

Review Article

Nathalie Korf*, Paul Martin Mähltitz, and Vera Susanne Rotter*

Round robin tests of secondary raw materials: A systematic review of performance parameters

<https://doi.org/10.1515/revac-2022-0033>

received September 24, 2021; accepted November 29, 2021

Abstract: An improved management of secondary raw materials (SRM) is a crucial contribution for a circular economy and necessitates knowledge about the composition of wastes and SRM. However, this information is scarce and has to be determined with chemical analysis (CA). CA of SRM faces challenges, which can be approached by using round robin tests (RRT) to identify deviations from the “true value” of an element/molecule content. An RRT is a testing approach, which involves multiple labs to analyze one or more samples and evaluates the lab results with regard to the goal of the RRT. This article presents a systematic literature review and investigates which purposes and which performance parameters (PP) are commonly applied in RRT of SRM. The examined literature shows that the two main purposes applied are assessment of method performance and assessment of lab performance. PP can be categorized into trueness performance parameters (TPP; assessing the deviation of a value from a reference value) and precision performance parameters (PPP; describing the variability of a data set). The main TPP identified are *z* score and relative deviation, the main PPP identified are standard deviation and relative standard deviation. These results offer the conclusions that RRT can be used as a bespoke method to deal with analytical effects and that the selection of PP for an RRT could be based on simplicity.

Keywords: round robin test, interlaboratory comparison, proficiency test, chemical analysis, secondary raw materials, performance parameters

Abbreviations

(C)RM	(certified) reference material
CA	chemical analysis
CP	conference proceedings
ILC	interlaboratory comparison
MAD	median absolute deviation
P	not-peer-reviewed paper
PP	performance parameter
PPP	precision performance parameter
PRA	peer-reviewed article
PT	proficiency test
QA/QC	quality assurance/quality control
R	report
RRT	round robin test
RSD	relative standard deviation
SD	standard deviation
SRM	secondary raw material
TPP	trueness performance parameter

1 Introduction

More and more resources are difficult to obtain and denoted as critical [1] due to their relevance, scarcity, supply risk, and a recycling and waste management system, which mainly focuses on bulk materials. Consequently, an improved management of secondary raw materials (SRM) contributes to achieving a true circular economy [2] and hence to climate protection [3]. This management of resources demands knowledge about their composition and quantity. Exploration and prospection of primary raw materials base on well-known, developed methods and techniques to determine location, quantity, and grade of the target materials. However, the exploration and prospection of waste materials and

* **Corresponding author: Nathalie Korf**, Technische Universität Berlin, Chair of Circular Economy and Recycling Technology, Straße des 17. Juni 135, 10623 Berlin, Germany, e-mail: n.korf@tu-berlin.de

* **Corresponding author: Vera Susanne Rotter**, Technische Universität Berlin, Chair of Circular Economy and Recycling Technology, Straße des 17. Juni 135, 10623 Berlin, Germany, e-mail: vera.rotter@tu-berlin.de

Paul Martin Mähltitz: Technische Universität Berlin, Chair of Circular Economy and Recycling Technology, Straße des 17. Juni 135, 10623 Berlin, Germany

SRM are still in their fledgling stages. Usually there is no information about the elemental composition of SRM or even waste materials, due to complex manufacturing chains, technological and appearance trends as well as a mix of waste products with differing age in the collected waste flows. But knowledge of the composition of a product or material is necessary to manage its recycling and recovery.

The content of target element(s) and/or impurities can be determined with chemical analysis (CA), the goal of which is to estimate the (unknown) “true value” of element contents by applying analytical methods [4]. The results of CA can serve different objectives, such as to assess a material with regard to target elements and impurities (i.e., to assess the material composition) as well as to check compliance with limit values of restricted substances. To assure the quality of the data resulting from CA, good laboratory practice as well as quality assurance and quality control (QA/QC) measures are applied, such as using standard reference methods for specified sample materials and (certified) reference materials (CRM) to check the fitness-of-purpose of the selected analytical method [4,5]. However, CA of SRM and waste materials faces challenges due to complexity (element types and their number as well as element specification and molecule types), heterogeneity (constitutional and distributional [6]), and high-paced innovation cycles [7–9]. These challenges are paired with a lack of reference methods and (certified) reference materials, which leads to deviations from the “true value” in the form of random and/or systematic effects. These analytical effects may be caused by, e.g., the condition and composition of the sample, chemical reactions between the sample and reagents, digestion processes, or measurement interferences.

In the absence of certified reference materials, round robin tests (RRT), i.e., interlaboratory comparisons (ILC), are one approach to identify and eliminate these analytical effects for SRM [7,8]. RRT is a testing approach with a specific, pre-defined purpose, focusing on one or more samples and one or more measurement or testing methods, with the involvement of multiple labs and one organizing and coordinating institution. The organizing institution gives out the samples and instructions, receives the lab results after CA, and evaluates the lab results with regard to the purpose of the RRT (e.g., assessment of lab performance). RRT can achieve the goal of identifying and eliminating analytical effects by collating multiple analysis results, oftentimes generated with different analytical methods, showing a dispersed set of values and hence identifying potential analytical effects. RRT are oftentimes used synonymously with ILC or proficiency tests (PT). DIN EN ISO/IEC 17043 “Conformity assessment – general

requirements for proficiency testing” defines ILC as the “organization, performance, and evaluation of measurements or tests on the same or similar items by two or more laboratories in accordance with predetermined conditions,” and further PT as the “evaluation of participant performance against pre-established criteria by means of ILC” [10]. In our work, we use the term “round robin test” as an ILC study with CA of one or more samples using one or more analytical methods.

Typical purposes for RRT include, inter alia, performance evaluation of laboratories for specific analytical methods, evaluation of analytical methods, assignment of reference values (i.e., assigned values) to reference materials, and identification of interlaboratory differences [10]. Hence, possible applications and benefits of RRT are manifold and can be strategically designed and applied for specific research questions.

An important part of RRT are the performance parameters (PP), which evaluate the performance of labs and/or methods in comparison with a pre-defined value (i.e., reference, assigned, or consensus value [10]) and/or other RRT participants. Common RRT PP are absolute/relative deviation and z score [10]. Every parameter has its own informative value and has to be selected with regard to the research question of the RRT. However, a large variety of PP exist and can make it difficult to select appropriate parameter(s) for a specific purpose, e.g., to overcome analytical effects.

Considering the complexity and heterogeneity of waste and SRM, as stated before, RRT and their evaluation using PP are a key tool to increase quality acceptance of SRM and thus contribute to a circular economy. Up to the publication of this study and to the best of the authors’ knowledge there is neither a comprehensive and exhaustive overview over purpose and application of RRT for waste and SRM samples nor an overview and categorization of PP commonly applied in RRT for CA of waste and SRM samples yet.

Hence, the goal of this article is to present insights into the use of RRT PP to researchers, practitioners, and data scientists who aim to identify and eliminate analytical effects and consequently to generate reliable CA results for unknown, complex waste materials and SRM. This goal will be achieved with a systematic literature review following two research questions:

- 1) Why are RRT conducted, i.e., what are purposes/goals of RRT for SRM?
- 2) How are RRT evaluated, i.e., what are common PP for RRT?

This review is complimented by elaborate supportive information [11], describing, inter alia, RRT in general,

relevant norms, and all identified PP as well as giving a detailed overview over the results from the systematic review.

Due to the frequent use of soil reference materials and soil methods for CA of waste materials and SRM, the scope of this article is widened to solid environmental samples (including soils and similar sample materials) to achieve a better overview.

2 Methodology of the systematic literature review

The systematic literature review follows the PRISMA guidelines [12]. First, relevant publications were identified and collected. Second, the collected publications were screened with regard to their eligibility. Third, the relevant information for the systematic literature review was extracted and collated. Fourth, the extracted information was assessed with respect to the two research questions. The following sections will describe this process in more detail and are accompanied by a comprehensive flow chart in ref. [11].

2.1 Identification

The domain of interest for this literature review is the multi-/interdisciplinary field between analytical chemistry, environmental technology, waste management, and recycling technologies. To cover all areas and prevent search bias, higher level search databases were used, i.e., Google Scholar, Web of Science, Scopus, as well as the publication exchange platform ResearchGate.

A set of predefined keywords were used (Table 1). Three synonyms for RRT (RRT, ICL, PT) have been combined with a selection of sample materials as well as analysis parameters. Based on our previous studies on the CA of printed circuit boards, battery ash, and mining waste [7,8], the keywords “printed circuit boards,” “batteries,” and “mining waste” were included in the literature research.

Every combination of method, sample materials, and parameters was searched for on each of the search platforms mentioned above. For every keyword combination, the first 100 search results were examined. No distinction was made between primary publications (search keywords are included in title, abstract, and/or keywords of the publication) and secondary publications (search keywords are included in the text body of the publication). A publication was selected, if it presented results (of an RRT) of CA of

Table 1: Keywords for the literature research on RRT for CA of specified sample matrices. The left column denotes the (synonymous) terms for RRT, the middle column denotes the sample material, and the right column denotes the analysis parameter

Method	Sample materials	Parameters
Interlaboratory comparison	Solid waste Anthropogenic ores	Element content
Round robin test	Residues (Waste) printed circuit boards Batteries	Elemental composition
Proficiency test	Ore Mining waste Soil	Metal content

solid environmental samples. For the first selection, the mentioning of RRT (or ILC, PT) was not critical with the aim to prevent exclusion of relevant publications.

The collected publications were collated in a Citavi project.

2.2 Screening

In the second step, all collated publications were screened with regard to eligibility for this specific literature review. Publications were denoted as eligible if they presented an RRT of solid environmental samples. RRT with a focus on liquid samples and food or feed samples were excluded. Duplicates were removed.

2.3 Extraction

In the third step, relevant information was extracted and collated in an MS Excel table, to answer the two research questions. The relevant information is defined as follows:

- publication year
- author(s) and reference
- purpose of the published RRT
- PP used in the RRT
- sample material
- analysis type

2.4 Assessment

In the fourth step, the extracted information is assessed with the aim to answer the two research questions. To

achieve this, the extracted information regarding purpose and PP are depicted and discussed.

3 Results

The following sections present the results of the systematic review. For more detail, please refer to the supporting information [11].

3.1 Overview over the review results

In total, we collected 350 publications (peer-reviewed articles (PRA), reports, papers in conference proceedings (CP), and not-peer-reviewed papers) with the specified keyword combinations, after removal of duplicates. Subsequent to the screening for eligibility, 88 publications remained, which comprise 62 PRA [13–74], 23 reports (R) [75–95], two CP [96,97], and one not-peer-reviewed paper (P) [98].

Figure 1 shows all eligible publications by publication type. The earliest found publication is a PRA from 1979, the newest publication is a PRA from 2021. Between 1979 and 1990, only seven publications in total can be found ($n_{\text{PRA}} = 5$, $n_{\text{R}} = 2$). Starting in the 1990s, the number of published RRT increased, resulting in a publication rate per year of one to a maximum of seven (in 2017). Reports were found from mainly two organizations (excluding the two reports in the 1980s).

Regarding sample materials used in the RRT, we identified four main categories: solid waste material, mineral (waste) material, soil, and other material. The

category “other material” comprises solid samples with origin or characteristics similar to solid waste, mineral material, or soil, such as air particulates collected on a filter, construction materials, fly ash from incineration processes, soil improvers, or microplastics. The sample type “soil” was identified in 41 publications, “solid waste material” in 18 publications, and “mineral (waste) material” in 16. The collective category “other material” includes 23 publications. With the keyword combination used in the search, no RRT for complex (waste) materials were found, such as printed circuit boards or batteries.

The main analysis type presented in all publications was elemental content (70), followed by molecule content (16). All other analysis types were summarized in the category “other method” (31) and include methods such as leaching, thermogravimetric analysis, or heating value. The results for sample and analysis type are presented in Figure 2. Detailed tables with all publications and the identified sample materials and analysis types are compiled in ref. [11].

3.2 Purpose of RRT

DIN EN ISO/IEC 17043 details a list of purposes for ILC and states that their application is increasing internationally [10]. The purposes can roughly be categorized into purposes addressing lab performance and method performance. Purposes addressing lab performance listed are: (i) assessment and monitoring of lab performance, with regard to specific measurements or tests; (ii) improvement of QA and QC by identifying problems in labs and initiating actions to improve the analytical steps; (iii) increasing quality assurance by educating participating labs based

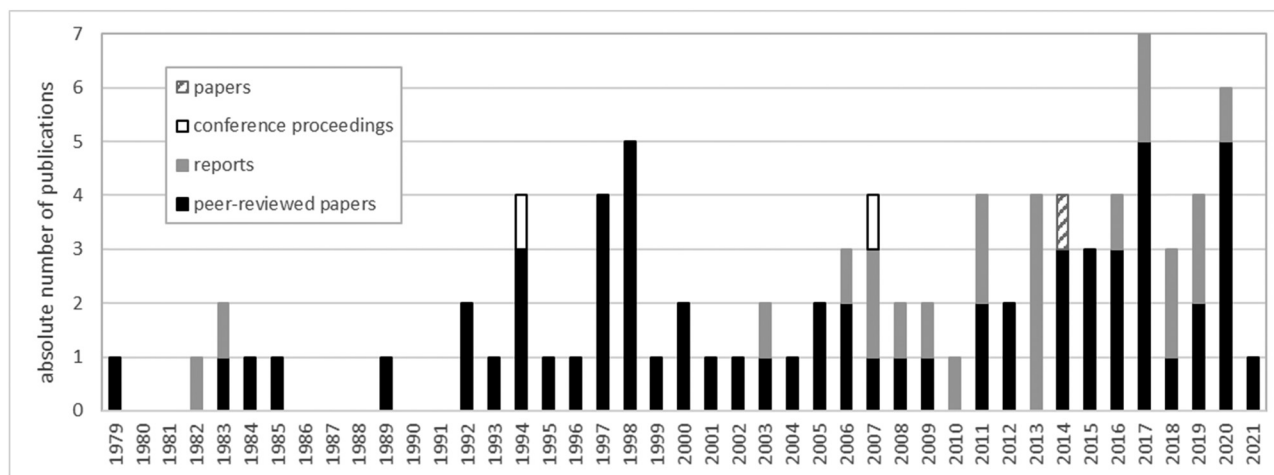


Figure 1: Number of publications per year and publication type, based on the review of refs [13–98].

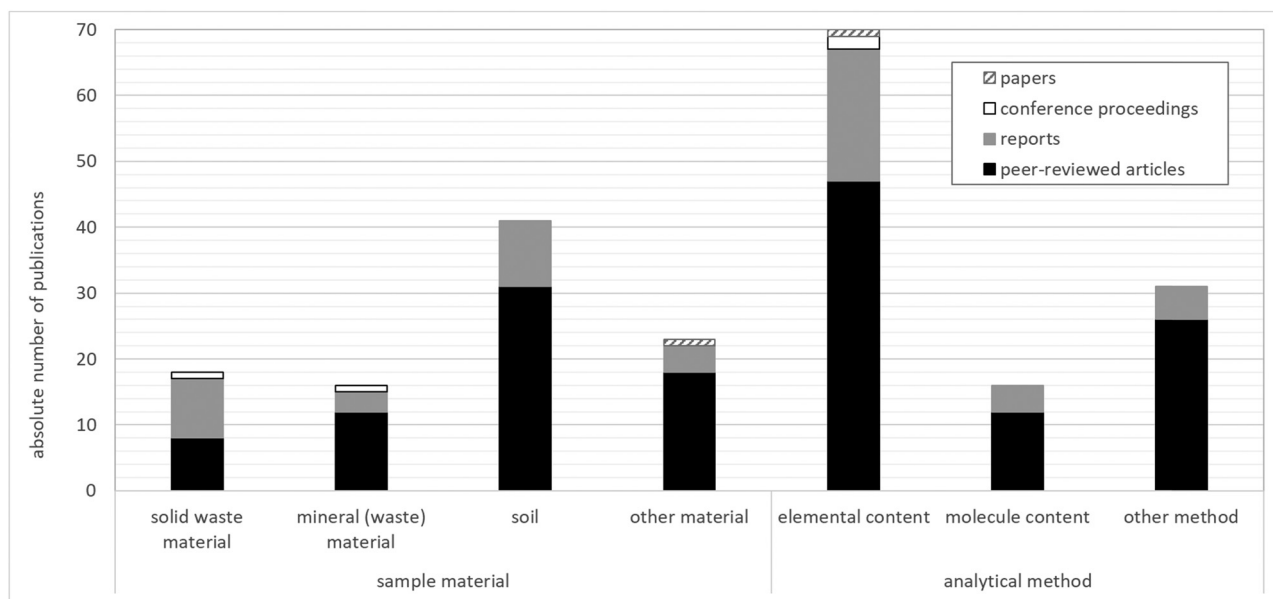


Figure 2: Number of identified sample materials (left) and analysis type per publication type (right), based on the review of refs [13–98].

on the results of the RRT; (iv) increasing confidence of customers in the lab competency, and (v) identifying interlaboratory differences. Purposes addressing method performance encompass: (i) validating analytical methods by testing effectiveness and comparability of specific measurement methods; (ii) testing and validating claims of uncertainty for specific measurements or methods; (iii) testing and evaluating performance characteristics of a specific method or measurement; and (iv) testing, evaluating, and supporting the equivalents of measurement methods.

The first research question focuses on the purpose of the RRT. Accordingly, the extraction template was designed to account for different RRT purposes, based on DIN EN ISO/IEC 17043 [10]. The pre-defined RRT purposes are

- lab (performance) assessment
- method validation
- method (performance) assessment
- method development
- (certified) reference material testing and development
- establishment of assigned values based on a reference method
- assignment of a consensus value based on the RRT lab results
- material characterization
- the investigation of a new/unknown sample

Figure 3a shows the result of the assessment of RRT purpose in all four publication types. Overall, the main two goals of an RRT are the assessment of an analytical method (applied 46 times) and the assessment of lab

performance (applied 34 times). Due to the fact that some publications address more than one purpose, the numbers will not add up to the number of publications in each publication type category.

Focussing on the two main publication types, PRA and R, there is a clear distinction with regard to purpose. Reports mainly present RRT with the goal to assess the lab performance (20) or the method performance (3). Method validation and assignment of a consensus value was each found once, the other purposes could not be identified in the collected reports. PRA, on the other hand, show a wider variety in purposes applied in the following order: method assessment (40), lab assessment (14), method validation (10), (C)RM testing (10), establish assigned value(s) (3), material characterization (3), method development (1), assign consensus value (1), and new/unknown sample (1). The two CP identified applied method assessment (2) and method development (1); the not-peer-reviewed paper addressed method assessment (1).

This shows that RRT for external QA/QC (published in reports) are mostly used to test the performance of either a lab or a specific method. The application of RRT in research is less focused, i.e., the purpose of the RRT complies with the research goal/question of the respective study.

3.3 PP used in RRT

In CA, PP are used in QA/QC to assess and assure a high data quality of the analysis results. In general, there are two

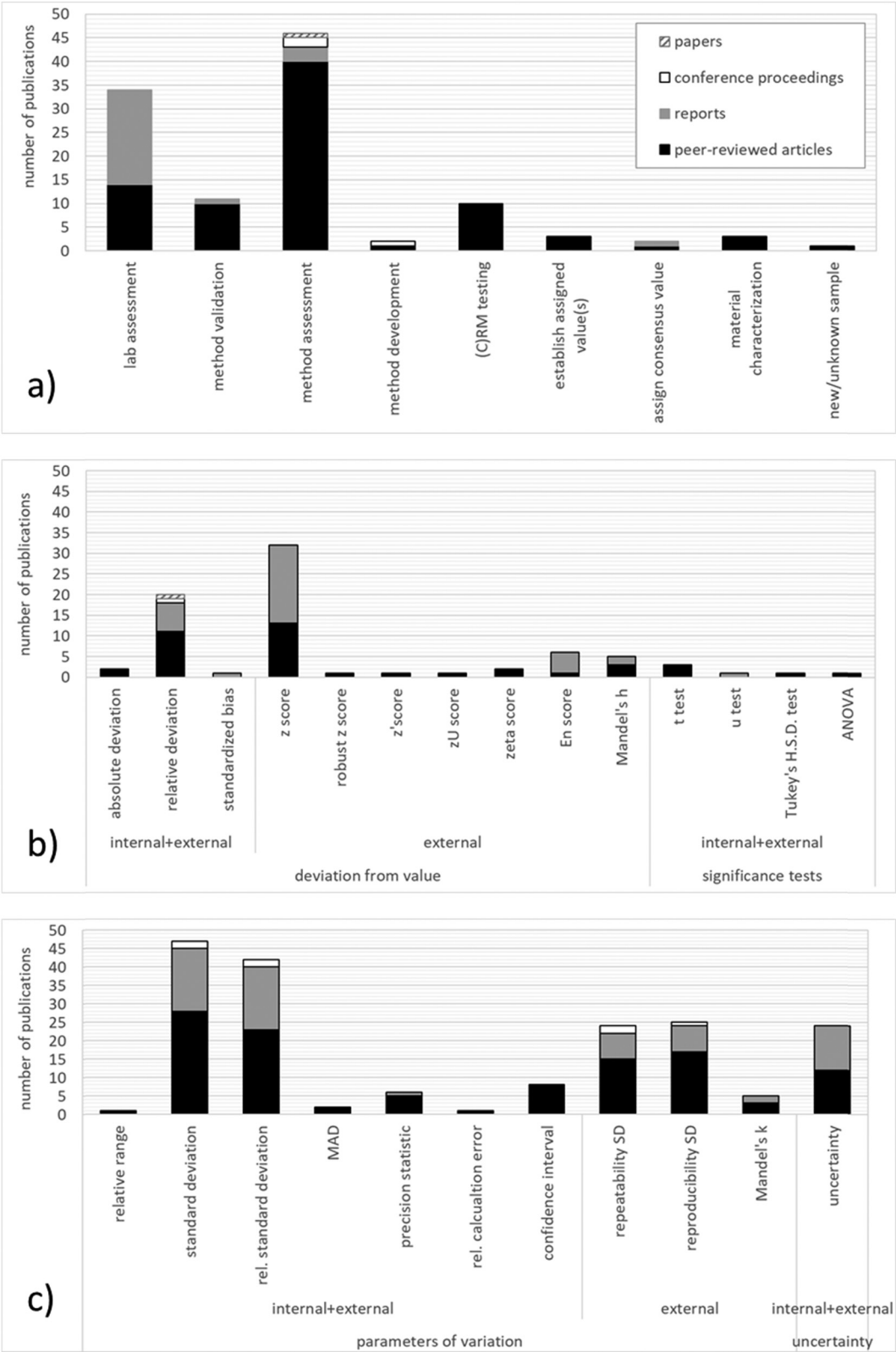


Figure 3: Overview of extracted information from all four publication types for (a) applied purposes per publication type, (b) use of TPP in all publications, and (c) use of PPP in all publications. The legend for the publication types is in the top right corner. Based on the review of refs [13–98].

main categories of PP: (i) parameters assessing the trueness, expressed as bias, i.e., the deviation of the lab result from any kind of reference value; and (ii) parameters assessing the precision, i.e., the variation between comparable measurements (e.g., replicates in one lab or comparable analysis results between labs). These PP can be applied in internal quality control within one lab, or in external quality control in an RRT. Annex B in DIN EN ISO/IEC 17043 [10] gives an overview over the most important and most commonly used statistical parameters to evaluate RRT lab results and is accompanied by DIN EN ISO/IEC 13528 “Statistical methods for use in proficiency testing by interlaboratory comparisons” [99]. These norms specify the determination of each parameter, without giving recommendations for their application and can hence not directly be used as, e.g., a standard operating procedure. An overview of all involved standards and norms is given in ref. [11].

The second research question addresses the use and application of PP in RRT of environmental samples. After extracting the information regarding PP, we were able to further categorize the parameters. Trueness performance parameters (TPP) can be classified into parameters indicating the deviation from a specified value (e.g., reference value from a CRM, assigned value from an RRT) as well as significance tests. Precision performance parameters (PPP) can be classified in parameters of variation (describing the spread of the data) and uncertainty parameters (i.e., parameters characterizing “the dispersion of the values that could reasonably be attributed to the measurand” [100]). Table 2 summarizes all PP collected in the extraction step, classified in their respective categories and with an indication, if they can be used solely for external QA/QC or for internal and external QA/QC. A detailed description of all PP is given in ref. [11].

A unifying, consistent aspect of all TPP is the difference D between the analysis result of the test lab $x_{k,j}$ (the index k denotes the element, j denotes the lab) and a reference or assigned value X_k (Eq. 1):

$$D = (x_{k,j} - X_k) \quad (1)$$

This is complemented by different denominators, i.e., D is expressed in different “units”. TPP require an acceptance level against which the calculated parameter is checked. For example, for z scores (standardized PP and the most common TPP in RRT [4]), lab results with a z score between -2 and 2 are acceptable, $2 < |z| < 3$ functions as a warning signal, and lab results with $|z| \geq 3$ are considered unacceptable [99].

PPP are more divers in their implementation. However, they all indicate some version of variability (i.e., spread) of the data set under investigation.

Table 2: Categorization of PP

Parameter category	Parameter type	Internal quality control	External quality control
Trueness	Deviation from reference/assigned value	Absolute deviation ^a Relative deviation ^a (Standardized) bias ^b z score ^a Robust z score ^c z' score ^a z_U score ^d Zeta score ^a E_n score ^a Mandel's h ^e t test ^f u test ^g Tukey's HSD test ^h ANOVA ^f	
	Significance test		
Precision	Parameters of variation	Range ^f Relative range SD ^f RSD ^f MAD ⁱ	SD for RRT ^a Mandel's k ^e Repeatability SD ^a Reproducibility SD ^a
	Uncertainty	Precision ^{g,i,k,l,m,n} Confidence interval ^f (Expanded, combined) uncertainty ^o	

References for the determination of the parameters are a – [99], b – [75], c – [55], d – [103], e – [104], f – [101], g – [81], h – [105], i – [102], j – [13], k – [28], l – [37], m – [53], n – [64], and o – [100].

The results for TPP and PPP are depicted in Figure 3b and c, respectively.

3.3.1 TPP

The TPP identified in this review are absolute deviation, relative deviation, (standardized) bias, z score, robust z score, z' score, z_U score, zeta score, E_n score, Mandel's h , t test, u test, Tukey's HSD test, and analysis of variance (ANOVA). Absolute and relative deviation describe D in absolute and relative numbers, respectively; the (standardized) bias sets D in relation to a pooled standard deviation (SD) and the number of labs and replicates; the z score expresses D in the unit of an SD; robust z , z' , z_U , zeta, E_n scores as well as Mandel's h are variations of the z score with varying degrees of information included in the denominator; t , u , Tukey's HSD, and ANOVA test are statistical significance tests testing the lab results against a reference value or their respective variances. Calculations for all TPP can be found in ref. [11].

Although there is a variety of different TPP, the two main parameters found are z score (applied 32 times, $n_{\text{PRA}} = 13$, $n_{\text{P}} = 19$) and relative deviation RD (applied 20 times, $n_{\text{PRA}} = 11$, $n_{\text{P}} = 7$, $n_{\text{CP}} = 1$, $n_{\text{P}} = 1$). E_n score was found six times in total, followed by Mandel's h statistic (5) and t statistic (3). With the exception of absolute deviation and zeta score (both applied twice) all other TPP were found only once in the eligible literature sources.

Notably, z score is indeed the most used and hence best known TPP in RRT. The other identified external TPP assessing the deviation of a value are progressively complex versions of this z score. They all describe the absolute deviation D of the measurement results from a pre-defined value (reference, assigned, consensus) and express it in different parameters of variation (for RRT so-called “target range” [10]), e.g., different types of SD or uncertainties. This affects the magnitude of the resulting score, i.e., the smaller the denominator (SD or uncertainty), the larger the score and hence the more difficult it is to comply with a fixed acceptance level. Simply put, the more variation we allow our lab results, the easier they can be assumed acceptable.

In contrast, the relative deviation of a lab result from a pre-defined value, expressed in percent of the pre-defined value, does not include the spread of the data. Here, it is central which location parameter is used (mean, median, etc.), but not how large the variation of the data set is.

Significance tests go even a step further. Here, D is set in relation not only to the SD of lab results and comparison value, but also to the number of replicate measurements (which build the used mean value). Additionally, they consider the accepted probability α of a type I error (i.e., the probability of a false positive) [101]. A type I error describes the situation in which a value is erroneously identified as significantly different from the reference value [101]. This designation as significantly different requires an acceptance criterion for the TPP, i.e., D or any other TPP by itself gives no indication how large and/or significant the deviation is. But a TPP exceeding a pre-defined acceptance criterion is consequently designated as significantly different from the reference value. For the z score, the acceptance criteria are given above [99], for t in a t test the acceptance criterion is set with α [101].

These results suggest that despite of the wide variety of available TPP, mainly the simplest, most comprehensible parameters are used, which increases the chance for correct application and interpretation of the parameter, i.e., a very precise yet complicated parameter (such as z' , zeta, or E_n score) might give more insight, but could potentially cloud this very same insight due to its complexity.

3.3.2 PPP

The PPP identified in this review are range, relative range, SD, relative standard deviation (RSD), median absolute deviation (MAD), SD for RRT, Mandel's k , repeatability SD, reproducibility SD, precision, confidence interval, and (expanded, combined) uncertainty. Range and relative range give the spread between the minimum and the maximum value; SD and RSD are descriptive statistics giving a mean deviation between all values and the mean value; MAD is the robust analog of the SD and gives the median deviation of the difference between all values and the median; SD for RRT is not universally defined but is the so-called “target range” set in an RRT by the organizing institution; Mandel's k is the relation of SD in one lab against the SD between all labs; repeatability SD is the mean SD of all labs; reproducibility SD combines the repeatability SD and the between-lab SD; the precision is not consistently defined and mainly describes a version of D ; confidence interval is here defined as the range between the 2.5% and 97.5% quantile; and (expanded, combined) uncertainty is a parameter describing the measurement uncertainty of a method. Calculations for all PPP can be found in ref. [11].

For PPP, the review result is not as clear as for TPP. The most frequently used PPP are SD (found 47 times, $n_{\text{PRA}} = 28$, $n_{\text{R}} = 17$, $n_{\text{CP}} = 2$) and RSD (found 42 times, $n_{\text{PRA}} = 23$, $n_{\text{R}} = 17$, $n_{\text{CP}} = 2$), both of which can be used for internal QA/QC as well as in RRT. SD and RSD are very well-known descriptive statistics describing the variation of a data set; hence their ubiquitous use is expected. They also build the basis for further PPP, such as repeatability SD and consequently reproducibility SD, which are incidentally the next most frequent PPP found in this review. Reproducibility SD was identified 25 times ($n_{\text{PRA}} = 17$, $n_{\text{R}} = 7$, $n_{\text{CP}} = 1$), repeatability SD 24 times ($n_{\text{PRA}} = 15$, $n_{\text{R}} = 7$, $n_{\text{CP}} = 2$). The repeatability SD describes the mean SD over all labs for a certain analysis result and is an indicator for the precision of a method, reproducibility SD expresses the variation based on the repeatability SD and the variation between laboratories. Both are parameters used in the evaluation of RRT. Another relevant PPP is the (combined or expanded) uncertainty [100], which was identified 24 times ($n_{\text{PRA}} = 12$, $n_{\text{R}} = 12$). These PPP are followed by the confidence interval (found 8 times, only in PRA), precision statistic (found 6 times, $n_{\text{PRA}} = 5$, $n_{\text{R}} = 1$), and Mandel's k statistic ($n_{\text{PRA}} = 3$, $n_{\text{R}} = 2$). However, the number of identified precision statistics has to be regarded with caution, because its determination is not consistent among the publications. The MAD [11,102] was found twice in PRA, all remaining PPP were identified once in this review.

Evidently, the main PPP all base on the concept of the SD (i.e., the shape parameter for the normal distribution

[101]), which is also an indication that the simplicity of a PP is important for its correct use. However, in case of not normally distributed data all PPP basing on the SD are biased.

3.3.3 PP in relation to purpose

Figure 4 depicts the relation between the RRT purposes and the most frequent (a) TPP (z score, RD, E_n score,

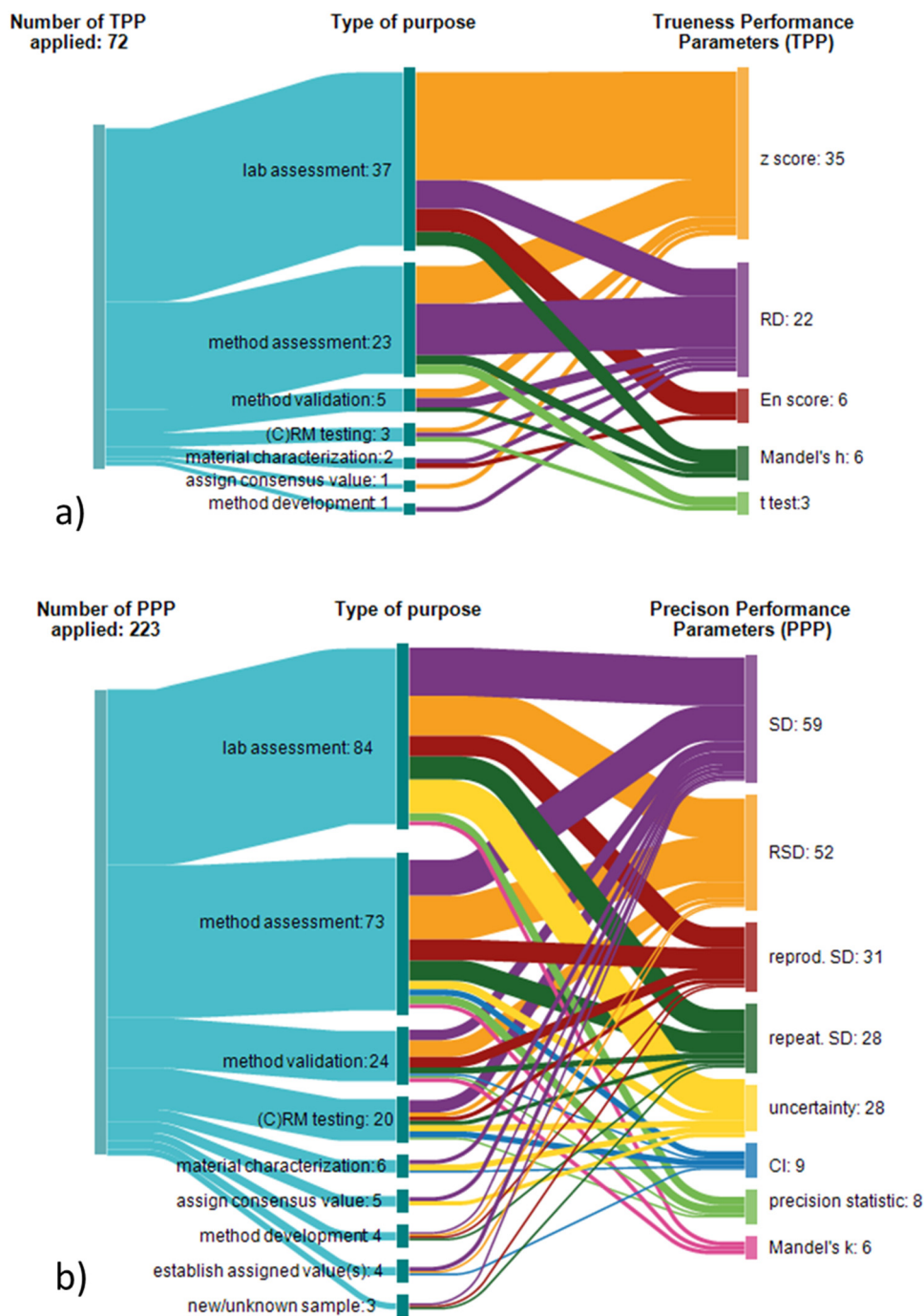


Figure 4: Relationship between (a) RRT purpose and TPP and (b) RRT purpose and PPP. The data in the figures are build on the review of refs [13–98].

Mandel's h statistic, and t test) as well as (b) PPP (SD, RSD, reproducibility SD, repeatability SD, uncertainty, precision statistic, Mandel's k statistic, and confidence interval). To do this, we used a Sankey diagram, which is commonly used in material flow analyses to depict the magnitude of flows between processes as a function of the width of the arrow [106]. Figure 4 shows the number of PP used in relation to the respective type of purpose for which it was used. The color of the bars linking type of purpose and PP highlights the specific PP for more clarity, the width of the bars shows the frequency/number of uses of a certain PP for a certain type of purpose. The proportions of the purposes and the PP among each other may vary in comparison to Figure 3a–c, because not all parameters are included. In addition, some publications follow more than one purpose, and all publications use more than one PP.

There is no distinct correlation visible between a PP and a purpose, neither for TPP nor for PPP, with the exception of the application of z scores for the assessment of lab performance and, to a lesser extent, RD for the assessment of method performance. These correlations are owed to the fact that the z score is the main TPP for assessing the lab performance, used as a standard PP by institutions organizing RRT for external QA/QC, and RD is the simplest TPP to assess the trueness of a method.

Overall, fewer TPP are used for each purpose than PPP (see absolute numbers of uses in Figure 4a and b). For “establish assigned value(s)” and “new/unknown sample” no TPP were used at all, solely descriptive statistics, and PPP [11]. This coincides with the description in Section 3.2 “Purpose of RRT” that the review result is not as clear for PPP as it is for TPP. The PPP identified in this review show a broader range of information content, i.e., the presented PPP differ more between each other than the presented TPP. Where all TPP base on the principle of the absolute deviation D of the lab result from a pre-defined value, the PPP describe alternative versions of the data variability (i.e., mainly of the SD).

4 Conclusions and outlook

RRT, in general for all sample materials as well as in particular for waste materials and SRM, are a common instrument for external QA/QC and are frequently executed by organizing institutions, but not necessarily published. Within external QA/QC, they mainly follow the purpose to assess either lab performance or method performance. In research, RRT are applied in a much wider

range and hence pursue a variety of purposes, depending on the underlying research goal. This means that RRT can be used as a bespoke method to potentially identify and eliminate analytical effects, by defining (i) a research goal; (ii) the evaluation of the RRT, including the selection of PP; and (iii) the data requirements to execute the evaluation. Consequently, RRT is a versatile tool to advance trueness and precision of analytical procedures and hence data quality for waste and SRM. Standardization and harmonization of QA/QC of CA of SRM will widely increase confidence in and acceptance of SRM, which ultimately supports and improves a circular economy.

The results of this systematic review suggest that PP for an RRT of waste materials and SRM are selected based on the criterion of simplicity. Although there are more complex PP, which can include information about, e.g., measurement uncertainty, mainly simpler PP are used to assess the lab data. The easier to use and to interpret a PP, the more likely it is to have a useful effect. Hence, established PP such as z score and RD as well as suitable versions of SD can give useful information on analytical effects for unknown, complex waste materials, and SRM. However, lab managers and assistants as well as customers and other users of lab data are well advised to look more closely on PP and their informative value to get a better insight into the accuracy (trueness as well as precision) of the lab data. The findings of this study may contribute to the use of more science-based assessment criteria.

The two conclusions – (1) use of RRT as a bespoke method for dealing with unknown, complex samples and (2) simplicity as an assumed criterion for PP selection – have to and will be further investigated. Based on previous studies [7,8], we conducted a Europe-wide RRT for three complex solid sample materials, battery ash, mineral waste, and printed circuit boards (as used in the referenced studies). The results from this RRT and its evaluation will be published in future articles.

Acknowledgements: The authors would like to thank Matthias Rösslein of Empa, Particles-Biology Interactions Laboratory for the many fruitful discussions about RRT and their evaluation.

Funding information: The authors state no funding involved.

Author contributions: Nathalie Korf: conceptualization, methodology, data curation, investigation, formal analysis, visualization, writing – original draft, writing – review and editing; Paul Martin Mähltz: methodology, visualization, writing – review and editing; Vera Susanne Rotter:

supervision, resources, funding acquisition, writing – review and editing.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The raw data for this systematic review can be found in ref. [11].

References

- [1] European Commission. Study on the EU's list of Critical Raw Materials: Final report; 2020. https://ec.europa.eu/growth/sectors/raw-materials/specific-interest/critical_en.
- [2] European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A new Circular Economy Action Plan; 2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN>.
- [3] Witt M, Hoogzaad J, von Daniels C. The circularity gap report 2020; 2020. <https://www.circularity-gap.world/2020>.
- [4] Prichard E, Barwick V. Quality assurance in analytical chemistry. 1st edn. Teddington, UK: Wiley; 2007.
- [5] Deutsches Institut für Normung e.V. DIN EN ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories; 2018.
- [6] Gy P. Sampling of discrete materials – a new introduction to the theory of sampling. *Chemometr Intell Lab.* 2004;74:7–24. doi: 10.1016/j.chemolab.2004.05.012.
- [7] Korf N, Løvik AN, Figi R, Schreiner C, Kuntz C, Mählietz PM, et al. Multi-element chemical analysis of printed circuit boards – challenges and pitfalls. *Waste Manage.* 2019;92:124–36. doi: 10.1016/j.wasman.2019.04.061.
- [8] Mählietz PM, Løvik AN, Figi R, Schreiner C, Kuntz C, Korf N, et al. Characterizing the urban mine - challenges of simplified chemical analysis of anthropogenic mineral residues. *Resources.* 2019;8(132):1–21. doi: 10.3390/resources8030132.
- [9] Løvik AN, Figi R, Schreiner C, Rösslein M, Widmer R, Bunge R, et al. Variability and bias in measurements of metals mass fractions in automobile shredder residue. *Recycling.* 2019;4(34):1–19. doi: 10.3390/recycling4030034.
- [10] Deutsches Institut für Normung e.V. DIN EN ISO/IEC 17043 Conformity assessment - General requirements for proficiency testing; 2010.
- [11] Korf N, Mählietz PM, Rotter VS. Supporting information for 'Round robin test of secondary raw materials: a systematic review of performance parameters'. DepositOnce, repository of TU Berlin. Berlin, Germany: Technische Universität Berlin; 2022. doi: 10.14279/depositonce-15189.
- [12] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6(7):1–28. doi: 10.1371/journal.pmed.1000100.
- [13] Dreesen DR, Gladney ES, Owens JW. Interlaboratory comparison of arsenic, molybdenum and selenium analyses from uranium mill tailings. *J Water Pollut Control Fed.* 1979;51:2447–56.
- [14] Domalski ES, Abramowitz S. Evaluation of data on calorific content values determined during ASTM round robin testing of RDF-3. *Resour Conserv Recy.* 1983;8(3):233–52. doi: 10.1016/0166-3097(83)90027-5.
- [15] Nair PS, Logan TJ, Sharpley AN, Sommers LE, Tabatabai MA, Yuan TL. Interlaboratory comparison of a standardized phosphorus adsorption procedure. *J Environ Qual.* 1984;13:591–5. doi: 10.2134/jeq1984.00472425001300040016x.
- [16] Steger HF, Bowman WS, McKeague JA. Additional certified values for CCRMP Soils SO-1 to -4. *Geostandard Newslett.* 1985;9:213–8. doi: 10.1111/j.1751-908X.1985.tb00451.x.
- [17] LaFleur LE, Dodo GH. An interlaboratory comparison of analytical procedures for the measurement of PCDDs/PCDFs in pulp and paper industry solid wastes. *Chemosphere.* 1989;18(1–6):77–84. doi: 10.1016/0045-6535(89)90107-0.
- [18] Blankenhorn I, Meijer D, van Delft RJ. Inter-laboratory comparison of methods used for analysing polycyclic aromatic hydrocarbons (PAHs) in soil samples. *Fresen J Anal Chem.* 1992;343:497–504. doi: 10.1007/BF00322157.
- [19] Desaulles A, Lischer P, Dahinden R, Bachmann HJ. Comparability of chemical analysis of heavy metals and fluorine in soils: results of an interlaboratory study. *Commun Soil Sci Plan.* 1992;23:363–77. doi: 10.1080/00103629209368595.
- [20] Crosland AR, McGrath SP, Lane PW. An interlaboratory comparison of a standardised EDTA extraction procedure for the analysis of available trace elements in two quality control soils. *Int J Environ An Ch.* 1993;51:153–60. doi: 10.1080/03067319308027620.
- [21] Kimbrough DE, Wakakuwa J. Interlaboratory comparison of instruments used for the determination of elements in acid digestate of solids. *Analyst.* 1994;119:383–8. doi: 10.1039/AN9941900383.
- [22] de Groot GJ, Hoede D. Validation of Dutch standard leaching tests using NEN-ISO 5725. *Stud Environ Sci.* 1994;60:305–14. doi: 10.1016/S0166-1116(08)71467-5.
- [23] Quevauviller P, Rauret G, Muntau H, Ure AM, Rubio R, López-Sánchez JF, et al. Evaluation of a sequential extraction procedure for the determination of extractable trace metal contents in sediments. *Fresen J Anal Chem.* 1994;349:808–14. doi: 10.1007/BF00323110.
- [24] Argyraki A, Ramsey MH, Thompson M. Proficiency testing in sampling: pilot study on contaminated land. *Analyst.* 1995;120:2799–803. doi: 10.1039/AN9952002799.
- [25] Quevauviller P, Lachica M, Barahona E, Rauret G, Ure A, Gomez A, et al. Interlaboratory comparison of EDTA and DTPA procedures prior to certification of extractable trace elements in calcareous soil. *Sci Total Environ.* 1996;178:127–32. doi: 10.1016/0048-9697(95)04804-9.
- [26] Bögershausen W, Ciccirelli R, Gercken B, König E, Krivan V, Müller-Käfer R, et al. Pure graphite as a reference material for the determination of trace elements – an interlaboratory collaborative study. *Fresen J Anal Chem.* 1997;357:266–73. doi: 10.1007/s002160050151.
- [27] Beck T, Joergensen RG, Kandeler E, Makeschin F, Nuss E, Oberholzer HR, et al. An inter-laboratory comparison of ten

- different ways of measuring soil microbial biomass C. *Soil Biol Biochem.* 1997;29:1023–32. doi: 10.1016/S0038-0717(97)00030-8.
- [28] Kučera J, Sychra V, Horáková J, Soukal L. Use of INAA in the preparation of a set of soil reference materials with certified values of total element contents. *J Radioanal Nucl Ch.* 1997;215:147–55. doi: 10.1007/BF02109891.
- [29] Cook JM, Gardner MJ, Griffiths AH, Jessep MA, Ravenscroft JE, Yates R. The comparability of sample digestion techniques for the determination of metals in sediments. *Mar Pollut Bull.* 1997;34:637–44. doi: 10.1016/S0025-326X(96)00186-5.
- [30] Hinnners TA, Hughes R, Outridge PM, Davis WJ, Simon K, Woolard DR. Interlaboratory comparison of mass spectrometric methods for lead isotopes and trace elements in NIST SRM 1400 Bone Ash. *J Anal Atom Spectrom.* 1998;13:963–70. doi: 10.1039/A803373K.
- [31] Maaskant JFN, Boekholt AH, Jenks PJ, Rucinski RD. An international interlaboratory study for the production of a sewage sludge certified reference material for routine use in inorganic quality control. *Anal Chem.* 1998;360:406–9. doi: 10.1007/s002160050723.
- [32] Kučera J, Sychra V, Koubek J. A set of four soil reference materials with certified values of total element contents and their extractable fractions. *Fresen J Anal Chem.* 1998;360:402–5. doi: 10.1007/s002160050722.
- [33] Quevauviller P. Operationally defined extraction procedures for soil and sediment analysis I. Standardization. *Trend Anal Chem.* 1998;17(5):289–98. doi: 10.1016/S0165-9936(97)00119-2.
- [34] Nejedlý Z, Campbell JL, Teesdale WJ, Dlouhy JF, Dann TF, Hoff RM, et al. Inter-laboratory comparison of air particulate monitoring data. *J Air Waste Manage.* 1998;48:386–97. doi: 10.1080/10473289.1998.10463698.
- [35] Butler OT, Howe AM. Development of an international standard for the determination of metals and metalloids in workplace air using ICP-AES: evaluation of sample dissolution procedures through an interlaboratory trial. *J Environ Monitor.* 1999;1:23–32. doi: 10.1039/A807526C.
- [36] Ottner F, Gier S, Kuderna M, Schwaighofer B. Results of an inter-laboratory comparison of methods for quantitative clay analysis. *Appl Clay Sci.* 2000;17(5–6):223–43. doi: 10.1016/S0169-1317(00)00015-6.
- [37] Balaram V. Assessment of the ICP-MS method using the interlaboratory QA study of two Polish soil RMs. *Accredit Qual Assur.* 2000;5:325–30. doi: 10.1007/s007690000163.
- [38] Kleinman PJA, Sharpley AN, Gartley K, Jarrell WM, Kuo S, Menon RG, et al. Interlaboratory comparison of soil phosphorus extracted by various soil test methods. *Commun Soil Sci Plan.* 2001;32:2325–45. doi: 10.1081/CSS-120000376.
- [39] Becker R, Koch M, Wachholz S, Win T. Quantification of total petrol hydrocarbons (TPH) in soil by IR-spectrometry and gas chromatography - conclusions from three proficiency testing rounds. *Accredit Qual Assur.* 2002;7:286–9. doi: 10.1007/s00769-002-0476-9.
- [40] Hall GEM, Oates CJ. Performance of commercial laboratories in analysis of geochemical samples for gold and the platinum group elements. *Geochem-Explor Env A.* 2003;3:107–20. doi: 10.1144/1467-787303-006.
- [41] Cools N, Delanote V, Scheldeman X, Quataert P, de Vos B, Roskams P. Quality assurance and quality control in forest soil analyses: a comparison between European soil laboratories. *Accredit Qual Assur.* 2004;9:688–94. doi: 10.1007/s00769-004-0856-4.
- [42] Tahir SS, Naseem R, ul Haq A, Saeed K. An inter-laboratory comparison study for the determination of copper and lead from the wastewater of printed circuit board manufacturing industry in Pakistan. *Accredit Qual Assur.* 2005;10:362–88. doi: 10.1007/s00769-005-0022-7.
- [43] de Kreijl C, Wever G. Proficiency testing of growing media, soil improvers, soils, and nutrient solutions. *Commun Soil Sci Plan.* 2005;36:81–8. doi: 10.1081/CSS-200042971.
- [44] Davidson CM, Urquhart GJ, Ajmone-Marsan F, Biasioli M, Da Costa Duarte A, Díaz-Barrientos E, et al. Fractionation of potentially toxic elements in urban soils from five European cities by means of a harmonised sequential extraction procedure. *Anal Chim Acta.* 2006;565:63–72. doi: 10.1016/j.aca.2006.02.014.
- [45] Tirez K, Scharf H, Calzolari D, Cleven R, Kisser M, Lück D. Validation of a European standard for the determination of hexavalent chromium in solid material. *J Environ Monitor.* 2007;9:749–59. doi: 10.1039/b706724k.
- [46] Kalbe U, Berger W, Eckardt J, Simon F-G. Evaluation of leaching and extraction procedures for soil and waste. *Waste Manage.* 2008;28:1027–38. doi: 10.1016/j.wasman.2007.03.008.
- [47] Balzamo S, de Zorzi P, Barbizzi S, Calabretta E, Potalivo M, Rosamilia S. Measurement of PAHs in environmental matrices: results from an interlaboratory comparison on the different steps of the measurement procedure. *Accredit Qual Assur.* 2009;14:487–95. doi: 10.1007/s00769-009-0576-x.
- [48] Gerboles M, Buzica D, Brown R, Yardley RE, Hanus-Illyar A, Salfinger M, et al. Interlaboratory comparison exercise for the determination of As, Cd, Ni and Pb in PM10 in Europe. *Atmos Environ.* 2011;45:3488–99. doi: 10.1016/j.atmosenv.2010.12.020.
- [49] Wragg J, Cave M, Basta N, Brandon E, Casteel S, Denys S, et al. An inter-laboratory trial of the unified BARGE bioaccessibility method for arsenic, cadmium and lead in soil. *Sci Total Environ.* 2011;409:4016–30. doi: 10.1016/j.scitotenv.2011.05.019.
- [50] Ikonomou MG, Kelly BC, Blair JD, Gobas FAPC. An interlaboratory comparison study for the determination of dialkyl phthalate esters in environmental and biological samples. *Environ Toxicol Chem.* 2012;31:1948–56. doi: 10.1002/etc.1912.
- [51] Buczek U, Kuchenbuch RO, Übelhör W, Nätscher L. Assessment of sampling and analytical uncertainty of trace element contents in arable field soils. *Environ Monit Assess.* 2012;184:4517–38. doi: 10.1007/s10661-011-2282-5.
- [52] Bürger S, Boulyga SF, Peñkin MV, Bostick D, Jovanovic S, Lindvall R, et al. Quantifying multiple trace elements in uranium ore concentrates: an interlaboratory comparison. *J Radioanal Nucl Ch.* 2014;301:711–29. doi: 10.1007/s10967-014-3224-9.
- [53] Harrington JM, Nelson CM, Weber FX, Bradham KD, Levine KE, Rice J. Evaluation of methods for analysis of lead in air particulates: an intra-laboratory and inter-laboratory comparison. *Environ Sci-Proc Imp.* 2014;16:256–61. doi: 10.1039/c3em00486d.

- [54] Vittori Antisari L, Bianchini G, Dinelli E, Falsone G, Gardini A, Simoni A, et al. Critical evaluation of an intercalibration project focused on the definition of new multi-element soil reference materials (AMS-MO1 and AMS-ML1). *J Environ Qual.* 2014;15:41–64. doi: 10.6092/ISSN.2281-4485/4553.
- [55] Da Silva Dias R, Aparecida de Abreu C, Ferreira de Abreu M, Paz González A. Statistical methods for evaluating results from soil micronutrient analyses in interlaboratory programs. *Commun Soil Sci Plan.* 2015;46:57–71. doi: 10.1080/00103624.2014.988091.
- [56] Nudi AH, Wagener A, Gabardo IT, Lourenço RA, Scofield A. Interlaboratory comparison of quantifying hydrocarbons and trace elements in sediment samples from a tropical estuary. *J Brazil Chem Soc.* 2015;26:297–309. doi: 10.5935/0103-5053.20140280.
- [57] Reis AT, Duarte AC, Henriques B, Coelho C, Lopes CB, Mieirol CL, et al. An international proficiency test as a tool to evaluate mercury determination in environmental matrices. *Trends Analyt Chem.* 2015;64:136–48. doi: 10.1016/j.trac.2014.08.015.
- [58] Bachmann HJ, Bucheli TD, Dieguez-Alonso A, Fabbri D, Knicker H, Schmidt H-P, et al. Toward the standardization of biochar analysis: the COST action TD1107 interlaboratory comparison. *J Agric Food Chem.* 2016;64:513–27. doi: 10.1021/acs.jafc.5b05055.
- [59] Geurts R, Spooren J, Quaghebeur M, Broos K, Kenis C, Debaene L. Round robin testing of a percolation column leaching procedure. *Waste Manage.* 2016;55:31–7. doi: 10.1016/j.wasman.2016.06.010.
- [60] Yatkin S, Belis CA, Gerboles M, Calzolari G, Lucarelli F, Cavalli F, et al. An interlaboratory comparison study on the measurement of elements in PM₁₀. *Atmos Environ.* 2016;125:61–8. doi: 10.1016/j.atmosenv.2015.10.084.
- [61] Gartiser S, Heisterkamp I, Schoknecht U, Burkhardt M, Ratte M, Ilvonen O, et al. Results from a round robin test for the ecotoxicological evaluation of construction products using two leaching tests and an aquatic test battery. *Chemosphere.* 2017;175:138–46. doi: 10.1016/j.chemosphere.2017.01.146.
- [62] Santoro A, Held A, Linsinger TPJ, Perez A, Ricci M. Comparison of total and aqua regia extractability of heavy metals in sewage sludge: the case study of a certified reference material. *Trends Analyt Chem.* 2017;89:34–40. doi: 10.1016/j.trac.2017.01.010.
- [63] Jagustyn B, Plis A, Mastalerz M, Hrabak J, Ściężko M. Investigation of homogeneity and stability of items for proficiency testing of solid recovered fuels (SRF) analysis. *Accredit Qual Assur.* 2017;22:355–60. doi: 10.1007/s00769-017-1283-7.
- [64] Qiao J, Salminen-Paatero S, Rondahl S, Bourgeois-Goget M, Roos P, Lagerkvist P, et al. Inter-laboratory exercise with an aim to compare methods for 90Sr and 239,240Pu determination in environmental soil samples. *J Radioanal Nucl Ch.* 2017;314:813–26. doi: 10.1007/s10967-017-5385-9.
- [65] Raven MD, Self PG. Outcomes of 12 years of the Reynolds cup quantitative mineral analysis round robin. *Clay Clay Miner.* 2017;65:122–34. doi: 10.1346/CCMN.2017.064054.
- [66] Qiao J, Lagerkvist P, Rodushkin I, Salminen-Paatero S, Roos P, Lierhagen S, et al. On the application of ICP-MS techniques for measuring uranium and plutonium: a Nordic inter-laboratory comparison exercise. *J Radioanal Nucl Ch.* 2018;315:565–80. doi: 10.1007/s10967-018-5697-4.
- [67] Isobe A, Buenaventura NT, Chastain S, Chavanich S, Cózar A, DeLorenzo M, et al. An interlaboratory comparison exercise for the determination of microplastics in standard sample bottles. *Mar Pollut Bull.* 2019;146:831–7. doi: 10.1016/j.marpolbul.2019.07.033.
- [68] Kalbe U, Lehnik-Habrink P, Bandow N, Sauer A. Validation of European horizontal methods for the analysis of PAH, PCB and dioxins in sludge, treated biowaste and soil. *Environ Sci Eur.* 2019;31(29):1–10. doi: 10.1186/s12302-019-0211-3.
- [69] Anca-Couce A, Tsekos C, Retschitzegger S, Zimbardi F, Funke A, Banks S, et al. Biomass pyrolysis TGA assessment with an international round robin. *Fuel.* 2020;276:118002. doi: 10.1016/j.fuel.2020.118002.
- [70] Hafner SD, Fruteau de Lacroix H, Koch K, Holliger C. Improving inter-laboratory reproducibility in measurement of biochemical methane potential (BMP). *Water.* 2020;12:1–30. doi: 10.3390/w12061752.
- [71] Pellikka T, Kajolinna T. Emission measurements of heavy metals with the European standard reference methods EN 14385 and EN 13211 - observations from an interlaboratory comparison (ILC) measurements performed at waste-to-energy plant in Finland. *J Air Waste Manag Assoc.* 2020;70(10):990–1008. doi: 10.1080/10962247.2020.1797926.
- [72] Yatkin S, Trzepla K, Hyslop NP, White WH, Butler O, Ancelet T, et al. Comparison of a priori and interlaboratory-measurement-consensus approaches for value assignment of multi-element reference materials on PTFE filters. *Microchem J.* 2020;158:105225. doi: 10.1016/j.microc.2020.105225.
- [73] Damastuti E, Santoso M, Yusuf S, Yatu WNSW. An interlaboratory comparison of INAA analytical method for coal fly ash elemental characterization. *J Phys Conf Ser.* 2020;1436:1–9. doi: 10.1088/1742-6596/1436/1/012138.
- [74] Delvigne C, Guihou A, Schuessler JA, Savage P, Poitrasson F, Fischer S, et al. Silicon isotope analyses of soil and plant reference materials: an inter-comparison of seven laboratories. *Geostand Geoanal Res.* 2021;45(3):525–38. doi: 10.1111/ggr.12378.
- [75] Johnston JW, Daniel JL. Summary report for the interlaboratory round robin on the MCC-1 static leach test method. United States: U.S. Department of Energy Office of Scientific and Technical Information; 1982. doi: 10.2172/5430287.
- [76] Kennedy PC, Roser BP, Hunt JL, Daly BK. A New Zealand interlaboratory comparison of analytical data for the CSSC reference soil: New Zealand Soil Bureau Scientific Report 59. 1983. p. 44. <http://digitallibrary.landcareresearch.co.nz/digital/collection/p20022coll2/id/161>.
- [77] Cools N, Delanote V, de Vos B, Quataert P, Roskams P, Scheldeman X. Quality Assurance and Quality Control in Forest Soil Analysis: 3rd FSCC Interlaboratory Comparison; 2003. <https://purews.inbo.be/ws/files/275633/184556.pdf>.
- [78] Ingham MN, Gowing CJB, Harrison HM. The validation of the determination of trace elements by energy dispersive polarised x-ray fluorescence spectrometry. Laboratory Operations Internal Report OR/07/012; 2007. <http://nora.nerc.ac.uk/id/eprint/7196>.
- [79] Mäkinen I, Vaajasaari K, Järvinen O, Sara-Aho T, Ivalo R, Tervonen K, et al. SYKE proficiency test 4/2018: leading

- testing of a solid waste - the one stage and the two stage batch leaching test; Reports of the Finnish Environment Institute; 2008. https://helda.helsinki.fi/bitstream/handle/10138/39823/SYKE_32_2008.pdf?sequence=1.
- [80] Leivuori M, Korhonen K, Järvinen O, Näykki T, Sara-Aho T, Tervonen K et al. SYKE proficiency test 4/2009: metals in waters and soils. Reports of the Finnish Environment Institute; 2009. https://helda.helsinki.fi/bitstream/handle/10138/39684/SYKE_28_2009.pdf?sequence=1.
- [81] Shakhshiro A, Toerwenyi A. Latin American regional proficiency test on the determination of trace elements and radionuclides in algae, soil and spiked water; 2010. https://nucleus.iaea.org/rpst/Documents/report_rla_pt.pdf.
- [82] Leivuori M, Korhonen K, Sara-Aho T, Näykki T, Järvinen O, Tervonen K et al. SYKE proficiency test 3/2010: metals in waters and sediment. Reports of the Finnish Environment Institute; 2011. https://helda.helsinki.fi/bitstream/handle/10138/39997/SYKE_1_2011.pdf?sequence=4.
- [83] Leivuori M, Korhonen-Ylönen K, Sara-Aho T, Näykki T, Tervonen K, Lanteri S et al. Proficiency test SYKE 3/2011: metals in water and sludge. Reports of the Finnish Environment Institute; 2011. https://helda.helsinki.fi/bitstream/handle/10138/39762/SYKE_22_2011.pdf?sequence=1.
- [84] Björkjöf K, Korhonen-Ylönen K, Kaasalainen M, Leivuori M, Väntsi S, Lanteri S. Proficiency Test SYKE 10/2012: Leaching testing of solid waste samples. Reports of the Finnish Environment Institute; 2013. <http://hdl.handle.net/10138/39609>.
- [85] Korhonen-Ylönen K, Nuutinen J, Leivuori M, Ilmakunnas M. Proficiency Test SYKE 8/2012: volatile organic compounds in water and soil. Reports of the Finnish Environment Institute; 2013. <https://helda.helsinki.fi/handle/10138/38618>.
- [86] Korhonen-Ylönen K, Nuutinen J, Leivuori M, Ilmakunnas M. Proficiency Test SYKE 9/2012: Oil hydrocarbons in water and soil. Reports of the Finnish Environment Institute; 2013. <http://hdl.handle.net/10138/38619>.
- [87] Leivuori M, Korhonen-Ylönen K, Sara-Aho T, Näykki T, Tervonen K, Lanteri S et al. SYKE Proficiency Test 5/2012: Metals in waters and soil. Reports of the Finnish Environment Institute; 2013. <http://hdl.handle.net/10138/41753>.
- [88] Koivikko R, Leivuori M, Näykki T, Sara-Aho T, Tervonen K, Lanteri S et al. Interlaboratory Proficiency Test 08/2015: Metals in waste water and sludge. Reports of the Finnish Environment Institute; 2016. <http://hdl.handle.net/10138/160817>.
- [89] Koivikko R, Leivuori M, Kaasalainen M, Tervonen K, Lanteri S, Ilmakunnas M. Interlaboratory Proficiency Test 12/2016: Leaching test for solid waste samples: two stage batch leaching test. Reports of the Finnish Environment Institute; 2017. <https://helda.helsinki.fi/handle/10138/181510>.
- [90] Leivuori M, Koivikko R, Sara-Aho T, Näykki T, Tervonen K, Lanteri S, et al. Interlaboratory Proficiency Test 10/2016: Metals in waste water and recycled material. Reports of the Finnish Environment Institute; 2017. <http://hdl.handle.net/10138/176910>.
- [91] Maunukela L, Pelkonen A, Björkjöf K, Ilmakunnas M, Kartio M, Leivuori M. Interlaboratory Comparison test 15/2018: Soil improver maturity test. Reports of the Finnish Environment Institute; 2018. https://www.ruokavirasto.fi/globalassets/laboratoriopalvelut/vertailulaboratoriotointa/julkaisuja-pdf/sykere_25_2018.pdf.
- [92] Koivikko R, Leivuori M, Kaasalainen M, Tervonen K, Lanteri S, Ilmakunnas M. Interlaboratory Proficiency Test 14/2017: Leaching behavior test for solid waste material: up-flow percolation test. Reports of the Finnish Environment Institute; 2018. <https://helda.helsinki.fi/handle/10138/236326>.
- [93] Koivikko R, Leivuori M, Kaasalainen M, Tervonen K, Lanteri S, Ilmakunnas M. Interlaboratory Proficiency Test 13/2018: leaching behavior test for solid waste material: one stage batch leaching test. Reports of the Finnish Environment Institute; 2019. <https://helda.helsinki.fi/handle/10138/301217>.
- [94] Oosterlaken-Buijs GA. Results of proficiency test total metals in polymers; 2019. <http://www.iisnl.com/pdf/iis19P06.pdf>.
- [95] Koivikko R, Leivuori M, Kaasalainen M, Tervonen K, Lanteri S, Ilmakunnas M. Interlaboratory Proficiency Test 04/2020: leaching behavior test for waste material: two stage batch leaching test; 2020. https://helda.helsinki.fi/bitstream/handle/10138/319540/SYKE_40_2020.pdf?sequence=1&isAllowed=y.
- [96] van der Sloot HA, van der Wegen GJL, Hoede D, de Groot GJ. Intercomparison of leaching tests for stabilized waste. Proceedings of the WASCON '94 Conference, 1994 June 1–3; Maastricht, The Netherlands: International Society for Construction with Alternative Materials.
- [97] Saric MM, Grzunov J. Proficiency testing: experiences from Croatia on the issue of heavy metals determination in marine sediments. Proceedings of The First International Proficiency Testing Conference. Sinaia, Romania: SC CEPROCI SA, Bucharest. 2007 Oct 11–13.
- [98] Hesbach P, Beck M, Eick M, Daniels WL, Burgers C, Greiner A, et al. Inter-laboratory comparison of leaching methods; 2014. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.162.2381&rep=rep1&type=pdf>.
- [99] Deutsches Institut für Normung e.V. DIN ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparisons; 2020.
- [100] JCGM - Joint Committee for Guides in Metrology. Evaluation of measurement data - Guide to the expression of uncertainty in measurement; 2008. <https://www.bipm.org/en/committees/jc/jcgm/publications>.
- [101] Wilcox RR. Basic Statistics: Understanding Conventional Methods and Modern Insights. New York: Oxford University Press; 2009.
- [102] Wilcox RR. Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy. 2nd edn. New York: Springer; 2010.
- [103] Deutsches Institut für Normung e.V. DIN 38402-45 German standard methods for the examination of water, waste water and sludge - General information; 2014.
- [104] Mandel J. The validation of measurement through interlaboratory studies. *Chemometr Intell Lab.* 1991;11:109–19. doi: 10.1016/0169-7439(91)80058-x.
- [105] Tukey JW. Exploratory data analysis. 1st edn. Reading, Mass.: Addison-Wesley; 1977.
- [106] Schmidt M. The sankey diagram in energy and material flow management part II: methodology and current applications. *J Ind Ecol.* 2008;12(2):173–85. doi: 10.1111/j.1530-9290.2008.00015.x.