Metrol. Meas. Syst., Vol. 33 (2026) No. 1, pp. 1–14
DOI: 10.24425/mms.2026.158363

PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS
Index 330930, ISSN 0860-8229
www.metrology.wat.edu.pl

M&MS

CERTIFICATION OF MULTIPARAMETER PHYSICOCHEMICAL REFERENCE MATERIALS. PART 1: HOMOGENEITY STUDY

Ewa Malejczyk[1)], Wojciech Hyk[2)]

1) Office of Measures, Physical Chemistry and Environment Department, ul. Elektoralna 2, 00-139 Warsaw, Poland
2) University of Warsaw, Faculty of Chemistry, Biological and Chemical Research Centre, ul. Ludwika Pasteura 1, 02-093 Warsaw, Poland (✉ *wojhyk@chem.uw.edu.pl*)

Abstract

Certified reference materials (CRMs) are indispensable tools in calibration, validation, and quality control across various fields, including chemical analysis, environmental monitoring, and industrial production. Their proper certification ensures the metrological traceability and reliability of the measurement results in laboratories around the world. This study presents a new application of an established statistical test (equality of coefficients of variation) to support the homogeneity assessment of multiparameter physicochemical CRMs. The method enables statistically justified indirect inferences about the homogeneity of untested properties based on the behaviour of tested ones, supporting more efficient and statistically defensible certification processes. The approach is demonstrated using CRMs that reproduce the density, refractive index, surface tension, and kinematic viscosity. The procedure is consistent with ISO 17034 and addresses the gaps identified in ISO 33405 regarding the assessment of homogeneity in multiparameter materials. By extending homogeneity assessment beyond conventional univariate methods, this work contributes a practical and statistically reliable tool for CRM producers developing complex, multiparameter materials.

Keywords: CRM, homogeneity, multiparameter, physicochemical.

1. Introduction

The production of *reference materials* (RMs) is one of the key elements to ensure the metrological traceability and validity of the test results in laboratories around the world. RMs are used, among other standards, for calibration of measuring instruments, quality control of test and measurement results, proficiency testing, and validation of measurement methods. The certification process of a reference material consists of determining its metrological characteristics and issuing a certificate, which contains information about the value of the certified property with its associated uncertainty, as well as the validity date and metrological traceability. The basic terminology and concepts related to RMs and their certification are defined in the *International Vocabulary of Metrology* (VIM) [1], which provides a common language for metrological concepts used throughout this study.

Article history: received July 14, 2025; revised October 13, 2025; accepted November 27, 2025; available online February 18, 2026.

Producers of *certified reference materials* (CRMs) are required to comply with the requirements of ISO 17034 [2], which presents the general requirements for the competence of *reference material producers* (RMP) and the consistent production of materials suitable for their intended use. This standard ensures the reliability and global comparability of CRMs. ISO 33405 [3] provides additional guidance on the characterisation of RM and the assessment of homogeneity and stability. However, while ISO 33405 allows for the evaluation of homogeneity of all properties of interest by testing a limited number of selected properties, it does not specify statistical tools for demonstrating the homogeneity of untested properties based on measurements of related ones.

This paper addresses this gap by applying a statistical test for the equality of coefficients of variation proposed by Miller [4] in a novel context, which can guide CRM producers in designing efficient homogeneity testing strategies. The proposed approach identifies groups of properties within a single CRM that are sufficiently interrelated, allowing the homogeneity of untested properties to be inferred from tested ones. The method also facilitates the evaluation of statistical correlations between different CRMs for a given property. This enables predictive homogeneity assessments for future batches of similar materials, thereby simplifying the certification process.

Multiparameter physicochemical CRMs can play a critical role in addressing current challenges in instrument calibration. Several studies have demonstrated that the accuracy of density measurements can be significantly affected by the physicochemical properties of the sample. For example, Furtado et al. [5] observed discrepancies in the density readings of viscoelastic samples measured with oscillation-type density metres, highlighting limitations of standard calibration approaches for complex fluids. Moura *et al.* [6] investigated the effect of surface tension on density measurements, while Furtado *et al.* [7] evaluated the robustness of such instruments for non-Newtonian fluids. These works underscored the influence of multiple interacting properties on measurement performance, but did not explore the use of CRMs as a potential solution.

The present study builds on these findings by proposing multiparameter CRMs that can be widely used for device calibration in non-standard measuring ranges, such as high-viscosity materials. In addition, these materials will enable quick and cost-effective calibration and verification of automated systems equipped with multiple measuring sensors.

A detailed procedure for the homogeneity assessment of candidate materials is presented as part of the RM certification process. The proposed statistical approach offers an ISO-compliant and practical tool for RMPs, filling a methodological gap in [3] and contributing to a more efficient production of reliable multiparameter CRMs. By allowing for indirect assessment of homogeneity across multiple physicochemical properties, this approach enhances CRM certification practices.

2. Materials and methods

2.1. Batch preparation

Eight candidate materials were selected to cover a wide range of density values (from approximately 600 to 1600 kg/m^3) and to represent different levels of the other target physicochemical properties. This selection was intended to ensure the applicability of the materials in a variety of calibration scenarios. The selected candidate materials are listed in Table 1, together with their nominal values of a property of interest at a temperature of 20°C. These reference materials are candidate CRMs for four physicochemical properties: density, refractive index, surface tension, and kinematic viscosity.

The following chemicals were used in this study:

1. isooctane, EMSURE® ACS, Reag. Ph Eur, ≥ 99.5%, source: Supelco (Merck), cat. No. 1.04727,

2. nonane, ReagentPlus®, 99%, source: Sigma-Aldrich (Merck), cat. No. N29406,
3. cyclohexane, ACS reagent, ≥ 99%, source: Sigma-Aldrich (Merck), cat. No. 179191,
4. chlorobenzene, synthesis grade, source: Sigma-Aldrich (Merck), cat. No. 8.01791,
5. 2,4–dichlorotoluene, 99%, source: Sigma-Aldrich (Merck), cat. No. 145009,
6. tetrachloroethylene, analytical grade, pure p.a., source: Chempur, cat. No. 112759209,
7. OM-5 and OM-140 that are mineral oils mainly consisting of unspecified base oil (content < 99.997%), registered under CAS No. 64742-54-7 and EC No. 265-157-1. Both materials are derived from petroleum refining processes.

Table 1. Candidate materials.

Material	**Density**	**Refractive index**	**Surface tension**	**Kinematic viscosity**
	kg/m^3	–	**mN/m**	**mm^2/s**
isooctane	692	1.391	18.8	–
nonane	718	1.406	22.7	–
cyclohexane	779	1.426	24.7	–
chlorobenzene	1106	1.525	33.0	–
2,4–dichlorotoluene	1250	1.546	35.2	–
tetrachloroethylene	1623	1.506	30.6	–
OM-5	806	1.455	25.7	5
OM-140	892	1.482	32.0	130

The selection of materials was based on their coverage in a wide density range, from approximately 0.6 kg/m^3 to 1.6 kg/m^3. Additionally, two liquids with kinematic viscosities of approximately 5 mm^2/s and 140 mm^2/s were included, as these represent viscosity levels that, based on practical experience, may pose challenges for viscosity correction algorithms in oscillation-type density metres.

The stability of the measured physicochemical properties during use was prioritised over nominal purity. Therefore, the liquids listed in points 1 to 6 above were stabilised as follows: the original manufacturer's bottles were opened and left loosely sealed for several days before being packed into the target packages, allowing equilibration with the surrounding air.

The handling of the mineral oils differed slightly: the oils were filtered using a Büchner funnel.

All materials were stored in amber glass bottles, in a light-free area under laboratory conditions (temperature between 17°C and 25°C).

Afterward, each RM was prepared as a batch of units. The candidate material was mixed, divided, and distributed from a 2.5 l bottle into smaller units. The batch size was 60 units for mineral oils (OM-5 and OM-140) and 50 units for other candidate materials.

The selection of the appropriate packaging was very important because the packaging might affect the stability of the RM. The packaging chosen and used in this analysis was 50 ml clear glass vials (by La-Pha-Pack, item No. 20 09 0289) with rubber/PTFE septum lids with crimped aluminium caps. The vials were washed with laboratory glassware detergent (a 1–4% aqueous solution of Trilux, by Analab). After rinsing the vials with tap water, they were rinsed three times with ultrapure water and then dried at 80°C under vacuum. Each vial contained 30 ml of liquid.

2.2. Selection of samples

A representative set of vials, representing at least 10% of a given batch, was selected for the analysis of homogeneity between vials (as recommended in [3]). The selection of vials for each batch was carried out as follows: a batch of vials arranged in packing order was divided into equal parts, and then one vial was randomly selected from each part. Other possible procedures for picking units from a batch are described in [3].

The minimum number of samples taken from each vial for between-unit homogeneity analysis was estimated based on the statistical power analysis with 95% test significance, an intended power of 80% and effect sizes anticipated from prior experience with similar materials. Power calculations for *analysis of variance* (ANOVA) were made using a free statistical power analysis program G*Power [8, 9].

For materials expected to show lower variability or better homogeneity (*e.g.*, low-density hydrocarbons), four samples per vial were taken to ensure sufficient sensitivity of the homogeneity test.

For materials with a larger expected effect size, three samples per vial were determined to be sufficient, particularly when measurement time was a limiting factor.

The design ensured that the ANOVA used for within-unit homogeneity assessment met the minimum requirement of five degrees of freedom for a reliable estimation of the variance components [3].

To meet all these requirements, the following experimental parameters were established for the measurements:

1. Number of units selected from the batch, *a*: Five units were analysed for isooctane, nonane, 2,4–dichlorotoluene, and tetrachloroethylene, whereas six units were examined for the remaining candidate materials.
2. Number of samples per unit, *b*: In the case of the determination of density, viscosity, and refractive index, three samples were drawn from each vial for materials with a higher expected variance, which was sufficient to achieve the desired statistical power (cyclohexane, chlorobenzene, 2,4–dichlorotoluene and tetrachloroethylene), while also reducing the total measurement time. For materials expected to show lower variability and smaller effect sizes (isooctane, nonane, OM-5, OM-140), four samples per vial were taken to ensure sufficient statistical sensitivity. In the case of surface tension measurements, the entire vial content was used without further subsampling for all materials.
3. Number of replicate measurements per sample, *n*: Each sample was subjected to five replicate measurements for viscosity and ten replicates for density, refractive index, and surface tension.

This approach ensured a balanced design that satisfied both statistical and practical requirements. The sampling procedure is shown schematically in Fig. 1 [10].

The analysis of between-unit homogeneity was intended to provide information on property differences due to the inhomogeneity of the batch of material, but also on possible trends due to the process of packaging the material into vials, which may be the cause of the inhomogeneity. Therefore, the vials were measured in a different order from the packaging order. This approach enabled the identification of a possible trend related to the drift of the measuring instrument during measurements and the application of an appropriate correction to eliminate its influence on the measurement results, if such a trend was detected. The order of the measurements was determined randomly by mixing the selected vials. If the measurement order and packaging order were too similar, the mixing was repeated.

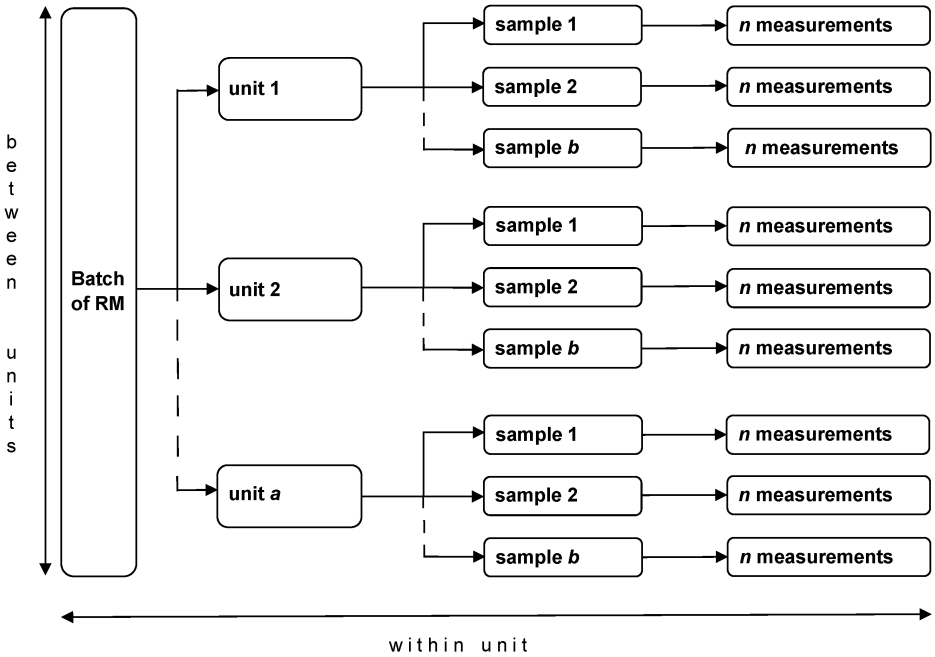

Fig. 1. Sampling procedure for a between-unit and within-unit homogeneity study.

2.3. Experimental methodology

2.3.1. Measurement methods

The following methods were used to measure the property values:

- density was measured using an oscillation-type density meter (type DMA 5000 for all candidates except n-nonane, OM-5 and OM-140, which were measured using type DMA 5001, by Anton Paar) [11, 12],
- the refractive index was measured using a photoelectric refractometer operating at a wavelength of 589 nm (sodium D-line) (type Abbemat Performance Plus 550, by Anton Paar),
- surface tension was measured using a force tensiometer with a Du Noüy ring and the Harkins-Jordan corrections (type K100, by Krüss) [6, 13],
- the kinematic viscosity was measured using an automatic kinematic viscometer (type SVM3001, by Anton Paar).

All samples were measured at 20°C and ambient pressure (0.10 MPa). The standard uncertainty of temperature for the measuring instruments used was $2.9 \cdot 10^{-3}$ °C for oscillation-type density meters and for the automatic kinematic viscometer, $1.3 \cdot 10^{-2}$ °C for the photoelectric refractometer and $1.5 \cdot 10^{-2}$ °C for the force tensiometer.

The measurements for homogeneity analysis were carried out under repeatability conditions: within one day, by one operator, and using one measuring instrument. This approach, in accordance with the recommendations of [3], allows variability due to material inhomogeneity to be isolated by eliminating the influence of external factors. As a result, the sources of variability are limited to two effects: between-unit and within-unit inhomogeneity. This enables the use of one-way or

two-way ANOVA to assess the homogeneity of the CRM in a statistically sound manner. Other sources of uncertainty, such as operator influence, measurement method, or time-related effects, are considered at later stages of the certification process, *e.g.* during material characterization.

The diagram in Figure 1 shows the sampling procedure, which involves taking a set of a units from a batch of material, then taking b samples from each unit and making n replicate measurements of each sample. The sample size taken for each instrument was predetermined by the dimensions of the measuring cell of the measuring instruments used and was 1.5 ml for the oscillation-type density meters, 2 ml for the automatic kinematic viscometer, 1 ml for the photoelectric refractometer and 30 ml for the force tensiometer.

All measurements were performed at the *Central Office of Measures* (GUM), which plays the role of the National Metrology Institute in Poland. All measuring instruments used in this study ensure traceability to the *International System of Units* (SI).

2.3.2. Data analysis methods

To identify potential trends in the data, a simple linear regression analysis was performed. The significance of the resulting slope coefficient was assessed using a t-test, which allowed evaluation of whether the observed variation was random or indicative of a systematic drift [3].

To correct for instrumental drift observed during the measurement series, a linear correction was applied. All measurements were arranged chronologically and indexed with a time variable. A linear regression was fitted to the dataset to estimate the drift slope. Each value was then corrected by subtracting the estimated drift relative to the first measurement. This procedure removed the systematic trend while preserving the internal variability of each sample.

To quantitatively assess the homogeneity of the material, analysis of variance (ANOVA) methods were employed. For surface tension measurements, the entire material from each unit was used without subdivision into smaller samples; therefore, within-unit homogeneity estimation was not applicable. In this case, the standard uncertainty associated with between-unit homogeneity (u_{bu}) was estimated using the one-way ANOVA function available in Microsoft Excel [3, 14].

For the remaining properties, multiple samples were taken from each unit, necessitating analysis of within-unit homogeneity. A two-way nested ANOVA was applied to evaluate the uncertainty components related to both between-unit (u_{bu}) and within-unit homogeneity (u_{wu}). Specifically, this method was used to estimate the uncertainty contributions for density, kinematic viscosity, and refractive index [2, 15].

The law of propagation of uncertainty was used for the combination of measurement uncertainty contributions associated with within-unit, u_{wu}, and between-unit homogeneity, u_{bu}, to determine the standard uncertainty associated with homogeneity, u_{hom} [3, 16].

A test of the equality of the coefficients of variation was performed to determine whether different materials possess the same relative variability associated with heterogeneity for the same physicochemical property. This test was also applied to determine whether different physicochemical properties show the same relative variability associated with heterogeneity for the same material. The null hypothesis was set, which assumes that the coefficients of variation of normal populations are equal: H_0: $CV_1 = \cdots = CV_k = CV$, where CV is the coefficient of variation, defined as the ratio of the standard uncertainty associated with homogeneity, u_{hom}, to the mean value, and k is the number of parameters compared. The null hypothesis was set against the alternative hypothesis, which assumes that at least one relative variability differs from the others.

The value of the test statistic was determined according to the approach presented by Miller [4]. To estimate the degrees of freedom for each uncertainty, which is a combination of mean squares derived from ANOVA, Satterthwaite's formula was applied [17].

The experimental value of the test statistics was related to the critical value χ^2_{crit} from the χ^2 table (a one-sided test variant) for a 95% (or 99.7%) confidence level. The rejection area R_α was represented by the interval $\langle \chi^2_{\text{crit}} ; +\infty)$.

3. Results and discussion

3.1. Trend analysis

First, the results for individual units were analysed in measurement order. To confirm the random nature of the scatter in the measured property values, a statistical significance analysis of the slope coefficient of the linear relationship fitted to the measurement data was carried out.

The values of the calculated t-statistics obtained for each material and property are presented in Table 2. The bold values in the table indicate cases where the t-statistic falls within the rejection area, indicating a statistically significant drift of the measuring instrument.

Table 2. Summary of t-statistics from the linear trend analysis of instrumental drift in physicochemical measurements of candidate materials.

Material	Density	Refractive index	Surface tension	Kinematic viscosity
isooctane	1.741	1.008	–1.874	–
nonane	–0.641	4.147	–3.714	–
cyclohexane	–1.345	–10.591	–3.462	–
chlorobenzene	–1.032	0.518	–	–
2,4–dichlorotoluene	–2.856	–0.968	–0.613	–
tetrachloroethylene	–1.792	–4.051	0.554	–
OM-5	0.020	0.193	–1.291	–0.489
OM-140	–1.417	1.804	0.051	1.353

In those cases where the test confirmed that the results did not show a trend related to the drift of the measuring instrument, they could be subjected to further homogeneity analysis. The existence of measurement drift generally results in an overestimation of the variance associated with the variation in the results, so where drift existed, further analysis was more complex and required the inclusion of an appropriate correction.

Figure 2 presents an example in which drift correction was applied. This case refers to refractive index measurements of a cyclohexane batch, where such drift was observed (see Table 2). Figures 2a and 2b show the raw data plotted in the measurement and packaging orders, respectively. Figure 2a shows the uncorrected data in the order of measurement, where a clear trend is visible across successive measurements. The significance of this trend was demonstrated statistically. In Fig. 2b, the same uncorrected data are arranged in packaging order.

Figures 2c and 2d show corrected data in which the effect of instrument drift has been eliminated by applying drift correction. Figure 2c presents the corrected data in measurement order, while Fig. 2d shows the same corrected dataset in packaging order. All plots use the same vertical scale, enabling direct visual comparison. Comparing graphs 2b and 2d, it can be seen that the scatter of results between the vials is reduced once the instrument drift has been eliminated.

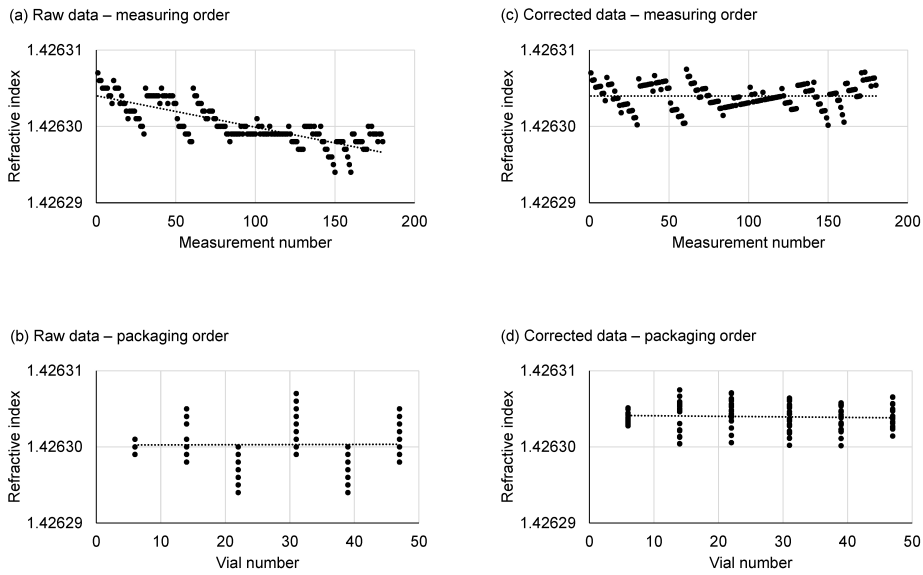

Fig. 2. Trend analysis of a batch of cyclohexane as a CRM candidate for the refractive index.

Similar corrections were applied in other cases where the statistical analysis indicated a significant trend. In the case of surface tension, a comparable but less pronounced trend was observed relative to that found for the refractive index. No statistically significant trend was detected for the remaining two physicochemical properties.

In addition to the instrumental drift, the data plotted in Fig. 2 reveal a consistent negative slope within individual samples. This trend may indicate the material's sensitivity to environmental exposure during the measurement process. Based on this observation, it is reasonable to recommend that CRM should be stored in tightly sealed vials, protected from external factors, and used once, immediately after opening.

3.2. Homogeneity assessment

Table 3 shows the results of the homogeneity analysis obtained for the candidate materials in question.

The uncertainty component associated with homogeneity, u_{hom}, is one of the elements included in the estimation of the combined uncertainty assigned to the certified property value. It was calculated by combining the uncertainty components associated with between-unit homogeneity, u_{bu}, and within-unit homogeneity, u_{wu} [3].

Target uncertainty, u_{trg}, represents the best measurement capability currently achieved for single-parameter CRMs, produced by the Central Office of Measures.

The results obtained are generally promising, with a few exceptions. The surface tension results for chlorobenzene and tetrachloroethylene are not good enough. Chlorobenzene exhibited instability during the measurements, which prevented accurate determination. This liquid is likely too sensitive to air humidity. Since the measurements were performed in an open vessel, surface processes occurring at the liquid–air interface may have caused significant changes in the measured surface tension.

For tetrachloroethylene, the results were more stable overall; however, the data exhibited excessive scatter, resulting in a between-unit uncertainty exceeding the target value specified for the CRM.

Also, the homogeneity of the refractive index and viscosity for OM-5 is too high.

Although the tested substances are chemically pure or well-mixed liquids, homogeneity remains a relative characteristic that depends on the sample size. Even in apparently homogeneous materials, small-scale effects, such as surface phenomena, local concentration gradients, or temperature fluctuations, may cause measurable variability when small sample intakes are used. Therefore, all homogeneity assessments in this study are valid in the context of the applied measurement methods and the volumes of sample intake.

Table 3. Uncertainties associated with homogeneity.

		Density	**Refractive index**	**Surface tension**	**Kinematic viscosity**
		kg/m^3	–	**mN/m**	**mm^2/s**
u_{bu}	isooctane	$1.3 \cdot 10^{-3}$	$4.8 \cdot 10^{-7}$	$1.9 \cdot 10^{-3}$	–
	nonane	$9.6 \cdot 10^{-4}$	$6.1 \cdot 10^{-7}$	$1.5 \cdot 10^{-3}$	–
	cyclohexane	$7.7 \cdot 10^{-4}$	$1.8 \cdot 10^{-6}$	$8.2 \cdot 10^{-3}$	–
	chlorobenzene	$1.6 \cdot 10^{-3}$	$1.8 \cdot 10^{-6}$	unstable	–
	2,4–dichlorotoluene	$3.1 \cdot 10^{-3}$	$2.3 \cdot 10^{-6}$	$3.1 \cdot 10^{-2}$	–
	tetrachloroethylene	$1.6 \cdot 10^{-3}$	$7.8 \cdot 10^{-7}$	1.8·10-1	–
	OM-5	$8.4 \cdot 10^{-3}$	$1.5 \cdot 10^{-5}$	$1.4 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$
	OM-140	$1.9 \cdot 10^{-4}$	$2.4 \cdot 10^{-6}$	$9.7 \cdot 10^{-3}$	$1.7 \cdot 10^{-2}$
u_{wu}	isooctane	$1.4 \cdot 10^{-3}$	$9.5 \cdot 10^{-7}$	–	–
	nonane	$8.6 \cdot 10^{-4}$	$5.4 \cdot 10^{-7}$	–	–
	cyclohexane	$6.1 \cdot 10^{-4}$	$1.5 \cdot 10^{-6}$	–	–
	chlorobenzene	$9.0 \cdot 10^{-4}$	$1.4 \cdot 10^{-6}$	–	–
	2,4–dichlorotoluene	$2.6 \cdot 10^{-3}$	$1.9 \cdot 10^{-6}$	–	–
	tetrachloroethylene	$2.5 \cdot 10^{-3}$	$6.3 \cdot 10^{-7}$	–	–
	OM-5	$1.8 \cdot 10^{-3}$	$1.7 \cdot 10^{-6}$	–	$3.1 \cdot 10^{-4}$
	OM-140	$1.9 \cdot 10^{-4}$	$7.4 \cdot 10^{-7}$	–	$2.2 \cdot 10^{-2}$
u_{hom}	isooctane	$1.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-6}$	$1.9 \cdot 10^{-3}$	–
	nonane	$1.3 \cdot 10^{-3}$	$8.1 \cdot 10^{-7}$	$1.5 \cdot 10^{-3}$	–
	cyclohexane	$9.8 \cdot 10^{-4}$	$2.3 \cdot 10^{-6}$	$8.2 \cdot 10^{-3}$	–
	chlorobenzene	$1.8 \cdot 10^{-3}$	$2.3 \cdot 10^{-6}$	noncertified	–
	2,4–dichlorotoluene	$4.1 \cdot 10^{-3}$	$3.0 \cdot 10^{-6}$	$3.1 \cdot 10^{-2}$	–
	tetrachloroethylene	$3.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-6}$	noncertified	–
	OM-5	$8.6 \cdot 10^{-3}$	$1.6 \cdot 10^{-5}$	$1.4 \cdot 10^{-2}$	$9.6 \cdot 10^{-3}$
	OM-140	$2.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-6}$	$9.7 \cdot 10^{-3}$	$2.7 \cdot 10^{-2}$
u_{trg}		$2.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-5}$	$5.0 \cdot 10^{-2}$	0.1%*

*Expressed as relative uncertainty. The target uncertainty for OM-5 is $6.0 \cdot 10^{-3}$ mm^2/s, for OM-140 it is $1.3 \cdot 10^{-1}$ mm^2/s.

3.3. Relative variability associated with heterogeneity

To investigate whether heterogeneity contributes the same to different properties of the same material and the same contribution to the same property of different materials, a statistical test of the equality of the coefficients of variation was performed. For this purpose, the uncertainties associated with homogeneity, u_{hom}, were transformed into coefficients of variation, CV. The calculated values of the coefficients of variation for each physicochemical property of the tested materials, together with the estimated degrees of freedom, f, are presented in Table 4.

Table 4 provides the basis for evaluating whether the relative contribution of heterogeneity is consistent across properties and materials. The subsequent sections present two applications of the equality test of coefficients of variation: first, to compare variability of a given property across different materials, and second, to examine the consistency of variability across properties within each material.

Table 4. Coefficients of variation and associated degrees of freedom for physicochemical properties across tested materials.

	Density		**Refractive index**		**Surface tension**		**Kinematic viscosity**	
	CV/%	f	**CV/%**	f	**CV/%**	f	**CV/%**	f
isooctane	$2.81 \cdot 10^{-4}$	9	$7.68 \cdot 10^{-5}$	10	$1.01 \cdot 10^{-2}$	1	–	–
nonane	$1.79 \cdot 10^{-4}$	11	$5.79 \cdot 10^{-5}$	10	$6.68 \cdot 10^{-3}$	1	–	–
cyclohexane	$1.26 \cdot 10^{-4}$	7	$2.85 \cdot 10^{-4}$	5	$3.33 \cdot 10^{-2}$	3	–	–
chlorobenzene	$1.62 \cdot 10^{-4}$	7	$1.51 \cdot 10^{-4}$	7	–	–	–	–
2,4–dichlorotoluene	$3.26 \cdot 10^{-4}$	7	$1.92 \cdot 10^{-4}$	6	$8.68 \cdot 10^{-2}$	3	–	–
tetrachloroethylene	$1.83 \cdot 10^{-4}$	8	$6.66 \cdot 10^{-5}$	5	–	–	–	–
OM-5	$1.06 \cdot 10^{-3}$	5	$1.07 \cdot 10^{-3}$	5	$5.33 \cdot 10^{-2}$	4	1.87·10-1	5
OM-140	$3.04 \cdot 10^{-5}$	7	$1.67 \cdot 10^{-4}$	6	$3.03 \cdot 10^{-2}$	4	$2.09 \cdot 10^{-2}$	10

3.3.1. Relative variability associated with heterogeneity between materials

For each physicochemical property, the equality of coefficients of variation between the candidate materials tested was assessed using statistical analysis. Table 5 shows the coefficients of variation only for those liquids that formed statistically homogeneous groups, i.e., groups in which the coefficients of variation were not significantly different from each other at 95% or 99.7% confidence level. The values of coefficients of variation for those materials that were statistically different from all others for a given property and confidence level are not shown in the table. This selective presentation facilitates the identification of materials with comparable variability for specific physicochemical properties.

The statistical test showed that the coefficients of variation in density for isooctane, nonane, cyclohexane, chlorobenzene, 2,4–dichlorotoluene and tetrachloroethylene are statistically equal. These materials are characterized by the same amount of heterogeneity if it is evaluated on the basis of density measurements.

In the case of refractive index, all liquids apart from OM-5 and nonane have the same coefficients of variation with a 95% confidence level. Nonane has the lowest value of coefficients of variation, indicating the best homogeneity among the tested materials. However, its value is very close to those of tetrachloroethylene and isooctane, suggesting that these materials also exhibit similarly high levels of homogeneity for the refractive index. When the confidence level is increased to 99.7%, the test confirms the equality of coefficients of variation for all materials except OM-5, which means that the same amount of heterogeneity characterizes all of these materials.

Table 5. Coefficients of variation for physicochemical properties with statistically homogeneous groups of candidate materials, at 95% and 99.7% confidence levels.

Physicochemical property	**Density**	**Refractive index**	**Refractive index**	**Surface tension**
Confidence level	95%	95%	99.7%	95%
isooctane	$2.81 \cdot 10^{-4}$	$7.68 \cdot 10^{-5}$	$7.68 \cdot 10^{-5}$	$1.01 \cdot 10^{-2}$
nonane	$1.79 \cdot 10^{-4}$	–	$5.79 \cdot 10^{-5}$	$6.68 \cdot 10^{-3}$
cyclohexane	$1.26 \cdot 10^{-4}$	$2.85 \cdot 10^{-4}$	$2.85 \cdot 10^{-4}$	$3.33 \cdot 10^{-2}$
chlorobenzene	$1.62 \cdot 10^{-4}$	$1.51 \cdot 10^{-4}$	$1.51 \cdot 10^{-4}$	–
2,4–dichlorotoluene	$3.26 \cdot 10^{-4}$	$1.92 \cdot 10^{-4}$	$1.92 \cdot 10^{-4}$	$8.68 \cdot 10^{-2}$
tetrachloroethylene	$1.83 \cdot 10^{-4}$	$6.66 \cdot 10^{-5}$	$6.66 \cdot 10^{-5}$	–
OM-5	–	–	–	$5.33 \cdot 10^{-2}$
OM-140	–	$1.67 \cdot 10^{-4}$	$1.67 \cdot 10^{-4}$	$3.03 \cdot 10^{-2}$

All materials tested, except chlorobenzene and tetrachloroethylene, which were deemed unsuitable for the production of surface tension CRMs, have the same amount of surface tension heterogeneity with a 95% confidence level.

Although OM-140 is outside this group, it shows the best density homogeneity compared to other liquids (see Table 4), making it an attractive candidate for a high-viscosity density standard at this stage of the study.

3.3.2. Relative variability associated with heterogeneity between properties

The statistical test of equality of coefficients of variation was applied separately for each material to test the equality of heterogeneities contributed by the different physicochemical properties. Table 6 shows the coefficients of variation only for those liquids that formed groups in which the coefficients of variation were not significantly different from each other at the 95% or 99.7% confidence level within individual materials. Coefficient of variation values for those properties that were statistically different from the others at the given confidence levels are not shown in the table.

Table 6. Coefficients of variation for physicochemical properties forming statistically homogeneous groups within each material, at 95% and 99.7% confidence levels.

Material	**Confidence Level**	**Density**	**Refractive index**	**Surface tension**	**Kinematic viscosity**
isooctane	–	–	–	–	–
nonane	–	–	–	–	–
cyclohexane	99.7%	$1.26 \cdot 10^{-4}$	$2.85 \cdot 10^{-4}$	–	–
chlorobenzene	95%	$1.62 \cdot 10^{-4}$	$1.51 \cdot 10^{-4}$	–	–
2,4–dichlorotoluene	95%	$3.26 \cdot 10^{-4}$	$1.92 \cdot 10^{-4}$	–	–
tetrachloroethylene	99.7%	$1.83 \cdot 10^{-4}$	$6.66 \cdot 10^{-5}$	–	–
OM-5	99.7%	$1.06 \cdot 10^{-3}$	$1.07 \cdot 10^{-3}$	–	–
OM-5	99.7%	–	–	$5.33 \cdot 10^{-2}$	1.87·10-1
OM-140	95%	–	–	$3.03 \cdot 10^{-2}$	$2.09 \cdot 10^{-2}$

The test showed that the individual properties of each material have different degrees of inhomogeneity. The variation in the heterogeneity of surface tension and kinematic viscosity is much greater than the variation in density and refractive index.

However, it is possible to find some relationships within certain materials. The test results show that for chlorobenzene and 2,4–dichlorotoluene, the relative variation in density and refractive index is statistically equal with a confidence level of 95%. Also, the relative variability in surface tension and kinematic viscosity for OM-140 is statistically the same with a 95% confidence level.

The increase in the confidence interval to 99.7% reveals further similarities in the amount of heterogeneity. Table 6 shows that cyclohexane, tetrachloroethylene and OM-5 have consistent variability for density and refractive index. OM-5 has also the same amount of heterogeneity for surface tension and kinematic viscosity.

For other materials, there is no such correspondence between properties.

4. Conclusions

This study presents an experimental scheme for the certification of liquid reference materials that reproduce up to four physicochemical properties. For each certified property, a separate chain of metrological traceability is established. This distinguishes the produced materials from chemical CRMs that reproduce chemical properties (in particular, multi-element CRMs). In addition, a statistical procedure is proposed that allows a probabilistic assessment of when homogeneity can be considered independent of the type of material or certified property. These findings can be useful for multidisciplinary proficiency testing studies that evaluate the performance of laboratories across various disciplines by analysing a common set of test samples. In this case, an efficient way to evaluate the evaluation of sample homogeneity is of great importance.

This study proposes and validates the use of a statistical test for the equality of coefficients of variation as a tool to support the homogeneity assessment of multiparameter physicochemical CRMs. The method enables the identification of groups of properties and materials with statistically comparable variability, which allows inference about the homogeneity of untested properties based on related ones.

The results demonstrate the following:

1. Certain groups of candidate materials exhibit statistically equal coefficients of variation for specific physicochemical properties, such as density or refractive index, suggesting similar degrees of heterogeneity.
2. Within individual materials, some physicochemical properties also show statistically equal coefficients of variation, indicating consistent levels of heterogeneity in parameters such as surface tension and kinematic viscosity.
3. The degree of heterogeneity varies significantly between different properties, with surface tension and kinematic viscosity showing higher variability compared to density and refractive index.

The developed approach facilitates predictive homogeneity assessments for future batches of CRMs by providing statistical evidence of correlation between materials and between properties within a single material. This offers a promising strategy for optimising the certification process, especially for multiparameter physicochemical CRMs.

The presented methodology can be applied in the context of ISO 17034 and ISO 33405 requirements and contributes to the development of efficient, reliable and cost-effective procedures for CRM production and certification. The novelty of this work lies in the way the statistical tools, commonly used in the certification process of reference materials, are integrated and applied to

the certification of multiparameter physicochemical CRMs. The statistically supported approach presented in this study demonstrates how to reliably assess the homogeneity of multiple properties across materials, including unmeasured ones. It is based on testing the equality of coefficients of variation, which allows CRM producers to make statistically justified inferences about the homogeneity of untested properties based on tested ones, an aspect not explicitly covered in current ISO standards. The proposed methodology provides a practical ISO-compliant tool that supports CRM producers in designing efficient homogeneity tests and predicting homogeneity in future production batches.

5. Outlook

The homogeneity assessment presented in this study represents the first stage of the certification process for the proposed multiparameter physicochemical CRMs. Future work will address subsequent stages, including stability studies and full characterization and certification of the materials. These forthcoming studies will provide further insights into the long-term behaviour, reliability, and applicability of the developed CRMs.

Acknowledgements

This work was supported by the Ministry of Science and Higher Education of the Republic of Poland (grant number DWD/5/0417/2021).

References

[1] JCGM. (2012). *International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM)* (JCGM 200:2012, 3rd ed.). https://doi.org/10.59161/JCGM200-2012

[2] International Organization for Standardization. (2016). *ISO 17034:2016 General requirements for the competence of reference material producers*. https://www.iso.org/standard/29357.html

[3] International Organization for Standardization. (2024). *ISO 33405:2024 Reference materials – Approaches for characterization and assessment of homogeneity and stability*. https://www.iso.org/standard/84226.html

[4] Miller, G. E. (1991). Asymptotic test statistics for coefficients of variation. *Communication in Statistics – Theory and Methods*, *20*(10), 3351–3363. https://doi.org/10.1080/03610929108830707

[5] Furtado, A., Pereira, J., Quendera, R., Schiebl, M., Lenard, E., Malejczyk, E., Alic, A., Alisic, S., Rauch, J., Lorenz, F., Bescupschii, A., Ciubara, A., Laky, B., & Amsüss, R. (2020). First density comparison on viscoelastic samples by oscillation-type densimetry. *ACTA IMEKO*, *9*(5), 79. https://doi.org/10.21014/acta_imeko.v9i5.943

[6] Moura, S. R., Furtado, A., Pellegrino, O., Lenard, E., Malejczyk, E., Sariyerli, G., Akcadag, U., & Sousa, J. (2023). Surface tension measurements – A comparative study. *ACTA IMEKO*, *12*(4), 1–6. https://doi.org/10.21014/actaimeko.v12i4.1418

[7] Furtado, A., Pereira, J., Quendera, R., & Cidade, M. T. (2019). Robustness studies of oscillation-type density meters with viscoelastic fluids. In *19th International Congress of Metrology* (Vol. 7, p. 13002). https://doi.org/10.1051/metrology/201913002

[8] Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

[9] Bausell, R. B., & Li, Y. (2002). Power Analysis for experimental research. In *Cambridge University Press eBooks*. https://doi.org/10.1017/cbo9780511541933

[10] Malejczyk, E., & Hyk, W. (2019). Certyfikacja materiałów odniesienia na przykładzie wybranego ciekłego wzorca gęstości. Cz. I. Ocena jednorodności materiału. *Analityka: Nauka i Praktyka*, *4*, 22–28. https://www.malamut.pl/analityka/numery#2019-4 (in Polish)

[11] Prochaska, L., Rauch, J., Furtado, A., Benga, F., Bartos, P., Bebic, J., Lenard, E., Alisic, S., Alic, A., Neuvonen, P. T., Sariyerli, G. S., Akcadag, Ü. Y., Ciubara, A., & Bescupschii, A. (2024). Final report on EURAMET key comparison EURAMET. M.D-K2.2 (1523). *Metrologia, 61*(1A), 07007. https://www.bipm.org/documents/d/guest/euramet-m-d-k2-2

[12] Neuvonen, P. T., Furtado, A., Moura, S., Laky, B., Malejczyk, E., & Elżbieta, L. (2022). Good practice guide for the measurement of the density of liquids in industry. In *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.6560044

[13] Harkins, W. D., & Jordan, H. F. (1930). A method for the determination of surface and interfacial tension from the maximum pull on a ring. *Journal of the American Chemical Society*, *52*(5), 1751–1772. https://doi.org/10.1021/ja01368a004

[14] Van Der Veen, A., & Pauwels, J. (2000). Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance. *Accreditation and Quality Assurance*, *5*(12), 464–469. https://doi.org/10.1007/s007690000237

[15] Guthrie, W. F. (2020). NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151) [Dataset]. In *National Institute of Standards and Technology (NIST)*. https://doi.org/10.18434/m32189

[16] JCGM. (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (JCGM 100:2008). https://doi.org/10.59161/JCGM100-2008E

[17] Gaylor, D. W., & Hopper, F. N. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics*, *11*(4), 691–706. https://doi.org/10.1080/00401706.1969.10490732

Ewa Malejczyk received her M.Sc. in chemical technology from Warsaw University of Technology in 2010. Since 2014, she has worked at the Laboratory of Physicochemical Standards at the Central Office of Measures in Warsaw. She represents Poland in the EURAMET TCM SC Density and Viscosity and TCF SC Volume, serves as a technical expert in density, volume, and certified reference material production for the national accreditation body, and is a member of KT 322 on reference materials under the national standardization body. Her research interests include national density standards, precise density and volume measurements, method development, and uncertainty evaluation.

Wojciech Hyk is an associate professor at the Faculty of Chemistry, University of Warsaw. He received his Ph.D. degree in analytical chemistry in 2000. The habilitation degree was granted to him in 2013. His research interests include exploration of mass transport phenomena to micro- and nanoelectrodes, design and characteristics of multifunctional polymeric materials and metallic nanostructures, development of green technologies for the recovery of critical raw materials form e-waste and development of new statistical strategies and software for quality assurance in analytical measurements. He is the author/co-author of about 55 papers published in peer reviewed journals and 2 patents.