Calibration campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram, part II: evolution of the BIPM as-maintained mass unit from the 3rd periodic verification to 2014

To cite this article: Estefanía de Mirandés et al 2016 Metrologia 53 1204

View the article online for updates and enhancements.

Related content
- Maintaining and disseminating the kilogram following its redefinition
  M Stock, S Davidson, H Fang et al.
- Calibration campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram part I: comparison of the international prototype with its official copies
  Michael Stock, Pauline Barat, Richard S Davis et al.
- A brief history of the unit of mass: continuity of successive definitions of the kilogram
  Richard S Davis, Pauline Barat and Michael Stock

Recent citations
- Avogadro constant measurements using enriched $^{28}$Si monocrystals
  K Fujii et al
- Evaluation of the accuracy, consistency, and stability of measurements of the Planck constant used in the redefinition of the international system of units
  Antonio Possolo et al
- Maintaining and disseminating the kilogram following its redefinition
  M Stock et al

This content was downloaded from IP address 54.70.40.11 on 05/01/2018 at 03:34
Calibration campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram, part II: evolution of the BIPM as-maintained mass unit from the 3rd periodic verification to 2014

Estefania de Mirandés, Pauline Barat, Michael Stock and Martin J T Milton

1 Bureau International des Poids et Mesures (BIPM), Pavillon de Breteuil, F-92312 Sèvres Cedex, France

E-mail: estefania.de.mirandes@bipm.org

Received 31 March 2016, revised 28 July 2016
Accepted for publication 17 August 2016
Published 29 September 2016

Abstract
In 2014 the Bureau International des Poids et Mesures (BIPM) carried out a calibration campaign using the international prototype of the kilogram (IPK). This is the second part in a series of publications describing the results of that campaign. As reported (Metrologia 52 310–6), following the comparisons between the IPK and its official copies, it was found that the BIPM ‘as-maintained mass unit’ was offset by 35 μg from the mass of the IPK in 2014. We report here the results of an investigation into this offset that has considered all data available from internal BIPM mass comparisons carried out between 1992 and 2014. This has enabled us to model the evolution of the offset in the as-maintained mass unit and to identify some possible reasons why it has developed. We also report how the model has been used to estimate corrections to all 1 kg mass calibration certificates issued by the BIPM during this period.

Keywords: international prototype of the kilogram, periodic verification, mass calibration

1. Introduction

An Extraordinary campaign of calibrations has been carried out at the Bureau International des Poids et Mesures (BIPM) to provide improved traceability to the international prototype of the kilogram (IPK) for those National Metrology Institutes (NMIs) that are carrying out experimental determinations of the Planck constant to support the forthcoming redefinition of the kilogram [2–4]. The first phase of the calibration campaign was started at the end of 2013 by carrying out mass comparisons between the IPK and its official copies. This work was completed in 2014 and the results have been published [1].

A second key set of measurements carried out during the first phase is the subject of this work. These were mass comparisons among the IPK and the ten BIPM Pt–Ir working standards. These working standards have been used to provide traceability to the IPK during 1993–2014 when the IPK was not accessed directly. This relied on the fact that some of these BIPM working standards had been calibrated directly against the IPK during the 3rd periodic verification (1988–1992) and since then all of the working standards have been intercompared to quantify any changes in mass. A detailed description of the procedure that has been followed at the
revealed an offset of 35 μg between the BIPM as-maintained mass unit and the mass of the IPK. Before these measurements, the IPK had been cleaned and washed twice, as required to realize the kilogram [1, 8].

Following these experimental findings, the BIPM conducted a new analysis of all the available data from internal mass comparisons among the ten BIPM working standards since the 3rd periodic verification in order to investigate the origin and evolution of this offset. A detailed description of this analysis is given here, together with a discussion of the consequences for calibrations of national prototypes and mass standards provided by the BIPM since the third verification.

This paper is organized as follows: section 2 describes the ten Pt–Ir working standards used by the BIPM during the period 1993–2014; section 3 presents the experimental results of the direct and indirect mass comparisons of the IPK and the ten working standards carried out in 2014; sections 4 and 5 introduce and describe a least-squares model developed at the BIPM to investigate the evolution of the offset; sections 6 and 7 describe the consequences of this mass offset on the calibration certificates provided by the BIPM during this period.

2. Pt–Ir mass prototypes and mass standards in use at the BIPM

Since 1889 the BIPM has been the custodian of the IPK [9], a cylinder with a diameter and height of 39 mm made of an alloy of 90% platinum and 10% iridium by mass. Since 1889 the mass of the IPK has defined the kilogram, the SI unit of mass.

The international prototype is stored in a safe at the BIPM along with six other nearly identical Pt–Ir cylinders, called the official copies or témoins. Access to the IPK and the copies is rare and is only possible under the authorization of the Comité international des poids et mesures (CIPM, International Committee for Weights and Measures). Since 1889, the IPK and its copies have only been used for comparisons in 1889–1911 (the first periodic verification [10], in which the international prototype was not used), in 1939–1953 (the second periodic verification [6]), in 1988–1992 (the third periodic verification [11]) and in 2013–2014 (the Extraordinary Calibrations [1]).

Table 1 lists all BIPM 1 kg Pt–Ir mass standards relevant to this work. These include: the IPK, the six official copies, and ten working standards. The latter consist of two Pt–Ir prototypes reserved ‘for special use’ and the working standards for current use. The first date when each mass standard was used is also given. As mentioned above, the international prototype of the kilogram and its official copies can only be accessed on rare occasions. Note also that prototypes nos. 25 and 73 are standards designated by the CIPM ‘for special use’, which were used less often as the standards for current use. Prototypes 9 and 31 are the only ones that participated in the three periodic verifications (and in the Extraordinary Calibrations) and have been regularly monitored in between. Prototypes nos. 63 and 73 and working standards nos. 42 and 650 were not part of the 3rd periodic verification, but their masses were determined at the same time and have been monitored routinely since then. Prototypes nos. 88 and 91 were added to into the group more recently (2003).

During the period 1889–1990 very few mass comparisons were carried out. The number of comparisons has been increasing progressively since the 3rd periodic verification in 1992, boosted by the availability of automatic mass comparators.

3. Calibration of the ten BIPM working standards against the international prototype of the kilogram during 2014 (the Extraordinary Calibration campaign)

Reference [1] presents extensive information about the experimental weighing protocol followed by the BIPM during the first phase of the 2014 Extraordinary Calibration campaign of the mass standards for current use. The first date when each mass standard was used is also given. As mentioned above, the international prototype of the kilogram and its official copies can only be accessed on rare occasions. Note also that prototypes nos. 25 and 73 are standards designated by the CIPM ‘for special use’, which were used less often as the standards for current use. Prototypes 9 and 31 are the only ones that participated in the three periodic verifications (and in the Extraordinary Calibrations) and have been regularly monitored in between. Prototypes nos. 63 and 73 and working standards nos. 42 and 650 were not part of the 3rd periodic verification, but their masses were determined at the same time and have been monitored routinely since then. Prototypes nos. 88 and 91 were added to into the group more recently (2003).

During the period 1889–1990 very few mass comparisons were carried out. The number of comparisons has been increasing progressively since the 3rd periodic verification in 1992, boosted by the availability of automatic mass comparators.

Table 1. The mass standards referred to in this work with the date of their first comparison.

<table>
<thead>
<tr>
<th>Standard Role</th>
<th>Date of first comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPK Int. prototype of the kilogram</td>
<td>1889</td>
</tr>
<tr>
<td>K1 Official copy</td>
<td>1889</td>
</tr>
<tr>
<td>7 Official copy</td>
<td>1889 (off. copy since 1925)</td>
</tr>
<tr>
<td>8(41) Official copy</td>
<td>1889 (off. copy since 1905)</td>
</tr>
<tr>
<td>32 Official copy</td>
<td>1889 (off. copy since 1905)</td>
</tr>
<tr>
<td>43 Official copy</td>
<td>1939</td>
</tr>
<tr>
<td>47 Official copy</td>
<td>1939</td>
</tr>
<tr>
<td>9 Working standard</td>
<td>1889</td>
</tr>
<tr>
<td>25 Working standard</td>
<td>1889 (for special use)</td>
</tr>
<tr>
<td>31 Working standard</td>
<td>1889</td>
</tr>
<tr>
<td>42 Working standard</td>
<td>1976</td>
</tr>
<tr>
<td>63 Working standard</td>
<td>1975</td>
</tr>
<tr>
<td>73 Working standard</td>
<td>1991 (for special use)</td>
</tr>
<tr>
<td>77 Working standard</td>
<td>1997</td>
</tr>
<tr>
<td>88 Working standard</td>
<td>2003</td>
</tr>
<tr>
<td>91 Working standard</td>
<td>2003</td>
</tr>
<tr>
<td>650 Working standard</td>
<td>1993</td>
</tr>
</tbody>
</table>

Note: Note that for the present work only data from comparisons carried out since the 3rd periodic verification and until the end of 2014 have been considered. In this table the third column refers to the date when the mass standard was first compared, whereas in table 1 of [1] the second column refers to the date from which the standard has belonged officially to the BIPM, as an official copy or as a working standard.

The term ‘prototype’ refers to the 1 kg mass standards in platinum–iridium (90% Pt/10% Ir) that were fabricated by the BIPM (or under its direct supervision) and which meet the criteria of prototypes as specified in the Metre Convention, i.e. the mass value must respect the adjustment tolerance of 1 kg ± 1 mg. The term ‘mass standard’ is more general and refers to 1 kg artefacts in Pt–Ir or stainless steel as well as 1 kg silicon spheres that may serve as mass standards.

2
the kilogram. It also discusses the results of repeated cleaning—washing procedures on the IPK, the official copies and BIPM working standards. These data will not be described again in this work but have been included in the global data analysis presented.

During the first phase of the 2014 Extraordinary Calibration campaign, the ten BIPM working standards were calibrated with respect to the IPK. Prior to this, the IPK had been cleaned and washed twice, as it defines the kilogram only in its cleaned state. The cleaning—washing operations were performed by following strictly the BIPM procedure for cleaning—washing [12]. The official copies were also cleaned and washed in the same way. By contrast, eight of the ten BIPM working standards did not undergo any cleaning—washing during the 2014 campaign. Only prototypes nos. 25 and 73 (for ‘special use’) were cleaned and washed and compared to the IPK before and after. This allows us to compare their masses (with respect to the cleaned IPK in 2014) with the mass they had at the end of the 3rd periodic verification in their cleaned state (with respect to the clean IPK at the time).

The calibrations of the ten working standards with respect to the IPK also provide data to verify if the mass unit had been maintained accurately by the BIPM since the end of the 3rd periodic verification. During the years 1993–2014, the BIPM working standards were calibrated with respect to the IPK, and the number of weighings since the end of the 3rd periodic verification and in 2014 at the time of the Extraordinary Calibrations.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Mass difference from IPK at 3rd PV (1992) (mg)</th>
<th>Mass difference from IPK at 2014 (mg)</th>
<th>m (2014)–m (1992) (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.158</td>
<td>0.139</td>
<td>–0.019</td>
</tr>
<tr>
<td>73</td>
<td>0.428</td>
<td>0.395</td>
<td>–0.033</td>
</tr>
<tr>
<td>9</td>
<td>0.312</td>
<td>0.289</td>
<td>–0.023</td>
</tr>
<tr>
<td>31</td>
<td>0.131</td>
<td>0.106</td>
<td>–0.025</td>
</tr>
<tr>
<td>42</td>
<td>0.061</td>
<td>–0.021</td>
<td>–0.082</td>
</tr>
<tr>
<td>63</td>
<td>0.362</td>
<td>0.274</td>
<td>–0.088</td>
</tr>
<tr>
<td>77</td>
<td>–1.318</td>
<td>–1.342</td>
<td>–0.024</td>
</tr>
<tr>
<td>88</td>
<td>Not measured at 3rd PV</td>
<td>0.125</td>
<td>—</td>
</tr>
<tr>
<td>91</td>
<td>Not measured at 3rd PV</td>
<td>–0.147</td>
<td>—</td>
</tr>
<tr>
<td>650</td>
<td>–2.258</td>
<td>–2.318</td>
<td>–0.060</td>
</tr>
</tbody>
</table>

Note: The repeatability of the 2014 weighing measurements was typically better than 0.5 μg.

It is important to note that this analysis is based on the underlying assumption that the mass of the IPK after two cleaning—washing operations in 2014 was the same as it was after two cleaning—washing operations in 1992. This assumption is what is stated by the present definition of the kilogram. The good agreement of the masses of the official copies (with respect to the IPK) observed in 2014 with their masses obtained during the 3rd periodic verification [11] provide some evidence for the stability of the IPK.

Table 2 summarizes the results from calibrations of the ten BIPM working standards against the IPK (in its cleaned state) during the 3rd periodic verification, as well as results from the calibrations performed in 2014.

Surprisingly, with no exception, it was found that all BIPM prototypes and working standards had lost mass with respect to the IPK during the period 1992–2014. The amount of mass loss was not the same for all standards but ranged from 19 to 88 μg. The mass overestimation is not to be confused with the absolute mass loss of the standards, which was different for each standard depending on its usage.

It should be noted that for some working standards their cleaning state in the 1992 mass comparison against the IPK was not the same as their cleaning state in the 2014 calibration, whereas the IPK underwent two cleaning—washing operations on both occasions. If a correction for the different cleaning states is applied, considering that a Pt–Ir standard experiences a mass gain of typically 1 μg yr⁻¹, the net mass loss would even increase further. However, prototypes nos. 25 and 73 had been cleaned and washed during the 3rd PV and during the 2014 campaign. Therefore, assuming that the mass of the IPK (in its cleaned state) has remained constant during 1992–2014, some unexpected mass loss mechanism has been affecting all of the BIPM working standards.

These results suggested that correlations might exist between the measured mass loss and some relevant experimental parameter. Indeed, a certain correlation was found between the mass loss experienced by the working standard and the number of weighings since the end of the 3rd periodic verification. The relationship between the mass loss and the number of weighings for each BIPM working standard is shown in figure 1.
This graph suggests a marked influence of the number of weighings on the mass loss experienced by the working standards. However, note that the total mass loss data considered for this graph are not corrected for the different cleaning states of the standards in 2014 with respect to their cleaning state in 1993. Weighings carried out in different mass comparators are not distinguished in this graph. A detailed mathematical analysis has therefore been conducted to separate the contributions to the total mass loss from all known phenomena affecting the mass of a working standard in order to establish proper correlations. This model is presented in section 4.

4. Modelling the drift of the BIPM as-maintained mass unit from 1992 to 2014

All data from internal BIPM mass comparisons of the IPK, its official copies and working standards between the 3rd periodic verification and the Extraordinary Calibrations in 2014 have been reexamined. The total number of independent mass differences measured during this period amounts to \( N = 546 \).

The mass evolution of each of the standards is assumed to follow a given mathematical equation (see equation (1)) which includes several terms, each of them taking into account mass changes due to a particular phenomenon. The general equation assumed for the temporal evolution of the mass \( m \) of a mass standard \( i \) is the following:

\[
m_i(t) = m_i(t = 0) + \alpha_i t + \nabla_i \sqrt{t - t_{CW}} + \sum_{\text{mass.comp.}} \omega_i \text{mass.comp.} N_{\text{mass.comp.}}(t)
\]  

(1)

where \( m_i(t = 0) \) is the initial mass of each mass standard at \( t = 0 \) (this being the time of the end of the third verification), \( \alpha_i \), the parameter characterizing the linear irreversible drift of the cleaned mass with time (e.g. due to irreversible contamination or outgassing), \( \nabla_i \) the parameter characterizing the reversible increase in mass experienced by a mass standard after a cleaning–washing procedure, \( t_{CW} \) is the time of the last cleaning–washing of mass standard \( i \) prior to time \( t \) and \( \omega_i \text{mass.comp.} \) the wear coefficient characterizing the mass loss/gain experienced by a mass standard due to the weighing process in a given mass comparator. \( N_{\text{mass.comp.}}(t) \) is the number of weighings of mass standard \( i \) in a given mass comparator until time \( t \). The choice of equation (1) is based on the knowledge of previous long-term behaviour of mass standards [11–13].

The known input quantities in this equation are the time \( t \) of each measurement, all cleaning–washing times \( t_{CW} \), the number of measurements \( N_{\text{mass.comp.}} \) involving each mass standard in each mass comparator and the experimental mass differences \( m_i(t) - m_j(t) \). The output parameters in this equation, to be estimated by the least squares adjustment are: \( m_i(t = 0) \), \( \alpha_i \), \( \nabla_i \) and \( \omega_i \text{mass.comp.} \), for all mass standards \( i \) and for the mass comparators.

In order to respect the current definition of the kilogram, equation (1) when applied to the IPK has been constrained with

\[
m_{\text{IPK}}(t = 0) = 0, \quad \alpha_{\text{IPK}} = 0
\]  

(2)

When writing equation (1) for all 546 mass differences measured at the BIPM since 1992, a global least squares adjustment can be carried out and the output parameters of the model for all the working standards can be determined. The total number of mass standards involved in the 546 mass comparisons is 17: the IPK, the six official copies and the ten BIPM working standards described in table 1.

As mentioned, the fourth term on the right hand side of equation (1) takes into account the mass change experienced by a mass standard due to a weighing measurement in a given mass comparator. Since the 3rd periodic verification, four different mass comparators have been used at the BIPM, with
most of the measurements carried out using two of them: a Mettler Toledo HK1000 and a Mettler Toledo Metrotec mass comparator. This is shown in figure 2.

It should be noted that both the HK1000 and Metrotec mass comparators were slightly modified at the BIPM and therefore they have not been used in their original configuration after purchase. In the following, the HK1000 mass comparator will be designated as comparator 1 and the Metrotec as comparator 2.

Taking into account these considerations, from the general equation (1) four different equations have been derived using each a particular subset of the terms. In all of them the square root term responsible for the mass gain after a cleaning–washing procedure has been maintained. What changes are the terms potentially responsible for the large mass losses that are observed experimentally. For each of them, a global least squares analysis has been carried out with the objective of finding out which of the four models best describes the ensemble of experimental data.

Model 1 assumes a linear dependence of the irreversible mass change on time and a reversible mass increase after cleaning–washing on the number of sets of weighings, assuming a specific wear coefficient for each of the four mass comparators while considering that different standards weighed in a given mass comparator share the same \( \omega \) coefficient:

\[
m_i(t) = m_i(t = 0) + \alpha t + \gamma \sqrt{t - t_0} + \omega_i \sqrt{t - t_0} - t_0
\]

Model 2 assumes, in addition to the terms of model 1, a dependence of the number of sets of weighings, assuming different wear coefficients per mass comparator, but only considering the two that were used the most: a Mettler Toledo HK1000 and a Mettler Toledo Metrotec mass comparator.

\[
m_i(t) = m_i(t = 0) + \gamma_i \sqrt{t - t_0} + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0
\]

Model 4 is a combination of model 1 and 3, it assumes a linear dependence of the mass with time plus a dependence of the mass on the number of sets of weighings, with one wear coefficient per mass standard and per mass comparator, again limited to the two that were used the most:

\[
m_i(t) = m_i(t = 0) + \alpha t + \gamma_i \sqrt{t - t_0} + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0
\]

A global least squares analysis has been carried out using the data from the 546 experimental mass difference measurements. The four models have been tested this way. As described below, model 3 has been chosen as the most appropriate. The number of free parameters in this model is 51: 16 \( m_i(t = 0) \), 10 \( \omega_i \sqrt{t - t_0} - t_0 \) and 15 \( \gamma_i \).

In our analysis, the off-diagonal elements of the covariance matrix have been assumed to be zero because there is no a priori correlation between different input mass difference measurements. The diagonal elements of the covariance matrix are identical as we assume the same uncertainty for all the input mass differences. This uncertainty is adjusted iteratively until \( \chi^2 \) is equal to its expected value, i.e. \( \chi^2 \leq N - n_p \) being \( n_p \) the number of free parameters in the model. This process has been carried out for each of the four models. At the end, for each model the standard deviation of the residuals of the fit of the mass differences \( u_{\text{fit}} \) (also called standard error of the fit) has been computed.

Model 3 assumes a linear dependence of the mass on the number of sets of weighings but no linear dependence of the mass on the time. In addition, it includes a wear coefficient per mass standard and per mass comparator, but only considering comparator 1 and 2 which were used much more intensely than the other two comparators (see figure 2):

\[
m_i(t) = m_i(t = 0) + \gamma_i \sqrt{t - t_0} + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0 + \omega_i \sqrt{t - t_0} - t_0
\]

A global least squares analysis has been carried out using the data from the 546 experimental mass difference measurements. The four models have been tested this way. As described below, model 3 has been chosen as the most appropriate. The number of free parameters in this model is 51: 16 \( m(t = 0) \), 10 \( \omega_i \sqrt{t - t_0} - t_0 \) and 15 \( \gamma_i \).

In our analysis, the off-diagonal elements of the covariance matrix have been assumed to be zero because there is no a priori correlation between different input mass difference measurements. The diagonal elements of the covariance matrix are identical as we assume the same uncertainty for all the input mass differences. This uncertainty is adjusted iteratively until \( \chi^2 \) is equal to its expected value, i.e. \( \chi^2 \leq N - n_p \) being \( n_p \) the number of free parameters in the model. This process has been carried out for each of the four models. At the end, for each model the standard deviation of the residuals of the fit of the mass differences \( u_{\text{fit}} \) (also called standard error of the fit) has been computed.

Note that \( u_{\text{fit}} \) is normalized with respect to the number of parameters of the model \( n_p \), therefore a decrease in \( u_{\text{fit}} \) should be interpreted as a better compliance of the model with the data to be described, regardless of the number of parameters. Table 3 shows the values of \( u_{\text{fit}} \) for the four models presented above. As soon as wear coefficients associated with the

---

3 The BIPM is not suggesting in any way that these comparators are intrinsically conceived in a way that would damage the standards but rather considering the possibility that the particular comparators used at the BIPM, in their particular configuration, did.
weighing process are introduced into the models (four wear coefficients $\omega$ in model 2 and 20 wear coefficients $\omega$ in models 3 and 4), $u_{\text{fit}}$ decreases, indicating that the existence of a real wear effect on the BIPM working standards is highly probable. This explanation is therefore better supported than a simple mass evolution of the working standards due to a mass drift depending exclusively on time (due to outgassing and/or irreversible mass accumulation) which would be the rationale supporting model 1. It is particularly interesting to compare models 1 and 3 because they consider respectively a mass dependence uniquely on time and uniquely on the number of measurements. Model 3 clearly fits the data better. A supplementary step has been added in Model 3 to ensure that the experimental differences are strictly respected.

Figure 3 compares results from the four models. The mass values deduced from the fit as well as the mass values originally assigned by the BIPM are shown for the case of working standard no. 42′ which is among those having experienced the biggest mass loss and the largest number of weighing measurements. Results from the four models are compared. The fit curve from model 1 is not able to follow the mass evolution substructures shown by the data. By contrast, models 2, 3 and 4 are, in this order, more able to fit those substructures. The same conclusions can be drawn from the results obtained for the other working standards. This supports the theory that a wear phenomenon is at the origin of the mass evolution profiles shown by the BIPM working standards.

During the years 1993–2014 all the BIPM working standards were frequently inter-compared. Therefore, the mass offset (the difference between the assumed mass of the standard and its mass with respect to the IPK, respectively the red dots and the blue diamonds of the figures above) experienced by a particular working standard propagated rapidly to the rest of the working standards. This implies that during this period, calibrations of national prototypes were affected by a mass offset which depended on the date of the calibration but not on the particular working standards used as references in the calibration.

By comparing the results of the four models, we chose to stick to model 3 for the following reasons: As shown in figure 3, adjustments based on models 3 and 4 better describe the experimental data. In model 4 the wear coefficients $\omega$ and the linear temporal drift coefficients $\alpha$ for a given working standard were found to be highly correlated, this impedes a proper distinction between the two. Model 4 was believed to have too many fit parameters and to be slightly over-fitting the
data. Model 3 was further refined as described in the following section.

5. Optimization of model 3

A supplementary least squares adjustment has been carried out varying the starting date for the wear effect of comparators 1 and 2 separately in model 3 since the wear effect might have started at a particular moment of time. The result of this optimization for the starting date of the comparator 1 wear effect, imposing no wear effect due to comparator 2, is shown in figure 4. It has been found that the standard deviation of the residuals is significantly reduced when the starting date approaches the end of 2004. In 2004, indeed, some in-house modifications on the weight handler of comparator 1 where carried out.

The same optimization process has been carried out to determine the starting date of the comparator 2 wear effect, when imposing no wear due to comparator 1. In that case no minimum of the standard deviation of the residuals has been found. The starting date of the wear effect of comparator 2, if any, seems to coincide with the date where the comparator entered into service. Model 3, which takes into account the aforementioned starting dates, will in the following be identified as model 3b.

Figures 5(a) and (b) compare models 3 and 3b on standard 42'.

Figure 6 shows the mass offset experienced by BIPM working standards according to model 3b, computed as the difference between mass assignments made by the BIPM at the time (red circles in the figures above) and mass assignments made by the models at present (blue diamonds in figures 3 and 5). Note that all working standards experienced a very similar offset. In model 3b the offset of the BIPM as-maintained mass unit begins to become significant around 2003, increases rapidly between 2005 and 2010 and stays almost constant after 2010, which is the date when comparator 1 was taken out of service. Indeed, model 3b attributes essentially all the wear effect to comparator 1 and practically none to comparator 2. This is explicit in figure 7 which shows the output fit parameters from model 3b. The wear coefficients of comparator 2 are very close to zero. The comparator 1 wear coefficients are significant for all the working standards except for prototypes nos. 9 and 31 where they are compatible with zero. This is essentially due to the fact that this comparator was only used on a few occasions with prototypes nos. 9 and 31 after the starting date of the wear effect which leads to a very large uncertainty of $\omega_{\text{comp1}}$ and $\omega_{\text{31,comp1}}$. Notice that for the standards showing the largest mass decrease, the average wear coefficient of comparator 1 is found to be about 0.12 $\mu$g/set of weighings while the average comparator 2 wear coefficient is found to be about 0.02 $\mu$g/set of weighings.

Correlations between output fit parameters have been evaluated for model 3b. It has been found that the wear coefficients of comparators 1 and 2 for each working standard are highly anti-correlated. The correlation coefficients range between $-0.60$ and $-0.94$.

These large anti-correlation coefficients indicate that the average value of the $\omega_{\text{comp1}}$ and $\omega_{\text{comp2}}$ for a given mass standard is well known but their difference is not.

In addition, recent experimental research on comparator 2 conducted at the BIPM after these findings were made, indicates that this comparator (which is still in service at the BIPM) does not result in any significant mass change to the mass standards compared when using it. In particular, two BIPM Pt–Ir 500 g standards, were used one on top of the other forming a stack, which was placed on one of the four-position weight handler of comparator 1. The other three positions

![Figure 4. Standard deviation of the residuals of the fit of the experimental mass differences, $u_{\text{fit}}$, as a function of the starting date of the comparator 1 wear effect, using model 3, when imposing no wear effect due to comparator 2. The inset graph shows the 2002–2006 period in greater detail.](image-url)
were filled with stainless steel standards. In this configuration, two separate wear tests were conducted. In the first test, 130 sets of weighings were carried out in the same conditions than in a normal calibration process, except for the fact that the stack stayed all the time on the weight handler of comparator 2. In the second test the stack was removed and put back in the comparator, and finally centered by 2 cycles of loading/unloading, this process was repeated 44 times; the stack was kept in hand between placements and was handled with gloves made of Lycra. After each test, the sum and the difference of the mass of the two 500 g standards were measured. No significant mass change was detected in the standards after any of these tests.

Similar tests will be conducted at the BIPM with comparator 1 after it is reassembled. Note that comparator 2 is equipped with a four-position turntable, each adjustable independently, while comparator 1 has a non adjustable turntable.

6. Corrections to past mass calibration certificates issued by the BIPM for national prototypes

The offset detected in 2014 in the BIPM as-maintained mass unit was significantly larger than the estimated mass uncertainty and therefore calibration certificates of national prototypes issued by the BIPM after the drift of the masses began, needed to be revised. As shown in figure 6 the exact offset
affecting past mass assignments to BIPM working standards depends mainly on the date at which the calibration was performed and very little on the working standards themselves. In the light of the present findings, the BIPM has re-evaluated all mass calibration certificates of national prototypes that were issued in the period 1995–2014, taking into account the date of the calibration measurements as well as the particular BIPM working standards used as references at the time. This reevaluation has been carried by using the results obtained from model 3b.

Figure 8 shows the overestimation in mass made by the BIPM in the calibration of each of the national prototypes that have been calibrated since the 3rd periodic verification. Each dot corresponds to one calibration of a national prototype.

Figure 8 has been presented at the meeting of the Consultative Committee for Mass and Related Quantities (CCM) in February 2015. The CCM approved the BIPM proposal that calibration certificates issued between 2003 and 2013 should be amended. The BIPM provided each Member State concerned with amended calibration certificates for Pt–Ir prototypes (69 amendments) and in May 2015 for stainless steel standards (92 amendments), stating the corrected mass assignments, accompanied by an explanatory note presenting the experimental findings of the 2014 Extraordinary Calibration campaign and their consequences.

All institutes involved in the determination of the Planck constant and the Avogadro constant had already been informed in December 2014 of the corrections to be applied to previous mass calibrations, to enable them to provide corrected Planck constant and Avogadro constant results for the 2014 CODATA fundamental constants adjustment.

7. Uncertainty associated with mass corrections of past mass calibration certificates for national prototypes

Before detailing the uncertainty analysis, it should be noted that all corrections assigned to past mass certificates have one intrinsic uncertainty component which comes from the assumption of the correctness of model 3b. The evaluation of this uncertainty is nevertheless difficult because the extent of worse models, potentially leading to significantly different corrections, is very large. Therefore the combined uncertainty does not include a contribution related to the choice of the model. Model 3b is assumed to be correct.

The uncertainty components of the revised mass values, which have been taken into account, can be broken down into two groups, see equation (7), one related to the actual weighing measurement used to calibrate the NMI prototype, which is essentially a statistical uncertainty, $u_{\text{stat}}$, and another associated with the knowledge of the mass of the BIPM working standards used as references during the mentioned calibration, $u_{\text{ref}}$:

$$u = \sqrt{u_{\text{stat}}^2 + u_{\text{ref}}^2}$$

$u_{\text{stat}}$ has been conservatively estimated to be $1 \mu g$ for all past calibrations of national prototypes after the year 2000. The uncertainty $u_{\text{ref}}$ of the reference consists of several components: (a) the uncertainty of the least squares adjustment, (b) a contribution related to the reproducibility of the cleaning–washing procedure and (c) the uncertainty related to the evolution of the mass of the reference standard. These three components will be discussed in the following.

7.1. Uncertainty of the fit

The fit uncertainty $u_{\text{fit}}(m)$ refers to the uncertainty of the mass $m$ of the BIPM working standard (the blue diamonds of figures 3 and 5) that was used as a reference at the time of the calibration after being properly corrected for the offset of the BIPM as-maintained mass unit. By propagating the uncertainty of the fit parameters this uncertainty component has been found to be about $2.4 \mu g$, relatively independent of the time of the calibration.

7.2. Reproducibility of the cleaning–washing procedure of the IPK

The end point of the curve representing the $35 \mu g$ offset of the BIPM as-maintained mass unit shown in figure 6 depends uniquely on the experimental mass differences measured in 2014 between the BIPM working standards and the clean IPK and on the assumption that the mass of the IPK (in its cleaned
state) in 2014 was identical to its mass during the 3rd periodic verification. This offset is completely independent of the model chosen to fit the data.

By contrast, the exact value of this end point influences the shape of the curve on figure 8 and therefore the values of the corrections for past mass calibration certificates. It might be questioned if the cleaning–washing of the IPK during the 3rd periodic verification and in 2014 had exactly the same efficiency, in particular because the operators were not the same. In order to evaluate the possible uncertainty in the value assigned to the end point of the correction curve of figure 8, at the end of the first phase of the Extraordinary Calibrations the BIPM carried out two supplementary trials of the cleaning–washing procedure on two official copies: nos. 7 and K1. The trials were carried out by a different, but experienced, BIPM operator in order to test the dependence of the BIPM cleaning–washing procedure on the particular operator. From these results we estimated that the possible uncertainty of the knowledge of the end point of the correction curve of figure 8 was 2 μg. To estimate the uncertainty of the corrections applied to previous mass calibration certificates, the end point of the correction curve was varied by ±2 μg with respect to the original. The resulting change in the correction for each mass calibration certificate has been assumed to be the best estimate of $u_{cw}$ and this is shown in figure 9 for each calibration certificate.

7.3. Mass drift of reference standards

During each internal comparison among BIPM working standards, when the mass differences were found to have changed, mass values were re-assigned to them, based on assumptions of which standards had been the most stable. However, when these working standards are subsequently used to calibrate a national prototype, a period of time will have passed since their latest internal comparison and therefore the mass of the working standards may have changed slightly. Note that the mass of the working standards needs to be estimated at the exact time when the working standards are used as references.

Figure 10 shows the values of $u_{\text{drift,Ref}}$ evaluated for each calibration of national prototypes.

The total uncertainty $u$ is computed as

$$u = \sqrt{u_{\text{stat}}^2 + u_{\text{cw}}^2 + u_{\text{fit}}^2(m) + u_{\text{drift,Ref}}^2} \quad (8)$$

A combined standard uncertainty of 3 μg was finally assigned to the amended mass values for all calibration certificates.

The final curve showing the mass overestimation error that was committed during each past BIPM calibration of a national prototype, along with the associated combined uncertainty is shown in figure 8.

8. Summary

During the 2014 Extraordinary Calibration campaign with the IPK we report here, it was discovered that the BIPM as-maintained mass unit was offset by 35 μg. Consequently, we started a review of all internal mass comparison data recorded in the period 1992–2014 in order to investigate the origin and evolution of the offset since the end of the 3rd periodic verification.

A least squares analysis of all the available data was carried out. The mathematical model which best described the data suggests that a wear phenomenon associated with one of the BIPM mass comparators, which had been modified in-house, was responsible for the mass loss experienced by the
This mass loss occurred principally during the years 2003–2010. Since the IPK was not accessible between the 3rd periodic verification and 2014, a part of this collective mass loss went undetected, which explains the fact that it was only observed in 2014. This mass comparator has already been taken out of service in 2010.

As a consequence of the undetected mass changes, mass values attributed in mass calibrations during this period were generally overestimated by up to 35 μg. The mathematical model allows the evolution of this mass offset over time to be determined. This has allowed the BIPM to calculate corrected mass values which have been provided for all calibrations since 2003. Institutes involved in precision measurements of the Planck constant and the Avogadro constant have been informed, so that they can provide updated measurement results for the CODATA 2014 fundamental constants adjustment.

Following this analysis, the BIPM has entirely revised its weighing procedures leading to the establishment of a more robust hierarchy among the BIPM mass standards.

Acknowledgments

The authors thank Richard Davis for the collection of crucial data from historical BIPM mass comparisons and for helpful discussions. The BIPM acknowledges the pioneering work of Lars Nielsen in the framework of the CCM for suggesting the pertinence of a deterministic model [13] to describe the evolution of BIPM working standards and for the initial analysis of BIPM data he carried out. The BIPM wishes to express its gratitude to the members of the CCM support group for the Extraordinary Calibrations for their advice and guidance.

References

[3] Stock M 2013 Watt balance experiments for the determination of the Planck constant and the redefinition of the kilogram Metrologia 50 R1–16
[4] Azuma Y et al 2015 Improved measurement results for the Avogadro constant using a 35Si-enriched crystal Metrologia 52 360–75