



Evaluation of comparisons of absolute gravimeters using correlated quantities: reprocessing and analyses of recent comparisons

Vojtech Pálinkáš¹ · Hartmut Wziontek² · Miloš Vaško¹ · Petr Křen³ · Reinhard Falk²

Received: 4 May 2020 / Accepted: 7 September 2020 / Published online: 31 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Due to the lack of a natural reference for the absolute acceleration of free fall and gravity, absolute gravimeters are compared at international comparisons, where the gravity reference is realized based on a set of precise absolute measurements and the functional model for their processing. The main estimates of comparisons, specifically, the reference values, deviations of gravimeters (biases) and associated uncertainty estimates, are obtained by the method of constrained least squares. In this paper, based on data from six comparisons, we demonstrate and discuss the importance of several aspects that significantly influence the estimated parameters, such as the type of the constraint, the weighting schemes, the correlations between measurements or outlier detection. Following the state-of-the-art key comparisons in metrology, we are applying correlations between measurements of a particular gravimeter for the first time, showing the ability to obtain appropriate uncertainty estimates, strictly applying the law of error propagation. The processing of comparisons by the method of least squares is described step by step to provide a background for analyses of absolute gravimeter comparisons in the future, reflecting all specifics of these comparisons. Estimated parameters from uniformly elaborated comparisons represent an interesting statistical data set that has been used to analyse the significance of differences between FG5, FG5X and other types of gravimeters. We believe that the described method of processing, as well as the published results, will be especially useful in the realization of the International Gravity Reference System.

Keywords Absolute gravimeter · Comparison · Bias · Degree of equivalence · Correlation

1 Introduction

The first International Comparison of Absolute Gravimeters of the International Association of Geodesy (IAG) was hosted by the International Bureau for Weights and Measures (BIPM) in 1981 (Boulanger et al. 1983), initiating a 40-year successful cooperation between geoscientific and metrology communities developed especially within working groups of IAG and Consultative Committee for Mass and Related Quantities (CCM). The comparisons took place at regular

four-year interval in Paris up until 2009 and beginning in 2001 were augmented in particular by the so-called European comparisons in Walferdange (Francis et al. 2012) and North-American comparisons in Boulder (Schmerge et al. 2012). More frequent comparisons had become necessary after extensive spreading of the commercial FG5 (Niebauer et al. 1995) and FG5X (Niebauer et al. 2011) absolute gravimeters (AG), which have been able to demonstrate repeatability better than $1.6 \mu\text{Gal}$ as it was shown, e.g. in Van Camp et al. (2005), Rosat et al. (2009) or Pálinkáš et al. (2010). Nevertheless, determination of bias changes over time for a particular gravimeter (e.g. after maintenance) or bias differences between instruments have to be known at the microgal level ($1 \mu\text{Gal} = 10 \text{ nm/s}^2$) for ensuring consistent gravity measurements needed in geophysics (e.g. Van Camp et al. 2017), geodynamics (e.g. Olsson et al. 2016) and geodesy (Pálinkáš et al. 2013). In metrology, the highest requirements on g -values are related to the realization of the kilogram by the Kibble balance (Stock 2013; Robinson and Schlamminger 2016) where the absolute gravity has to be known at

✉ Vojtech Pálinkáš
vojtech.palinkas@pecny.cz

¹ Geodetic Observatory Pecný, Research Institute of Geodesy, Topography and Cartography, 25165 Ondřejov 244, Czech Republic

² Federal Agency for Cartography and Geodesy (BKG), Frankfurt/Main and Leipzig, Germany

³ Czech Metrology Institute, Okružní 31, 63800 Brno, Czech Republic

the centre of the test mass with accuracy better than 5 μGal (e.g. Jiang et al. 2013).

With the establishment of a new absolute gravity reference system (Wilmes et al. 2016), the international comparisons will gain importance as a backbone of its realization. The absolute gravimeters contributing to the realization of the system (International Gravity Reference Frame) should participate in the comparisons or validate the measurements at the reference stations established within the frame. Combination of results from comparisons and monitoring of temporal gravity changes at reference stations equipped by superconducting gravimeters should allow to provide information on biases of a particular AG.

Since at a given time and a given site, only one gravimeter can measure, the comparisons were organized at stations with several sites allowing simultaneous measurements of different AGs, taking care to optimize the measurement schedule and also including corrections due to the time variable gravity field during comparisons lasting from a few days to a few months. The reference values and biases at comparisons were determined by an adjustment procedure that was significantly progressing over time, applying different weighting schemes and uncertainty estimates. Moreover, since 2009 (Jiang et al. 2012), as a consequence of the Mutual Recognition Arrangement of the International Committee for Weights and Measures (CIPM MRA¹), the international comparisons of AGs have been split up into the key comparison (KC) and the pilot study or recently additional comparison (Marti et al. 2014) to adjust to the established rules in metrology. As a consequence, different subsets of gravimeters have to be considered in comparisons with different contributions to the comparison reference values. While gravimeters having metrological background (assigned to a National Metrology Institute or Designated Institute, NMI/DI), enclosing a complete uncertainty budget with their results, take part at the KC and can contribute to the definition of the absolute reference values, the majority of gravimeters operated by geodetic or geophysical institutions (non-NMI/DI) can be treated in the adjustment procedure only as relative meters. At first glance, this separation seems to be illogical especially when both groups of institutions are using the same type of gravimeters (mostly FG5 or FG5X); however, it is a typical situation in metrology where the NMI/DIs have responsibilities to maintain the national standards and should declare Calibration and Measurements Capabilities (CMC) published in the Key Comparison Database (KCDB²) that comprises a review of laboratories and declared CMCs (including uncertainties). Therefore, in an ideal case, the measurements of NMI/DIs should be associated with such an uncertainty budget, where all the possible error sources influencing measurements are

taken into account, and therefore, the declared uncertainties represent realistic estimates of the measured gravity values. Consequently, from a metrological point of view, the main goal of key comparisons is the validation of CMCs published in the KCDB and in other words, to verify whether the determined bias (degree of equivalence) is consistent with declared uncertainties. On the other hand, users from geodesy and geophysics plan to use the determined biases for correcting their measurements (Olsson et al. 2016; Pálinkáš et al. 2013) and by this way to obtain consistent time series and utilize the excellent long-term repeatability of FG5 and FG5X (FG5/X) gravimeters. Nevertheless, the real accuracy of the gravity reference itself must always be taken into account when the “absolute g ” is under discussion. This is relevant not only for the Kibble balance experiments, but also for long-term gravity time series in geosciences, since the realization of the gravity reference depends on the cloud of participating gravimeters. If one type of AG is dominating (FG5/X since 1997, cf. Table 1), possible common systematic errors might affect the reference. Therefore, it is necessary to investigate systematic errors of the individual technologies, as in case of FG5/X gravimeters mainly the diffraction effect (Van Westrum and Niebauer 2003; Křen and Pálinkáš 2018) and the distortion effect (Křen et al. 2016) and in case of cold-atom gravimeters (e.g. Pereira and Bonvalot 2016) the wavefront aberration (Schkolnik et al. 2015). Of course, comparisons might serve as in Jiang et al. (2012) also for an experimental determination of biases between different technologies.

As it can be seen, the comparisons and its processing play a key role in the determination of the absolute gravity reference. Although the processing strategies have been evolving over time, at present still several open questions exist. In this paper, we present and discuss the results of reprocessing of data obtained in the recent International and European comparisons since 2009, see Table 1. The least squares approach is used for obtaining estimates for the reference values and biases of the comparisons. However, for the first time in gravimetry, it takes into account correlations between measurements. Such an approach has been used also for other comparisons in metrology (Sutton 2004; Woolliams et al. 2006). As discussed in White (2004) and Koo and Clare (2012), it allows to obtain appropriate uncertainty estimates for the determined biases. Since the comparisons of gravimeters have several specifics as, e.g. existence of two groups of gravimeters (NMI/DI and non-NMI/DI), more reference values or several measurements from particular gravimeters, in Sect. 2, we show and explain, step by step, the important aspects that influence the elaboration of comparisons as (1) harmonization of uncertainties, (2) definition of the constraint, (3) link of the regional comparisons, (4) correlation of measurements, (5) consistency check and outlier detection, (6) uncertainty estimates for results of the comparison. In the

¹ <https://www.bipm.org/en/cipm-mra/cipm-mra-text/>.

² <https://www.bipm.org/kcdb/>.

Table 1 List of key comparisons of absolute gravimeters organized by the Consultative Committee for Mass and Related Quantities (CCM) and EURAMET since 2009

Comparison Year	CCM.G-K1 2009	EURAMET.M.G-K1 2011	CCM.G-K2 2013	EURAMET.M.G-K2 2015	CCM.G-K2.2017 2017	EURAMET.M.G-K3 2018
Station	Sèvres	Walferdange	Walferdange	Belval	Beijing	Wetzell
# Sites	5	15	15	9	9	4
# Measurements	63	65	73	55	112	48
Duration/days	31	35	104	10	17	59
Comparison ref. height/m	1.00	1.30	1.30	1.25	1.25	1.25
# AGs total–NMI/DI	21–11	22–6	25–10	17–4	28–13	16–5
# FG5/X total–NMI/DI	13–6	18–4	19–7	15–3	18–12	14–4
# Linking AGs	None	4	None	4	None	4

Parameters as number of measured sites, number of measurements, duration of the comparison, comparison reference height, number of absolute gravimeters (in total and those belonging to NMI/DIs), number of FG5/X type gravimeters are showed, as well as the number of gravimeters used to realize the link between CCM and EURAMET comparison

published reports of AG comparisons, some of these aspects have not been treated consistently and the final results have been associated with uncertainty estimates that do not always follow the law of error propagation. Acknowledging the enormous efforts in the preparation of these reports, the presented elaboration of comparisons described in Sect. 2 should not be understood as a criticism of the published results, but as an effort aiming to contribute to the discussion on elaboration of comparisons in general, and finding a consensus on the elaboration of AG comparisons in the future.

In Sect. 3, results of reprocessing are presented by two solutions for each comparison. The first solution, labelled as KCN, is taking into account the division of gravimeters as belonging to NMI/DIs that can contribute to the definition of KCRV (key comparison reference values). The second solution, labelled as ICN, is treating all gravimeters at the same level—as appropriate to contribute to the definition of reference values. According to the Strategy paper of CCM and IAG (Marti et al. 2014), the already published KC solutions are the official ones. Therefore, KCN solutions provide improved estimates of the official estimates.

In Sect. 4, we are analysing the results of the selected comparisons as a whole, to identify the significance of differences between different types of gravimeters and comparison sites. Further, variabilities of biases for individual gravimeters are shown.

All the uncertainties reported in this study are representing the standard uncertainties with coverage factor $k = 1$, while expanded uncertainties ($k = 2$) are used in the KCDB. Accordingly, we are using symbol \pm for expressing the standard uncertainty.

2 Elaboration of comparisons

Each gravimeter participating in a comparison is operated at several sites (usually 3, but 4 in 2017) and reporting results of the absolute gravity measurements g_{raw} associated with the standard uncertainty u_{raw} . These measurements are representing the mean acceleration of free fall at a given site at the specific measurement height of a gravimeter (Pálinkáš et al. 2012) corrected for defined geophysical effects (tides, atmospheric mass variations, polar motion) and all known instrumental effects. The reported values (g_{raw}, u_{raw}) are then transferred to a common comparison reference height using vertical gravity gradients, determined from measurements with relative gravimeters. If a superconducting gravimeter is operated at the station, also corrections due to residual temporal gravity variations are applied (Francis et al. 2012). Every measurement made by the gravimeter “ i ” (with a bias δ_i) at the site “ j ” during the comparison at the given comparison reference height may be described by the observation

$$g_{ij} = g_j + \delta_i + \varepsilon_{ij} \tag{1}$$

where ε_{ij} is the random error associated with the measurement distributed around zero mean $E(\varepsilon) = 0$ with a variance s_i^2 . The input g -values g_{ij} are associated with standard uncertainties u_{ij} that besides u_{raw} also include error contributions due to vertical transfer and temporal variations, too.

Equation (1) may be rewritten in matrix form as

$$y = X\beta + e \tag{2}$$

where y is the column vector of all measured g -values, e is the column vector of random errors, X is the design matrix representing the functional relationship between observations,

sites and gravimeters, and $\beta = \begin{pmatrix} g \\ \delta \end{pmatrix}$ represents a column vector of the unknown reference values g at each site and biases δ for each gravimeter. It is clear that since the measurements are realized by different gravimeters at different times and sites, the input covariance matrix V associated with y will not be composed of equal diagonal elements, and weighted least squares have to be applied for the solution of (2). The corresponding weight matrix is $W = V^{-1}$, assuming the standard error of unit weight is equal to one.

As the set of observation equations has no unique solution, a constraint, which can be interpreted as the definition of the consensus reference values, is required (White 2004). The reference values in absolute gravimetry, similarly as for comparisons where the true value of the artefact is unknown, are obtained by constraining biases of participating laboratories/gravimeters as

$$\sum_{i=1}^n w_i \delta_i = l, \tag{3}$$

where w_i are the weights assigned to each of the n gravimeters and l is the linking converter, discussed in Sect. 2.3. The constraint (3) can be described in matrix form using the column vector of weights w and consequently by the column vector of constraint coefficients c by one linear equation

$$\begin{pmatrix} 0 & w^T \end{pmatrix} \begin{pmatrix} g \\ \delta \end{pmatrix} = c^T \beta = l \tag{4}$$

The linearly constrained weighted least squares problem that minimizes $e^T W e$ and fulfils the constraint (4) can be solved by the Lagrangian approach in the following form

$$\begin{pmatrix} X^T W X & c \\ c^T & 0 \end{pmatrix} \begin{pmatrix} \beta \\ k \end{pmatrix} = \begin{pmatrix} X^T W y \\ l \end{pmatrix} \tag{5}$$

where k is the unknown Lagrange coefficient (a scalar quantity because only one constraint has been defined). In textbooks on adjustment theory, e.g. Reissmann (1976), it is shown that the inverse of the normal equations (the covariance matrix of unknown parameters) has the following configuration

$$\begin{pmatrix} X^T W X & c \\ c^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} V_{\beta\beta} & V_{\beta k} \\ V_{\beta k}^T & V_{kk} \end{pmatrix} \tag{6}$$

By this, the column matrix of adjusted g -values and biases can be expressed as the linear relation between unknowns and measurements

$$\beta = V_{\beta\beta} X^T W y + V_{\beta k} l = R y + V_{\beta k} l. \tag{7}$$

In our case, with one constraint defined by Eq. (3), $V_{\beta k}$ is a column vector containing the value of 1 and -1 for biases δ and reference values g , respectively. It is just because the linking converter only shifts the mean level by l . A positive shift of the biases by l causes a decrease in reference values by $-l$. Therefore, the covariance matrix associated with estimated unknowns is obtained as

$$\text{cov}(\beta) = R V R^T + V_{ll} I, \tag{8}$$

where V_{ll} is the variance of the linking converter, represented by the squared uncertainty of the linking converter, and I is the identity matrix.

2.1 Harmonization of uncertainties

Not all non-NMI/DI laboratories present the full uncertainty budget, and therefore, it is legitimate to assume that some error sources might have left unaccounted or underestimated. This fact becomes obvious from the declared measurement uncertainties of the FG5/X gravimeters at all the comparisons except for 2018, where non-NMI/DIs declared uncertainties were roughly lower by 20% for the same type of gravimeters. This fact is causing troubles especially in case of the weighted constraint, and therefore, we decided for a harmonization of the FG5/X gravimeters similarly as in Pálinkáš et al. (2017). Such a harmonization should ensure more realistic weighting of the g -values in Eq. (1) and especially in the constraint given by Eq. (3). We have determined the average uncertainty of 2.4 μGal from uncertainties declared by NMI/DIs for the FG5/X gravimeters. Declared uncertainties of those FG5/X gravimeters that have been lower than this value were changed to 2.4 μGal in the ICN solution. In case of the KCN solution, only the uncertainties of non-NMI/DIs were harmonized.

2.2 Construction of the input covariance/weight matrix

As discussed in Koo and Clare (2012), the special importance in analyses of comparison results is the construction of the input covariance matrix V associated with the measurements y . For comparison results documented in gravimetry, always a diagonal weight matrix $W = V^{-1}$ was applied, obtained from the diagonal covariance matrix V based on the squared declared uncertainties. It is important to note that the weight matrix has no impact on the definition of the mean level of the reference values (g -values in absolute sense), since it is only connected with Eq. (1), where all absolute gravity measurements are treated relative to the (unknown) reference values for each site and the systematic deviation of each gravimeter (bias) to this value. Similarly, it would be possible to include measurements of a relative gravimeter

to Eq. (1). Therefore, it is adequate to include the measurements of non-NMI/DI gravimeters in the official metrological key comparison solution and by this to improve the robustness of the gravity differences between reference values. The covariance or weighting matrix within the constrained least squares approach describes the stochastic component of the functional model and therefore only the capability of an absolute gravimeter to determine gravity values in a relative sense, as observed by a relative gravimeter. Therefore, strictly speaking, the covariance matrix should reflect the measurement repeatability instead of the uncertainty and this is even obvious from the definition of ϵ in Eq. (1). However, excluding the contribution of systematic error components from uncertainties, error propagation leads to inappropriate covariance estimates from Eq. (8) for the unknowns. Nevertheless, results published in Koo and Clare (2012) show that there is no difference in the estimate of the unknown parameters, regardless of whether V is constructed only as diagonal matrix from the measurement repeatability s^2 or from the measurement uncertainty u^2 , once the correlations between measurements will be accounted in non-diagonal elements of V as $cov = u^2 - s^2$. Therefore, the same covariance matrices can be used for the determination of the unknowns in Eq. (7) and their error estimates by Eq. (8).

As pointed out in White (2004), the uncertainties reported by the laboratories should be separated into two parts: "...that characterizing the laboratory repeatability and that characterizing the range of values that may reasonably be attributed to the laboratory bias". Do we have enough information to provide such a separation in absolute gravimetry? Certainly yes, at least for the most common FG5/X gravimeter of which the repeatability is known from combination of measurements with superconducting gravimeters at reference stations as mentioned in Sect. 1. Further, repeatability of different gravimeters has been computed from the scatter of measurements of a particular gravimeter with respect to the reference values at comparisons. From the published results (Francis et al. 2012, 2015; Jiang et al. 2012; Pálinkáš et al. 2017), the repeatability of an FG5/X is about $s = 1.2 \mu\text{Gal}$, and it means roughly half of the typical FG5/X uncertainty ($u = 2.4 \mu\text{Gal}$). Generally, this information means that measurements of a particular gravimeter are correlated which should be reflected in the input covariance matrix V by introducing non-diagonal elements at least for the measurements with the same gravimeter. For the given example of FG5/X gravimeter with uncertainty of $2.4 \mu\text{Gal}$ and repeatability of $1.2 \mu\text{Gal}$, the covariance will be $cov = u^2 - s^2 = 4.32 \mu\text{Gal}^2$, describing the joint variability of an FG5/X due to errors common to all measurements of the laboratory expressed by the correlation coefficient of $\rho = (u^2 - s^2)/u^2 = 0.75$. Therefore, we used this correlation coefficient to determine the covariances for a particular FG5/X "i" from the harmonized uncertainties u_{ij} as $cov = 0.75 u_{ij, \min}^2$, where $u_{ij, \min}$ is the minimum of

all u_{ij} . This approach can be understood as if the measurements of a particular gravimeter carried out within a few days are affected by a group of systematic errors that remain the same.

For a few other types of instruments as A-10, CAG-01, a similar ratio between repeatability and declared uncertainty has been published (Falk et al. 2012; Karcher et al. 2018); therefore, we decided to apply the approach described for FG5/X for all other type of gravimeters, except the rise and fall type of gravimeter IMG-02 (D'Agostino et al. 2008), where random errors are dominating in the error budget (A. Prato, personal communication), and covariances have been set to zero.

A comparison between the standard approach (V is diagonal with elements u^2) and our approach where covariances $cov = u^2 - s^2$ with the correlation coefficient of 0.75 appear in off-diagonal elements showed that the influence on the estimates of unknowns is negligible, with differences below $0.1 \mu\text{Gal}$. However, as expected, the error estimates are significantly different as shown in Fig. 1. Here, the variances of the estimated biases are roughly twice as large as those obtained without accounting for the correlations.

2.3 Construction of the constraint

The constraint given by Eq. (3) is defining the consensus value of comparisons in gravimetry, since the true g -values are unknown. The weighted constraint was used for processing of the comparisons in 2009, 2015, 2017 and 2018 (Jiang et al. 2012; Pálinkáš et al. 2017; Wu et al. 2020; Falk et al. 2020). On the other hand, a non-weighted constraint was used for processing comparisons in 2011 and 2013 (Francis et al. 2012, 2015).

The linking converter l in Eq. (3) is conventionally taken to be zero in CCM key comparisons. However, in regional comparisons that have to be linked to a CCM comparison, rigorously l should be computed as (weighted) average of biases achieved at this CCM comparison for those gravimeters that provide the link. In such a case, the sum of respective weights in Eq. (3) must be $\sum w_i = 1$.

In de Viron et al. (2011), it was recommended to use a constraint which minimizes the L1 norm of the biases instead of imposing zero mean of biases.

As it can be seen, several possibilities have to be considered in the construction of the constraint. Due to its importance, the following aspects are elaborated:

- A. Derivation of the weights. Laboratories participating at comparisons are treated as independent, even when they are using the same type of instrument. Of course, some of the error sources might be therefore common to these instruments, but it can be only hardly captured, and thus, the results of all the comparisons presented

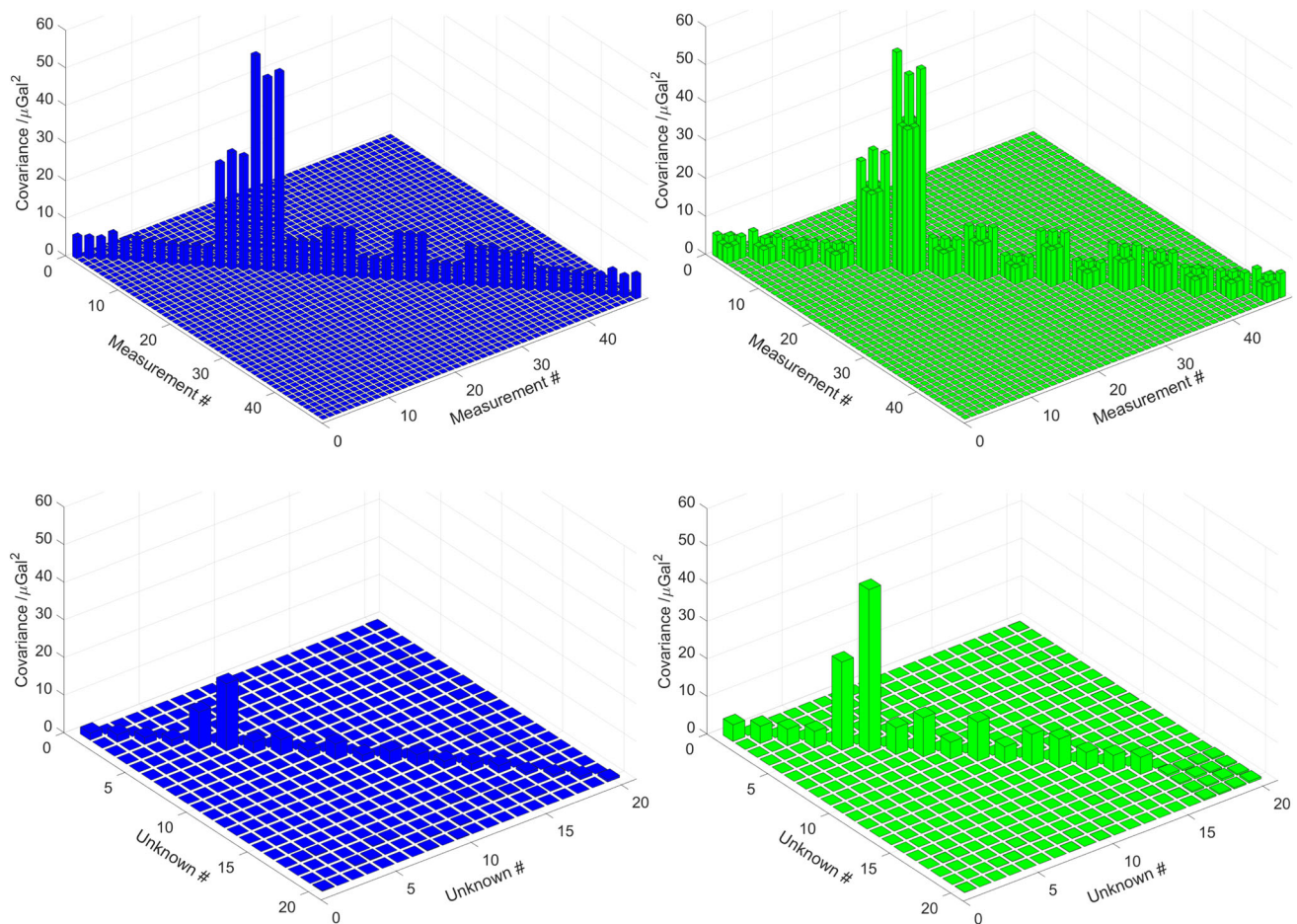


Fig. 1 Comparison of covariance matrices related to the measurements (Top) and the estimates of unknowns (Bottom) for comparison in 2018 with 48 measurements of 16 gravimeters (unknowns #1–#16) at 4 sites (unknowns #17–#20). Left: All the measurements are treated as

independent with diagonal elements computed as square of declared uncertainties. Right: Correlation with coefficient of $\rho = 0.75$ is taken into account for measurements of a particular gravimeter

here will be FG5/X dependent as it can be expected from the dominance of these meters in Table 1. The weighted constraint was used for all the results presented here, and the respective weights are computed from declared uncertainties in case of the KCN solution and from harmonized uncertainties (see Sect. 2.1) in case of ICN solutions. It turns to the question which uncertainty estimate u_i from several measurements of a particular gravimeter should be associated with the weight $w_i = u_0^2/u_i^2$ in the constraint? Usually, all the declared uncertainties for a particular gravimeter are the same within a range of 10% and commonly the root mean square of the declared uncertainties u_{ij} of each laboratory is used to compute u_i . Nevertheless, this approach fails in case of large discrepancies between uncertainty estimates between several observations of a particular gravimeter. Another possibility is to estimate the uncertainty of the mean value analogically to the standard error

of the weighted mean as $u_i = \sqrt{\frac{1}{\sum u_{ij}^{-2}}}$, that is, however, correct only for uncorrelated observations, which is not applicable for the majority of gravimeters with dominating contributions from systematic error sources. Therefore, the rigorous way has to account for correlations as discussed in Sect. 2.2. It means separating the error contributions that will not decrease with the number of measurements due to averaging. Nevertheless, such a computation is impractical and also inaccurate due to the limitations to estimate the true value of the correlation coefficient. Therefore, an appropriate and easy possibility is to use the *minimum* uncertainty from declared/harmonized uncertainties $u_i = \min(u_{ij})$ for a particular gravimeter, practically saying that the contribution of a gravimeter to the realization of the reference value will not be smaller than its best measurement having the lowest uncertainty.

B. Contribution of laboratories and linking converter. In case of the ICN solution, all the laboratories (except those showing incompatibility, see Sect. 2.4), will contribute to the constraint based on the harmonized uncertainties. Further l will be set equal to zero, which means no difference between the CCM and regional comparison will be considered, expecting that the cloud of participating gravimeters (at least 16 for all comparisons) defines a consistent reference independently. In case of the KCN solution, the contributions/weights of non-NMI/DIs have to be zero in the constraint. Further, the linking converter should be applied for regional comparisons. It is important to test the quality of the link by the differences between the biases obtained from both (CCM and regional) comparisons for those gravimeters providing the link, see Fig. 2. The variability of such differences is independent on the choice of the constraint and allows us to test whether the gravimeters were stable enough to keep their mutual deviations from the CCM to the regional comparison (organized few years later). In an ideal case, all the differences should stay within the repeatability of gravimeters. As it can be seen from Fig. 2, the appropriate link has been done only in 2018. In contrast, it can be clearly seen that the mutual differences between biases have changed for FG5-209 and FG5-215 between 2011 and 2009, similarly as for FG5-215 and FG5-221 between 2015 and 2013. Therefore, in the results presented in Sect. 3, we set the linking converters for KCN solutions to 0.0 μGal , 0.0 μGal and (-0.78 ± 1.26) μGal for the comparisons 2011, 2015 and 2018, respectively. These values differ from the original KC solution only for the comparison in 2015, where the linking converter of + 0.32 μGal was used, but as it is shown in Fig. 2, the quality of the link is weak and the approach as in 2011 should have been followed ($l = 0$).

C. The solution minimizing the L1 norm of biases was computed numerically. The ‘‘L1 norm’’ results have been achieved from ‘‘zero mean’’ results, by shifting the biases by a value δ_c in the range of ± 10 μGal with a step of 0.01 μGal . Finally, we determined such a δ_c for which:

$$\sum_{i=1}^n |w_i(\delta_i + \delta_c)| = \min. \tag{9}$$

2.4 Consistency check

According to Cox (2002), a Chi-square test should be used for testing the overall consistency of the results, before accepting

consensus reference values of the comparison. In case of AG comparisons, we have the possibility to test directly the biases

$$\chi_{\text{ave}}^2 = \sum_{i=1}^n \frac{\delta_i}{u_i^2} \tag{10}$$

or the complete set of deviations from reference values $d_{ij} = g_{ij} - g_j$ expressed by the column vector \mathbf{d} [Eq. (16)]. In this case, we suppose that the related Chi-square value has to be computed from the variance matrix to take into account correlations according to Woolliams et al. (2006)

$$\chi_{\text{obs}}^2 = \mathbf{d}^T \mathbf{V}^{-1} \mathbf{d}. \tag{11}$$

To specify the critical Chi-square value with probability of 5% for statistics given by Eqs. (10) and (11), the degrees of freedom have to be assigned.

The Chi-square test has not been used in published results of AG comparisons. Nevertheless, inconsistent measurements were investigated based on normalized deviations (Newell et al. 2017)

$$r_{ij} = \frac{(g_{ij} - g_j)}{u_{ij}} \tag{12}$$

or the compatibility index (Francis et al. 2015)

$$E_{ij} = \frac{(g_{ij} - g_j)}{\sqrt{u_{ij}^2 + u_j^2}}. \tag{13}$$

However, these equations are not in agreement with the error propagation after the adjustment, since measurements and reference values are correlated and this is not taken into account in the denominators of Eqs. (12) and (13). As, for example, given in Cox (2002), for a reference value g_{ref} computed as simple weighted mean, the deviation of institute i results in the variance $u^2(d_i) = u^2(g_i) - u^2(g_{\text{ref}})$. Similarly, to find out the correct uncertainty estimate for the model given by Eqs. (1) and (3), the covariance matrix of deviations has to be computed.

The matrix \mathbf{R} from Eq. (7) expresses the unknowns (biases and reference values) as the linear combination of measurements. The complete vector of measurements and estimated unknowns can be written (without accounting for the linking converter) as

$$\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{R} \end{pmatrix} \mathbf{y} \tag{14}$$

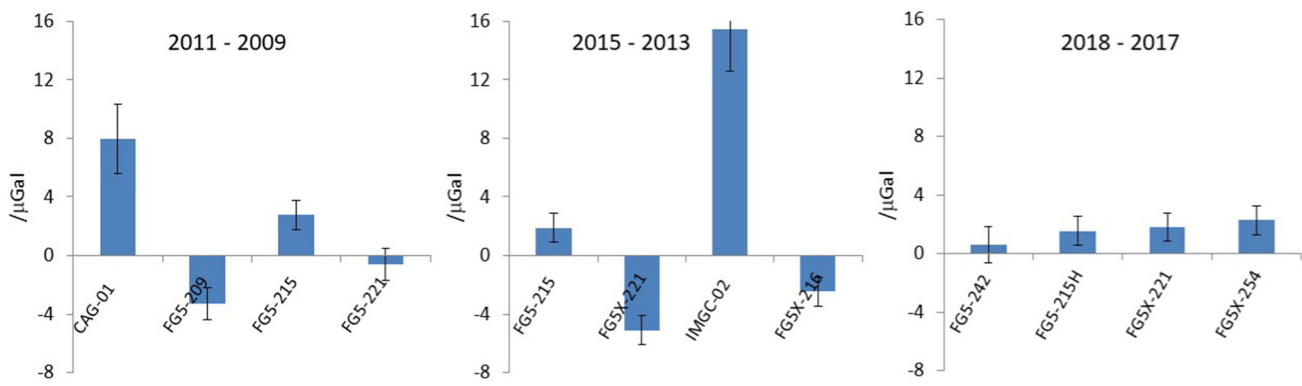


Fig. 2 Quality of the link carried out at comparisons in 2011, 2015 and 2018 is expressed as differences between biases (from the ICN solutions) at regional and CCM comparison (in 2009, 2013 and 2017) of those gravimeters that provide the link. Error bars have been determined

based on repeatabilities of gravimeters. *Note* Correction for systematic effects (self-attraction, diffraction) was not unified between comparisons

where I is the identity matrix with size (n_o, n_o) , and n_o is the number of measurements. The associated covariance matrix is then

$$\text{cov} \begin{pmatrix} y \\ \beta \end{pmatrix} = \begin{pmatrix} I \\ R \end{pmatrix} V \begin{pmatrix} I \\ R \end{pmatrix}^T = V_{all} \quad (15)$$

reflecting correlations between measurements and unknowns according to the functional model. Therefore, the matrix from Eq. (15) can be used to express covariances for any functional dependency between measurements and estimated unknowns, including the deviations from reference values: $d_{ij} = g_{ij} - g_j$. These deviations can be described by a design matrix A_d (composed of values 0, 1 and -1) as

$$d = A_d \begin{pmatrix} y \\ \beta \end{pmatrix} \quad (16)$$

with the associated covariance matrix

$$\text{cov}(d) = A_d V_{all} A_d^T + V_{ll} I, \quad (17)$$

where the variance of the linking converter V_{ll} should be included. Finally, the compatibility index is

$$E_{ij} = \frac{(g_{ij} - g_j)}{u(d_{ij})}, \quad (18)$$

where the uncertainty of deviations $u(d_{ij})$ is achieved as square root of the diagonal elements of $\text{cov}(d)$. Note that $\sqrt{u_{ij}^2 - u_j^2} \leq u(d_{ij}) \leq u_{ij}$ is always valid and outliers should be identified according to the level of confidence with the coverage factor k as $|E_{ij}| > k$.

If a gravimeter does not meet the compatibility criteria, it should not contribute to the definition of the reference value (weight = 0 within the constraint). The treatment of

incompatible measurements within the weighting matrix is, however, a different issue since the repeatability is a parameter that is important in this case, i.e., a gravimeter with a significant bias should not be automatically excluded from the weighting matrix. For those gravimeters i that showed incompatible measurements at the confidence level of 95% ($k = 2$), we have computed also the parameter of repeatability and following modifications have been done in the adjustment for the ICN solution: (1) if more than one measurement of a gravimeter showed incompatibility, such a gravimeter was excluded to contribute to the constraint, (2) if one measurement was incompatible, the harmonized uncertainties were increased by 50%, and (3) if the repeatability of measurements was higher than the harmonized uncertainty, the harmonized uncertainty was enlarged by 50% that consequently reduced the contribution of this gravimeter to the constraint and the parameter estimate too. The increase in uncertainties by 50% is therefore just a parameter of a cycle for progressive adaptation of uncertainties within the stochastic model. In case of the KCN solutions, we followed the rules applied in published KC solutions, non-compatible measurements of NMI/DIs have been excluded and the harmonized uncertainties of non-NMI/DIs were used in the variance matrix as in the ICN solution.

2.5 Degree of equivalence and related uncertainty estimates

In metrology, the degree of equivalence (DoE) of a measurement standard is quantitatively defined as the deviation from the key comparison reference value. Therefore, the biases determined from the adjustment should be equivalent with DoE. Nevertheless, the harmonization of uncertainties described in Sect. 2.1 and the treatment of inconsistent measurements explained in Sect. 2.4 mean that individual

deviations from the reference value of a particular gravimeter are not weighted exactly according to the assumptions of the operators. Therefore, those DoE should be computed according to Jiang et al. (2012) as weighted average of the deviations from reference values, using the formula

$$D_i = \left[\sum w_{ij}(g_{ij} - g_j) \right] / \sum w_{ij}, \quad (19)$$

where weights are computed based on the declared uncertainties $w_{ij} = 1/u_{ij}^2$. To derive the covariance estimates of DoEs, it is necessary to construct a design matrix \mathbf{A}_D that contains, in correspondence with Eq. (19), the ratios $w_{ij}/\sum w_{ij}$ and zeros. Further, analogically to Eq. (17), we have to derive the covariance matrix of individual deviations, however, in this case based on declared uncertainties, to determine the error propagation according to the declared estimates. To do so, it is necessary to construct a matrix \mathbf{V}_{dec} (similarly to \mathbf{V}) from squared declared uncertainties as diagonal elements and treating non-diagonal elements using the assumed correlation coefficients. Consequently, \mathbf{V}_{dec} instead of \mathbf{V} is propagated according to Eq. (15), achieving \mathbf{V}_{alldec} instead of \mathbf{V}_{all} and analogically to Eq. (17) resulting in $\text{cov}(\mathbf{d}_{dec}) = \mathbf{A}_d \mathbf{V}_{alldec} \mathbf{A}_d^T \{+V_{ll} \mathbf{I}\}$ to determine the covariance matrix of the DoE

$$\text{cov}(\mathbf{D}) = \mathbf{A}_D \text{cov}(\mathbf{d}_{dec}) \mathbf{A}_D^T \{+V_{ll} \mathbf{I}\}. \quad (20)$$

The discussion concerning uncertainties of individual deviations might be assumed as marginal and fruitless. Nevertheless, the main goal of a comparison is to achieve not only the deviations of the laboratories from the reference value but also the associated uncertainty estimates that indicate the laboratories if their measurements are consistent within the declared uncertainties.

In the past, different approaches were used to determine the uncertainty of DoE. Since no correlations were assumed for the covariance matrix of observations, mathematically correct estimates given by Eq. (8) were too optimistic and were therefore not used. As pointed out in Francis et al. (2015): “It can be shown that with increasing N (number of measurements for a particular gravimeter) the uncertainty of the DoE determined in this way decreases approximately in proportion to $1/\sqrt{N}$. Thus this uncertainty is not appropriate for assessing the compatibility of the DoE with the declared uncertainty of the gravimeter. Using it effectively implies an uncertainty model where with increasing N the DoE of a gravimeter should converge towards zero for the gravimeter to stay in equivalence”. Therefore, the DoE was linked to the RMS of the uncertainties of the differences between the gravimeter measurements and the KCRV by summing up variances of measurement and KCRV. However, this solution is incorrect for two reasons: (1) As already discussed in Sect. 2.4, the uncertainty of the difference should

be smaller than the uncertainty of the measurements itself because the correlation with reference values has to be taken into account. (2) The RMS cannot be a good estimate as it can be easily demonstrated when measurements of a particular gravimeter have different uncertainties, because the weighted average of a set of deviations cannot be associated with higher uncertainty than the lowest uncertainty from individual measurements.

In the final check of consistency, the square root of diagonal elements of $\text{cov}(\mathbf{D})$ given by Eq. (20) should be compared with DoE at a given confidence level.

2.6 Mean square error of the unit weight

As it can be seen from Eqs. (8) or (20), for obtaining uncertainty estimates for the unknown parameters, the covariance matrices have not been multiplied/scaled by the mean square error of the unit weight (MSE)

$$s_0^2 = \frac{\boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon}}{\text{DoF}} \quad (21)$$

with DoF degrees of freedom and $\boldsymbol{\varepsilon}$ as the vector of residuals from Eq. (1) estimated from the adjustment. It means that the given uncertainty estimates (square root of diagonal elements of the respective covariance matrix) are based directly on the declared or harmonized uncertainties, the assumed correlation between measurements and the error propagation through the adjustment, in agreement with White (2004) and Koo and Clare (2012). This approach (non-scaled covariance matrix) is not usual in geodesy. However, the main role of the comparison in metrology is to check whether the declared uncertainties are consistent with the measurement. Therefore, the input covariance matrix \mathbf{V} can be understood as built on known/a priori uncertainty estimates and the goal is not to obtain an a posteriori estimate based on dispersion of residuals, but to determine how the a priori errors are propagated through the functional model. The relevance of such an approach is demonstrated in Fig. 3, where both, scaled and non-scaled estimates of bias uncertainties for different choices of the correlation coefficient, are shown for the comparison in 2018. It is evident that the scaled estimates are significantly depending on the choice of the correlation factor since the estimates of s_0 are changing by 100% from 0.4 to 0.8 μGal , for a correlation coefficient of 0.0 and 0.75, respectively.

The MSE should reach $s_0^2 \cong 1$ for measurements associated with proper covariance matrix and a complete functional model. Therefore, s_0^2 can be understood as a measure for the goodness of fit of the model to the measurement results and can be used as an overall consistency test between model and measurements by comparing $\text{DoF} \cdot s_0^2$ with the expected

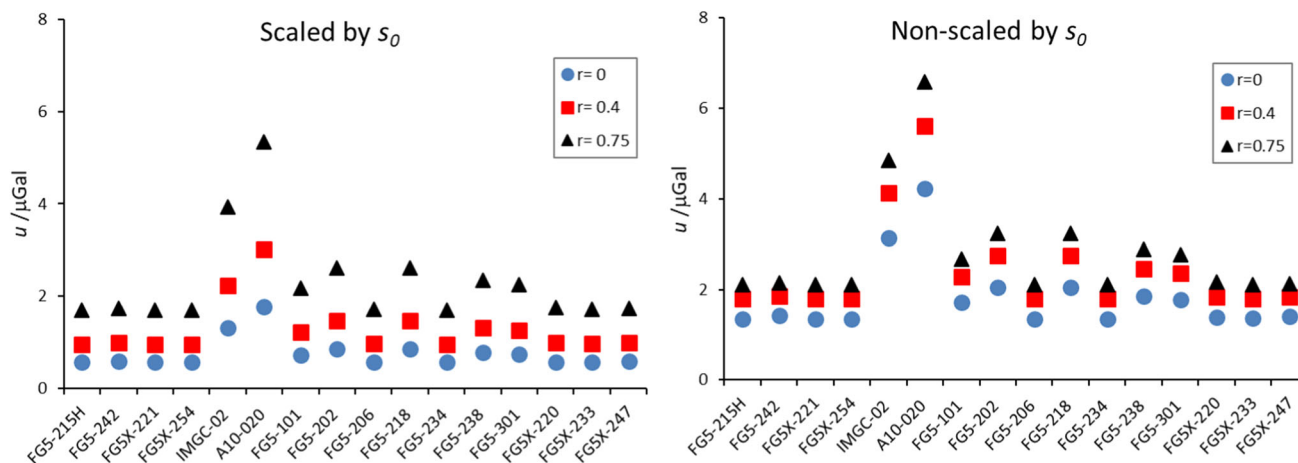


Fig. 3 Uncertainty estimates for the bias (ICN solution) of gravimeters participating in the comparison of 2018 depending on the choice of the correlation coefficient between measurements of the same gravimeter and scaling/non-scaling of the covariance matrix given by Eq. (8) by the mean square error s_0

value from the Chi-square distribution DoF with the standard deviation $\sqrt{2 \cdot \text{DoF}}$ (Sutton 2004).

3 Results of reprocessing

Tables 3, 4, 5, 6, 7, 8 and 9 in Appendix show the results of comparison reprocessing according to the approach described in Sect. 2. The estimated parameters of the adjustment are the gravimeter's biases and the reference gravity values. The bias estimates are related to the harmonized uncertainties and therefore are not so strongly influenced by a potential overestimation of the measurement accuracy of some laboratories. On the other hand, degrees of equivalence (DoE) are related to the declared uncertainties and should be used to validate the equivalence of the estimated deviations of a laboratory/gravimeter from the reference values, including all observations. If the harmonized uncertainties of a particular gravimeter are higher than the declared ones, the error estimates for biases are also higher than those for DoEs. It is exactly the reason, why a few non-NMI/DI gravimeters have been found to be inconsistent—due to declared uncertainties which are too low, see Tables 3, 4, 5, 6, 7 and 8. Generally, our approach of computing DoE uncertainties causes narrower limits to indicate a gravimeter inconsistency as the approaches published previously.

Uncertainty estimates of reference values of the ICN solution are (20–50)% smaller than those of the KCN solution. This is caused by the larger number of gravimeters used for constraining the adjustment, and it clearly indicates the advantage to include the non-NMI/DI gravimeters in the comparisons. The ICN solution cannot be used as official results since key comparisons are mandatory according to the Strategy paper (Marti et al. 2014). Nevertheless, it is a

valuable test of the robustness of the KCN solution and therefore very useful especially in regional comparisons, where the link may be uncertain (see Sect. 2.3). Except of the huge benefit from the non-NMI/DIs for the precise determination of relative ties between reference values, it is another argument to keep the concept of joint comparisons of all absolute gravimeters.

The main results of the reprocessing are summarized in Table 2. On average, the differences between official reference values (KC) and corresponding new reprocessing (KCN) are up to $0.5 \mu\text{Gal}$. The systematic difference between ICN and KCN solutions is due to the different definition of the constraint. Surprisingly, the largest difference of $1.5 \mu\text{Gal}$ can be seen for the comparison in 2018, where the link was done in a rigorous way, accounting for the linking converter with a precision of about $0.6 \mu\text{Gal}$. It therefore indicates that the reference level of the cloud of all gravimeters in 2017 and 2018 would be different by $1.5 \mu\text{Gal}$. This example shows how important an additional solution to the KC is to check the robustness and reliability of the results.

As expected, the difference between the approaches constraining zero-mean biases and minimizing the L1 norm of biases reaches values exceeding $1 \mu\text{Gal}$ only for the KCN solutions of regional comparisons, due to the small number of NMI/DI gravimeters that are contributing to the definition of the constraint. Naturally, the results of the ICN solution are therefore statistically more robust on the definition of the constraint. Generally, it is worth to apply also the L1 norm statistics to the comparisons, since it again brings additional information on the robustness of the results.

The mean square errors s_0 are ranging between 0.73 and $1.27 \mu\text{Gal}$ for all the comparisons. In case correlations between measurements of a particular gravimeter would not be accounted, s_0 would always drop below $0.7 \mu\text{Gal}$, which

Table 2 Main results of the reprocessing of comparisons

Comparison Year	CCM.G-K1 2009	EURAMET.M.G-K1 2011	CCM.G-K2 2013	EURAMET.M.G-K2 2015	CCM.G-K2.2017 2017	EURAMET.M.G-K3 2018
ICN/KCN/ μ Gal	+ 0.9	− 0.3	+ 0.4	+ 1.2	+ 0.1	− 1.5
KCN-KC/ μ Gal	− 0.1	− 0.4	− 0.2	+ 0.2	0.0	0.0
L1 Norm δ_c ICN/KCN/ μ Gal	+ 0.1/+ 0.1	0.0/+ 0.3	− 0.6/+ 0.2	+ 1.0/+ 1.6	− 0.2/+ 0.1	− 0.2/+ 1.1
s_0/μ Gal	1.12	1.09	1.27	0.87	1.27	0.73
DoF $\cdot s_0^2$	47.7	34.5	54.8	22.7	122.6	15.5
DoF $\pm \sqrt{2 \cdot \text{DoF}}$	38 \pm 8.7	29 \pm 7.6	34 \pm 8.2	30 \pm 7.7	76 \pm 12.3	29 \pm 7.6

ICN/KCN is the average difference between both new solutions with different contributions to the constraint. All gravimeters are used to determine the reference level in the ICN solution and only non-NMI/DIs in the KCN solution. KCN-KC is the average difference between KCN solution and the solutions of the official comparison report

s_0 is the root-mean-square error, *DoF* is the degree of freedom, and δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

clearly indicates an inconsistency between stochastic model and measurements. Table 2 contains also a Chi-square test of s_0^2 consistency according to Sect. 2.6. This test indicates a discrepancy between the measurement uncertainty (including correlations) and the consistency of the observations with the functional model. If the assumption about the input covariance matrix matches the dispersion of the residuals, the mean square errors s_0 should be close to 1. As it can be seen, the comparisons in 2013 and 2017 show a significant deviation from this assumption, reflected by the DoF and its limits. In contrast, for 2018, the consistency between observations and models is much better than reflected by the input covariance matrix. Both cases may be related to the harmonized uncertainties, the implicated repeatability and selected correlation coefficient, which may have caused an over- or underestimation in these cases.

4 Analyses of comparison results

Data achieved from reprocessing of 6 comparisons allowed us to investigate some interesting aspects of final results that helps to answer questions related to the significance of deviations between different types of gravimeters or the influence of noisy sites. Results from the ICN solutions have been used for this analysis.

4.1 Deviation between FG5 and FG5X gravimeters

Altogether, 97 biases of FG5/X gravimeters have been determined at 6 comparisons, while 6 of them have been excluded from all the next analyses due to the detected outliers at the 99% level of confidence. These outliers are related only to three gravimeters (FG5-102 in 2009, 2011 and 2013; FG5-230 in 2009 and 2011; and FG5X-247 in 2015). The data set of 91 biases seems to be normally distributed, see

Fig. 4. Nevertheless, is it correct to assume that FG5 and FG5X gravimeters share the same probability distribution of biases?

From the remaining 91 biases, 62 and 29 are related to FG5 and FG5X gravimeters, respectively. In case of FG5 biases, the mean value reaches $(0.07 \pm 0.27) \mu$ Gal with the sample standard deviation of 2.08μ Gal. In case of FG5X biases, the mean value reaches $(-0.21 \pm 0.39) \mu$ Gal with the sample standard deviation of 2.08μ Gal. Based on statistical *t* test and *F* test, we can conclude that both data sets show very high level of consistency and therefore can be described with a single distribution function of which the mean value is $(-0.02 \pm 0.22) \mu$ Gal and the sample standard deviation of 2.08μ Gal (with standard error of 0.16μ Gal). The achieved standard deviation of about 2.1μ Gal can be interpreted as an experimental estimate for the reproducibility of FG5/X gravimeters in general. It is a quite important result, since it is not related to a single instrument (as estimates given in Sect. 2), but it reflects the repeatability of the specific type of gravimeter. Further, this result shows that the experimental standard uncertainty of FG5/X gravimeters should be larger than 2.1μ Gal due to possible systematic errors of this gravimeter's type.

4.2 Comparison sites

It is shown in Sect. 4.1 that biases of FG5 and FG5X gravimeters might be treated as a random variable with a common Gaussian distribution. Nevertheless, also the aspect of invariability of the variances of biases between different comparisons has to be fulfilled for describing all the biases with a single distribution. This condition is not self-evident because comparisons have been carried out at different sites with different noise conditions. Standard deviations of FG5/X biases at particular comparisons are shown in Fig. 5. As it can be seen, the scatter of results is within their standard

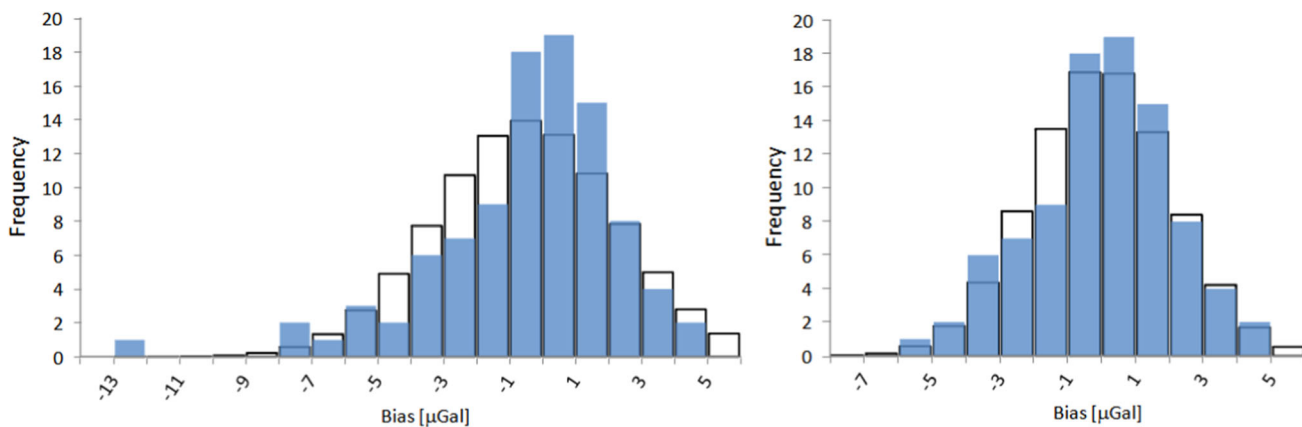


Fig. 4 Histogram of biases for FG5/X gravimeters at 6 comparisons. Left: 97 biases (filled columns) together with corresponding histogram for the normal distribution $N(-0.4, 2.8^2)$ depicted by empty columns.

Right: Histogram of 91 biases after excluding 6 outliers (filled column) together with corresponding histogram (empty columns) of the normal distribution $N(0, 2.1^2)$

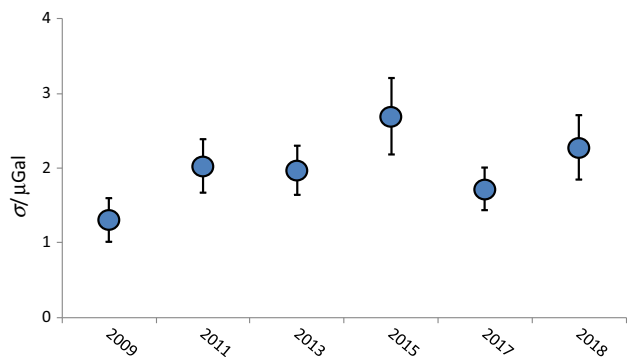


Fig. 5 Standard deviations of biases of FG5/X gravimeters achieved at 6 comparisons. Error bars represent the standard error of standard deviations

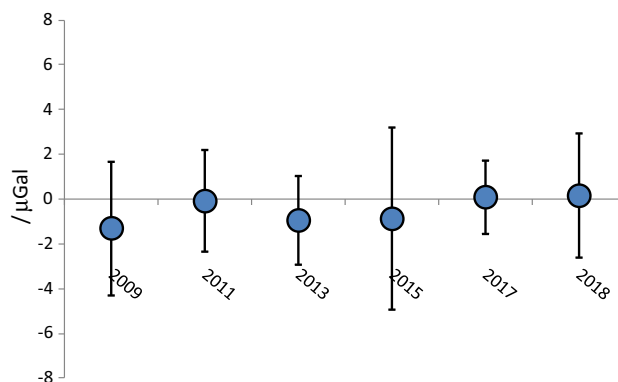


Fig. 6 Differences between average biases of FG5/X gravimeters and other type of gravimeters at particular comparisons. Error bars represent the standard uncertainty of the difference

errors associated with standard deviations. Nevertheless, the largest standard deviation is related to the comparison in 2015 in Belval, where the measurement conditions were significantly worse than for other comparisons, because the comparison sites were located close to sources of anthropogenic noise (traffic, construction works around). Further, the sites have been located on a “strong floor” (for load tests in civil engineering) that was not founded directly on the subsoil but supported by three 3 m high and 10 m long girders grounded on the building foundation. To test whether the variance at the Belval comparison is consistent with the other comparisons, the F -test statistics of 1.88 has been determined that corresponds to an 10.2% F critical value and the hypothesis on equality of standard deviations cannot be rejected at 95% confidence level so the data set might be described by a common distribution function of $N(0, 2.1^2)$. In fact, this finding shows that FG5/X gravimeters are able to reach consistent results also in case of sites with high noise level.

4.3 FG5/X bias

The determined results at comparisons are significantly influenced by FG5/X gravimeters due to their dominance and high weights in the constraint. To investigate whether there is a significant difference between FG5/X and other type gravimeters, we have divided the biases of gravimeters into these two groups and corresponding mean differences have been computed for all the comparisons. As it can be seen in Fig. 6, all these differences are within the respective standard uncertainties. Nevertheless, due to higher uncertainties of non-FG5/X gravimeters, the standard uncertainties of the mean differences are ranging from 1.7 μGal to 4.1 μGal for individual comparisons. The weighted average of all differences reaches $(-0.4 \pm 1.0) \mu\text{Gal}$.

4.4 Bias variability

Variability of biases from the ICN solutions for gravimeters participating at more than 3 comparisons is shown in Fig. 7. Significant changes are obvious for FG5-102 after its rebuilt to FG5X-102 in 2014. On the other hand, highly consistent results are documented for FG5-202 or FG5-242. Also, slight transient changes are visible for FG5-301 and FG5X-302 which barely exceed the uncertainty. In case of FG5-301, a change could also be related to service by the manufacturer in 2016.

5 Summary and conclusions

A general procedure to process and analyse comparisons of absolute gravimeters by the method of constrained least squares has been described, focusing on several details which have been treated differently in the past, addressing specifically the separation into groups of gravimeters operated by metrology laboratories and the geodetic/geoscience community. It has been shown that for a reliable and thorough evaluation of comparison results, it is necessary to distinguish between stochastic and systematic contributions within the frame of declared uncertainties of gravimeters. It allows to construct a realistic covariance matrix of measurements which allows to rigorously follow the law of error propagation from the observational data to the measure of compatibility. This approach was described mathematically in detail and applied for 6 comparisons where the covariances between measurements of a particular gravimeter have been accounted. Based on a ratio between repeatability and uncertainty, a correlation coefficient reaching up to 75% was applied. Compared to the common approach without accounting for correlations, the estimated unknown parameters, biases and reference values, were affected only slightly (within a range of 0.2 μGal). However, the estimated variances of these parameters are much more realistic, achieving roughly doubled values. The new approach for computing DoE uncertainties results in smaller values than the approach used in the past. Therefore, more gravimeters seem to be inconsistent with declared uncertainties as published previously. In majority, this is the case for non-NMI/DI gravimeters that are tending to underestimate their measurement uncertainties. Of course, the constraint might be influenced accordingly and in fact, from scientific point of view, this is the only argument against a solution where all gravimeters are treated equally. Nevertheless, as we showed, this can be solved by a harmonization of uncertainties.

As it can be anticipated, the main impact on the determination of reference values and biases is given by the definition of the constraint. We demonstrated that a solution (ICN) in which gravimeters are not separated into groups according to

metrological institutes or geodetic operators, yields parameter estimates which are naturally more robust because more gravimeters contribute to the constraint. This solution should be used in testing the robustness of official key comparisons results as well. It is needed especially for the regional (e.g. EURAMET) comparisons where the link to international (CIPM) comparisons is provided only by a few linking gravimeters. We showed that for EURAMET comparisons in 2011 and 2015, carried out two years after CIPM comparisons, 4 gravimeters providing the link were not able to keep their biases within the expected precisions. On the other hand, in 2018, when the EURAMET comparison has been organized one year after the CIPM comparison, the link has been realized with a precision well below 1 μGal .

Differences between the published and newly presented solutions (including those based on L1 norm) are within a range of 2 μGal for all the reprocessed comparisons. Such a difference is still in agreement with the presented uncertainty estimates for the reference values (0.7–1.5 μGal). Of course, the fact that all comparisons are affected by significant dominance of FG5/X gravimeters is not reflected in the error estimates, since all the gravimeters were treated as independent. Nevertheless, generally, the method of elaboration of comparisons we have shown allows to include the possible correlation related to the whole set of observations of gravimeters type.

The results of uniformly processed comparisons are used to verify the statistical significance of some interesting information included in results. We compared the biases achieved by FG5 and FG5X gravimeters showing that both can be described by the same normal distribution $N(0, 2.1^2)$. Such a finding can be interpreted as the different dropping chambers of these instruments neither causing systematic biases between instruments nor differences in the reproducibility. However, the zero mean of the normal distribution for FG5/X rather stands for the consistency of this particular group of gravimeters than for bias-free results or the ability to determine the bias of these types of gravimeters with respect to other gravimeters with an accuracy well below 1 μGal . On the other hand, the obtained standard deviation of 2.1 μGal of a normal distribution represents an experimentally documented reproducibility of FG5/X.

Comparisons of absolute gravimeters and their processing play an essential role for the determination of the absolute gravity reference, important specifically in geodesy and metrology in the frame of the realization of International Reference Gravity System and the Kibble balance experiment. With this study, we introduce a consistent treatment of comparisons carried out in the past, allowing for a thorough reanalysis with regard to stability of references values over time and compatibility of applied systematic corrections. We further provide a frame for the elaboration of future comparisons (KCN and ICN solutions) as a basis for the

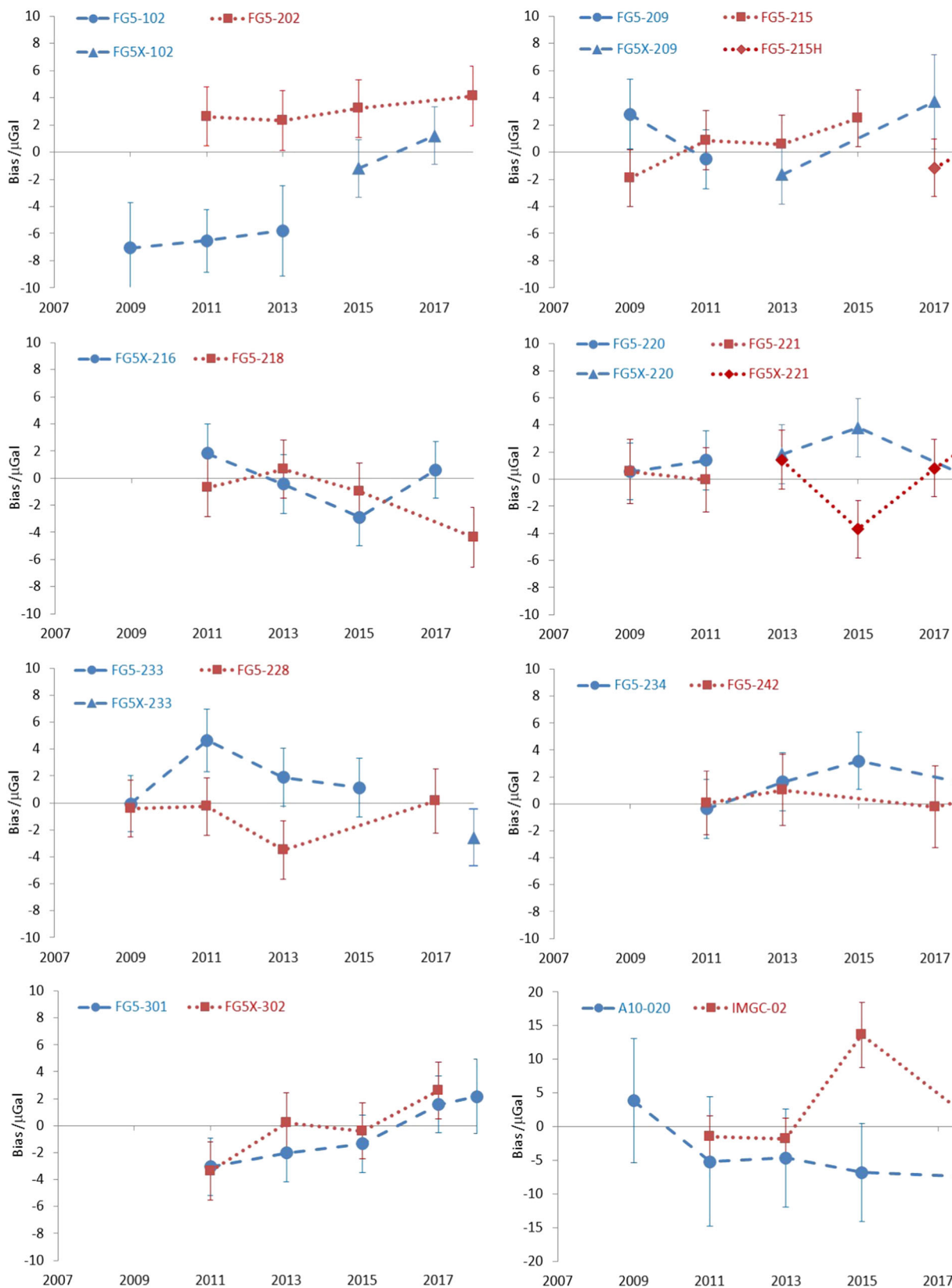


Fig. 7 Variability of biases for gravimeters participating at more than 3 elaborated comparisons. Error bars represent the standard uncertainty of determined biases. An upgrade of a gravimeter from FG5 to FG5X is

indicated by a gap in connecting lines and by different markers. In case of FG5-215, the gravimeter was not upgraded to FG5X but equipped with a new measurement system described in Křen et al. (2016)

establishment of a long-term stable gravity reference frame, compliant with the requirements in geodesy and metrology.

The official (KC) reference values published in the past differ on average by less than $0.5 \mu\text{Gal}$ from our equivalent KCN solutions, showing that adequate processing strategies have been applied in the past. The significant improvement, highlighted in this paper, is more realistic uncertainty estimates for evaluated parameters (reference values, biases, DoE). The solution including all gravimeters (ICN) which is not strictly following the rules in metrology has a huge benefit at least for testing robustness of the KC results that might be less reliable especially in regional comparisons with only a few linking gravimeters. From this point of view, the ICN solution, as the statistically most robust solution, should also be provided along with the official results and should be documented for the realization of the IGRS.

Presently, only 7 NMI/DIs have published Calibration and Measurement Capabilities in gravity, and hence, we strongly recommend to keep the model of joint comparisons of absolute gravimeters from metrology and geoscience communities for the future. It seems to us as a key factor to continue the 40-year successful cooperation between metrol-

ogy and geoscientific communities that boosted knowledge in absolute gravimetry.

Acknowledgements We would like to acknowledge the financial support from the project “Research related with International Gravity Reference System” funded by the Ministry of Education, Youth and Sports of the Czech Republic (MŠMT) as Grant No. LTT19008, and it is also supported by VÚGTK and ČMI. We are grateful to three anonymous reviewers for their helpful and constructive comments.

Author contributions V.P., H.W. and P.K. designed the research. H.W, M.V. and V.P. evaluated the data and prepared figures. V.P., H.W., R.F. and P.K. analysed the data and wrote the paper.

Data availability Data from the comparisons of absolute gravimeters are available in the Key Comparison Database (<https://www.bipm.org/kcdb/>) and cited literature. The datasets analysed in this study are available from the corresponding author on reasonable request.

Appendix

See Tables 3, 4, 5, 6, 7, 8 and 9.

Table 3 Results of comparisons in 2009 in μGal

Gravimeter	ICN solution		KCN solution		Repeatab.
	Bias	DOE	Bias	DOE	
<i>CCM.G-K1 (ICAG 2009)</i>					
NIM-2	9.0 ± 8.4	7.3 ± 5.7	5.4 ± 6.3	8.1 ± 5.6	8.6
CAG-01	-2.7 ± 8.3	-2.6 ± 5.6	-1.8 ± 5.5	-1.7 ± 5.5	9.0
JILAg-06	6.0 ± 6.7	6.5 ± 6.7	6.9 ± 6.6	7.4 ± 6.7	4.3
FGL-103	-2.7 ± 4.1	-2.7 ± 4.0	-1.9 ± 4.0	-1.9 ± 4.0	2.8
A10-005	-4.5 ± 7.3	-4.4 ± 4.5	-3.4 ± 4.4	-3.6 ± 4.4	2.8
FG5-105	0.6 ± 2.4	0.6 ± 2.4	1.5 ± 2.3	1.5 ± 2.3	1.4
FG5-209	2.8 ± 2.6	2.8 ± 2.6	3.7 ± 2.5	3.7 ± 2.5	1.4
FG5-213	-1.1 ± 2.2	-1.1 ± 2.2	-0.1 ± 2.1	-0.1 ± 2.1	0.8
FG5-215	-1.9 ± 2.1	-1.9 ± 2.1	-1.0 ± 2.0	-1.0 ± 2.0	0.2
FG5-221	0.6 ± 2.4	0.6 ± 2.4	1.5 ± 2.3	1.5 ± 2.3	1.1
FG5-224	-5.6 ± 2.6	-5.6 ± 2.6	-4.7 ± 2.4	-4.8 ± 2.4	0.6
A10-014	4.7 ± 7.3	4.7 ± 5.6	5.5 ± 7.4	5.5 ± 5.6	3.0
A10-020	3.9 ± 9.3	3.6 ± 9.3	4.8 ± 9.3	4.5 ± 9.3	4.1
MPG-2	7.8 ± 7.4	8.0 ± 7.4	8.7 ± 7.5	8.9 ± 7.5	6.6
FG5-101	-0.5 ± 2.1	-0.5 ± 1.7	0.5 ± 2.4	0.4 ± 2.0	1.8
FG5-102	-7.1 ± 3.4	-7.1 ± 2.3	-6.2 ± 3.4	-6.2 ± 2.4	3.1
FG5-220	0.6 ± 2.1	0.6 ± 2.1	1.5 ± 2.4	1.5 ± 2.4	1.0
FG5-228	-0.4 ± 2.1	-0.4 ± 2.0	0.4 ± 2.4	0.4 ± 2.2	0.5
FG5-230	-5.8 ± 2.3	-5.7 ± 2.2	-4.9 ± 2.4	-4.9 ± 2.3	1.2
FG5-233	-0.1 ± 2.1	-0.1 ± 2.1	0.8 ± 2.4	0.8 ± 2.4	0.2
FG5-238	1.4 ± 2.4	1.5 ± 2.4	2.2 ± 2.7	2.4 ± 2.7	1.6
s_0	1.12		1.26		
δ_c	+ 0.08		+ 0.14		
Site	ICN solution $g - 980920000$	KCN solution $g - 980920000$	ICN – KCN		KCN – KC ^a
B	8020.4 ± 0.8	8019.5 ± 1.0	0.9		- 0.1
B1	8013.9 ± 0.8	8012.8 ± 1.0	1.0		- 0.2
B2	7999.7 ± 0.8	7999.0 ± 1.0	0.8		0.1
B5	8021.8 ± 0.8	8020.9 ± 1.0	0.9		- 0.1
B6	8001.4 ± 0.9	8000.6 ± 1.0	0.8		- 0.1

Based on determined degree of equivalence (DOE), three/two gravimeters (bold italic format) showed inconsistency at 95% confidence with declared uncertainties for the ICN/KCN solution. Gravimeters in bold belong to NMI/DIs

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases
^aResults are compared with the KC solution labelled as A[^]KP in Jiang et al. (2012), where self-attraction and diffraction corrections have been applied to the measurements

Table 4 Results of comparisons in 2011 in μGal

Gravimeter	ICN solution		KCN solution		Repeatab.
	Bias	DOE	Bias	DOE	
<i>EURAMET.M.G-K1 (ECAG 2011)</i>					
CAG-01	5.2 ± 4.4	5.1 ± 4.4	5.0 ± 4.3	4.8 ± 4.3	2.4
FG5-209	- 0.5 ± 2.1	- 0.5 ± 2.1	- 0.8 ± 1.9	- 0.8 ± 1.9	0.7
FG5-215	0.9 ± 2.1	0.9 ± 2.2	0.6 ± 2.0	0.6 ± 2.0	1.0
FG5-221	0.0 ± 2.4	0.0 ± 2.4	- 0.3 ± 2.2	- 0.3 ± 2.2	1.5
FG5-242	0.1 ± 2.4	0.2 ± 2.4	- 0.2 ± 2.2	- 0.1 ± 2.2	1.7
IMGC-02	- 1.5 ± 3.0	- 1.5 ± 3.0	- 1.7 ± 3.0	- 1.7 ± 3.0	1.4
FG5-102	- 6.5 ± 2.3	- 6.5 ± 1.9	- 6.8 ± 2.5	- 6.8 ± 2.1	1.4
FG5-103	0.5 ± 2.2	0.5 ± 1.8	0.3 ± 2.5	0.3 ± 2.1	0.4
FG5-202	2.6 ± 2.2	2.6 ± 2.0	2.3 ± 2.5	2.3 ± 2.3	0.5
FG5-218	- 0.7 ± 2.2	- 0.7 ± 1.8	- 1.0 ± 2.5	- 0.9 ± 2.1	0.9
FG5-220	1.4 ± 2.2	1.4 ± 2.0	1.1 ± 2.5	1.1 ± 2.3	0.6
FG5-228	- 0.3 ± 2.2	- 0.3 ± 1.8	- 0.5 ± 2.5	- 0.5 ± 2.2	0.9
FG5-230	- 12.4 ± 2.3	- 12.4 ± 2.0	- 12.6 ± 2.5	- 12.6 ± 2.2	0.8
FG5-232	- 2.3 ± 2.2	- 2.4 ± 1.9	- 2.6 ± 2.5	- 2.6 ± 2.2	1.0
FG5-233	4.6 ± 2.3	4.6 ± 2.3	4.4 ± 2.5	4.4 ± 2.5	0.7
FG5-234	- 0.4 ± 2.2	- 0.4 ± 1.9	- 0.6 ± 2.5	- 0.6 ± 2.2	1.3
FG5-237	0.0 ± 2.2	0.0 ± 1.8	- 0.2 ± 2.5	- 0.3 ± 2.1	0.2
FG5-301	- 3.1 ± 2.1	- 3.1 ± 2.0	- 3.3 ± 2.5	- 3.3 ± 2.3	0.7
FG5X-216	1.8 ± 2.2	1.8 ± 1.9	1.6 ± 2.5	1.6 ± 2.3	1.4
FG5X-302	- 3.4 ± 2.2	- 3.4 ± 2.0	- 3.6 ± 2.5	- 3.6 ± 2.3	1.1
A10-020	- 5.2 ± 9.6	- 5.2 ± 9.6	- 5.5 ± 9.7	- 5.5 ± 9.7	11.4
T-1	5.5 ± 5.5	5.5 ± 5.5	5.2 ± 5.6	5.2 ± 5.6	1.2
s_0	1.09		1.09		
δ_c	+ 0.04		+ 0.31		
Site	ICN solution $g - 980960000$	KCN solution $g - 980960000$	ICN - KCN		KCN - KC
A1	4234.7 ± 1.0	4235.0 ± 1.4	- 0.3		- 0.6
A2	4222.5 ± 0.9	4222.8 ± 1.3	- 0.3		- 0.3
A3	4213.9 ± 0.9	4214.2 ± 1.3	- 0.3		- 0.3
A4	4200.1 ± 1.2	4200.4 ± 1.5	- 0.3		- 0.4
A5	4188.7 ± 0.9	4189.0 ± 1.3	- 0.3		- 0.4
B1	4084.0 ± 1.0	4084.2 ± 1.3	- 0.3		- 0.2
B2	4075.6 ± 0.9	4075.9 ± 1.3	- 0.3		0.5
B3	4075.1 ± 0.9	4075.3 ± 1.3	- 0.3		- 0.8
B4	4068.7 ± 0.8	4068.9 ± 1.2	- 0.3		- 0.7
B5	4054.8 ± 0.9	4055.0 ± 1.2	- 0.3		- 0.5
C1	3957.5 ± 0.9	3957.8 ± 1.3	- 0.3		- 0.6
C2	3953.3 ± 1.0	3953.6 ± 1.4	- 0.3		- 0.3
C3	3952.8 ± 1.0	3953.1 ± 1.4	- 0.3		0.3
C4	3954.1 ± 1.0	3954.4 ± 1.3	- 0.3		- 0.7
C5	3947.9 ± 0.9	3948.2 ± 1.2	- 0.3		- 0.4

Based on determined degree of equivalency (DOE), three/two gravimeters (bold italic format) showed inconsistency at 95% confidence with declared uncertainties for the ICN/KCN solution. Gravimeters in bold belong to NMI/DIs

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

Table 5 Results of comparisons in 2013 in μGal

Gravimeter	ICN solution		KCN solution		Repeatable
	Bias	DOE	Bias	DOE	
<i>CCM.G-K2 (ICAG 2013)</i>					
CAG-01	6.0±4.9	5.9±4.9	3.3±4.9	6.5±4.8	5.2
FG5-213	-3.6±2.3	-3.6±2.3	-1.7±2.3	-3.4±2.2	2.1
FG5-215	0.6±2.2	0.6±2.1	1.1±2.0	1.1±2.0	0.7
FG5-231	-1.8±2.2	-1.8±1.9	-1.5±1.8	-1.5±1.8	1.0
FG5-242	1.0±2.6	1.0±2.6	1.4±2.6	1.4±2.5	0.0
FG5X-104	-0.5±2.2	-0.5±1.8	-0.2±1.7	-0.2±1.7	0.4
FG5X-209	-1.7±2.2	-1.7±1.9	-1.0±1.8	-1.0±1.8	0.4
FG5X-221	1.4±2.2	1.4±2.1	2.0±2.0	2.0±2.0	0.3
IMGC-02	-1.8±3.1	-1.8±3.1	-1.8±3.1	-1.8±3.1	7.5
NIM-3A	1.0±4.5	1.2±4.5	1.0±4.5	1.2±4.5	4.1
A10-006	-3.2±9.8	-3.2±9.8	-2.6±9.8	-2.6±9.8	7.5
A10-020	-4.6±7.3	-4.7±4.8	-4.2±7.4	-4.3±4.9	2.4
FG5-102	-5.8±3.3	-5.8 ± 2.0	-5.9±3.4	-5.9 ± 2.1	2.1
FG5-202	2.3±2.2	2.3±1.9	2.8±2.4	2.8±2.1	0.8
FG5-206	-2.8±2.2	-2.8±2.2	-2.2±2.4	-2.2±2.4	1.5
FG5-218	0.7±2.2	0.7±1.7	1.4±2.4	1.4±1.9	1.3
FG5-223	2.0±2.2	2.0±1.9	2.3±2.4	2.3±2.1	1.2
FG5-228	-3.5±2.2	-3.5 ± 1.7	-3.1±2.3	-3.1±1.9	1.1
FG5-233	1.9±2.2	1.9±2.1	2.4±2.4	2.4±2.4	1.1
FG5-234	1.6±2.2	1.6±1.9	1.9±2.4	1.9±2.1	0.8
FG5-301	-2.0±2.2	-2.0±1.9	-1.6±2.3	-1.6±2.1	1.3
FG5X-216	-0.4±2.2	-0.4±1.9	-0.6±2.4	-0.6±2.1	1.7
FG5X-220	1.8±2.2	1.8±1.9	2.2±2.4	2.2±2.1	0.8
FG5X-302	0.2±2.2	0.2±1.9	0.1±2.4	0.1±2.1	0.5
T-2	8.5±4.6	8.5±4.6	8.9±4.6	8.9±4.6	3.4
s_0	1.27		1.12		
δ_c	-0.56		+0.17		
Site	ICN solution g - 980960000	KCN solution g - 980960000	ICN - KCN		KCN - ICN
A1	4228.7±1.0	4228.5±1.2	0.2		0.1
A2	4216.1±1.0	4215.4±1.1	0.8		-1.1
A3	4206.1±0.9	4205.7±1.0	0.4		-0.6
A4	4190.4±0.8	4190.3±1.0	0.1		0.6
A5	4183.1±0.9	4182.9±1.0	0.2		-0.2
B1	4077.4±0.9	4077.0±1.1	0.4		0.3
B2	4071.8±0.8	4071.4±1.0	0.5		-0.6
B3	4068.9±0.9	4067.9±1.1	1.0		-0.5
B4	4062.6±1.0	4062.2±1.1	0.4		-0.4
B5	4049.4±0.9	4050.2±1.2	-0.9		1.0
C1	3951.8±0.8	3950.9±1.0	0.8		-1.0
C2	3945.9±1.1	3945.4±1.2	0.5		0.4
C3	3948.3±0.9	3947.8±1.0	0.5		-0.2
C4	3946.5±0.9	3946.3±1.1	0.2		0.2
C5	3942.4±0.9	3942.1±1.1	0.3		-0.4

Based on determined degree of equivalency (DOE), two/one gravimeters (bold italic format) showed inconsistency at 95% confidence with declared uncertainties for the ICN/KCN solution. Gravimeters in bold belong to NMI/DIS

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

Table 6 Results of comparisons in 2015 in μGal

Gravimeter	ICN solution		KCN solution		Repeatable
	Bias	DOE	Bias	DOE	
<i>EURAMET.M.G-K2 (ECAG 2015)</i>					
FG5X-221	-3.7 ± 2.1	-3.7 ± 2.1	-2.4 ± 1.8	-2.4 ± 1.8	0.3
FG5-215	2.5 ± 2.1	2.5 ± 2.1	3.4 ± 1.9	3.7 ± 1.8	0.6
IMGC-02	13.7 ± 4.8	13.7 ± 4.8	12.8 ± 5.9	14.9 ± 4.8	4.7
FG5X-216	-2.9 ± 2.1	-2.9 ± 1.9	-1.6 ± 1.6	-1.6 ± 1.6	0.8
FG5X-102	-1.2 ± 2.1	-1.2 ± 1.7	0.0 ± 2.5	0.0 ± 2.2	0.9
FG5-202	3.2 ± 2.1	3.2 ± 1.9	4.4 ± 2.5	4.4 ± 2.3	0.4
FG5-218	-1.0 ± 2.1	-1.0 ± 1.7	0.2 ± 2.5	0.2 ± 2.1	0.3
FG5X-220	3.8 ± 2.1	3.8 ± 2.1	5.0 ± 2.5	5.0 ± 2.5	0.8
FG5X-229	-1.3 ± 2.1	-1.3 ± 1.7	-0.2 ± 2.5	-0.2 ± 2.2	1.4
FG5-230	-4.5 ± 2.2	-4.5 ± 1.8	-3.3 ± 2.5	-3.3 ± 2.2	1.0
FG5-233	1.1 ± 2.2	1.1 ± 2.2	2.3 ± 2.6	2.3 ± 2.6	0.9
FG5-234	3.2 ± 2.1	3.2 ± 1.8	4.3 ± 2.5	4.3 ± 2.2	1.0
FG5-238	1.9 ± 5.5	2.3 ± 5.6	3.1 ± 5.6	3.5 ± 5.7	6.8
FG5X-247	-7.5 ± 5.5	-9.5 ± 2.4	-6.3 ± 5.6	-8.3 ± 2.6	4.7
FG5-301	-1.3 ± 2.1	-1.3 ± 1.9	0.0 ± 2.5	0.0 ± 2.3	0.2
FG5X-302	-0.4 ± 2.1	-0.4 ± 1.7	0.9 ± 2.5	0.9 ± 2.1	0.7
A10-020	-6.8 ± 7.3	-6.7 ± 5.5	-5.6 ± 7.4	-5.5 ± 5.6	4.0
s_0	0.87		0.88		
δ_c	+ 0.99		+ 1.62		
Site	ICN solution $g - 980940000$	Site	KCN solution $g - 980940000$	ICN - KCN	KCN - KC
01	9049.3 ± 0.8		9048.1 ± 1.3	1.2	0.3
02	9045.1 ± 0.8		9044.0 ± 1.3	1.2	0.3
04	9036.3 ± 1.0		9035.2 ± 1.5	1.1	0.2
05	9029.4 ± 0.8		9028.3 ± 1.3	1.1	0.3
06	9066.7 ± 0.8		9065.4 ± 1.3	1.2	0.3
07	9063.5 ± 0.8		9062.0 ± 1.3	1.5	0.1
08	9060.2 ± 0.9		9059.1 ± 1.4	1.1	0.1
09	9056.8 ± 0.8		9055.6 ± 1.3	1.2	0.2
10	9048.6 ± 0.9		9047.3 ± 1.4	1.2	0.2

Based on determined degree of equivalency (DOE), three gravimeters (bold italic format) showed inconsistency at 95% confidence with declared uncertainties. Gravimeters in bold belong to NMI/DIs

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

Table 7 Results of comparisons in 2017 in μGal

Gravimeter	ICN solution		KCN solution		Repeatab.
	Bias	DOE	Bias	DOE	
<i>CCM.G-K2.2017 (ICAG 2017)</i>					
FG5-242	-0.2 ± 3.0	-0.2 ± 3.0	-0.1 ± 3.0	-0.1 ± 3.0	1.7
NIM-3A	0.3 ± 2.8	0.3 ± 2.7	0.4 ± 2.7	0.4 ± 2.7	3.8
FG5-215H	-1.2 ± 2.1	-1.2 ± 1.9	-1.0 ± 1.9	-1.0 ± 1.9	1.8
FG5X-221	0.8 ± 2.1	0.8 ± 2.0	0.9 ± 2.0	0.9 ± 2.0	1.3
FG5-228	0.1 ± 2.4	0.1 ± 2.4	0.2 ± 2.4	0.2 ± 2.4	2.1
FG5-213	-0.2 ± 2.8	-0.2 ± 2.8	-0.1 ± 2.8	-0.1 ± 2.8	1.8
FG5X-216	0.6 ± 2.1	0.6 ± 2.1	0.8 ± 2.1	0.8 ± 2.1	1.5
FG5X-104	-0.7 ± 2.1	-0.7 ± 1.9	-0.5 ± 1.8	-0.5 ± 1.8	1.4
FG5X-253	-0.6 ± 2.1	-0.6 ± 2.1	-0.4 ± 2.1	-0.4 ± 2.1	0.9
FG5X-209	3.7 ± 3.5	3.7 ± 3.5	3.8 ± 3.5	3.8 ± 3.5	0.8
FG5X-248	-0.8 ± 2.1	-0.9 ± 2.0	-0.8 ± 2.0	-0.8 ± 2.0	1.5
FG5X-254	-2.8 ± 2.1	-2.8 ± 2.1	-2.6 ± 2.1	-2.6 ± 2.1	1.0
FG5-204	1.0 ± 2.1	1.1 ± 1.6	1.3 ± 1.5	1.2 ± 1.5	0.7
NIM-AGRb	-2.5 ± 4.1	-2.5 ± 4.1	-2.4 ± 4.2	-2.4 ± 4.2	1.7
WAG-H5-1	-4.6 ± 8.8	-3.9 ± 8.9	-4.5 ± 8.9	-3.8 ± 8.9	6.9
ZAG-Atom	-11.7 ± 12.5	-11.6 ± 12.5	-11.6 ± 12.5	-11.5 ± 12.5	12.3
HUST-QG	1.1 ± 2.7	1.1 ± 2.7	1.2 ± 2.8	1.2 ± 2.8	2.5
A10-034	-1.7 ± 7.2	-1.7 ± 5.5	-1.6 ± 7.2	-1.6 ± 5.5	5.1
T-1	5.7 ± 12.0	5.2 ± 6.0	5.8 ± 12.0	5.4 ± 6.0	13.6
Age-110	6.1 ± 9.1	6.1 ± 9.1	6.3 ± 9.1	6.3 ± 9.1	6.2
FG5X-255	-3.6 ± 2.1	-3.6 ± 1.6	-3.5 ± 2.3	-3.5 ± 1.7	0.6
FG5-112	0.2 ± 2.1	0.2 ± 1.6	0.3 ± 2.3	0.3 ± 1.7	0.9
IGG-02	-0.3 ± 9.5	-0.3 ± 9.5	-0.2 ± 9.5	-0.1 ± 9.5	2.7
A10-022	0.5 ± 9.7	0.5 ± 9.7	0.6 ± 9.8	0.6 ± 9.8	5.1
FG5-214	0.6 ± 2.1	0.6 ± 1.7	0.7 ± 2.3	0.7 ± 1.8	1.2
FG5-301	1.6 ± 2.1	1.6 ± 2.0	1.7 ± 2.3	1.7 ± 2.2	1.2
FG5X-102	1.2 ± 2.1	1.2 ± 1.6	1.3 ± 2.3	1.3 ± 1.7	1.2
FG5X-302	2.6 ± 2.1	2.6 ± 1.6	2.7 ± 2.3	2.7 ± 1.7	1.3
s_0	1.27		1.29		
δ_c	-0.17		$+0.09$		
Site	ICN solution $g - 980120000$	KCN solution $g - 980120000$	ICN - KCN	KCN - KC	
01	2575.1 ± 0.7	2575.1 ± 0.8	0.1	0.0	
02	2568.9 ± 0.7	2568.9 ± 0.7	0.0	0.0	
03	2579.1 ± 0.7	2578.9 ± 0.8	0.2	-0.1	
04	2575.7 ± 0.7	2575.5 ± 0.7	0.2	-0.1	
05	2581.2 ± 0.7	2581.2 ± 0.8	0.0	0.1	
06	2649.6 ± 0.7	2649.5 ± 0.8	0.1	-0.1	
07	2641.9 ± 0.7	2641.8 ± 0.7	0.1	-0.1	
08	2654.1 ± 0.7	2654.0 ± 0.7	0.1	0.1	
09	2647.5 ± 0.7	2647.4 ± 0.8	0.1	0.1	

Based on determined Degree of Equivalency (DOE), one gravimeter (bold italic format) showed inconsistency at 95% confidence with declared uncertainties. Gravimeters in bold belong to NMI/DIs

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

Table 8 Results of comparisons in 2018 in μGal

Gravimeter	ICN solution		KCN solution ^a		Repeatable
	Bias	DOE	Bias	DOE	
<i>EURAMET.M.G-K3 (ECAG 2018)</i>					
FG5-215H	0.4 ± 2.1	0.4 ± 1.9	- 1.1 ± 2.1	- 1.1 ± 2.1	1.0
FG5-242	0.4 ± 2.1	0.4 ± 2.1	- 1.1 ± 2.3	- 1.1 ± 2.3	0.4
FG5X-221	2.6 ± 2.1	2.6 ± 2.0	1.2 ± 2.2	1.2 ± 2.2	0.5
FG5X-254	- 0.5 ± 2.1	- 0.5 ± 2.0	- 2.0 ± 2.2	- 1.9 ± 2.2	0.5
IMGC-02	0.6 ± 2.9	0.6 ± 2.9	- 0.9 ± 3.4	- 1.1 ± 3.6	6.1
A10-020	- 7.4 ± 7.3	- 7.5 ± 4.3	- 8.9 ± 7.5	- 8.9 ± 4.7	3.6
FG5-101	- 1.2 ± 2.7	- 1.2 ± 2.7	- 2.6 ± 3.2	- 2.6 ± 3.2	1.2
FG5-202	4.1 ± 2.2	4.1 ± 1.9	2.7 ± 2.7	2.7 ± 2.5	0.9
FG5-206	- 2.6 ± 2.1	- 2.6 ± 2.1	- 4.1 ± 2.7	- 4.1 ± 2.7	0.8
FG5-218	- 4.4 ± 2.2	- 4.4 ± 1.9	- 5.9 ± 2.7	- 5.9 ± 2.5	1.5
FG5-234	1.4 ± 2.1	1.4 ± 1.7	- 0.1 ± 2.7	- 0.1 ± 2.4	0.3
FG5-238	1.4 ± 2.9	1.4 ± 2.9	- 0.1 ± 3.4	- 0.1 ± 3.4	0.4
FG5-301	2.2 ± 2.8	2.2 ± 2.8	0.7 ± 3.3	0.7 ± 3.3	0.5
FG5X-220	0.0 ± 2.2	0.0 ± 2.1	- 1.5 ± 2.8	- 1.5 ± 2.8	0.8
FG5X-233	- 2.6 ± 2.1	- 2.6 ± 2.1	- 4.1 ± 2.7	- 4.1 ± 2.7	0.9
FG5X-247	0.2 ± 2.1	0.2 ± 1.9	- 1.3 ± 2.8	- 1.2 ± 2.5	0.3
s_0	0.73		0.73		
δ_c	- 0.20		+ 1.07		
Site	ICN solution $g - 980830000$	KCN solution $g - 980830000$	ICN - KCN		KCN - KC
CA	6951.2 ± 0.7	6952.7 ± 1.7	- 1.4		- 0.1
DA	6939.8 ± 0.7	6941.3 ± 1.7	- 1.5		0.0
EA	6948.0 ± 0.7	6949.5 ± 1.7	- 1.5		- 0.1
FA	6958.5 ± 0.7	6960.0 ± 1.7	- 1.5		0.1

Based on determined degree of equivalence (DOE), two/one gravimeters (bold italic format) showed inconsistency at 95% confidence with declared uncertainties for the ICN/KCN solution. Gravimeters in bold belong to NMI/DIs

s_0 is the root-mean-square error

δ_c is a shift of biases (Bias + δ_c) to reach a solution that corresponds with constraining L1 norm of biases instead of imposing zero mean of biases

^aThe KCN solution is linked to the comparison CCM.G-K2.2017 by the linking converter of (- 0.78 ± 1.26) μGal . Therefore, the uncertainties of reference values, biases and DoEs include the contribution from the uncertainty of the linking converter

Table 9 Information on the reprocessing of comparisons based on criteria described in Sect. 2.4

Comparison	Solution	Excluded from constraint	ICN: Downweighted in the constraint KCN: Measurement (at site) excluded	Harmonized uncertainty (increase in %)
ICAG-2009	ICN	FG5-102, FG5-230	FG5-224	FG5-102 (50%), NIM-2 (50%), CAG-01 (50%)
	KCN		NIM-2 (B)	FG5-102 (50%)
ECAG-2011	ICN	FG5-102, FG5-230	FG5-233	
	KCN			
ICAG-2013	ICN	FG5-102	FG5-213, T-2, CAG-01	FG5-102 (50%)
	KCN		FG5-213 (B5), CAG-01 (B3)	FG5-102 (50%)
ECAG-2015	ICN	FG5X-247	FG5-230, IMGC-02	FG5X-247 (250%)
	KCN		IMGC-02 (10), FG5-215 (7)	FG5X-247 (250%)
ICAG-2017	ICN	T-1		T-1 (100%)
	KCN			T-1 (100%)
ECAG-2018	ICN		FG5-202, FG5-218	
	KCN			

Two solutions are identified for all the comparisons: (1) ICN, where all gravimeters might contribute to the constraint. Depending on the consistency of measurements of a particular gravimeter, its contribution to the constraint is excluded (zero weight) or downweighted (corresponding to the decrease in uncertainty by 50%). Declared uncertainties are harmonized, and if the repeatability of a gravimeter does not correspond with harmonized uncertainty, they are increased by at least 50%. (2) KCN, where only NMI/DIs might contribute to the constraint. Depending on the consistency of individual measurements, the outliers are excluded from the adjustment. Declared uncertainties of non-NMI/DIs are harmonized, and if the repeatability of a gravimeter does not correspond with harmonized uncertainty, they are increased by at least 50%

References

- Boulanger YD, Arnautov GP, Scheglov SN (1983) Results of comparison of absolute gravimeters, Sévres, 1981. *Bull d'Information BGI* 52:97–124
- Cox MG (2002) The evaluation of key comparison data. *Metrologia* 39:589–595. <https://doi.org/10.1088/0026-1394/39/6/10>
- D'Agostino G, Desogus S, Germak A, Origlia C, Quagliotti D, Berrino G, Corrado G, Derrico V, Ricciardi G (2008) The new IMGC-02 transportable absolute gravimeter: measurement apparatus and applications in geophysics and volcanology. *Ann Geophys* 51:39–49. <https://doi.org/10.4401/ag-3038>
- de Viron O, Van Camp M, Francis O (2011) Revisiting absolute gravimeter intercomparisons. *Metrologia* 48:290–298. <https://doi.org/10.1088/0026-1394/48/5/008>
- Falk R, Müller J, Lux N, Wilmes H, Wziontek H (2012) Precise gravimetric surveys with the field absolute gravimeter A-10. In: Kenyon S, Pacino M, Marti U (eds) *Geodesy for planet earth*. International Association of Geodesy Symposia, vol 136. Springer, Berlin, pp 273–279. https://doi.org/10.1007/978-3-642-20338-1_33
- Falk R, Pálinkáš V, Wziontek H, Rülke A, Val'ko M, Ullrich Ch, Butta H, Kostelecký J, Bilker-Koivula M, Näränen J, Prato A, Mazzoleni F, Kirbaş Ç, Coşkun İ, Van Camp M, Castelein S, Bernard JD, Lothhammer A, Schilling M, Timmen L, Iacovone D, Nettis G, Greco F, Messina AA, Reudink R, Petrini M, Dykowski P, Sękowski M, Janák J, Papčo J, Engfeldt A, Steffen H (2020) Final report of EURAMET.M.G-K3 regional comparison of absolute gravimeters. *Metrologia* 57:07019. <https://doi.org/10.1088/0026-1394/57/1A/07019>
- Francis O, Klein G, Baumann H et al (2012) Final report of the regional key comparison EURAMET.M.G-K1: European Comparison of Absolute Gravimeters ECAG-2011. *Metrologia* 49:07014. <https://doi.org/10.1088/0026-1394/49/1A/07014>
- Francis O, Baumann H et al (2015) CCM.G-K2 key comparison. *Metrologia* 52:07009. <https://doi.org/10.1088/0026-1394/52/1A/07009>
- Jiang Z, Pálinkáš V, Arias FE, Liard J, Merlet S, Wilmes H, Vitushkin L, Robertsson L, Tisserand L et al (2012) The 8th International Comparison of Absolute Gravimeters 2009: the first Key Comparison (CCM.G-K1) in the field of absolute gravimetry. *Metrologia* 49:666–684. <https://doi.org/10.1088/0026-1394/49/6/666>
- Jiang Z, Pálinkáš V, Francis O, Baumann H, Mäkinen J, Vitushkin L, Merlet S, Tisserand L, Jousset P, Rothleitner C, Becker M, Robertsson L, Arias EF (2013) On the gravimetric contribution to watt balance experiment. *Metrologia* 50:452–471. <https://doi.org/10.1088/0026-1394/50/5/452>
- Karcher R, Imanaliev A, Merlet S, Pereira Dos Santos F (2018) Improving the accuracy of atom interferometers with ultracold sources. *New J Phys* 20:113041. <https://doi.org/10.1088/1367-2630/aaf07d>
- Koo A, Clare JF (2012) On the equivalence of generalized least-squares approaches to the evaluation of measurement comparisons. *Metrologia* 49:340–348. <https://doi.org/10.1088/0026-1394/49/3/340>
- Křen P, Pálinkáš V (2018) Two methods for determination of the effective wavenumber of Gaussian beams in absolute gravimeters. *Metrolog Meas Syst* 25:701–713. <https://doi.org/10.24425/mms.2018.124876>
- Křen P, Pálinkáš V, Mašika P (2016) On the effect of distortion and dispersion in fringe signal of the FG5 absolute gravimeters. *Metrologia* 53:27–40. <https://doi.org/10.1088/0026-1394/53/1/27>
- Marti U, Richard P, Germak A, Vitushkin L, Pálinkáš V, Wilmes H (2014) CCM—IAG Strategy for Metrology in Absolute Gravimetry. https://www.bipm.org/wg/CCM/CCM-WGG/Allowed/2015-meeting/CCM_IAG_Strategy.pdf. Accessed 8 July 2020
- Newell D, van Westrum D et al (2017) Regional comparison of absolute gravimeters SIM.M.G-K1 key comparison. *Metrologia* 54:07019. <https://doi.org/10.1088/0026-1394/54/1A/07019>
- Niebauer TM, Sasagawa GS, Faller JE, Hilt R, Klotz F (1995) A new generation of absolute gravimeters. *Metrologia* 32:159–180. <https://doi.org/10.1088/0026-1394/32/3/004>

- Niebauer TM, Billson R, Ellis B, Mason B, van Westrum D, Klopping F (2011) Simultaneous gravity and gradient measurements from a recoil-compensated absolute gravimeter. *Metrologia* 48:154–163. <https://doi.org/10.1088/0026-1394/48/3/009>
- Olsson PA, Engfeldt A, Ågren J (2016) Investigations of a suspected jump in Swedish Repeated Absolute Gravity Time Series. In: Freymueller JT, Sánchez L (eds) *International Symposium on Earth and Environmental Sciences for Future Generations. International Association of Geodesy Symposia*, vol 147. Springer, Cham. https://doi.org/10.1007/1345_2016_250
- Pálinkáš V, Kostecký J, Šimek J (2010) A feasibility of absolute gravity measurements in geodynamics. *Acta Geodyn Geomater* 7:61–69
- Pálinkáš V, Jiang Z, Liard J (2012) On the effective position of the free-fall solution and the self-attraction effect of the FG5 gravimeters. *Metrologia* 49:552–559. <https://doi.org/10.1088/0026-1394/49/4/552>
- Pálinkáš V, Lederer M, Kostecký J, Šimek J, Mojžeš M, Ferienc D, Csapó G (2013) Analysis of the repeated absolute gravity measurements in the Czech Republic, Slovakia and Hungary from the period 1991–2010 considering instrumental and hydrological effects. *J Geod* 87:29–42. <https://doi.org/10.1007/s00190-012-0576-1>
- Pálinkáš V, Francis O, Val'ko M et al (2017) Regional comparison of absolute gravimeters, EURAMET.M.G-K2 key comparison. *Metrologia* 54:07012. <https://doi.org/10.1088/0026-1394/54/1A/07012>
- Pereira F, Bonvalot S (2016) Cold atom absolute gravimetry. In: Grafarend EW (ed) *Encyclopedia of geodesy*. Springer, Basel. https://doi.org/10.1007/978-3-319-02370-0_30-1
- Reissmann G (1976) *Die Ausgleichsrechnung: Grundlagen u. Anwendungen in der Geodäsie*, 4th edn. Verlag für Bauwesen, Berlin
- Robinson IA, Schlamminger S (2016) The watt or Kibble balance: a technique for implementing the new SI definition of the unit of mass. *Metrologia* 53:A46. <https://doi.org/10.1088/0026-1394/53/5/A46>
- Rosat S, Boy JP, Ferhat G, Hinderer J, Amalvict M, Gegout P, Luck B (2009) Analysis of a 10-year (1997–2007) record of time varying gravity in Strasbourg using absolute and superconducting gravimeters: new results on the calibration and comparison with GPS height changes and hydrology. *J Geodyn* 48(3–5):360–365. <https://doi.org/10.1016/j.jog.2009.09.026>
- Schkolnik V, Leykauf B, Hauth M et al (2015) The effect of wavefront aberrations in atom interferometry. *Appl Phys B* 120:311–316. <https://doi.org/10.1007/s00340-015-6138-5>
- Schmerge D, Francis O, Henton J et al (2012) Results of the first North American comparison of absolute gravimeters, NACAG-2010. *J Geod* 86:591–596. <https://doi.org/10.1007/s00190-011-0539-y>
- Stock M (2013) Watt balance experiments for the determination of the Planck constant and the redefinition of the kilogram. *Metrologia* 50:R1–R16. <https://doi.org/10.1088/0026-1394/50/1/R1>
- Sutton CM (2004) Analysis and linking of international measurement comparisons. *Metrologia* 41:272–277. <https://doi.org/10.1088/0026-1394/41/4/008>
- Van Camp M, Williams SDP, Francis O (2005) Uncertainty of absolute gravity measurements. *J Geophys Res* 110:B05406. <https://doi.org/10.1029/2004JB003497>
- Van Camp M, de Viron O, Watlet A, Meurers B, Francis O, Caudron C (2017) Geophysics from terrestrial time-variable gravity measurements. *Rev Geophys* 55(4):938–992. <https://doi.org/10.1002/2017rg000566>
- Van Westrum D, Niebauer TM (2003) The diffraction correction for absolute gravimeters. *Metrologia* 40:258–263. <https://doi.org/10.1088/0026-1394/40/5/008>
- White DR (2004) On the analysis of measurement comparisons. *Metrologia* 41:122–131. <https://doi.org/10.1088/0026-1394/41/3/003>
- Wilmes H, Vitushkin L, Pálinkáš V, Falk R, Wziontek H, Bonvalot S (2016) Towards the definition and realization of a global absolute gravity reference system. In: Freymueller JT, Sánchez L (eds) *International symposium on earth and environmental sciences for future generations. International Association of Geodesy Symposia*, vol 147. Springer, Cham. https://doi.org/10.1007/1345_2016_245147
- Woolliams ER, Cox MG, Fox NP, Harris PM (2006) Final report of the CCPR K1-a key comparison of spectral irradiance 250 nm to 2500 nm. *Metrologia* 43:02003. <https://doi.org/10.1088/0026-1394/43/1A/02003>
- Wu S et al (2020) The results of CCM.G-K2.2017 key comparison. *Metrologia* 57:07002. <https://doi.org/10.1088/0026-1394/57/1A/07002>