Int J Thermophys (2013) 34:737–762
DOI 10.1007/s10765-012-1225-x

Analysis of Thermal-Conductivity Measurement Data from International Comparison of National Laboratories

B. Hay · R. Zarr · C. Stacey · L. Lira-Cortes · U. Hammerschmidt · N. Sokolov · J. Zhang · J.-R. Filtz · N. Fleurence

Received: 23 November 2011 / Accepted: 25 May 2012 / Published online: 19 June 2012

Abstract For the first time under the auspices of the Bureau International des Poids et Mesures (BIPM), seven national metrology institutes (NMIs) participated in an international interlaboratory comparison on thermal-conductivity measurements by the guarded hot-plate method. Measurements were conducted successively by all participants on the same set of specimens of insulating materials (mineral wool and expanded polystyrene) at temperatures ranging from 10 °C to 40 °C, according to the International Standard ISO 8302. This protocol aims to minimize issues of material

---

B. Hay (✉) · J.-R. Filtz · N. Fleurence
Scientific and Industrial Metrology Centre, Laboratoire National de Métrologie et d'Essais,
1 Rue Gaston Boissier, 75015 Paris, France
e-mail: bruno.hay@lne.fr

R. Zarr
Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive,
Gaithersburg, MD 20899-8632, USA

C. Stacey
Thermal Performance Group, Materials Division, National Physical Laboratory, Teddington,
Middlesex TW11 0LW, England, UK

L. Lira-Cortes
Centro Nacional de Metrología, Área Eléctrica, Division de Termometría, Km 4,5 Carretera a los Cues,
El Marques, C.P. 76241 Querétaro, Mexico

U. Hammerschmidt
Physikalisch-Technische Bundesanstalt, AG 1.74, Bundesallee 100, 38116 Braunschweig, Germany

N. Sokolov
D.I. Mendeleyev Institute for Metrology, Moskovsky Prospect 19, St. Petersburg 190005, Russia

J. Zhang
Heat Division, National Institute of Metrology, Bei San Huan Dong Lu 18, Beijing 100013, China

variability by circulating the same pairs of specimens among the laboratories following the strict format of a round-robin test program. This comparison is a pilot study which is intended as a first stage for future key comparisons between NMIs. The descriptive analysis of obtained results shows good agreement between laboratories for the mineral wool (MW) specimens and the thicker specimens of expanded polystyrene (EPS), with relative deviations within the uncertainties of measurement. A positive drift of thermal-conductivity values, which has appeared progressively during the comparison process, seems to be correlated with the size of the metering area of the guarded hot plates used. A statistical analysis was applied to repeated thermal-conductivity measurements at 23 °C, to identify anomalous and outlying data, to assess the within- and between-laboratory variability, and to evaluate the participant laboratories' performance.

Keywords Guarded hot plate · Insulating materials · Interlaboratory comparison · Thermal conductivity

1 Introduction

The guarded hot-plate (GHP) method is recognized worldwide as the primary technique for the measurement of thermal conductivity, λ, of low conducting materials. This steady state method, which is used especially for analyzing insulating materials for building applications, is standardized in the ISO 8302 [1] and ASTM C177 [2] standards. In principle, the values of thermal conductivities determined using a well-designed apparatus could be accurate to between 0.5 % and 1 %. But experience shows that the design, construction, and operation of a GHP apparatus are very challenging tasks, and that laboratories were not previously able to achieve agreement within the calculated combined uncertainties. Several intercomparisons of GHP measurements were performed in the past by accredited testing organizations, insulating materials manufacturers, and national metrology institutes (NMIs) [3–5] to assess the consistency of their measurements. The comparison of these GHP instruments is of prime interest to the energy usage and building communities. These apparatus are considered the primary methods for the determination of design heat transmission data for buildings. It follows that accurate instruments, protocols, and test methods are needed for accurate data and that comparisons at the highest metrology levels are needed to ensure integrity in the data.

For the first time, an international interlaboratory comparison on thermal conductivity of insulating materials was organized by the Bureau International des Poids et Mesures (BIPM) [6]. The main task of the BIPM is to ensure worldwide uniformity and comparability of measurements and their traceability to the International System of Units (SI). International comparisons between NMIs are thus carried out to compare and refine measurement procedures and evaluation of uncertainties, and to estimate the measurement capabilities of the NMIs involved. The interlaboratory comparison described in this paper was conducted by Working Group 9 "Thermophysical Properties" of the Consultative Committee for Thermometry (CCT). Seven NMIs are involved: Laboratoire National de Métrologie et d'Essais (LNE), National

Institute of Standards and Technology (NIST), National Physical Laboratory (NPL), National Institute of Metrology (NIM), Physikalisch Technische Bundesanstalt (PTB), Mendeleyev Institute for Metrology (VNIIM), and Centro Nacional de *Metrología* (CENAM). The thermal conductivities of two thermal insulating materials were measured at 10 °C, 23 °C, and 40 °C using GHP (single-specimen or two-specimen) apparatus.

The primary goal of this pilot study is to establish the state of the art for thermal-conductivity measurements by the GHP method in NMIs, by assessing in particular the variability and coherency of their thermal-conductivity measurements. This comparison was launched in 2007; the measurements were completed in 2010; and the resulting data were analyzed in 2011 by LNE.

This paper describes the interlaboratory comparison protocol, the different GHP apparatus used by the participants, as well as the test results and data analyses. The originality of this study lies in particular in the fact that

- this intercomparison involves for the first time only NMIs, which are assumed to have the best level of uncertainties for this type of measurements in their countries,
- the apparatuses used here were designed and built by the NMIs and represent the state of the art in thermal-conductivity metrology,
- the data analysis combines both descriptive analysis and statistical assessment.

By providing relevant information on the organization of a comparison (protocol, uncertainties of measurements…) and examples of data analysis, this work can be useful to a broad scientific audience wanting to perform such comparisons. In addition, the level of agreement that national laboratories can achieve compared with their calculated uncertainties would certainly be valuable information to researchers working in the field of thermal-conductivity metrology.

2 Organization of the Comparison

The comparison protocol was jointly drawn up by CCT Working Group 9 taking into account the major characteristics (specimen dimensions, temperature, thermal-conductivity ranges, etc.) of the GHP apparatus involved in this comparison. The thermal-conductivity measurements were carried out using GHP apparatus according to the International Standard ISO 8302 [1] by all laboratories except PTB. The main features of the instruments used (type and size, working temperature, specimen dimensions, among other factors) are presented in Table 1. Depending on the participant, the individual GHP apparatus operates either with a single specimen or a pair of specimens. The measurement uncertainties, estimated by each laboratory according to JCGM 100:2008 [7], are given as expanded uncertainties (coverage factor $k = 2$). For brevity, the laboratories are coded in the rest of this paper using the number assignments given in Table 1. These identification numbers correspond to the chronological order in which the NMIs performed their measurements.

Two different types of insulating material were selected for this comparison: MW and EPS. These materials were chosen because they best meet the selective criteria for thermal-conductivity measurements: low dispersion of density within a batch and long-term stability. The MW specimens are high-density glass-fiber boards (nominal

Table 1 Characteristics of GHP apparatus of various laboratories used in this intercomparison

Laboratory	NIST	LNE	NPL	VNIIM	NIM	CENAM	PTB
ID	1	2	3	4	5	6	7
GHP type	Double	Double	Single	Double	Single	Double	Single
Specimen dim. (mm)	Ø 1 016	610 × 610	610 × 610	Ø 330	Ø 337	Ø 305	Ø 100
Metering area (mm)	Ø 406.4	300 × 300	305 × 305	Ø 150	Ø 200	Ø 165	–
Mean temp. (°C)	7 to 65	0 to 50	5 to 40	−25 to 70	≥20	−5 to 60	−50 to 195
Specimen thick. (mm)	10 to 300	20 to 160	20 to 250	20 to 80	20 to 80	Up to 50	5 to 25
Temp. difference (K)	5 to 30	5 to 40	10 to 30	5 to 20	5 to 30	5 to 30	3 to 20
Thermal cond. ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$)	5 to 150	1.5 to 1500	≤100	20 to 200	50 to 2000	30 to 170	20 to 7000

density: 72 $\mathrm{kg \cdot m^{-3}}$) having a thickness of 35 mm. They come from a batch of a certified reference material named IRMM-440, whose properties were characterized by six European laboratories in a framework of a certification project initiated by the Institute of Reference Materials and Measurements [8].

Two special batches of EPS boards (35 mm and 70 mm thick) with a nominal density of 22 $\mathrm{kg \cdot m^{-3}}$ were specifically produced by Lafarge (France). It was a gray EPS containing graphite to avoid the "thickness effect" that is observed usually for normal white EPS (cf. EN 13163 standard [9]). A pair of disk-shaped specimens (with a diameter Ø of 1016 mm) was prepared by LNE from each of the three materials characterized above. In particular, these specimens are identified as MW35-1, MW35-2, EPS35-1, EPS35-2, EPS70-1, and EPS70-2. Two other pairs of EPS specimens of thicknesses 20 mm and 25 mm (identified, respectively, as EPS20-1, EPS20-2, EPS25-1, and EPS25-2) were specially machined for laboratory 7 because of their variant type of GHP apparatus.

In the case of a two-specimen apparatus, the mean thermal conductivity, λ, of the pair of specimens is determined at steady state conditions using

$$\lambda = \frac{\Phi L}{2A\Delta T}, \tag{1}$$

where $\Phi/2$ is the measured rate of heat flow (W) passing through two surfaces of the metering area for the specimen pair; A ($\mathrm{m^2}$) is the metering cross-sectional area; $\Delta T = T_\mathrm{h} - T_\mathrm{c}$ (K) is the measured temperature difference between each of the specimens' hot (T_h) and cold surfaces (T_c); and L (m) is the mean thickness of the pair of specimens. Values of λ are indicated for the mean specimen temperature, $T_\mathrm{m} = (T_\mathrm{h} + T_\mathrm{c})/2$.

In a single-specimen apparatus, the second specimen is replaced by a guard plate. In this case, Eq. 1 is modified slightly by removal of the constant coefficient 2. Depending on each individual apparatus, each participant performs either just a part of the following program or the whole on each pair of specimens.

- four successive runs at a fixed temperature of 23 °C with a temperature difference of 20 K over a short period of time. After each run, the specimens are removed

Table 2 Summary of the measurement program

Laboratory	1	2	3	4	5	6	7
Temperatures (°C) and repetition							
10 (1)	✓	✓	✓	✓		✓	
23 (4)	✓	✓	✓	✓	✓	✓	✓
40 (1)	✓	✓	✓	✓	✓	✓	✓
Materials and thickness (mm)							
EPS35	✓	✓	✓	✓	✓	✓	
EPS70	✓	✓	✓			✓	
MW35	✓	✓	✓	✓	✓	✓	
EPS20							✓
EPS25							✓

from the apparatus and then reassembled. This procedure yields information on the repeatability.

- one run at each of the two mean test temperatures of 10 °C and 40 °C under a temperature difference of 20 K.

Table 2 summarizes all individual measurement programs. The same set of specimens was circulated between the different NMIs and was thus successively measured by the participants (with the exception of laboratory 7). This protocol aims to minimize issues of material variability. Each participant sends the specimens back to the pilot laboratory (laboratory 2) after having performed their series of measurements. The pilot laboratory arranges machining of the specimens according to the requests of each successive participant. Initially, the specimens are measured by laboratory 1 that needed 1016 mm diameter specimens. Then, the specimens are stepwise cut down, first to 610 mm × 610 mm and finally to 330 mm diameter. Each cutting process has to leave the central part of the specimen undisturbed. All cutting scraps are marked and retained at the pilot laboratory, to reassemble, as closely as possible, 610 mm × 610 mm specimens at the end of the comparison process. This procedure enables the pilot laboratory to check the stability of the specimens by measuring their thermal conductivity at the beginning and at the end of the comparison.

Before performing thermal-conductivity measurements, participants are requested to

- condition specimens at (23 ± 2) °C and (50 ± 5) %RH for a minimum of 5 days, to reach thermal equilibrium with the environment. This equilibrium is judged as having been obtained when two successive mass measurements within a 24 h interval do not differ by more than ±0.5 %,
- measure the thickness of the specimens inside the apparatus (under a clamping pressure of about 1000 Pa). In the case of MW, adequate spacers are used to avoid thickness variation (and density change) during the thermal-conductivity measurements,
- calculate the specimen density from measurements of the mass and volume.

3 Presentation and Descriptive Analysis of Results

The test data (bulk density and thermal conductivity) reported to LNE by the participants were first checked for consistency and analyzed using graphical exploration techniques. This descriptive analysis of results is inspired from that performed by Zarr and Filliben [5] for a previous comparison.

3.1 Repeated Measurements of Thermal Conductivity at 23 °C

Table 3 summarizes the measurements and data (thermal conductivity, λ, and expanded uncertainties U on each individual measurement) reported to LNE by the participant laboratories for the MW and for the EPS.

All laboratories performed four consecutive thermal-conductivity measurements at 23 °C on the two specimens of each pair they studied according to the measurement program given in Table 2, except for laboratories 3 and 5. Laboratory 5 reported data for only one specimen per pair (MW35-1 and EPS35-1). Laboratory 3, which determined the thermal conductivity of each specimen in the single-sided mode, performed four replicate measurements for one specimen per pair (MW35-1, EPS35-1, and EPS70-1) and only one measurement for the other specimen of a given pair (MW35-2, EPS35-2, and EPS70-2). For comparison purposes for laboratory 3, the four replicate measurements were recalculated for each pair (MW35, EPS35, and EPS70).

Figure 1 plots the thermal-conductivity measurements at 23 °C versus laboratory (in chronological order from laboratories 1 to 6) for the three sets of specimens (MW35, EPS35, and EPS70). These individual values are given along with their uncertainty bars representing the expanded uncertainty U estimated by each one of the participants with a coverage factor of $k = 2$.

The grand mean $\bar{\bar{\lambda}}$ calculated from all individual measurements (30 for MW35 and EPS35, 20 for EPS70) is shown as a horizontal line for each pair of specimens. The white dots plotted for laboratory $2'$ correspond to the additional measurements performed by laboratory 2 at the end of the comparison process (cf. Sect. 3.3.1). Figure 1 seems to demonstrate that there is no laboratory–material interaction. The behavior of the results of the various laboratories does not change from one set of specimens to another, since the relative position of the results of each laboratory (for both mean values and dispersions) is constant on the three graphs.

In a first analysis, these graphs show also good agreement between the individual values obtained by the laboratories 1 to 4 for MW35, EPS35, and EPS70. The differences between the results of these four laboratories are within the measurement uncertainties claimed (except for EPS35). In the case of EPS35, the discrepancy between results from laboratories 1 to 6 is significantly higher than for MW35 and EPS70. Tables 4 and 5 give, respectively, the mean values $\bar{\lambda}$ and the standard deviations SD(λ) for the four repeated measurements of thermal conductivity performed by laboratories 1 to 7 on the different sets of specimens. The grand mean value of thermal conductivity, $\bar{\bar{\lambda}}$, the grand SD, and the range are also indicated for each pair of specimens in Tables 4 and 5. Each value of SD(λ) shown in Table 5 represents the within-laboratory

Table 3 Repeated measurements at 23 °C for the MW and expanded polystyrene

Lab.	Specimen	λ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	Specimen	λ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	Specimen	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$
1	MW35-1 +	31.54	0.40	EPS35-1 +	31.88	0.30	EPS70-1 +	31.90	0.60
	MW35-2	31.52	0.40	EPS35-2	31.90	0.30	EPS70-2	31.90	0.60
		31.52	0.40		31.89	0.30		31.91	0.60
		31.50	0.40		31.87	0.30		31.90	0.60
2	MW35-1 +	31.50	0.32	EPS35-1 +	31.88	0.32	EPS70-1 +	32.02	0.32
	MW35-2	31.61	0.32	EPS35-2	31.91	0.32	EPS70-2	32.01	0.32
		31.54	0.32		31.92	0.32		32.04	0.32
		31.50	0.32		31.87	0.32		32.02	0.32
3	MW35-1	31.81	0.48	EPS35-1	32.29	0.48	EPS70-1	32.33	0.48
		31.80	0.48		32.26	0.48		32.39	0.49
		31.79	0.48		32.29	0.48		32.36	0.49
		31.78	0.48		32.27	0.48		32.40	0.49
	MW35-2	31.84	0.48	EPS35-2	32.33	0.48	EPS70-2	32.32	0.48
4	MW35-1 +	31.93	0.20	EPS35-1 +	32.69	0.20			
	MW35-2	31.93	0.20	EPS35-2	32.71	0.20			
		31.95	0.20		32.70	0.20			
		31.94	0.20		32.69	0.20			
5	MW35-1	32.70	0.70	EPS35-1	34.20	0.50			
		32.30	0.70		34.00	0.50			
		32.80	0.70		34.30	0.50			
		32.20	0.70		34.50	0.50			

Table 3 continued

Lab.	Specimen	λ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	Specimen	λ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	Specimen	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	$U(\lambda)$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$
6	MW35-1 +	32.13	0.39	EPS35-1 +	32.78	0.59	EPS70-1 +	32.83	0.66
	MW35-2	31.97	0.38	EPS35-2	33.08	0.60	EPS70-2	32.51	0.65
		32.09	0.39		32.84	0.59		32.91	0.66
		32.19	0.39		32.06	0.58		32.66	0.65
7				EPS20-1	33.10	0.99	EPS25-1	30.40	0.91
					33.30	1.00		30.30	0.91
					32.70	0.98		30.20	0.91
					33.30	1.00		31.00	0.93
				EPS20-2	31.40	0.94	EPS25-2	30.30	0.91
					31.30	0.94		29.40	0.88
					32.00	0.96		31.00	0.93
					31.30	0.94		30.60	0.92

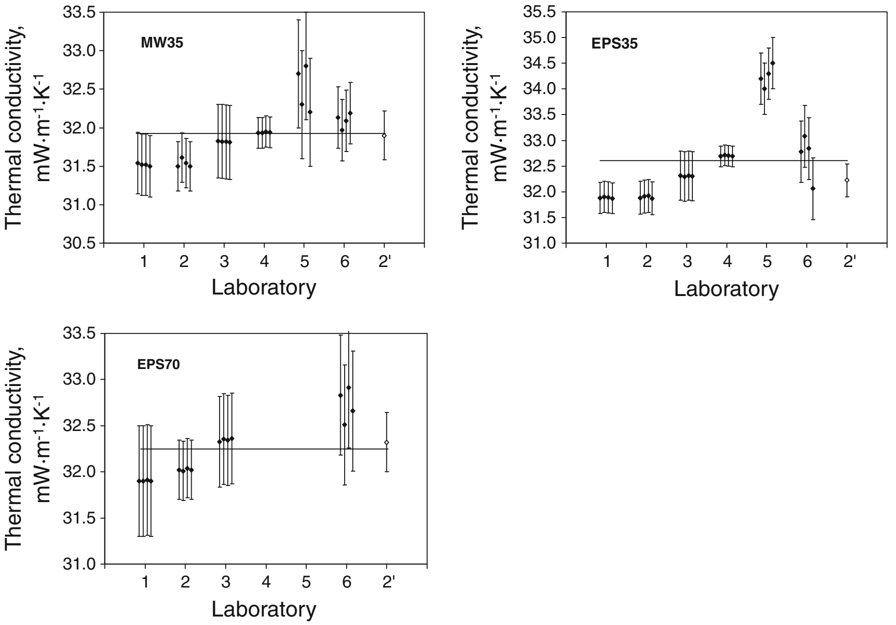

Fig. 1 Repeated measurements at 23 °C as a function of laboratory (uncertainty *bars* the expanded uncertainty estimated with a coverage factor of 2; *white dots* the additional measurements performed by laboratory 2 at the end of the comparison)

Table 4 Means values $\bar{\lambda}$ ($\mathrm{mW} \cdot \mathrm{m}^{-1} \cdot \mathrm{K}^{-1}$) of the repeated measurements of thermal conductivity at 23 °C

Laboratory	MW35	EPS35	EPS70	EPS20	EPS25
1	31.52	31.89	31.90		
2	31.54	31.90	32.02		
3	31.82	32.30	32.35		
4	31.94	32.70			
5	32.50	34.25			
6	32.10	32.69	32.73		
7				33.10/31.50	30.48/30.33
Grand mean $\bar{\bar{\lambda}}$	31.90	32.62	32.25	NA	NA
Range	1.30	2.63	1.01	NA	NA

NA not applicable

variability (for a laboratory and a pair of specimens) and the grand SD for each pair of specimens includes both within- and between-laboratory variability.

The mean values and SDs determined for EPS20 and EPS25 were calculated from the thermal-conductivity measurements performed by laboratory 7 in the single-sided mode on the specimens EPS20-1, EPS20-2, EPS25-1, and EPS25-2. Mean values

Table 5 Standard deviations SD(λ) ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$) of the repeated measurements at 23 °C

Laboratory	MW35	EPS35	EPS70	EPS20	EPS25
1	0.02	0.01	0.01		
2	0.05	0.02	0.01		
3	0.01	0.01	0.02		
4	0.01	0.01			
5	0.29	0.21			
6	0.09	0.44	0.18		
7				0.28/0.34	0.36/0.68
Grand SD(λ)	0.36	0.84	0.34	NA	NA

NA not applicable

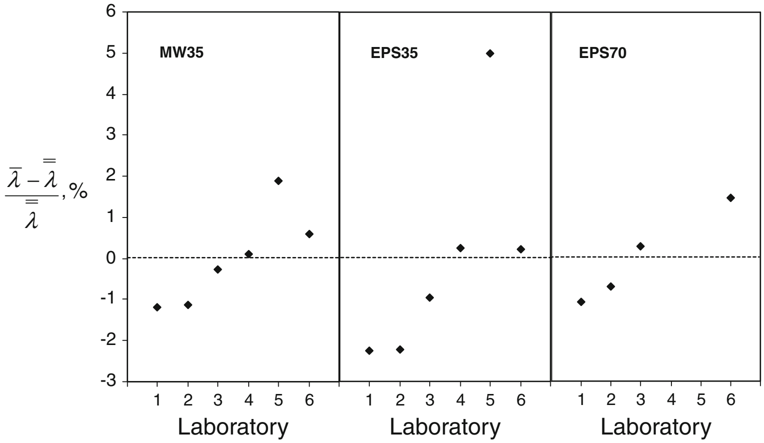

Fig. 2 Relative variation in mean thermal conductivity versus laboratory

of laboratory 7 achieved on thinner specimens are comparatively more scattered and deviate from those of laboratories 1 to 6.

Results given in Table 3 indicate that in the case of EPS, the relationship between the thermal conductivity and thickness is weak or negligible for measurements performed by laboratories 1 to 3, and 6 on 35 mm and 70 mm thick specimens, and negative for measurements carried out by laboratory 7 on 20 mm and 25 mm thick specimens. Figure 2 plots the relative differences between the mean values $\bar{\lambda}$ and the grand mean $\bar{\bar{\lambda}}$ (calculated from the values given in Table 3) as a function of laboratory. Figure 2 indicates that laboratory 5 is consistently higher than the other laboratories. With the exception of EPS35 where the grand mean is strongly influenced by the mean value of laboratory 5, most of the mean values of the laboratories are within 1.5 % of the grand mean for each material. It seems, however, that there is a positive drift of thermal-conductivity values measured chronologically by laboratories 1 to 6 for the three sets of specimens.

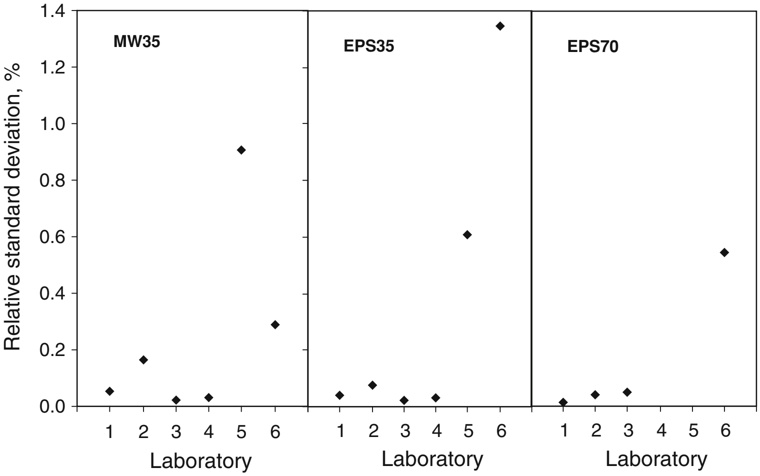

Fig. 3 Relative SD of repeated thermal-conductivity measurements at 23 °C versus laboratory

Figure 3 plots the relative SDs for each laboratory for the three pair of specimens. An examination of the plots reveals that the relative variability within a laboratory does not really change from one pair of specimens to another. It appears that results of repeated measurements at 23 °C of laboratories 5 and 6 are consistently more scattered than those from laboratories 1 to 4, with a relative SD between 0.5 % and 1.4 %, while those of laboratories 1 to 4 are less than 0.15 %. The apparent lack of repeatability of measurements for laboratories 5 and 6 (compared with laboratories 1 to 4) could be explained by divergences in the application of the test protocol. In particular in the case of laboratory 6, it seems indeed that four successive runs were performed, by measuring for each of them the thermal conductivity at the three test temperatures (10 °C, 23 °C, and 40 °C) in ascending order. The rest of participating laboratories carried out four runs at a fixed temperature of 23 °C by removing and placing again the specimen in the GHP, all other experimental conditions being kept constant.

The main conclusions of this preliminary analysis are the following:

- There are systematic differences between laboratories 1 to 6 for the three pairs of specimens, with the presence of a positive drift in the thermal-conductivity values (cf. Fig. 2). This point is discussed further in Sect. 3.3.
- Some individual measurements, mean values, and SDs of some laboratories are significantly different from those of the others (cf. Figs. 1, 2, 3). A statistical analysis of the repeated thermal-conductivity measurements at 23 °C is presented in Sect. 4 to identify anomalous and outlying data, and to exclude them if justified.

3.2 Thermal-Conductivity Measurements as a Function of Temperature

The results of thermal-conductivity measurements obtained by the participants on the three pairs of specimens for mean specimen temperatures of 10 °C and 40 °C are plotted

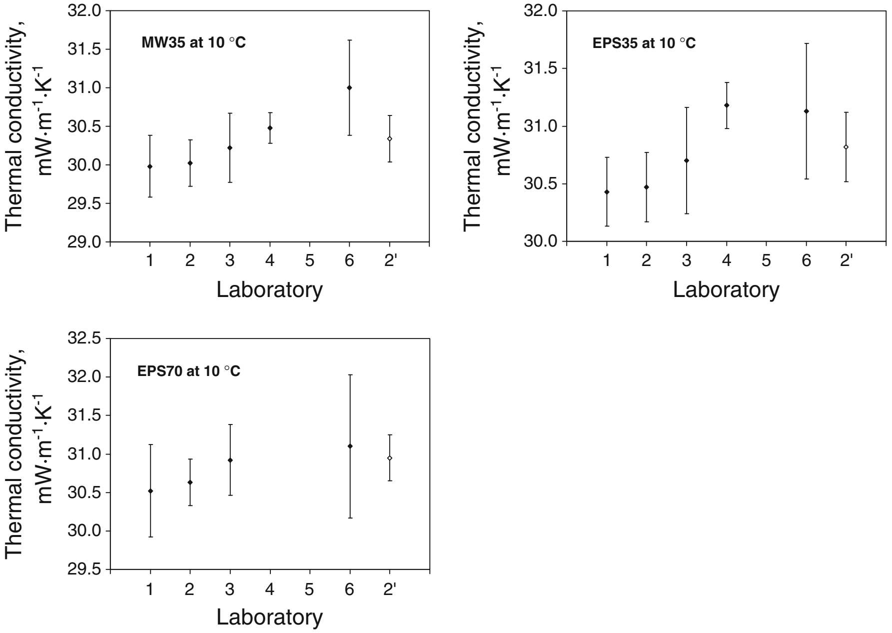

Fig. 4 Thermal-conductivity measurements at 10 °C as a function of laboratory (uncertainty *bars* the expanded uncertainty estimated with a coverage factor of 2; *white dots* the additional measurements performed by laboratory 2 at the end of the comparison)

as a function of laboratory in Figs. 4 and 5. These values are given with uncertainty bars representing the expanded uncertainty estimated by each of the participants.

Figures 4 and 5 demonstrate good agreement between the individual values obtained by laboratories 1 through 6 for MW35 and EPS70. The differences among the laboratory results were, for the most part, within the measurement uncertainties. As observed in Fig. 1 for the repeated runs at 23 °C, the measurements performed at 10 °C and 40 °C on EPS35 are more dispersed.

The single-point data at 10 °C and 40 °C preclude a rigorous statistical analysis from being performed for these two temperatures, as was conducted for the repeated measurements at 23 °C in Sect. 4.

3.3 Analysis of the Evolution of Thermal Conductivity

3.3.1 Long-Term Stability

The positive drift of thermal-conductivity values observed in Figs. 1 and 2 could be explained either by a modification of the tested materials during the comparison process or by systematic laboratory effects. In order to investigate this behavior, the pilot laboratory performed additional measurements on the three pairs of specimens at 10 °C, 23 °C, and 40 °C at the end of the comparison, after all participants had completed their measurements. The final specimens of 330 mm diameter (dimension

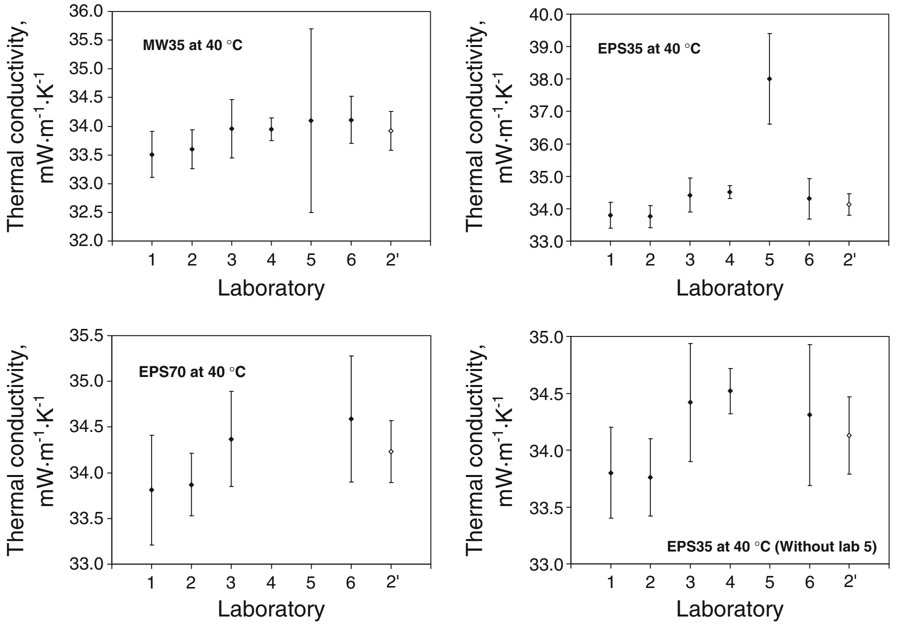

Fig. 5 Thermal-conductivity measurements at 40 °C as a function of laboratory (uncertainty *bars* the expanded uncertainty estimated with a coverage factor of 2; *white dots* the additional measurements performed by laboratory 2 at the end of the comparison)

required by laboratories 4, 5, and 6) were reassembled with the corresponding cutting scraps, which were retained at LNE, to obtain specimens having dimensions required by laboratory 2 (610 mm × 610 mm). The mean values obtained by laboratory 2 at the end of the comparison are approximately 1 % higher than those obtained 3 years earlier. These long-term variations are small compared to the differences observed between the results of the different laboratories, and are within the measurement uncertainties of laboratory 2. This might prove the stability of the materials and the specimens.

Other measurements were performed by laboratory 2 in March 2011 on new EPS and MW specimens obtained from the same batches as those studied in this comparison. These new specimens were tested before and after the cutting/reassembly process, to quantify the influence of the cutting on the thermal-conductivity measurements. The results of these measurements show a systematic increase of the thermal-conductivity values ranging from 0.5 % to 1 % depending on the pair of specimens. This could explain a part of the variation observed between thermal-conductivity values determined by laboratory 2 at the beginning and at the end of the comparison. This fact reinforces the previous conclusion about the generally good stability of the tested specimens in terms of their inherent thermal conductivity.

3.3.2 Influence of Density

The bulk density of the tested specimens was calculated by each participant from their measurements of mass, thickness, diameter or length, and width. All laboratories

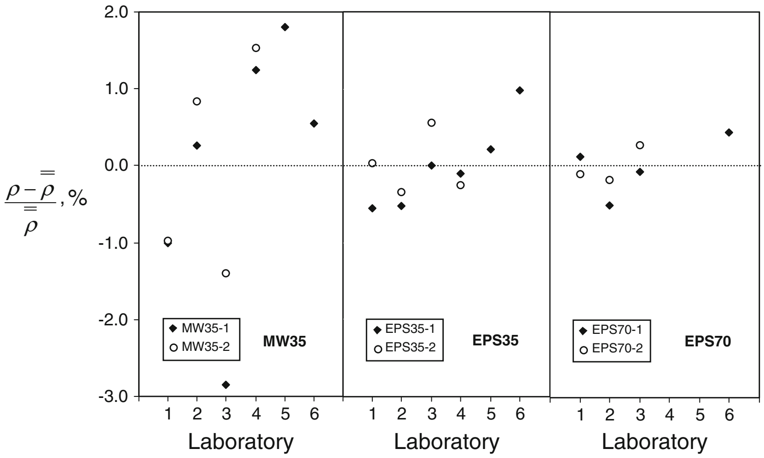

Fig. 6 Relative variation in specimen bulk density versus laboratory

performed these measurements on the two specimens of each pair used for thermal-conductivity measurements (cf. Table 2), except for laboratories 5 and 6 which reported data for only one specimen per pair (MW35-1, EPS35-1, and EPS70-1). The results indicate good agreement between the density measurements performed by all participants on the six specimens. Figure 6 plots the relative differences between bulk density measurements and the grand mean value $\overline{\overline{\rho}}$ (calculated for each specimen from the density measurements performed by all the participants) as a function of laboratory. The density measurements for MW35 specimens are more dispersed than those carried out on EPS, probably due to difficulties in accurately measuring the dimensions of MW35 specimens because of their compressibility. The relative variations in the density of MW35, EPS35, and EPS70 specimens vary, respectively, from −3 % to +2 %, −0.5 % to +1 %, and −0.5 % to +0.5 %.

Figure 6 shows that there is no correlation between the density measurements and the laboratories, because no systematic behavior appears for a laboratory for all specimens. For example, laboratory 3 measured the lowest density for the MW35-1 and MW35-2 specimens, and the highest ones for EPS35-2 and EPS70-2. In conclusion, the bulk density of each specimen can be considered reasonably stable with time, even if in one case (EPS35-2), it seems that there is drift of density values with time or with the laboratories (these factors being correlated).

Figure 7 plots the mean thermal-conductivity values of each pair of specimens at 23 °C as a function of their bulk density. It does not show any relationship between thermal conductivity and bulk density for these particular specimens. Hence, the deviations in thermal-conductivity results found here cannot be ascribed to density variations.

3.3.3 Influence of the Dimensions of the Metering Area

One possible explanation of these apparent drifts is that the dimensions of the metering area and of the specimens (both the parameters being obviously correlated) may

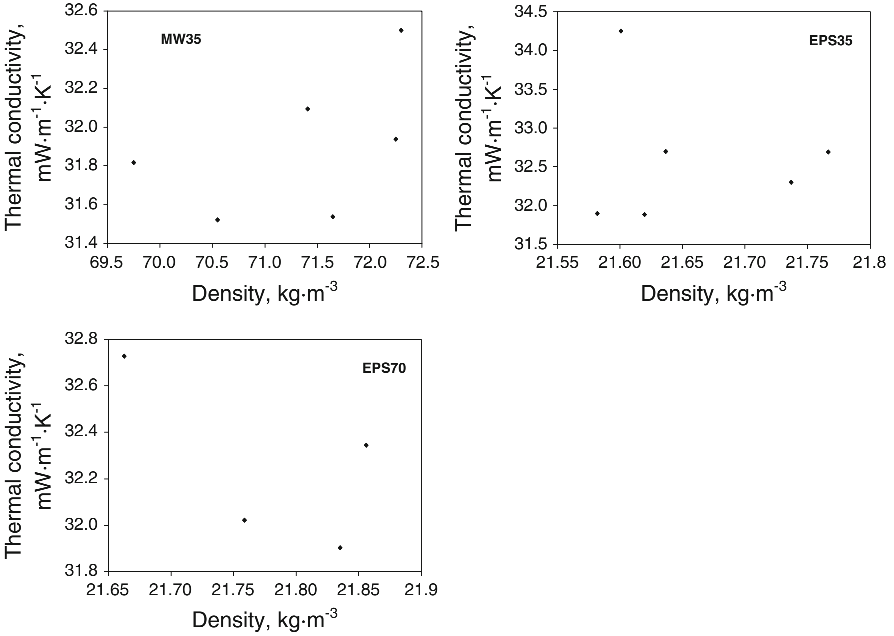

Fig. 7 Mean thermal conductivity of EPS and MW specimens at 23 °C versus density

influence the measurement results for the three pairs of specimens. These parameters change indeed strongly from one laboratory to another (up to a factor of 10 between laboratories 1 and 4). Figure 8 plots the mean thermal-conductivity values of each pair of specimens at 23 °C as a function of the metering area. The white dots correspond to the additional measurements performed by laboratory 2. Linear regression lines are determined for the three sets of specimens from the experimental data (except for one value for EPS35 which is undoubtedly an outlier—this point is discussed in Sect. 4), and are superimposed on the corresponding graphs. Figure 8 indicates a negative relationship between thermal-conductivity values and the metering area of the GHPs used in the different laboratories (the same observations can be done with the size of the specimen).

It is interesting to note that the slopes of the linear regression lines have the same order of magnitude for EPS and MW specimens and are very close for the two pairs of EPS specimens (EPS35 and EPS70). In a previous comparison, Zarr and Filliben [5] also noticed a significant correlation between thermal-conductivity results and the size of the metering area for a MW. As we do not dispose the detailed technical features of all the guarded hot plates involved in this study, it is difficult to perform a relevant interpretation of this correlation. One possible explanation could nevertheless be the increase of the sensitivity of thermal-conductivity measurements to potential errors on heat-flow measurements when the size of the metering area decreases. The GHP method assumes unidirectional heat flow through the test specimens in the region of the metering area. The accuracy of this method depends upon the validity of this assumption. Distortions of the heat-flow lines may however appear because of thermal

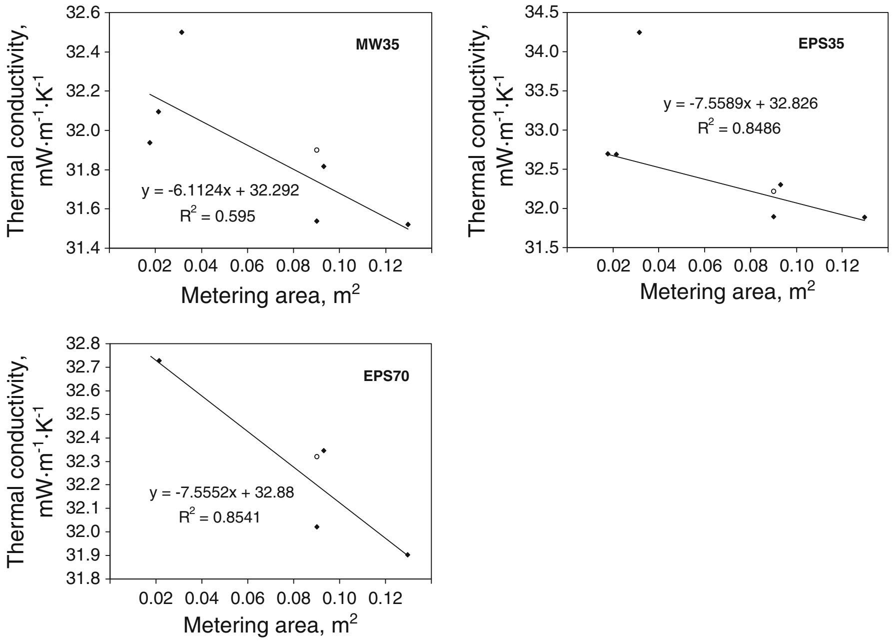

Fig. 8 Mean thermal conductivity of EPS and MW specimens at 23 °C versus metering area (*white dots* the additional measurements performed by laboratory 2 at the end of the comparison)

unbalance between the hot plate and the guard, heat flow through the edge of the specimen, or the presence of the gap separating the hot plate and the guard. It was demonstrated in many papers [10–13] that errors due to parasitic heat flows related to thermal unbalance and edge effects are correlated to the dimensions of the metering size (and to the ratio of the length of the metering area to the specimen thickness) and that they increase as plate size decreases.

4 Statistical Analysis of the Repeated Thermal-Conductivity Measurements at 23 °C

A statistical analysis of the repeated thermal-conductivity measurements at 23 °C was performed in two complementary steps

- determination of the performance of the GHP method; and
- evaluation of participant laboratories’ performance.

The first step aims to examine the consistency of the results, to identify anomalous and outlying data, and to estimate the general mean and variances of the experiments. The presence of individual laboratories or values that appear to be inconsistent with all other laboratories or values may change these estimates and the interpretation of results. As all laboratories 1 to 6 have carried out their thermal-conductivity measurements by applying the same standard measurement method (ISO 8302), the performance of the method can be assessed according to the ISO 5725 standard [14] through quantified parameters: the repeatability standard deviation, s_r, and the reproducibility standard deviation, s_R. Statistical methods used in proficiency testing are then applied to the results of this interlaboratory comparison in accordance with the ISO 17043 [15] and ISO 13528 [16] standards in a second step, to evaluate the participants' performance.

4.1 Scrutiny of Results for Consistency and Outliers

The following two statistical approaches are used to examine the consistency of the results and identify outliers: graphical consistency technique and numerical outlier tests. The results presented hereafter are based on an assumption that most of the variability between laboratories is random.

4.1.1 Graphical Consistency Technique

Two measures known as Mandel's h and k statistics, which are, respectively, a between-laboratory consistency statistic and a within-laboratory consistency statistic, are used here as one tool to assess the consistency of the data from the different laboratories participating in this comparison. It may be noted that, as well as describing the variability of the measurement method, these help in laboratory evaluation. Figures 9 and 10 plot Mandel's h and k statistics as a function of laboratory for the three pairs of specimens. These results are compared with the indicators for Mandel's h and k statistics at the 1 % significance level and at the 5 % significance level (given in the ISO 5725 standard).

- if the test statistic is less than or equal to its 5 % critical value, the item tested is accepted as correct,
- if the test statistic is greater than its 5 % critical value and less than or equal to its 1 % critical value, the item tested is called a straggler,
- if the test statistic is greater than its 1 % critical value, the item is called a statistical outlier.

The dashed and solid horizontal lines correspond to the critical values of Mandel's indicators h and k at the 1 % and 5 % significance levels, respectively. The h graph (Fig. 9) shows clearly that laboratory 5 obtained higher results than all other laboratories for all materials. In the case of EPS35, the mean value for laboratory 5 is very close, but lower than the 1 % level, it is thus considered as a straggler value. The k graph (Fig. 10) exhibits rather large variability between repeated test results for laboratories 5 (for MW35) and 6 (for EPS35 and EPS70). The corresponding results from laboratories 5 and 6 appear as outlier values according to these Mandel's statistics.

4.1.2 Numerical Outlier Technique

A statistical analysis based on the tests of Cochran and Grubbs is also recommended by ISO 5725 to identify stragglers or outliers. The first one is a test of the within-laboratory variabilities (i.e., a test on standard deviations) and the second is primarily a test of between-laboratory variability (i.e., a test on mean values).

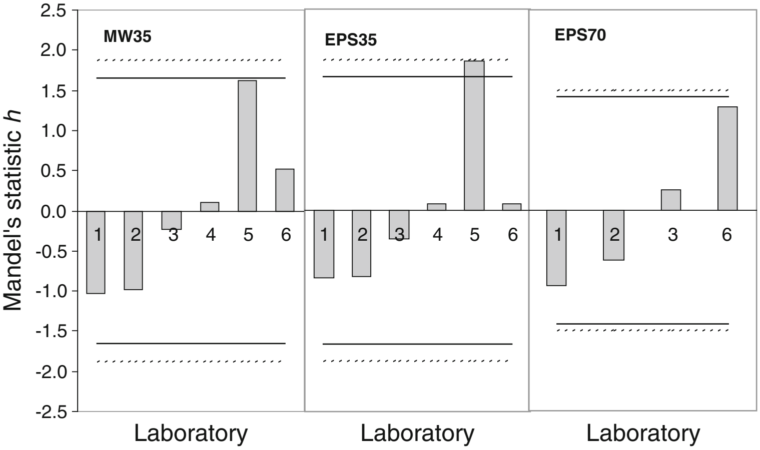

Fig. 9 Mandel's between-laboratory consistency *h*

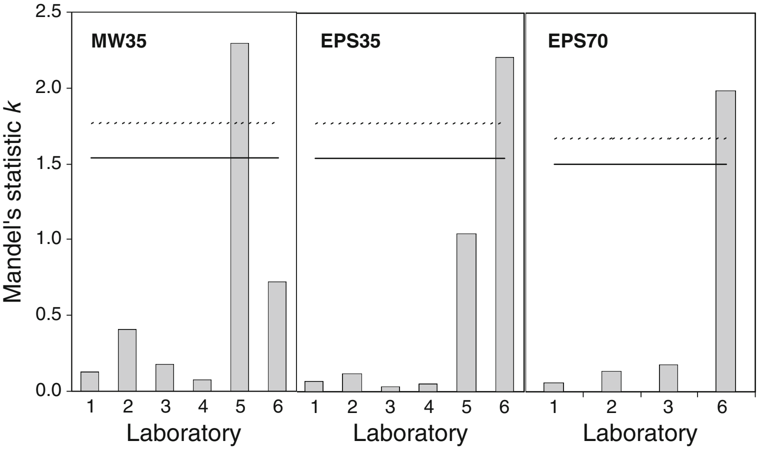

Fig. 10 Mandel's within-laboratory consistency *k*

4.1.2.1 Application of Cochran's Test Tables 6, 7, and 8 present the application of Cochran's test to the thermal-conductivity data obtained by all participants on the three sets of specimens. The Cochran's test statistic C is compared with its 5 % and 1 % critical values (as obtained from the statistical table in the ISO 5725 standard). The rules of identification of straggler and outlier values are similar to those described above for Mandel's statistics. Cochran's criterion tests only the highest value in a set of standard deviations. If the highest standard deviation is classified as an outlier, then the value is omitted and Cochran's test is repeated on the remaining values. Application of Cochran's test yields the following results.

- The values 0.876 and 0.707 of Cochran's test statistic for laboratories 5 and 6 for MW35 clearly indicate that the corresponding experimental results are outliers.

Table 6 Application of Cochran's test to MW35 thermal-conductivity data

	$p = 6; n = 4$	$p = 5; n = 4$	$p = 4; n = 4$	$p = 3; n = 4$
Critical values for Cochran's test				
$C_{1\,\%}$	0.626	0.696	0.781	0.883
$C_{5\,\%}$	0.532	0.598	0.684	0.798
Values of Cochran's test statistic C				
With all labs	0.876			
Without lab 5		0.707		
Without labs 5 and 6			0.752	
Without labs 5, 6 and 2				0.597

p number of laboratories, n number of replicates within each laboratory

Table 7 Application of Cochran's test to EPS35 thermal-conductivity data

	$p = 6; n = 4$	$p = 5; n = 4$	$p = 4; n = 4$
Critical values for Cochran's test			
$C_{1\,\%}$	0.626	0.696	0.781
$C_{5\,\%}$	0.532	0.598	0.684
Values of Cochran's test statistic C			
With all labs	0.814		
Without lab 6		0.980	
Without labs 6 and 5			0.643

Table 8 Application of Cochran's test to EPS70 thermal-conductivity data

	$p = 4; n = 4$	$p = 5; n = 4$
Critical values for Cochran's test		
$C_{1\,\%}$	0.781	0.883
$C_{5\,\%}$	0.684	0.798
Values of Cochran's test statistic C		
With all labs	0.987	
Without lab 6		0.577

The value of 0.752 means that results of laboratory 2 for MW35 can be considered as straggler values.

- The results of the Cochran's test statistic for EPS35 and EPS70 show that the data coming from laboratories 6 and 5 are statistical outliers.

Grubb's tests were also applied to the individual results obtained by laboratories 5 and 6 on MW35, EPS35, and EPS70 specimens to investigate whether the high standard deviations observed for these two laboratories could be due to a single outlier result, which could be corrected. No single stragglers or outliers were found by applying Grubb's tests. This observation confirms the previous conclusions of Cochran's test.

Table 9 Application of Grubbs' test to the mean values obtained by laboratories 1 to 6 on MW35, EPS35, and EPS70 specimens

	MW35		EPS35		EPS70	
	Single high	Single low	Single high	Single low	Single high	Single low
Grubbs' critical values						
$G_{1\%}$	1.973	1.973	1.973	1.973	1.496	1.496
$G_{5\%}$	1.887	1.887	1.887	1.887	1.481	1.481
Grubbs' test statistics G	1.621	1.032	1.861	0.840	1.022	1.218

4.1.2.2 Application of Grubbs' Test Table 9 presents the application of the Grubbs' test to the mean values of thermal conductivity (from Table 4) calculated for each participant on the three pairs of specimens. The Grubbs' test enables one to determine for each pair of specimens whether the largest or the smallest mean value of thermal conductivity is a straggler or an outlier value. The Grubbs' test statistic G is compared with its 5 % and 1 % critical values. The rules of identification of straggler and outlier values are similar to those of Mandel's and Cochran's statistics. Due to the small number of laboratories, only the simple test of Grubbs is applied.

No straggler or outlier values are found by applying Grubb's tests, even if in the case of EPS35, the mean value of laboratory 5 is very close to the 5 % level and can be thus suspected to be a straggler value. These observations are close to the conclusion of the graphical consistency technique (see Fig. 9).

4.2 Calculation of the General Mean and Variances

The objective of this step is to estimate separately for each pair of specimens (MW35, EPS35, and EPS70) according to ISO 5725 the following parameters, which give information on the performance of the GHP method:

- the general mean m
- the repeatability variance s_{r}^2
- the between-laboratory variance s_{L}^2
- the reproducibility variance $s_{\mathrm{R}}^2 = s_{\mathrm{r}}^2 + s_{\mathrm{L}}^2$.

On the basis of the above statistical analysis, the data reported to LNE by laboratories 5 and 6 for the three pairs of specimens are considered to be outlier values. The MW35 data for laboratory 2 are found to be straggler values. As is usually done in such an analysis, the outlier values are rejected and the straggler values are retained in calculations of general means and variances.

Table 10 summarizes the values of the general mean m and SDs s_{r}, s_{L}, and s_{R}. It is interesting to note that the expanded uncertainties ($k = 2$) which can be calculated as twice as the reproducibility of standard deviations s_{R} are of the same order of magnitude for MW35 ($0.42\,\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$) and EPS70 ($0.47\,\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$) as the measurement uncertainties claimed by most of the participant laboratories (see Table 3). In the case of EPS35, the obtained value ($0.77\,\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$) is slightly

Table 10 Computed values of general mean m and standard deviations s_r, s_L, and s_R

	MW35	EPS35	EPS70
Number of laboratory (p)	4	4	3
General mean m ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$)	31.71	32.20	32.09
Repeatability standard deviation s_r ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$)	0.028	0.013	0.014
Between-laboratory standard deviation s_L ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$)	0.208	0.386	0.233
Reproducibility standard deviation s_R ($\mathrm{mW \cdot m^{-1} \cdot K^{-1}}$)	0.210	0.386	0.233

higher than the uncertainties given by the laboratories, probably due to the unusually high dispersion of results.

4.3 Evaluation of Participant Laboratories' Performance

The participant laboratories' performance can be evaluated using the z score, which is the standardized measure of laboratory bias, calculated with Eq. 2 using the assigned value X and the standard deviation for proficiency assessment $\hat{\sigma}$,

$$z = \frac{x - X}{\hat{\sigma}} \tag{2}$$

The assigned values are the "reference" values of thermal conductivity attributed to each pair of specimens. These assigned values and the associated standard deviations correspond here to the robust average x^* and robust standard deviation s^*, which are calculated from the thermal-conductivity mean values using Algorithm A of ISO 13528. The z scores are calculated for each participant and each pair of specimens from these robust averages x^* and robust standard deviations s^*.

- $|z| \leq 2.0$ indicates "satisfactory" performance and generates no signal,
- $2.0 < |z| < 3.0$ indicates "questionable" performance and generates a warning signal,
- $|z| \geq 3.0$ indicates "unsatisfactory" performance and generates an action signal.

Table 11 summarizes for each pair of specimens the results of the z score analysis performed with all the mean values of thermal conductivity at 23 °C, including the outlier values. The standard uncertainty $u_X (k = 1)$ of the assigned value X is estimated by Eq. 3 according to [16] (where p is the number of laboratories).

$$u_X = 1.25 \times s^* / \sqrt{p}. \tag{3}$$

Table 11 shows satisfactory performance for all participants for the three pairs of specimens, with the exception of laboratory 5 for EPS35 ("questionable" performance). The results of these proficiency assessments are in agreement with those of the between-laboratory consistency statistic of Mandel (see Fig. 9).

This result is logical because the two approaches analyze only the results in terms of mean values. As Mandel's within-laboratory consistency and Cochran's test clearly

Table 11 Computed values of robust average x^*, robust standard deviation s^*, and z score (calculation with outlier values)

Laboratory	MW35		EPS35		EPS70	
	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score
1	31.52	−0.91	31.89	−0.86	31.90	−0.83
2	31.54	−0.87	31.90	−0.84	32.02	−0.54
3	31.82	−0.20	32.30	−0.29	32.35	0.23
4	31.94	0.09	32.70	0.25		
5	32.50	1.43	34.25	**2.36**		
6	32.10	0.46	32.69	0.24	32.73	1.14
Robust average x^*	31.90		32.51		32.25	
Robust SD s^*	0.419		0.735		0.419	
Std. uncert. $u_x(k=1)$	0.214		0.375		0.262	

Bold value represents "questionable" and "unsatisfactory" performances of the Participant Laboratories

Table 12 Computed values of robust average x^*, robust standard deviation s^*, and z score (calculation without outlier values)

Laboratory	MW35		EPS35		EPS70	
	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score	$\bar{\lambda}$ ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	z score
1	31.52	−0.78	31.89	−0.71	31.90	−0.72
2	31.54	−0.70	31.90	−0.68	32.02	−0.26
3	31.82	0.49	32.30	0.25	32.35	0.98
4	31.94	1.00	32.70	1.14		
5	32.50	**3.39**	34.25	**4.68**		
6	32.10	1.67	32.69	1.13	32.73	**2.46**
Robust average x^*	31.70		32.20		32.09	
Robust SD s^*	0.235		0.439		0.260	
Stand. uncert. $u_x(k=1)$	0.147		0.275		0.187	

Bold values represent "questionable" and "unsatisfactory" performances of the Participant Laboratories

demonstrate that the data of laboratories 5 and 6 can be considered as outlier values (see Fig. 10; Tables 6 , 7, 8), due to the scatter in their measurements, a second z score analysis was carried out, for comparison, using only the mean values retained in the analysis of consistency and outliers. Table 12 shows satisfactory performance for all participants for the three pairs of specimens, except for laboratory 5 for MW35 and EPS35 ("unsatisfactory" performance) and laboratory 6 for EPS70 ("questionable" performance).

Another criterion for performance evaluation is the normalized deviation number E_n. This criterion gives information on the validity of the expanded uncertainty estimate associated with each result. E_n numbers are calculated using

Table 13 Computed values of the normalized deviation number E_n (calculation with outlier values)

Laboratory	MW35		EPS35		EPS70	
	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n
1	31.52	−0.65	31.89	−0.78	31.90	−0.44
2	31.54	−0.68	31.90	−0.76	32.02	−0.37
3	31.82	−0.13	32.30	−0.24	32.35	0.13
4	31.94	0.08	32.70	0.24		
5	32.50	0.69	34.25	**1.88**		
6	32.10	0.33	32.69	0.17	32.73	0.56

Bold value represents “questionable” and “unsatisfactory” performances of the Participant Laboratories

Table 14 Computed values of the normalized deviation number E_n (calculation without outlier values)

Laboratory	MW35		EPS35		EPS70	
	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n	$\bar{\lambda}$ $(\mathrm{mW \cdot m^{-1} \cdot K^{-1}})$	E_n
1	31.52	−0.37	31.89	−0.50	31.90	−0.27
2	31.54	−0.38	31.90	−0.47	32.02	−0.14
3	31.82	0.20	32.30	0.15	32.35	0.41
4	31.94	0.66	32.70	0.86		
5	32.50	0.98	34.25	**2.66**		
6	32.10	0.79	32.69	0.54	32.73	0.83

Bold value represents “questionable” and “unsatisfactory” performances of the Participant Laboratories

$$E_\mathrm{n} = \frac{x - X}{\sqrt{U_\mathrm{lab}^2 + U_\mathrm{ref}^2}} \tag{4}$$

- $|E_\mathrm{n}| \leq 1.0$ indicates “satisfactory” performance and generates no signal,
- $|E_\mathrm{n}| > 1.0$ indicates “unsatisfactory” performance and generates an action signal.

U_lab is the expanded uncertainty on the mean value of the thermal conductivity estimated for each pair of specimens and each participant, and U_ref is the expanded uncertainty of the assigned value X (calculated from the standard uncertainty u_X). U_lab is calculated from the expanded uncertainty on individual values reported by the participants (see Table 3), and the standard deviation on the four successive independent measurements (see Table 5).

Whatever the assumption (with or without the outlier values), Tables 13 and 14 demonstrate that the uncertainties estimated by the participants are consistent with the assigned values X and their expanded uncertainties U_ref, except for laboratory 5 for EPS35.

Similar observations can be made by examining Fig. 11, which plots for each pair of specimens the mean thermal-conductivity values with their expanded uncertainties

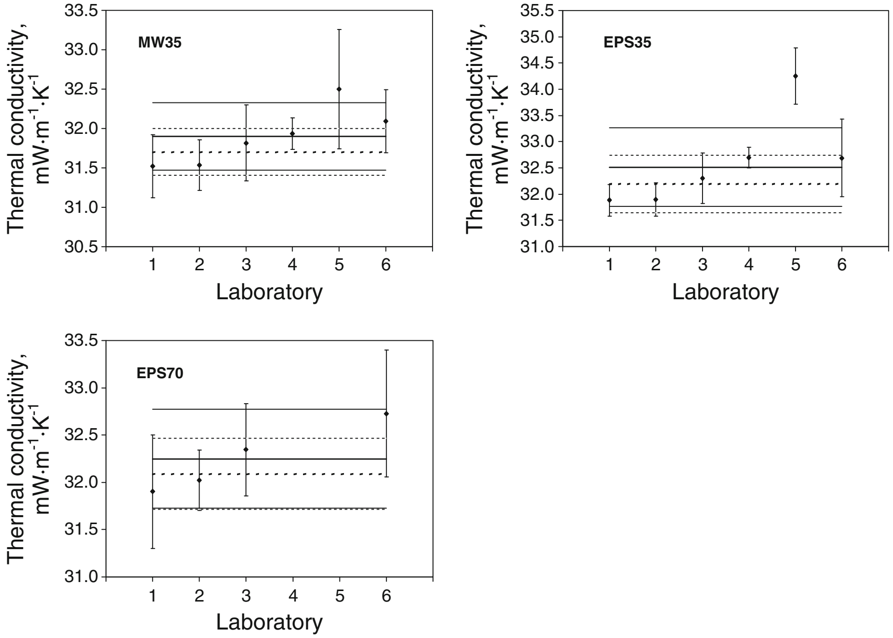

Fig. 11 Comparison between thermal-conductivity mean values from each laboratory and the assigned thermal-conductivity values, X (uncertainty *bars* the expanded uncertainty estimated with a coverage factor of 2)

compared to the assigned values X. The assigned values estimated in taking into account the outlier data are represented by solid bold lines; those calculated without the outlier values are plotted by dashed bold lines. The other solid and dashed lines represent the upper and lower limits of the expanded uncertainties of the assigned values X, respectively. It can be noticed however that in the case of EPS35, there is no overlap between the uncertainty bar of laboratory 4 and those of laboratories 1 and 2.

5 Conclusion

This international interlaboratory comparison investigated the agreement and variability in thermal-conductivity measurements performed by seven NMIs using the GHP method. All laboratories (except for laboratory 7) measured the same sets of specimens of insulating materials (MW and EPS) to avoid any potential influence of specimen heterogeneity from sampling.

The results obtained at 10 °C, 23 °C, and 40 °C on the three pairs of specimens (MW35, EPS35, and EPS70) indicate that there is no laboratory–material interaction. These results show, however, a systematic behavior of the laboratories for the three sets of specimens, with the appearance of a positive drift of thermal-conductivity values during the chronological progress of the comparison. It has been demonstrated that the deviations in thermal-conductivity results can not be ascribed to density variations or to an ageing phenomenon. This behavior could be due to an effect of the size of the metering area (or of the specimen), that progressively decreases from laboratory 1 to 6.

With the exception of laboratory 5 for EPS35, mean thermal-conductivity values obtained by laboratories 1 to 6 are in good agreement, with maximum differences among laboratories of 3.1 % for MW35, 7.3 % for EPS35, and 2.6 % for EPS70. These maximum relative differences decrease to 1.8 % for MW35 and 2.5 % for EPS35 if values from laboratory 5 are excluded. These deviations are less than the imprecision levels specified in ISO 8302.

Repeated measurements at 23 °C show that the results from laboratories 5 and 6 are more scattered (relative SD between 0.5 % and 1.4 %) than those of laboratories 1 to 4 (relative SD less than 0.15 %). The statistical analysis of these repeated thermal-conductivity measurements identified data from laboratories 5 and 6 as outlier values, in particular, due to their high scatter.

By retaining only data from laboratory 1 to 4 in accordance with the above comment, the repeatability SD and the reproducibility SD of the method are estimated to be less than 0.1 % and 1.2 %, respectively.

The expanded uncertainties of the "reference" values of thermal conductivity estimated in the proficiency assessment, as determined during the evaluation performance of the GHP method, are consistent with the measurement uncertainties estimated by the majority of the participants (except for laboratory 5 for EPS35). However, the absence of any overlap between uncertainty bars for the results from laboratories 1, 2, and 4 for EPS35 at the three temperatures indicates that there may be laboratory effects that cause the laboratories' means to disagree in this case, or that one or more of these uncertainties have been underestimated.

This pilot study is an essential step for the organization of a future Key Comparison. It highlights that improvements must be made in the organization of future interlaboratory comparisons, particularly for the following points:

- repeated measurements should be performed at all investigated temperatures to be able to apply a consistent statistical analysis to all obtained results,
- an unambiguous test protocol should be developed, and respected as far as possible by all participants,
- all participants should provide a detailed uncertainty budget.

The graphical and statistical analyses presented in this paper are meant to serve as a tool for progress for improvements of the measurement procedures of the participants, to minimize the effects of factors that may introduce bias between laboratories, and to reduce the level of random measurement error inherent in the method.

Acknowledgments The authors are grateful to Alexandre Allard and Yann Garcia from LNE for their contribution to this study.

References

1. ISO 8302, Thermal Insulation—Determination of Steady-State Thermal Resistance and Related Properties—Guarded Hot Plate Apparatus (2001)
2. ASTM C 177, Standard Test Method for Steady-State Heat Flux Measurements and Thermal Transmission Properties by Means of the Guarded Hot Plate Apparatus (2004)

3. S. Quin, J. Hameury, High Temp. High Press. **29**, 649 (1997)
4. D.R. Smith, Int. J. Thermophys. **18**, 1557 (1997)
5. R.R. Zar, J.J. Filliben, International Comparison of Guarded Hot Plate Apparatus Using National and Regional Reference Materials. NIST Technical Note 1444 (2002)
6. B. Hay, L. Cortes, B. Doucey, J.-R. Filtz, U. Hammerschmidt, N. Sokolov, C. Stacey, R. Zarr, J. Zhang, International Comparison on Thermal Conductivity Measurements of Insulating Materials by Guarded Hot Plate Preliminary Results, *Thermal Conductivity*, vol. 30 (DEStech Publications, Lancaster, PA, 2010), pp. 378–385, IBSN 978-1-60595-015-0
7. JCGM 100:2008 (GUM 1995 with Minor Corrections), Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (2008)
8. S. Quin, G. Venuti, F. DePonte, A. Lamberty, Certification of a Resin-Bonded Glass Fibre Board for Thermal Conductivity Between $-10\,^\circ$C and $+50\,^\circ$C—IRMM440, Report EUR 19572 EN (2000)
9. EN 13163, Thermal Insulation Products for Buildings—Factory Made Products of Expanded Polystyrene (EPS)—Specification (2008)
10. W. Woodside, A.G. Wilson, Unbalance Errors in Guarded Hot Plate Measurements, Special Technical Publication No. 217 (ASTM, West Conshohocken, PA, 1957), pp. 32–46
11. W. Woodside, Analysis of Errors due to Edge Heat Loss in Guarded Hot Plates, Special Technical Publication No. 217 (ASTM, West Conshohocken, PA, 1957), pp. 49–62
12. H.W. Orr, A Study of the Effects of Edge Insulation and Ambient Temperatures on Errors in Guarded Hot Plate Measurements, NRCC 10680, Research Paper No. 398, reprinted from *Proceedings of 7th Conference on Thermal Conductivity*, in NBS Special Publication No. 302 (1969), pp. 521–526
13. Q.T. Pham, C.G. Smith, Rev. Sci. Instrum. **57**, 99 (1986)
14. ISO 5725-2, Accuracy (Trueness and Precision) of Measurement Methods and Results—Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Requirements Method (1994)
15. ISO 17043, General for Proficiency Testing (2010)
16. ISO 13528, Statistical Methods for use in Proficiency Testing by Interlaboratory Comparisons (2005)