et al., 2013, Palermo et al., 1985). Exchangeable potassium (K_{exch}) was determined by 1 M ammonium acetate (NH_4OAc) extraction at pH 7 and subsequent atomic emission readings (Jackson, 1964) using an ICP-OES (Inductively coupled plasma mass spectrometry) when all bases (Mg, K, Ca, Na) were determined or by flame emission spectrophotometry using a PerkinElmer AAnalyst 700 system when solely K_{exch} was determined (Núñez and Morón, 2017).

2.4. Method comparison for SOC

After the installation of the LECO equipment in 2007, INIA conducted an internal validation and comparison campaign in 2007 and 2008 by analyzing 1927 soil samples using both methods: Tinsley and LECO. This method comparison resulted in a correction factor of 0.81 using the regression coefficient without intercept for recalculating LECO into Tinsley SOC values. In 2011, the Tinsley method was officially replaced, and the correction factor of 0.81 to convert from LECO to Tinsley values was not reported anymore (Beretta-Blanco et al., 2019). To verify this internal validation procedure, a methodological comparison between Tinsley and LECO method was conducted in the present study. A total of 45 archived LTE soil samples with three different SOC concentration ranges were reanalyzed in 2019 (supplementary Table A.5) by dry combustion (LECO) and compared with the original SOC values obtained with wet combustion (Tinsley). IRM2 was introduced in 2018 for the evaluation of dry combustion method, which explains the lower number of repetitive analysis for this reference material (supplementary Table A.2). Between 2011 and 2017, no IRM was used for SOC determination with LECO. The LECO equipment was always operated by the same operator.

2.5. Annual soil survey procedure in the LTE

Annual LTE soil survey data were used to crosscheck laboratory precision and trueness of analytical soil data obtained from IRM for the evaluation of long-term trends in LTE. Soil core samples for LTE soil surveys were collected annually in autumn since 1964 to a depth of 15 cm at 25–30 random pits and composited in each experimental unit. Soil samples were dried for 48 h at 40 °C, milled to 2 mm, and analyzed for different soil properties at INIA's soil laboratory. For the present study, soil survey data of the LTE from the particular years of the IRM sampling were used to proof soil data precision of the LTE soil data.

2.6. Statistical analysis

Statistical analyses were conducted with Excel and the R software package (version 3.6.0 of 2019, www.cran.r-project.org). Standard procedures of laboratory testing for method accuracy of IRM data were applied, using first the Grubbs Test ($\alpha = 0.05$) for outlier detection, and the Shapiro-Test and histograms for normality tests. Parameters with non-normally distributed residuals were tested with Kruskal-Wallis for significant differences ($P < 0.05$) between years and between operators. Parameters with normal distribution were tested with a one-way ANOVA (da Silva Dias et al., 2015; González et al., 2010). Differences between laboratory methods of SOC and phosphorus were analyzed with a paired *T*-Test of equal variance.

For each soil laboratory method, the IRM with the highest number of replicates (analyses) was selected to test for variation between years and operators to obtain high quality error estimates (van Leeuwen et al., 2021, supplementary Table A.2). For the comparison between operators, the number of replications was reduced. If an operator performed less than ten analytical measurements per calendar year for a specific soil property, thus observations were not considered in this analysis. Different performance parameters were applied to validate laboratory methods. The relative standard deviation (RSD, Eq. 1) was calculated to detect imprecision. William Horwitz studied the relationship between variability of chemical measurements and the concentration of the analyte and established the basis for current analytical procedures (Albert and Horwitz, 1997; Workman and Mark, 2006). From this, the Horwitz trumpet was developed (Horwitz, 1982) in order to quantify the relationship of the laboratory coefficient of variation (CV, in %) and the analyte concentration which was applied to calculate the HorRat values, which are used here to evaluate method precision (RSD_H, Eq. 2). The HorRat value combines the actual RSD of the reference material with the Horwitz equation that includes the mean concentration of the analyte (Eq. 3) and gives a rough classification of data precision and acceptability for a chemical method (Zhang et al., 2009; supplementary Table A.3).

$$RSD_{R=}Relative\ standard\ deviation\ of\ the\ reference\ analyte = \frac{Standard\ deviation}{Reference\ analyte\ mean}*100 \tag{1}$$

$$RSD_H = Horwitz\ equation = 2^{(1-0.5logC)} \tag{2}$$

(C = mean concentration of the reference analyte)

$$HorRat = Horwitz\ ratio = \frac{RSD_R}{RSD_H} \tag{3}$$

For data trueness, the *Z*-Score was calculated (Eq. 4) from interlaboratory results and classified subsequently (supplementary Table A.4). The interpretation of the Z-Score is based on the assumption that laboratory results follow a normal distribution, implying that laboratory results can be found within two standard deviations (SD) with a probability of 95% and within three SD with a probability of 99.7%

Table 1
Relative standard deviation (RSD%), average concentration (Av), and number of observations (n) for six laboratory methods applied on selected internal reference soil materials (IRM) between 2001 and 2019 at INIA's soil laboratory, La Estanzuela, Uruguay.

	SOC-Tinsley				SOC-LECO				Bray-P				Citric acid-P				pH				K_{exch}				Bray-P			
	IRM1				IRM2				IRM3				IRM4				IRM4				IRM4				IRM4			
Year	RSD%	Av (g kg^{-1})		n	RSD%	Av (g kg^{-1})		n	RSD%	Av (ppm)		n	RSD%	Av (ppm)		n	RSD%	Av		n	RSD%	Av (cmol kg^{-1})		n	RSD%	Av (ppm)		n
2001	1.7	14.6	a	14																								
2002	1.3	14.9	b	40																								
2003	1.4	15.0	bc	40																								
2004	1.6	14.9	b	40																								
2005	1.2	15.1	c	15																								
2009									3.1	17.7	ab	12																
2010									6.4	17.4	a	123																
2011									6.1	18.0	b	145																
2012									4.5	18.8	c	109																
2013									4.1	19.4	d	58	8.6	28.0	a	73	2.9	5.9	a	37	3.4	0.78	ab	18	4.0	36.8	a	49
2014													4.8	28.2	a	22	2.8	5.9	a	46	3.4	0.77	ab	37	3.3	37.0	a	53
2015													6.2	28.2	a	27	1.1	5.9	a	35	4.5	0.76	a	44	3.4	37.2	a	55
2016													5.1	32.5	b	22	1.8	5.8	a	55	4.9	0.77	ab	46	7.9	44.0	b	51
2017													7.2	27.6	a	19	1.6	5.8	a	28	5.6	0.79	b	28	7.2	44.3	b	47
2018					4.8	17.9	a	47					10.6	26.5	a	23	5.2	5.8	a	14	9.0	0.76	a	26	9.6	44.0	b	32
2019					6.6	18.0	a	42																				
P-value		<0.0001				0.3872				<0.0001				<0.0001				0.0623				0.0362				<0.0001		
N		149				89				447				186				215				199				287		

(SOC: soil organic carbon; Bray-P: available soil phosphorous with Bray; citric acid-P: available soil phosphorous with citric acid; K_{exch}: exchangeable soil potassium. Means with the same letter are not significantly different by least significant difference test at $P < 0.05$)

Table 2
Relative standard deviation (RSD%), average concentration (Av), and number of observations (n) for five laboratory methods applied on selected internal reference soil materials (IRM) by up to five different operators (T) at INIA's soil laboratory, La Estanzuela, Uruguay.

	SOC-Tinsley				Bray-P				Citric acid-P				pH				K_{exch}				Bray-P			
	IRM1				IRM3				IRM4				IRM4				IRM4				IRM4			
	RSD%	Av (g kg^{-1})		n	RSD%	Av (ppm)		n	RSD%	Av (ppm)		n	RSD%	Av		n	RSD%	Av (cmol kg^{-1})		n	RSD%	Av (ppm)		n
T1	2.1	14.9	a	33	6.8	18.0	ab	157	7.1	27.7	a	64	2.1	5.9	c	131	4.6	0.79	a	162	11.4	40.2	a	50
T2	1.4	15.0	a	95	4.8	19.0	c	32	11	29.0	c	77	2.2	5.8	b	11	8.6	0.77	a	33	7.1	38.9	a	22
T3	1.1	15.0	a	10	4.4	18.9	c	50	7.7	27.4	a	19	3.4	5.8	b	49					10.7	40.5	a	117
T4					6.1	17.5	ab	35	8.1	28.1	b	17	2.2	5.7	a	15					11.3	41.2	a	44
T5					6.1	18.2	b	88													9.2	38.8	a	28
P-value		0.055				<0.0001				0.0191				<0.0001				0.1103				0.1094		
N		138				362				177				206				195				261		

SOC: soil organic carbon; Bray-P: available soil phosphorous with Bray; citric acid-P: available soil phosphorous with citric acid; K_{exch}: exchangeable soil potassium. Means with the same letter are not significantly different by least significant difference test at $P < 0.05$).

Table 3
Validation parameters of reference material for six different soil laboratory methods averaged over all operators at INIA's soil laboratory, La Estanzuela, Uruguay (IRM: internal reference soil materials, n: number of observations, RSD_R: relative standard deviation of the reference material for repeatability in %, Av: average concentration of the reference material in ppm, RSD_H: result obtained from Horwitz equation (predicted relative standard deviation for reproducibility in %); Max RSD_R: maximum acceptable standard deviation (=RSD_H x2) in %; HorRat: Horwitz ratio (see Table 2)).

Lab method (ID, years in use)	n	RSD_R	Av (ppm)	RSD_H	Max RSD_R	HorRat
SOC-Tinsley (IRM1, 5)	149	1.5	14,940	3.8	7.6	0.4
SOC-LECO (IRM2, 2)	89	5.7	17,900	3.7	7.4	1.5
Bray-P (IRM3, 5)	447	6.7	18.2	10.3	20.6	0.7
Citric acid-P (IRM4, 6)	186	9.5	28.3	9.7	19.4	1.0
pH* (IRM4, 6)	215	2.8	5.8	12.3	24.6	0.2
K_{exch}** (IRM4, 5)	199	5.2	0.77	16.6	33.2	0.3

(SOC: soil organic carbon; Bray-P: available soil phosphorous with Bray; citric acid-P: available soil phosphorous with citric acid; K_{exch}: exchangeable soil potassium; *soil pH was not transformed in negative logarithm of the hydrogen ion concentration to obtain ppm concentration, but remained in pH scale; ** in cmol kg^{-1})

(Steel and Torrie, 1960; Valcárcel, 2000).

$$Z-Score=\frac{Laboratory\ result-overall\ mean\ obtained\ by\ all\ participating\ laboratories}{standard\ deviation} \tag{4}$$

(after Belli et al., 2009)

3. Results and discussion

3.1. Analytical method validation: precision

∼Table 1.
∼Table 2.
∼Table 3.

3.1.1. SOC by Tinsley wet combustion

The factor year of analysis (Table 1) significantly affected SOC concentrations of the internal reference soil material 1 (IRM1), but operator did not (Table 2). The maximum acceptable RSD for wet combustion by the Tinsley method, according to the Horwitz equation, is 7.6% (Table 3). Still, calculated values were always under 2%, following the reported CV by Chatterjee et al. (2009). The SOC concentration of IRM1, used over five years, ranged between 14.6 and 15.1 g kg^{-1} (Table 1). Firstly, this confirmed a minor variation in concentration and method accuracy, pointing towards a low error of the analytical method on SOC evolution over time, and hence a low impact on a possibly measurable treatment effect. Secondly, the results showed that SOC concentration was affected by long-term storage time after it has been dried, becoming slightly higher from year to year. This finding was not in accordance with results of other studies (Blake et al., 2000; Kühnel et al., 2019) and needs to be crosschecked when stored soil samples from LTE archives are (re) analyzed. However, the unbalanced number of samples in two deviant years 2001 and 2005 weakened the statistical analysis and does not provide a robust evidence for concrete storage failure as RSD remained very low. To minimize storage effects on laboratory precision, the inclusion of a certified standard could solve the problem.

3.1.2. SOC by LECO dry combustion

The SOC concentration of IRM2 analyzed with LECO did not show strong variations. The RSD for the IRM2 analyzed with LECO was 4.8% in 2018 and 6.6% in 2019 (Table 1) indicating an acceptable data precision. The IRM2 was analyzed for a minimum time span of 12 months to ensure the suitability as reference material (Jochum and Brueckner, 2008). However, it was not possible to conduct a robust method validation because there were only two years of LECO analysis observations; significantly less than Tinsley IRM data (> five years).

3.1.3. Comparison between methods Tinsley and LECO method

To compare data obtained by Tinsley's and LECO's procedure, samples of INIAs soil archive were reanalyzed. A *t*-test of paired samples

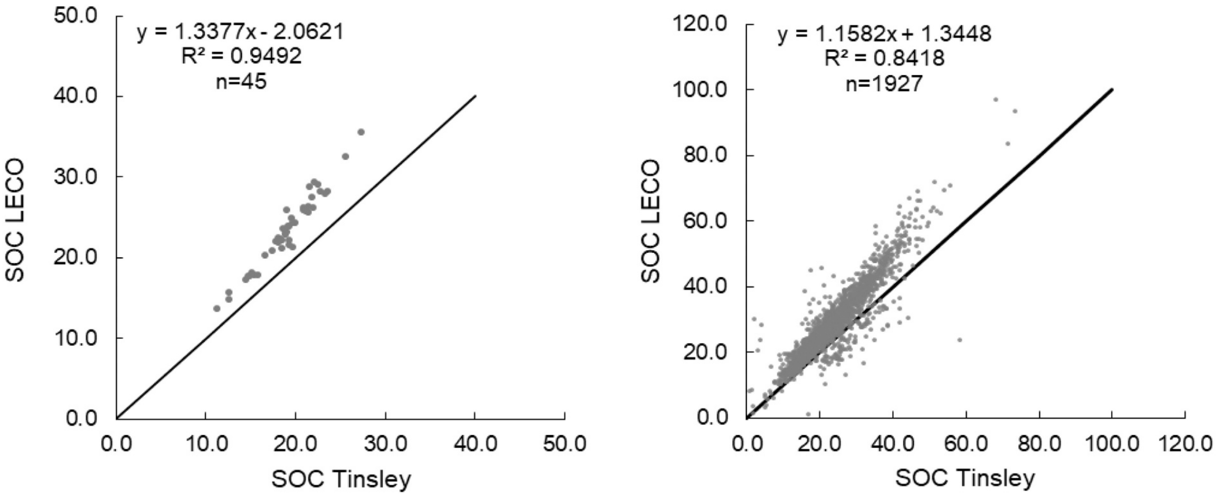

Fig. 1. Method comparison for soil organic carbon (SOC) concentrations (g kg^{-1}) with Tinsley and LECO against the 1:1 line for (A) Reanalyzed samples of the LTE in 2019 and (B) Soil samples for internal method validation (collection of data provided by Alejandro Morón) in 2007 and 2008 at INIA's soil laboratory, La Estanzuela, Uruguay.

showed significant differences ($P < 0.0001$) between the data obtained from Tinsley's and LECO's procedure, respectively (Fig. 1A). This finding agrees with the past internal method validation using almost 2000 randomly collected soil samples from various soil types and broader SOC concentration levels. For both sample sets, the same correction factor of 0.81 was obtained (Fig. 1B).

Internal method validation showed that the correction factor and correlation between both methods depend on the sample's SOC concentration. Samples with SOC content below 15 g kg^{-1} showed a low correlation ($R^2 = 0.25$, $n = 281$) compared with SOC above 35 g kg^{-1} ($R^2 = 0.50$, $n = 196$). Hence, the correction factor to convert LECO to Tinsley values was 0.79 for soil samples with low SOC, and 0.85 for samples with high SOC concentrations. This was similar to results found by Chan et al. (2011), who used correction factors based on a country scale survey conducted by Skjemstad et al. (2000) in Australia. For a site with high initial SOC content, the correction formula for conversion of Walkey-Black (W—B, like Tinsley) SOC to LECO-SOC was $C_{LECO} = 1.074\ C_{W\text{-}B} + 0.112$ ($R^2 = 0.95$). This differentiated conversion of SOC values is required when LTE studies used several SOC methods over time and handle a wide range of SOC concentrations within the data set. According to De Vos et al. (2007), SOC variation in a forest soil was explained by up to 29% due to the laboratory method (here especially the reaction temperature) and to a much lesser extent by texture or tree species.

However, in our study, the wet digestion technique showed higher precision (Table 1, RSD) compared with dry combustion, although extraction rates of SOC were around 20% higher with dry combustion which was in line with findings from other studies (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 H_2SO_4 dilution is less than that externally supplied (Mikhailova et al., 2003). This incomplete reaction explains that only 75% to 80% of all SOC was measured (Piper, 1944).

3.1.4. Extractable phosphorus determined by Bray I and citric acid extractants

The Bray-P and citric acid-P extraction of different IRMs was performed over five and six years, respectively. The IRM3 and the IRM4 were analyzed with the Bray-P and citric acid-P extraction methods, respectively (Supplementary Table A.2). Significant differences for Bray-P concentrations in IRM3 were found for year and operator. An average increase in Bray-P concentration in IRM3 of 0.5 ppm per year over five years was observed. These numerical differences must be treated with caution and do not necessarily imply agronomically relevant effects (Table 1). Yearly variations in RSD were low and ranged from 3.1% to 6.4%. The RSD for all operators ranged between 4.4 and 6.8% and was much lower than the reported maximum acceptable RSD of 20.6% (Table 3). In contrast, no significant differences between operators were found in IRM4 with Bray I extraction (Table 2). The IRM4 also showed a sharp concentration increase of 7 ppm in the fourth and fifth year of Bray-P repetitive analysis (Table 1). It is most likely that this happened due to the method change which was implemented in January 2016: an increase of the soil/solution ratio from 1/7 to 1/10 and the flask shaking was modified from vertical to horizontal stirring.

Table 4
Results of interlaboratory trials (*Z*-Score, compare Table 2) for three different soil laboratory methods over 11 years at INIA's soil laboratory, La Estanzuela, Uruguay (n: number of yearly interlaboratory samples analyzed by INIA's laboratory, SOC: soil organic carbon; K_{exch}: exchangeable soil potassium).

		Z-Score		
Year	n	SOC	Soil pH	K_{exch}
2006	12	1.3	0.9	1.5
2007	4	−0.4	0.3	0.9
2008	8	0.5	2.0	1.9
2009	8	−0.5	2.0	0.7
2010	20	−0.4	1.0	0.4
2012	4	1.6	1.8	1.6
2015	20	0.1	1.0	1.3
2016	20	−0.8	1.0	2.0
2017	20	−0.1	0.6	2.7
2018	20	−1.2	0.5	2.1
2019	12	−0.2	0.9	−0.4

The IRM3 was not dried again after its preparation and stayed in a not closely sealed plastic container exposed to air moisture. The same reference material (IRM3) was reanalyzed during routine daily analyses between July and December 2019. It resulted in average Bray-P concentrations of 30 ppm ($n = 14$, RSD = 7.9%). The Bray-P concentration in IRM3 had an annual linear increase of 1.1 ppm from 2009 to 2019 ($R^2 = 0.93$, data not shown). We assume that sample storage containers for IRM3 did not prevent daily or seasonal dry-wetting cycles. Thus, the observed increase in Bray-P might be associated to P mobilization by microbes as reactivated and increased with time due to absent repetitive drying of the reference material. These results are in contrast with De Nobili et al. (2006), who found no significant change in soil microbial biomass after 12 years of storage in recently air-dried and rewetted soils. Houba and Novozamsky (1998) have shown that extractable nutrients like P in soil samples only increased significantly after six to 24 months when stored at 70 °C. Authors mentioned as possible explanation that the physical structure of organic matter may change with time which in turn affects attached nutrients. Therefore, a subsequent analysis within one calendar year or airtight and cool storage of soil samples for Bray-P determination is recommended. For soils, alternate wetting and drying have a considerable effect on phosphorus solubility (Worsfold et al., 2005). Overall, to the best of our knowledge there are no previous studies that evaluate the effects of long-term soil storage on soil available P. In this matter, Benton Jones (2001) reported limited knowledge of the effects of storage time in soil chemical indicators.

Phosphorus concentrations by citric acid extraction were significantly affected by year and operator (Tables 4 and 5). The average citric acid-P of the IRM4 was 28 ppm in 2013, 2014, and 2015, respectively, increased to 32 ppm in 2016 and decreased to 27 ppm in 2017 and 2018. However, no sample storage time effects were observed on P extraction with citric acid, but variations occurred for the particular year 2016.

Differences between methods in RSD were not addressed statistically, but extractable P with citric acid and Bray had the highest RSD values (9.5 and 6.7, respectively, Table 3). Essentially, these operationally-defined methods are *incomplete* extractions, where the amount of P extracted depends on the extraction conditions (i.e. shaking time, air temperature, soil:solution ratio, extraction solution concentration). Therefore, the precision of the size of this P pool depends on the precision of these extraction conditions. These methods have inherently more sources of error than other methods such as SOC combustion, where a sample is subjected to digestion under conditions *in excess*. HorRat values for both methods were in an acceptable range (Table 3) and coincided with HorRat values reported for the Mehlich-3 test for extractable P, which range between 0.36 and 1.07 (Zhang et al., 2009).

Since the IRM4 was used for both P extraction methods between 2013 and 2018 (supplementary Table A.2), analytical differences and extraction behavior are presented. The P extraction of the same IRM was higher with Bray-P (40.2 ppm, $n = 287$) compared with citric acid-P (28.5 ppm, $n = 186$). Considering that both methods use different extraction solutions and that suitability to represent available P vary with soil type and properties (Abdu, 2006; Wuenscher et al., 2015), correlation with plant P extractions should be evaluated.

3.1.5. Soil pH

Significant differences in pH were found between operators (Table 2). This finding reinforces the importance of using stable and homogenous IRM in each sample batch for proper quality control as recommended by international suppliers of proficiency tests like WEPAL (WEPAL, 2019). However, no significant differences in pH were found

Table 5

Comparison between internal reference soil materials (IRM) originated from LTE and annual soil survey from LTE for a particular year and plot at INIA's soil laboratory, La Estanzuela, Uruguay (n = number of observations in the year of IRM analysis, Av: average, SD = standard deviation, LTE value: result of composite sample originating from annual soil survey in long-term experiment corresponding to Year of IRM analysis).

	SOC-Tinsley (IRM1, g kg^{-1})	SOC-LECO (IRM2, g kg^{-1})	Bray-P (IRM3, ppm)	Citric acid-P (IRM4, ppm)	pH (IRM4)	K_{exch} (IRM4, cmol kg^{-1})
Year of IRM sampling	2000	2016	2003	2012	2012	2012
Year of IRM analysis[a]	2002	2018	2010	2013	2013	2014
n (IRM)	40	47	123	73	37	37
Av (IRM)	14.9	17.9	17.4	28	5.9	0.77
SD	0.2	0.9	1.1	2.4	0.2	0.03
LTE value[b]	14.9	16.9	21.2	32.6	6.2	0.59

[a] Year of highest number of analysis.

[b] Value of the composite sample taken in the year of IRM sampling in the same plot during the annual soil survey of the LTE.

between years (Table 1). The IRM4 was used over six years without any significant differences in analytical results of soil pH. The latest changes in the pH methodology were implemented in 2014, reducing the stirring period from 15 min to 3 min (Beretta-Blanco et al., 2014), which did not affect the pH of the IRM4. Stable pH values were observed over a storage time of six years. This was in disagreement with Falkengren-Grerup (1995) results who reported a gradual decrease in pH that averaged 0.8 units in soil samples after 40 years of storage. Still, the evaluation period was not sufficient to draw conclusions on long-term pH changes in stored soil samples. The applied method of soil pH reached the highest precision, which contributed to a sound data interpretation of LTE values after 55 years conducted in Grahmann et al. (2020).

3.1.6. *Exchangeable K*

The effect of the year on K_{exch} was relatively weak ($p = 0.0362$, Table 1). The method determines a very small range of concentration, which automatically increases the acceptable variation (Table 3). All years, except 2018, exhibited excellent precision for K_{exch}. No significant differences for K_{exch} in the IRM4 were found between the two operators that worked with this method (Table 2). Data suggest that a limited number of selected operators conducting a particular method could lead to higher precision with higher repeatability between individuals (Table 2). Blake et al. (2000) found that K_{exch} concentrations in long-term stored samples analyzed in 1959 and again in 1991 were significantly different. Our data do not support this trend, although only six years were evaluated.

3.2. *Analytical method validation: trueness*

Interlaboratory performance for SOC, soil pH, and K_{exch} was tested in 11 years to check the trueness of data. The *Z*-Scores for SOC content were found to vary in such interlaboratory performance tests between −1.2 and 1.6 (Table 4). Similar results with a Z-Score between −1.5 and 1.5 were found in a Mexican interlaboratory trial with eight participating laboratories (Guerrero Peña et al., 2019). For soil pH, and K_{exch} the Z-Score varied between 0.3 and 2.0 and − 0.4 to 2.7, respectively. Note that most method validation parameters of pH measurement remain low by definition as they correspond to a log scale. Therefore, the Z-Score is a minor indicator to identify the variation of pH analysis between laboratories. For pH it is more relevant to look at the absolute range of variation (Suvannang and Hartmann, 2019).

An exception was observed in the analysis of K_{exch} in 2017 and 2018 (Table 4), when the Z-Score exceeded the critical value of 2 which points towards unsatisfactory laboratory performance. The fact that the participating laboratories used different standard operation procedures for K extraction (amount of extractant, concentration of extractant, amount of soil sample, etc.) might explain the observed differences (Linsinger et al., 1998; Suvannang and Hartmann, 2019).

3.3. *Implications for the data accuracy of the LTE annual soil surveys*

Following the Horwitz curve (Horwitz and Albert, 2006), the acceptable RSD increased with decreasing target concentrations of the reference material (Table 3). Essential elements had the lowest predicted RSD (=CV%), for example 3.8% for SOC determined with Tinsley, whereas macronutrients with smaller target concentration like exchangeable K had a predicted RSD of 16.6%. In commercial laboratories, a maximum acceptable variation of 3% is quite common to ensure high data quality which increases at low concentration ranges (König et al., 2013).

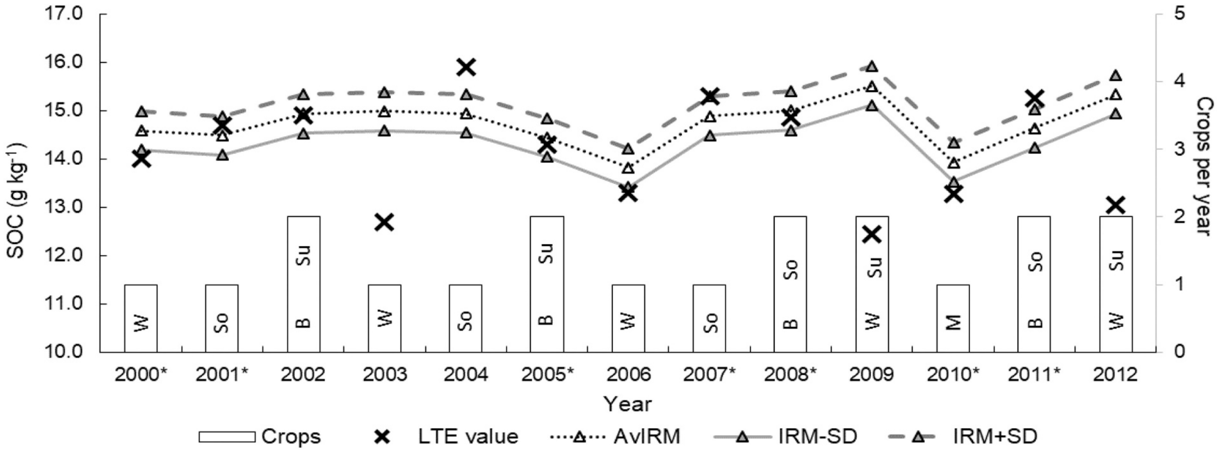

Fig. 2. Comparison of soil organic carbon (SOC) concentrations (g kg^{-1}) between annually-averaged internal reference soil materials (AvIRM) and annual soil survey value over 13 years (IRM-SD and IRM + SD present maximum tolerable concentration range of internal reference soil materials) and respective grown crops in the LTE treatment CA+ (W: wheat, B: barley, So: sorghum, Su: sunflower, M:maize; Supplementary Table 1). Stars indicate year of new IRM batch sampling in plot 5 (similar treatment CA+) of the LTE (Supplementary Table 3).

Marginal differences between the yearly internal reference soil material average (Av IRM) and the corresponding annual LTE value ("LTE value", Fig. 2, details on sampling explained in section 2.5 and Grahmann et al., 2020) imply that for most soil chemical properties, annual soil surveys had high data quality with accurate values (Table 5). An additional use of certified material could underline these findings, but was mostly absent in this laboratory. These kinds of high quality datasets can be used for model development, model fitting and robust statistical analysis of long-term treatment effects (Johnston and Poulton, 2018) although soil data originate from a non-certified soil laboratory. As mentioned before, the precision for all methods was within the acceptable range provided by the HorRat value. K_{exch} had considerable differences between LTE and IRM value (Table 5). Most likely, the differences between IRM and LTE value correspond to spatial plot variability or sampling error. There was no K fertilization in the entire experiment over 50 years, and the sampled LTE plots still demonstrate a natural K within-field variation. The IRM4 was taken from the outer parts of the plot with less biomass production due to plot border effects and hence reduced K soil removal, leading to higher K concentrations in the reference soil sample. In the remaining central plot area, where the LTE soil survey was conducted annually, K was depleted by long-term crop removal (Ryan et al., 2012). The LTE soil survey values of P, extracted either by Bray I or citric acid, were higher compared with the average IRM values. P is applied each year as mineral fertilizer in the experimental plots. It is possible that the fact that fertilizer spreaders made better P distribution in the plot center where annual soil samples were collected explained lower values of IRM taken in the plot edge.

High fluctuations in SOC from 12.7 g kg^{-1} in 2003 to 15.9 g kg^{-1} in 2004 were observed in the annual soil survey in plot 5 of the LTE ("LTE value", Fig. 2). Assuming a bulk density of 1.25 g cm^{-3} in the top 0–15 cm soil depth, this variation represents an increase of 6.00 Mg ha^{-1} which is unrealistic to reach in such a short period for any agricultural system under these conditions. According to a meta-analysis conducted by Jian et al. (2020), the mean rate of carbon sequestration using cover cropping (higher values of C accumulation) across all studies was 0.56 Mg ha^{-1} yr^{-1}. Fig. 2 considered the yearly averaged IRM concentration of seven batches, including their minimum and maximum acceptable RSD of the laboratory method. The IRM batches (IRM1-IRM1f; supplementary Table A.6), all originating from plot 5 of the LTE, were collected between 2000 and 2012 and used for two to five years. For the previously reported SOC laboratory precision analysis (section 3.1.1), only IRM1 was used. In three years (2003, 2009, 2012), the measured SOC concentration of the LTE value was outside the confidence interval resulting from the IRM (Fig. 2). If soil samples were taken only every three to six years, which is quite common in other LTE studies, conclusions would diverge significantly from actual values as SOC increase and decrease rates were much higher compared with years that were within the confidence interval. This sharp fluctuation of SOC in the annual soil campaigns can mainly be attributed to a sampling error or high spatial variability and less to a significant treatment effect caused by crop rotation effects (which is the focus of the studied LTE). Also, we can exclude that laboratory accuracy was responsible for these concentration leaps in 2003, 2009 or 2012 as the IRM curves were flatter (Fig. 2). Intrinsic spatial variation of each plot needs to be considered in sporadic, punctual or short-term studies taking place in LTE and could mask the expected long-term treatment effects. This background variability of experimental units in LTE was already discussed and is of outstanding importance to account for initial soil differences within plots (Poeplau et al., 2016; -Siri-Prieto et al., 2006; Zhang et al., 2020). Poeplau et al. (2016) described that heterogeneity of SOC values within a plot can be as significant as the SOC difference obtained over many decades. We suggest that these particular data points outside the IRM confidence interval should be regarded as outliers or taken with caution in the LTE data base. The validity of studies that analyze SOC changes in LTE considering only sporadic soil analysis is therefore limited. Also, in many LTE studies that analyze temporal changes in soil quality, nutrient dynamics or C stocks, the quality of soil data is not considered. We found that uncertainty estimates and method validation parameters like RSD or HorRat values for different soil property analyses are rarely included in the materials and methods section of the reviewed references which was also observed by van Leeuwen et al. (2021). Hence, the reader is not informed about precision performance of the laboratory analysis in LTE studies. To avoid the risk of neglecting the importance of this information regarding subsequent proper data analysis and interpretation, we encourage LTE scientific coordinators and institute managers to establish and publish long-term method validation schemes, particularly for non-certified soil laboratories. Furthermore, it is recommended that soil laboratories document all relevant laboratory data and information on method validation meticulously over the LTE running time. Internationally endorsed protocols to harmonize and standardize soil laboratory procedures for LTE studies need to be developed (Jandl et al., 2014) and should adhere to the framework of the Global Soil Laboratory Network (GLOSOLAN; Suvannang and Hartmann, 2019).

4. Conclusions

This paper proposes a complete and straightforward procedure for the non-certified soil laboratory in Uruguay to evaluate data quality of different soil chemical parameters from samples taken in the oldest LTE in Latin America. It describes various statistical procedures and performance parameters that can be used to ensure precise and true soil data over many years, which is a prerequisite to answer scientific questions in long-term trials. The study showed that repeatability (year dimension) and reproducibility (operator dimension) tests for laboratory methods are crucial to further understand LTE soil data and to improve precision of laboratory techniques over time. We recommend that soil laboratories involved in LTE studies should provide yearly quality method validation reports using internal, or rather preferable, certified reference soil materials.

Results of this study demonstrated that laboratory analysis over many years were of high precision for most soil analytes, despite the lack of certification or accreditation. However, the overall concept of accuracy including trueness could not be confirmed for all soil properties due to the lack of interlaboratory trials for P methods and absent certified reference material. This is a fundamental prerequisite to use the obtained LTE soil data for study comparisons, modelling approaches and meta-analysis of LTE soil studies around the globe. Especially for SOC, one of the most often analyzed analytes in soils, data accuracy of the LTE values was very high. Inappropriate IRM soil sample storage showed low variability over the years for soil pH, K_{exch} and SOC, but resulted in concentration increases for Bray-P, which we attributed to rewetting and demands for an improvement of storage conditions.

A detailed protocol for soil sampling and their subsequent laboratory analysis is critical in research, especially when it aims to detect potential differences in soil properties generated by different treatments over time. If soil samples were sent to different laboratories during different sampling campaigns, researchers need to request internal quality control data to verify data accuracy and analytical bias for each laboratory. We also suggest a meticulous materials and methods section in publications when soil data from LTE are presented. Particularly method details like extraction time, amount of used soil or mL of extractant contribute to making values from different LTE more comparable, e.g. for meta-analysis. We recommend soil laboratories to analyze various method validation parameters in regular intervals, as only the combination of these performance tools (RSD, HorRat, *Z*-Score and statistical significance tests) enables to decide about the suitability and reliability of a laboratory method and its subsequent data utilization in LTE studies.

The aggregate of data demonstrated that, for the oldest agricultural LTE in Latin America, internal reference soil material and interlaboratory data, were useful to test method precision, trueness and data-set reliability of several soil properties evaluated at the site over a decade. These results add to the evidence of the importance of chemical soil laboratory method validation in LTE research and should be considered in other LTE soil dynamic studies. Proficiency testing in non-certified soil laboratories using internal reference soil material contributes to an improved outreach of long-term studies in international soil science but requires further advancement and efforts in the use of certified reference material.

Funding information

The National Institute of Agricultural Research (INIA), Uruguay, funded research that supports the results presented in this publication.

Compliance with ethical standards

Research involving human participants and/or animals.

This article does not contain any studies involving animals and human participants performed by any of the authors.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Open Access

Acknowledgments

We thank the institute's field team (Emiliano Barolin, Gualberto Soulier, Hector Vergara) for current LTE trial maintenance, and the soil laboratory team (Daniel Bassahun, Deborah Torres, Oscar Torres, Raquel Musselli, Leticia García, Rosario Oten, Irene Purtscher) for the analysis of soil samples over many decades, as well as a detailed record-keeping of internal quality control results. We also acknowledge previous keepers of the trial, José L. Castro, Roberto Díaz-Rosello, Alejandro Morón and Jorge Sawchik.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2022.e00487.

References

Abdu, N., 2006. Soil-phosphorus extraction methodologies: a review. Afr. J. Agric. Res. 1, 159–161.

Albert, R., Horwitz, W., 1997. A heuristic derivation of the Horwitz curve. Anal. Chem. 69, 789–790. https://doi.org/10.1021/ac9608376.

Belli, M., de Zorzi, P., Sansone, U., Shakhashiro, A., Gondin da Fonseca, A., Trinkl, A., Benesch, T., 2009. A soil sampling intercomparison exercise for the ALMERA network. J. Environ. Radioact. 100, 982–987. https://doi.org/10.1016/j.jenvrad.2009.08.002.

Benton Jones, J.J., 2001. Laboratory Guide for Conducting Soil Tests and Plant Analysis. CRC Press, Boca Raton, FL, USA.

Beretta-Blanco, A., Bassahum, D., Musselli, R., 2014. ¿Medir el pH del suelo en la mezcla suelo: agua en reposo o agitando? Agrociencia Uruguay 18, 90–94.

Beretta-Blanco, A., Pérez, O., Carrasco-Letelier, L., 2019. Soil quality decrease over 13 years of agricultural production. Nutr. Cycl. Agroecosyst. 3 https://doi.org/10.1007/s10705-019-09990-3.

Berhe, A.A., Ghezzehei, T.A., 2020. Race and racism in soil science. Eur. J. Soil Sci. 1–6 https://doi.org/10.1111/ejss.13078.

Berti, A., Marta, A.D., Mazzoncini, M., Tei, F., 2016. An overview on long-term agro-ecosystem experiments: present situation and future potential. Eur. J. Agron. 77, 236–241. https://doi.org/10.1016/j.eja.2016.01.004.

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. Eur. J. Soil Sci. 51, 345–353. https://doi.org/10.1046/j.1365-2389.2000.00307.x.

Boone, R.D., Grigal, D.F., Sollins, P., Ahrens, R.J., Armstrong, D.E., 1999. Soil sampling, preparation, archiving, and quality control. In: Standard Soil Methods for Long-Term Ecological Research. Oxford University Press, pp. 3–28.

Bray, R.H., Kurtz, L.T., 1945. Determination of total, organic and available forms of phosphorus in soils. Soil Sci. 39–45.

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 Res 320–328. https://doi.org/10.1071/SR10185.

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M.Z., Ebinger, M.H., 2009. Evaluation of different soil carbon determination methods. CRC Crit Rev Plant Sci 28, 164–178. https://doi.org/10.1080/07352680902776556.

Congreves, K.A., Hayes, A., Verhallen, E.A., Van Eerd, L.L., 2015. Long-term impact of tillage and crop rotation on soil health at four temperate agroecosystems. Soil Tillage Res. 152, 17–28. https://doi.org/10.1016/j.still.2015.03.012.

da Silva Dias, R., Aparecida de Abreu, C., Ferreira de Abreu, M., Paz González, A., 2015. Statistical methods for evaluating results from soil micronutrient analyses in Interlaboratory programs. Commun. Soil Sci. Plant Anal. 46, 57–71. https://doi.org/10.1080/00103624.2014.988091.

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 Biol. Biochem. 38, 2871–2881. https://doi.org/10.1016/j.soilbio.2006.04.044.

De Oliveira, A., Carneiro Amado, T., Rice, C.W., Briedis, C., 2018. Driving factors of soil carbon accumulation in Oxisols in long-term no-till systems of South Brazil. Sci. Total Environ. 735/742 https://doi.org/10.1016/j.scitotenv.2017.12.019.

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 Manag. 23, 221–229. https://doi.org/10.1111/j.1475-2743.2007.00084.x.

Dyke, G.V., 1974. Long-term experiments. Comp Exp with F Crop 58–68. https://doi.org/10.2307/2983114.

Falkengren-Grerup, U., 1995. Effects of long-term storage on some chemical properties of forest soil samples. Ecological Bulletins 44, 129–132.

González, A.G., Herrador, M.Á., Asuero, A.G., 2010. Intra-laboratory assessment of method accuracy (trueness and precision) by using validation standards. Talanta 82, 1995–1998. https://doi.org/10.1016/j.talanta.2010.07.071.

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. Agric. Ecosyst. Environ. 287, 106710 https://doi.org/10.1016/j.agee.2019.106710.

Guerrero Peña, A., Jarquín Sánchez, A., Etchevers Barra, J.D., González Acuña, I.J., Hernández Nataren, L.D.C., Cueto Wong, J.A., Báez Pérez, A., Hidalgo Moreno, C.I., López Anaya, M.A., Martínez Vargas, M., 2019. Evaluación de la calidad de la medición de carbono total en suelo mediante ensayo de intercomparación. Rev Terra Latinoam 37, 273. https://doi.org/10.28940/terra.v37i3.377.

Hernández, J., Berger, A., Deambrosi, E., Lavecchia, A., 2013. Soil phosphorus tests for flooded Rice grown in contrasting soils and cropping history. Commun. Soil Sci. Plant Anal. 44, 1193–1210. https://doi.org/10.1080/00103624.2012.756000.

Hibbert, D.B., 2007. Systematic errors in analytical measurement results. J. Chromatogr. A 1158, 25–32. https://doi.org/10.1016/j.chroma.2007.03.021.

Horwitz, W., 1982. Evaluation of analytical methods used for regulation of foods and drugs. Anal. Chem. 54, 67A–76A. https://doi.org/10.1021/ac00238a765.

Horwitz, W., Albert, R., 2006. The Horwitz ratio (HorRat): a useful index of method performance with respect to precision. J. AOAC Int. 89, 1098–1109.

Houba, V.J.G., Novozamsky, I., 1998. Influence of storage time and temperature of air-dried soils on pH and extractable nutrients using 0.01 Mol/L CaCl2. Fresenius J. Anal. Chem. 360, 362–365. https://doi.org/10.1007/s002160050712.

IAEA, 2003. Development and use of reference materials and quality control materials. Int At Energy Agency 1, 1–113.

Jackson, M.L., 1964. Análisis Químico de Suelos. Ediciones Omega S.A, Barcelona.

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

Jian, J., Du, X., Reiter, M.S., Stewart, R.D., 2020. A meta-analysis of global cropland soil carbon changes due to cover cropping. Soil Biol. Biochem. 107735 https://doi.org/10.1016/j.soilbio.2020.107735.

Jochum, K.P., Brueckner, S.M., 2008. Reference materials in geoanalytical and environmental research - review for 2006 and 2007. Geostand. Geoanal. Res. 32, 405–452. https://doi.org/10.1111/j.1751-908X.2008.00916.x.

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. Eur. J. Soil Sci. 69, 113–125. https://doi.org/10.1111/ejss.12521.

Kay, B.D., VandenBygaart, A.J., 2002. Conservation tillage and depth stratification of porosity and soil organic matter. Soil Tillage Res. 66, 107–118. https://doi.org/10.1016/S0167-1987(02)00019-3.

Kennedy, V.H., Rowland, A.P., Parrington, J., 1994. Quality assurance for soil nutrient analysis: a case study. Commun. Soil Sci. Plant Anal. 25, 1605–1627. https://doi.org/10.1080/00103629409369139.

König, N., Cools, N., Derome, K., Kowalska, A., De Vos, B., Fürst, A., Marchetto, A., O'Dea, P., Tartari, G.A., 2013. Data quality in laboratories: methods and results for soil, foliar, and water chemical analyses. Develop. Environ. Sci. https://doi.org/10.1016/B978-0-08-098222-9.00022-4.

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. J. Plant Nutr. Soil Sci. 182, 772–781. https://doi.org/10.1002/jpln.201800529.

Linsinger, T.P.J., Kandler, W., Krska, R., Grasserbauer, M., 1998. The influence of different evaluation techniques on the results of interlaboratory comparisons. Accred. Qual. Assur. 3, 322–327. https://doi.org/10.1007/s007690050254.

Martínez, I., Chervet, A., Weisskopf, P., Sturny, W.G., Rek, J., Keller, T., 2016. Two decades of no-till in the Oberacker long-term field experiment: part II. Soil porosity and gas transport parameters. Soil Tillage Res. 163, 130–140. https://doi.org/10.1016/j.still.2016.05.020.

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 Manag. 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. Accred. Qual. Assur. 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. Commun. Soil Sci. Plant Anal. 34, 1853–1860. https://doi.org/10.1081/CSS-120023220.

Minasny, B., Fiantis, D., Mulyanto, B., Sulaeman, Y., Widyatmanti, W., 2020. Global soil science research collaboration in the 21st century: time to end helicopter research. Geoderma 373, 114299. https://doi.org/10.1016/j.geoderma.2020.114299.

Núñez, A., Morón, A., 2017. Potassium dynamics in Western Uruguayan agricultural Mollisols. Commun. Soil Sci. Plant Anal. 48, 2558–2572. https://doi.org/10.1080/00103624.2017.1416134.

Palermo, A., Pérez, M., Servetto, M., 1985. Índices de disponibilidad de nitrógeno y fósforo en suelos arroceros. Undergraduate Thesis, Facultad de Agronomía, Universidad de la República. Montevideo, Uruguay.

Piper, C.S., 1944. Soil and Plant Analysis. The University of Adelaide, Adelaide, Australia.

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.

Pravia, M.V., Kemanian, A.R., Terra, J.A., Shi, Y., Macedo, I., Goslee, S., 2019. Soil carbon saturation, productivity, and carbon and nitrogen cycling in crop-pasture rotations. Agric. Syst. 171, 13–22. https://doi.org/10.1016/j.agsy.2018.11.001.

Quevauviller, P., 1995. Quality assurance in environmental monitoring: sampling and sample pretreatment. VCH Verlagsgesellschaft 1–25.

Rayment, G.E., Miller, R.O., Sulaeman, E., 2000. Proficiency testing and other interactive measures to enhance analytical quality in soil and plant laboratories. Commun. Soil Sci. Plant Anal. 31, 1513–1530. https://doi.org/10.1080/00103620009370523.

Ryan, J., Singh, M., Masri, S., Ibrikci, H., 2012. Spatial variation in soil organic matter, available phosphorus, and potassium under semi-arid conditions: Research Station management implications. Commun. Soil Sci. Plant Anal. 43, 2820–2833. https://doi.org/10.1080/00103624.2012.716126.

Schut, A.G.T., Giller, K.E., 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. Geoderma 380, 114680. https://doi.org/10.1016/j.geoderma.2020.114680.

Siri-Prieto, G.S., Reeves, D.W., Shaw, J.N., Mitchell, C.C., 2006. World's oldest cotton experiment: relationship between soil chemical and physical properties and Papparent electrical conductivity. Commun. Agric. Appl. Biol. Sci. 37, 767–786.

Skjemstad, J.O., Spouncer, L.R., Beech, A., 2000. Carbon conversion factors for historical soil carbon data. National Carbon Accounting System Technical Report No 15, 28 pp.

Soil Survey Staff, 2014. Keys to soil taxonomy. Soil Conserv. Serv. 211–256. Chapter 12 Mollisols.

Steel, R.G., Torrie, J.H., 1960. Principles and Procedures of Statistics, with Special Reference to the Biological Sciences. McGraw-Hill Book Company, New York, Toronto, London.

Suvannang, N., Hartmann, C., 2019. First Inter-Laboratory Comparison Report of the Regional Soil Laboratory Network for Asia (SEALNET). Rome.

Tivet, F., de Moraes, Carlos, Sá, J., Borszowskei, P.R., Letourmy, P., Briedis, C., Ferreira, A.O., dos Santos, Burkner, Thiago Massao In, 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 Sci. Soc. Am. J. 76, 1048–1059. https://doi.org/10.2136/sssaj2011.0328.

Valcárcel, M., 2000. Principles of Analytical Chemistry, Analytical Chemistry. Springer, Berlin Heidelberg, Cordoba, Spain. https://doi.org/10.1021/ac50062a780.

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. Eur. J. Soil Sci. 1–17 https://doi.org/10.1111/ejss.13137.

Van Zoonen, P., Hoogerbrugge, R., Gort, S.M., Van De Wiel, H.J., Van't Klooster, H.A., 1999. Some practical examples of method validation in the analytical laboratory. TrAC - Trends Anal Chem 18, 584–593. https://doi.org/10.1016/S0165-9936(99)00159-4.

Verhulst, N., Kienle, F., Sayre, K.D., Deckers, J., Raes, D., Limon-Ortega, A., Tijerina-Chavez, L., Govaerts, B., 2011. Soil quality as affected by tillage-residue management in a wheat-maize irrigated bed planting system. Plant Soil 340, 453–466. https://doi.org/10.1007/s11104-010-0618-5.

WEPAL, 2019. International soil-analytical exchange programme. http://www.wepal.nl.

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.

Worsfold, P.J., Gimbert, L.J., Mankasingh, U., Omaka, O.N., Hanrahan, G., Gardolinski, P.C.F.C., Haygarth, P.M., Turner, B.L., Keith-Roach, M.J., McKelvie, I. D., 2005. Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils. Talanta 66, 273–293. https://doi.org/10.1016/j.talanta.2004.09.006.

Wuenscher, R., Unterfrauner, H., Peticzka, R., Zehetner, F., 2015. A comparison of 14 soil phosphorus extraction methods applied to 50 agricultural soils from Central Europe. Plant Soil Environ. 61, 86–96. https://doi.org/10.17221/932/2014-PSE.

Zhang, H., Kariuki, S., Schroder, J.L., Payton, M.E., Focht, C., 2009. Interlaboratory validation of the Mehlich 3 method for extraction of plant-available phosphorus. J. AOAC Int. 92, 91–102. https://doi.org/10.1093/jaoac/92.1.91.

Zhang, X., Wendroth, O., Matocha, C., Zhu, J., Reyes, J., 2020. Assessing field-scale variability of soil hydraulic conductivity at and near saturation. Catena 187, 104335. https://doi.org/10.1016/j.catena.2019.104335.