



More Thoughts on AG–SG Comparisons and SG Scale Factor Determinations

DAVID CROSSLEY,¹  MARTA CALVO,² SEVERINE ROSAT,³ and JACQUES HINDERER³

Abstract—We revisit a number of details that arise when doing joint AG–SG (absolute gravimeter–superconducting gravimeter) calibrations, focusing on the scale factor determination and the AG mean value that derives from the offset. When fitting SG data to AG data, the choice of which time span to use for the SG data can make a difference, as well as the inclusion of a trend that might be present in the fitting. The SG time delay has only a small effect. We review a number of options discussed recently in the literature on whether drops or sets provide the most accurate scale factor, and how to reject drops and sets to get the most consistent result. Two effects are clearly indicated by our tests, one being to smooth the raw SG 1 s (or similar sampling interval) data for times that coincide with AG drops, the other being a second pass in processing to reject residual outliers after the initial fit. Although drops can usefully provide smaller SG calibration errors compared to using set data, set values are more robust to data problems but one has to use the standard error to avoid large uncertainties. When combining scale factor determinations for the same SG at the same station, the expected gradual reduction of the error with each new experiment is consistent with the method of conflation. This is valid even when the SG data acquisition system is changed, or different AG’s are used. We also find a relationship between the AG mean values obtained from SG to AG fits with the traditional short-term AG (‘site’) measurements usually done with shorter datasets. This involves different zero levels and corrections in the AG versus SG processing. Without using the Micro-g FG5 software it is possible to use the SG-derived corrections for tides, barometric pressure, and polar motion to convert an AG–SG calibration experiment into a site measurement (and vice versa). Finally, we provide a simple method for AG users who do not have the FG5-software to find an internal FG5 parameter that allows us to convert AG values between different transfer heights when there is a change in gradient.

Key words: Superconducting gravimeters, absolute gravimeters, scale factor, calibration.

1. Introduction

1.1. The SG at Apache

The original work in this paper was done on the Apache Point (AP, New Mexico, USA) SG–AG data from 2011 to 2015, and later applied to data from the J9 installation in Strasbourg, France. It is necessary to briefly discuss the site situation at AP, which is a first class astronomical observatory that hosts one of the best lunar laser ranging (LLR) facilities in the world. In 2009 an SG was installed at the site to assist in constraining the displacement of the ground during the LLR experiments which use a 3.5 m optical dish attached to a solid pier. Due to logistic and financial considerations it was not possible to place the SG in its own isolated building as is common at other geodetic sites, and as a compromise the SG was located in the cone room, a small access room directly under the telescope housing. There are both advantages and disadvantages to this location, but the subsequent difficulty of providing a suitable environment for the AG instrument to calibrate the SG was not considered.

It was quickly discovered during the first calibration experiment in 2011 that the AG was subject to excessive disturbance during the nighttime operations, when the telescope was in constant use, and these disturbances severely compromised the quality of the AG data. There are two effects associated with the telescope motion, one being small self-correcting offsets in the SG data (at the level of 0.5 μGal or less) due to mass changes associated with the telescope position above the gravimeter. These can be removed by constructing a model using additional data from the telescope slews, but this is a time-consuming operation that has not been done systematically for all the AP data, and not for the SG data used in the

¹ Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO, USA. E-mail: crossley@slu.edu

² Observatorio Geofísico Central, IGN, Madrid, Spain.

³ Université de Strasbourg, CNRS, EOST, IPGS, UMR 7516, 67000 Strasbourg, France.

calibrations. For the AG, the cooling system beneath the telescope blows air directly into the cone room and onto the AG instrument itself which cause data disturbances that are not damped by the F5 super-spring. Unfortunately, there is no possibility to avoid this problem by moving the SG/AG to another location in the observatory complex, and the only remedy with the AG is to reject all the disturbed data.

Thus we are obliged to use the AG data as recorded, and cannot easily improve the situation. Coupled with this is the limitation on residence time for the AG. Site requirements permit the AG instrument to remain in the cone room for 5 days or less, except for the first experiment in 2011 where it was allowed to run over a weekend (and gave by far the best data), and this severely limits the amount of good data we can collect. Our site is, therefore, one of the noisiest and most challenging for an AG–SG calibration experiment. Ground accelerations induced by the movements of a nearby VLBI antenna have also been also detected in the SG recordings at Ishigakijima, Japan (Imanishi et al. 2018), therefore calibration experiments at such a site might encounter similar problems.

1.2. Motivation

One of the motivations for this paper is to share our experience with the calibration experiments at AP that were initially done without the collaboration with the Strasbourg group that later became available, and thus represents the situation that might face a less experienced team of SG–AG operators. For example, an initial assumption at AP was that both the SG series and the AG series must be compared without any AG corrections (i.e., for tides, ocean-tide loading, local pressure or polar motion), and so all such corrections were turned off in the FG5 setup. Later it became clear that, as is done routinely at many observatories, the experiment can be done with the standard AG corrections, and the FG5 settings can be changed to remove the corrections for the calibration and produce uncorrected files. At SG installations where there is no in-house or dedicated AG, one may need external assistance for the FG5 instrument, which for the case of AP is the National Geospatial-Intelligence Agency (NGA). Frequently, such FG5's

are in heavy demand which limits their availability. Further, as mentioned, site constraints at AP dictate not only a very small space for the SG and AG, in a room in the middle of a very complex building, but visitations by the FG5 are disruptive of local operations so only one or two AG measurements per year are preferred. This would be similar to an SG being located remotely (e.g., Syowa, Antarctica), or in a special underground environment for hydrological purposes. It is obvious that whenever we did an SG calibration we needed also to produce an AG site measurement, which is the normal 1–2 day occupation with all corrections turned on, unlike a calibration experiment which normally takes at least 5 days.

Thus, a major goal of the paper is to process the SG–AG calibration data using only the FG5 text files, without access to the software that is supplied with the Micro-g software (denoted 'g-software') <http://www.microglacoste.com/pdf/g9Help.pdf>. The g-software gives complete user control over the processing of the AG fringe data, and produces a set of internal binary files and a set of 3 ASCII files—the drop data file, the set data file, and the project file that summarize the results of the processing. For the AP station the binary files were available from NGA, as was some limited re-processing, but were not useful to us at AP; the situation in Strasbourg is of course entirely different. This limitation was perfectly anticipated in Sect. 2 of the paper by Palinkas et al. (2012) who noted that some users of the gravity data have no access to the g-software (see the “Appendix” for further information). We acknowledge that the g-software is available independently of the instrument from Micro-g, but perhaps the suggestions in our paper may help some users to avoid that necessity.

Accurate calibration of a superconducting gravimeter (SG) is fundamental in many geophysical and physical applications, for instance for the search for time-variability in the Earth's response to tides induced by internal process inside the Earth or by surface loading (Calvo et al. 2014), or the search for anisotropy in the Newtonian gravitational constant G (Warburton and Goodkind 1976). Many papers cite ocean tide loading as a prominent requirement for accurate SG scale factors (along with accurate phase

calibration) e.g., Boy et al. (2003) and Baker and Bos (2003). Although much of the initial work at AP could have been avoided using the *g*-software, we hope some of the results are still of interest to those who contract out AG measurements, or perform only occasional calibration or drift checks on their SG.

There are numerous papers on the use of an AG to calibrate an SG, summarized in Hinderer et al. (2015). Here we investigate some small issues that arise in this type of comparison. In order of presentation, these are: (a) a discussion on the merits of various ways to use drop or set SG data, (b) the effect of adding the data acquisition time delay and a local trend to the solution, (c) combining multiple determinations of the scale factor for a particular station, and (d) comparing the AG offset from a calibration experiment to regular determinations of the AG site gravity. We use data from the Apache Point (AP) station in New Mexico, USA, and from J9 station in Strasbourg (ST), France to demonstrate the various points. As mentioned, AP is a site with especially high nighttime site noise, perhaps the extreme end of stations that have high cultural noise during some part of the day. This has been encountered at some older SG installations, e.g., Wuhan, China or Vienna (Meurers 2012) or a recent one at Ishigaki (Imanishi et al. 2018), whereas ST is typical of a station with quiet and fairly constant site noise. Van Camp et al. (2016) make the useful suggestion that higher drop rates (e.g., every 5 s) should be used, and this would be beneficial in future for AP measurements during the undisturbed daytime recording.

1.3. Basic Equations

To begin, we assume a simultaneous measurement of AG and SG gravity over a time period $T \sim 5$ days, to be assured of reaching a reasonable convergence in the scale factor (e.g., Francis 1997; Meurers 2012). All the SG data come from either the raw 1 s data, or the filtered 1 min files available at GGP/IGETS (Crossley and Hinderer 2010; Voigt et al. 2016). Very little of our SG data at AP required corrections for SG-specific disturbances such as He refills, disturbances, or offsets but simple pre-processing was done where necessary. Likewise there was no problematic data (such as a large earthquake)

that would have affected both instruments (in different ways) and therefore, to be avoided. The common time period T was chosen to span a period during the largest diurnal tides at the station, which occur fortnightly. Pre-processing of some SG data from ST was done to avoid data disturbances, as described in Rosat et al. (2009).

The FG5 data at both stations, denoted by $y(t)$, was collected drop-by-drop every 10 s, and accumulated each 20 min as a set mean of 100 drops. In our original processing, the SG data $x(t)$ was 1 min smoothed from 1 s raw data by applying a low-pass filter, which avoids the problem of aliasing of the 5–10 s microseismic noise (Van Camp et al. 2016). This is still present even for station AP in the middle of the N. American continent, but less than ST in Central Europe. The SG data is normally cubic-splined to the AG drop or set times which are given at a sampling time t (Rosat et al. 2009). Later we also used 1 s data for comparison.

The AG data is composed of a constant mean value y_0 (over the time period T of the experiment) plus a time-varying part $y_1(t)$; similarly the SG data is composed of a constant part x_0 plus a time-varying part $x_1(t)$. We perform a least-squares (LSQ) fit of $y(t)$ (μGal , 10^{-8} m/s²) to the SG data $x(t)$ (volt) using

$$y(t) = \alpha + \beta \times x(t) + \gamma \times t + \varepsilon \quad (1)$$

where ε is assumed to be Gaussian random noise, and the sum of ε^2 is minimized. The parameters determined from the fit are α , the offset between the mean zero levels of the AG and SG data, β the scale factor SF (or calibration constant) of the SG (μGal or $\text{nm s}^{-2}/\text{volt}$), and γ is a trend to account for possible differential instrument drifts (see e.g., Imanishi et al. 2002). If no mean values are subtracted, then

$$\begin{aligned} x(t) &= x_0 + x_1(t) \\ y(t) &= y_0 + y_1(t) \end{aligned} \quad (2)$$

and after the LSQ fit (we use *lfit* from Numerical Recipes, but any similar code will do) for (α, β, γ) we can equate, within the errors, the constant and time-variable parts

$$y_0 = \alpha + \beta \times x_0$$

$$y_1(t) = \beta \times x_1(t) + \gamma \times (t - t_0) \quad (3)$$

where t_0 has been added to indicate the time of the first AG drop. We refer to the quantity y_0 as the *AG mean value*, which depends on the fitted offset α , the SG mean x_0 (which can be computed separately prior to the fit), and β —the SG scale factor. With Gaussian errors, if the standard deviations for (α, β) are σ_α and σ_β , then the variance of y_0 is

$$\sigma_0^2 = \sigma_\alpha^2 + x_0^2 \times \sigma_\beta^2 + 2x_0 \times \sigma_{\alpha\beta} \quad (4)$$

where $\sigma_{\alpha\beta}$ is the covariance of (α, β) .

A few points need to be mentioned about these equations. First, many authors do not consider the offset α , nor the mean values x_0 or y_0 , as being of sufficient interest to mention, and others ignore the trend γ , thus leaving β as the only parameter of interest. This is understandable if one chooses to get a regular AG site value at the site by reprocessing the AG data using the *g*-software. Later we show another method to do this based on y_0 . As for the offset α , Imanishi et al. (2002) discussed in some detail its variations for a month-long series of AG-SG measurements, and ascribed the cause to possible AG instrumental drift during the experiments. The possibility of such an effect is one reason we include the term $\gamma \times t$ in (1). Unfortunately we could not repeat the experiment of Imanishi et al. because we could only record at AP over a few days, but any linear trend in α will appear in the $\gamma \times t$ term in (1). Short-term effects over the time T of the calibration are distinct from the classic long-term SG drift, but it is reasonable that the latter be removed first from the SG data, though it is unnecessary here.

Wziontek et al. (2006) add explicitly a drift function to the SG component, and an offset to the AG data, but no AG trend. Although arbitrarily adding a drift parameter to (1) without being able to identify the reason might seem unjustified, Meurers (2012) clearly showed that a linear trend can perturb the amplitude ratio between the AG and SG data (which is the goal of the calibration), and so there is a good reason to include it. Many other papers also advocate a drift parameter, for instance Hinderer

et al. (1991) included drift when using an earlier JILA-5 instrument with twin laser drift problems, and a drift is explicitly included by Tamura et al. (2005) and Van Camp et al. (2016). Meurers (2002) explored the effect of unmodeled drift on the calibration factor using synthetic and real datasets. We also received a suggestion that He gas from the SG might leak into the room and affect the AG; this could have a preferential effect on one instrument and not the other (B. Meurers, editorial comment). This phenomenon has also been reported by Mäkinen et al. (2015), but it is not usually a problem for closed cycle SG's such as the current observatory SG or iGrav. Note that at the J9 station in Strasbourg, unlike at AP, the AG is recording in a separate room from that of the SG.

2. Drops or Sets?

Early calibration experiments in Strasbourg (Hinderer et al. 1991) used only 1-day experiments but used both drop and set data, and also considered both the L_1 (minimum absolute error) and L_2 (LSQ) norms when solving for the constants in Eq. (1). It appears the L_1 norm has not been widely used in recent years. Amalvict et al. (2001, 2002), however, showed that with good data there was little difference in the scale factor between drop and set methods, and noted that the errors (which they stated to be standard deviations) in both methods were similar despite the very large difference in the number of SG–AG pairs to be fitted (generally there are about 100 drops for each set) which should result in a smaller formal error using drop data. Although there has been a recent trend in SG–AG calibration processing towards the use of drop data rather than set data, obviously the less numerous AG set values are still less scattered than the drop values.

Several recent papers have covered similar ground to this study, and with which our results are consistent. Tamura et al. (2005) for example used AG drops, and found no evidence for scale factor changes at Esashi (Japan). Wziontek et al. (2006) identified AG offsets at station Bad Homburg (Germany) from calibration experiments using different FG5 instruments, and also treated mainly drops. Meurers (2012) used drops in a comprehensive assessment of many of

the factors in SG–AG processing, and Van Camp et al. (2016) also favored using drops, emphasizing the need not just for many drops, by increasing the drop rate, but measuring at high tides to improve accuracy.

We need to be clear about the difference between the two types of measurement. An AG drop results in a trajectory of a falling corner cube, whose flight is sampled by a number of fringe zero-crossing times of which there are a large (many thousand) number of fringes per drop (see e.g., Kren et al. 2016). The variance covariance matrix of the LSQ fit to the fringe crossings yields a statistically determined drop value and a scatter, or standard deviation, σ_d . When drops are processed in sets (often 100 drops, every 10 s) the set mean is the unweighted mean of the accepted drops averaged over a set. One can take the set sigma σ_s as the usual standard deviation of the set mean [Eq. (A2) in the “Appendix”] that reflects the drop to drop scatter (column labeled ‘Sigma’ in the set text files). This is our choice for most of this paper, except for the final two Figures. Alternatively one may choose the standard error of the set mean (SEM), which is σ_s/\sqrt{N} where N is the number of drops per set (unweighted), as in Tables 3 and 4. N is frequently close to 100, so the set SEM is about 10 x smaller than σ_s , and given by the column ‘Error’ in the set file. Drops are accepted or rejected by the g-software based on the usual 3- σ criterion, i.e., a drop outlier is rejected when more than 3- σ from the set mean. When using single drops, more flexibility is available to select drops in the solution, as we will see. Rather than using ‘set sigmas’, and ‘drop sigmas’, to avoid any ambiguity we frequently refer to the columns (Sigma, Error) in the drop files and (Sigma, Error) in the set files.

2.1. Tests on Set Data

For reasons that will be clear later, we wish to also find y_0 from the fit, and this requires an initial assessment of the SG mean value x_0 ; certainly x_0 can be ignored if the only goal is to find β . Various possibilities for the span of SG data were tried: (i) SG values starting at UT 0 the first day, and ending at midnight the last day, (ii) using SG values only at AG times, and (iii) using SG values starting at the first

AG drop time and ending at the last AG drop. The differences in the scale factor were (and expected to be) insignificant, but there was an affect at the μGal level on the AG mean value y_0 in Eq. (4). Obviously option (iii) is the logical choice of the time span for the SG data.

A test was also made to quantify the effect of the SG instrument time delay to the experiment, as mentioned previously. For SG 046 at Apache Point, an observatory-style gravimeter, the nominal time delay (lag) of the system is predominantly that of the GGP1 filter (Hinderer et al. 2015) which is 8.16 s, so this delay has to be incorporated in any calculation that returns the SG value at the AG times. We tested the shift in the scale factor for various time delays in Table 1, showing that for most stations, the effect is negligible even up to 30 s. This confirms similar results of Meurers (2012) and Van Camp et al. (2016).

Again using the AP 2011 data, another test was done to assess the effect of a relative drift between the SG and AG data, i.e., adding the term $\gamma \times t$ in the RHS of Eq. (1). This is not primarily to account for the known SG drift, but allows for other effects occurring preferentially in one of the instruments. The instrument drift of SG046 between 2009 and 2012 was 70 $\mu\text{Gal}/\text{year}$, or + 0.192 $\mu\text{Gal}/\text{day}$ —which is unusually large for an SG. For this reason, the sensor was replaced in 2013 by the manufacturer GWR (Goodkind–Warburton–Reineman, San Diego, California) with a significant decrease in drift. The results are shown in Table 2 where we give the scale factor, the trend, and the AG mean value, all with a time lag of 8.16 s. We have included the errors σ_β to show that although the trend can be larger and of opposite sign than the known SG drift, its effect on β is always smaller than σ_β . The same situation occurs for the AG mean value y_0 . Note, however, that σ_β is not determined very accurately by these set solutions at AP. We note that the trend can also be regarded as a diagnostic for possible problems in the data, e.g., the trend for AP2016 is – 0.55 $\mu\text{Gal}/\text{day}$, which is sufficient to perturb the offset and SF. In this case, we know the AG data quality for 2016 was rather low.

The argument for using AG set values for SG–AG calibration probably arises because it is the natural choice when determining an AG site value, where the

Table 1
Effect of SG time delay on scale factor ($\mu\text{Gal/V}$)

Time delay (s)	Scale factor β +/-0.407	% change
0 ^a	- 79.397	0.0
8.16	- 79.399	0.003
15.0	- 79.4015	0.005
30.0	- 79.4027	0.01

AP 2011 experiment, all set data

^aReference

Table 2
Effect on the scale factor and base value of adding a trend to the SG-AG fit

AP set experiment	Trend γ ($\mu\text{Gal/day}$)	Scale factor β ($\mu\text{Gal/V}$)	Change in β ($\mu\text{Gal/V}$)	Mean value y_0^a (μGal)	Change in y_0 (μGal)
July 2011	- 0.03 \pm 0.22	- 79.39 \pm 0.41	+ 0.004	5164.83 \pm 0.87	- 0.114
Nov 2013	+ 0.08 \pm 0.84	- 94.40 \pm 2.43	+ 0.104	5189.66 \pm 1.69	+ 0.135
Sept 2014	- 0.36 \pm 0.99	- 94.33 \pm 1.02	+ 0.098	5161.53 \pm 1.40	- 0.417
June 2016	- 0.55 \pm 0.78	- 94.96 \pm 1.15	+ 0.279	5136.14 \pm 2.35	- 1.476

After the July 2011 experiment the sensor at AP was changed, with a new scale factor. Uses JS3 pass 2 method and set data processed with the Sigma column of the set file; values include γ and changes are relative to $\gamma = 0$

^aAdd value to 9788050000 μGal

geophysical corrections are applied to get the site gravity. Among past papers, Rosat et al. (2009) used AG set values when doing SG-AG experiments, and there is some benefit in having a set average with a well-defined set sigma (σ_s) used in weighting the fit (as we will see). On the other hand, there are two arguments against using sets, the first being the inability of the set average to precisely track the top and bottom of the semi diurnal tides if the regular set averages are used. To combat this, Meurers (2012) suggested using a moving window average of both AG and SG data to help reduce the AG scatter and yet track the tidal signal more precisely. At each drop time an average of the AG and SG data is taken over the length of a set spanning the drop point, thus keeping the high number of drop values, but reducing the drop to drop scatter. This is related to the second point that in principle set averages, rejecting drop outliers, should work better on AG data where the geophysical corrections (principally tides and atmospheric pressure) have been subtracted, and the corrected signal has only a small scatter. The

situation is different when doing SG-AG calibrations that require the full tidal signal; the AG set averages are biased by the changing level of the large time-varying signal. The importance of this procedure is one of the options we test in our processing. Recent authors have tended to recommend the use of AG drops to generate the scale factor, ignoring the trivial increase in computer time over the set method.

2.2. Initial Attempts at Set and Drop Processing

We start with Fig. 1 showing the fit of SG-to-AG data between July 28 and Aug 3, 2011, which was the span of the first AG measurement at Apache Point Observatory after installation of SG046 in February 2009.

The fit is based on AG set values (100 drops/set, set interval 20 min, drop interval 10 s), where it is seen that the σ_s 's are considerably larger during the nighttime hours when the LLR telescope is active for various sky surveys. The set mean can nonetheless be acceptable if poor drops are rejected and de-

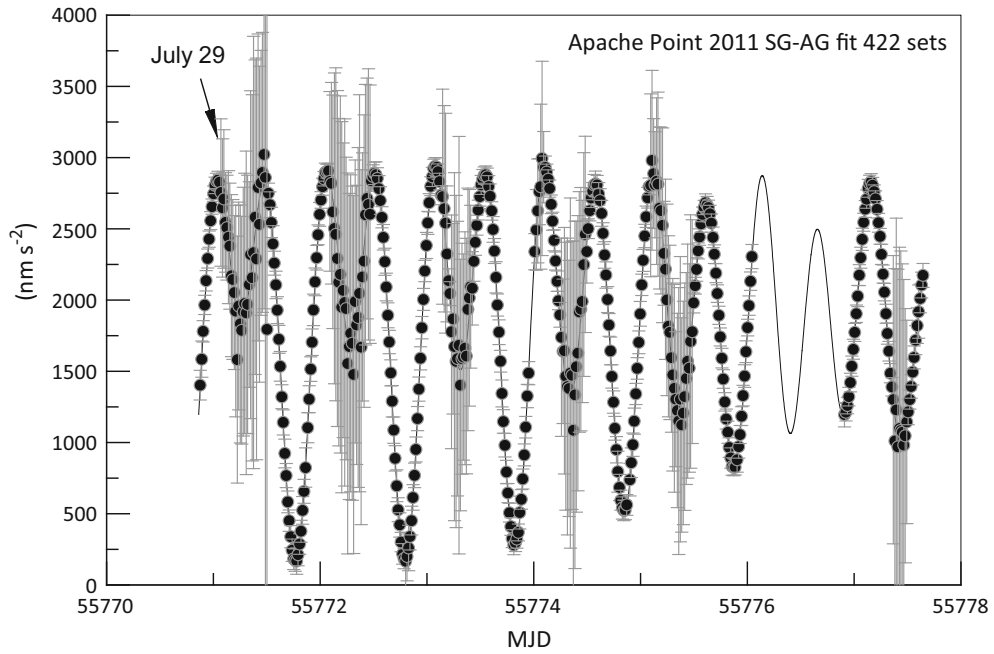


Figure 1

Fit of all Apache Point set data between July 28 (MJD 55771) and Aug 3, 2011. The noisiest data, indicated by large set standard deviations (the Sigma column in the set files), varies according to the telescope viewing schedule

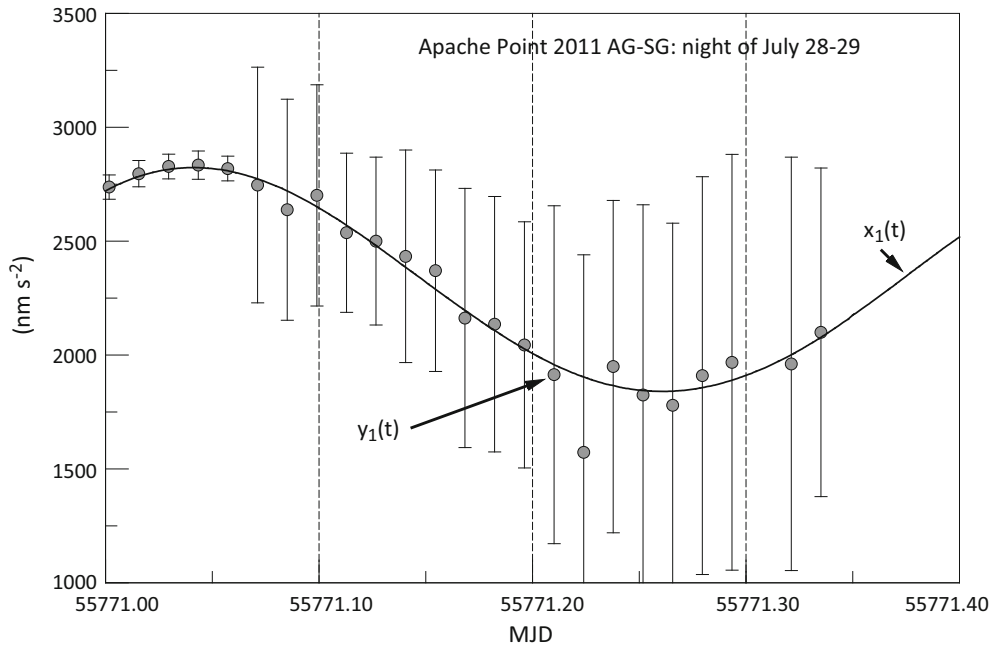


Figure 2

Zoom of Fig. 1 for night of July 28–29 (about the first half day of the experiment) showing the set means are also disturbed by the telescope activity. Labelled are the fitted SG and AG measurements $x_1(t)$ and $y_1(t)$, Eq. (4). Almost all such sets are discarded in the processing

emphasized in the set means by their large inverse variances (i.e., $1/\sigma^2$) in the SF fit. As can be seen in Fig. 2, zoomed to the first night of the calibration, not only are the set errors large, but the set means deviate from the SG values.

There were four SG–AG calibrations done for AP between 2011 and 2016, and for comparison we processed four datasets from ST between 2008 and 2011. We first assessed the histograms of the drop and set σ 's (the columns marked 'Sigma' in the FG5 text files) for all datasets used in this study and these are shown in Fig. 3a for AP, and Fig. 3b for ST. For AP the drop sigmas are clearly divided into two groups, for every experiment. There is a tight clustering of values around 16 μGal for AP2011 and AP2013, and similarly around about 25 μGal for AP2014, AP2016. Beyond 30 μGal , the drop sigmas extend to very high values (several 100 μGals) for badly disturbed data, and so we choose a maximum acceptable cutoff value of $\delta g_m = 30 \mu\text{Gal}$ for all AP data. For the AP set data, however, the sigmas vary from a peak around 5–7 μGal , to a scatter of values up to about 30 μGal , suggesting the latter is also a suitable set cutoff to avoid disturbed data. The situation is different for the ST data where there is very little disturbed data, but the histogram peaks also occur at different values depending on the experiment. The drop sigmas for ST divide between a low of about 10 μGal for ST2010 and ST2011 to between 25 and 30 μGal for the other 2 datasets. The ST set sigmas vary between 5 and about 27 μGal , so to include most data we choose for ST a rather large value of $\delta g_m = 35 \mu\text{Gal}$ as a cutoff, but this is not a critical value because there are very few sets above 25 μGal .

We processed the four SG–AG experiments at AP using both set and drop data. Initially we used all the set or drop data in the files, without prior selection, and compute the solution relying on the weights (inverse variances from the drop sigmas) to reject large set or drop σ 's. This initial processing is called JS0 (selection method 0) and yields the two rows 'JS0' in Table 3. To be explicit this procedure consisted of:

(a) Selecting only those drops or sets that had sigmas below the sigma cutoff discussed above (shown

in Fig. 3) and weighting the AG data according to their inverse variances, and

(b) For the SG data we used 1 min GGP data and interpolated the SG value to the drop time or set time that matched the AG times. This is the procedure discussed in Rosat et al. (2009).

The result is that much data was discarded, depending on the experiment, and in the worst case for the AP2013 experiment less than 30% of the data could be used. Notice in Table 3 that the drop SFs (scale factors) are sometime quite different to the set SFs, whereas the latter (JS0) are generally similar to the other 'better' JS solutions (described below). The reason is that the drop σ 's are not necessarily indicative of which drops are far from the SG curve, and so the weighting does not automatically diminish the influence of drop outliers that may have acceptable σ_d 's. This is not true of the set σ 's, as they are more robust against bad drop data, and sets with high σ 's are degraded in the solution against sets with small σ 's. Table 3 shows the results for JS1 and JS3 (see later) for passes 1 and 2, but note that pass 2 is not required for some of the set solutions if there are no set residuals outside the 3- σ criterion.

Considering the difficult AP data, namely the lack of agreement between the scale factor for sets and drops, we wondered if it was possible to improve the fit by further selecting the AG drops. Because most drops had about the same σ_d , they all equally contribute to the SG fit, but in reality many drops are far away from the SG curve; perhaps these drops degrade the scale factor fit and could be rejected? It is also clear that drops near the tidal peaks are more important in determining the scale factor that those midway between peaks. Rather than pursue such an approach, we changed strategy and decided to test a number of options presented in the AG–SG processing. For the moment we pass over sections (c) of Tables 3 and 4, and return to them later.

2.3. Improving the Algorithm for Rejecting Bad Data

When collaboration began with the Strasbourg group, and especially after the paper by Calvo et al. (2014), it became clear that better ways to reject bad data had been gaining acceptance, following the

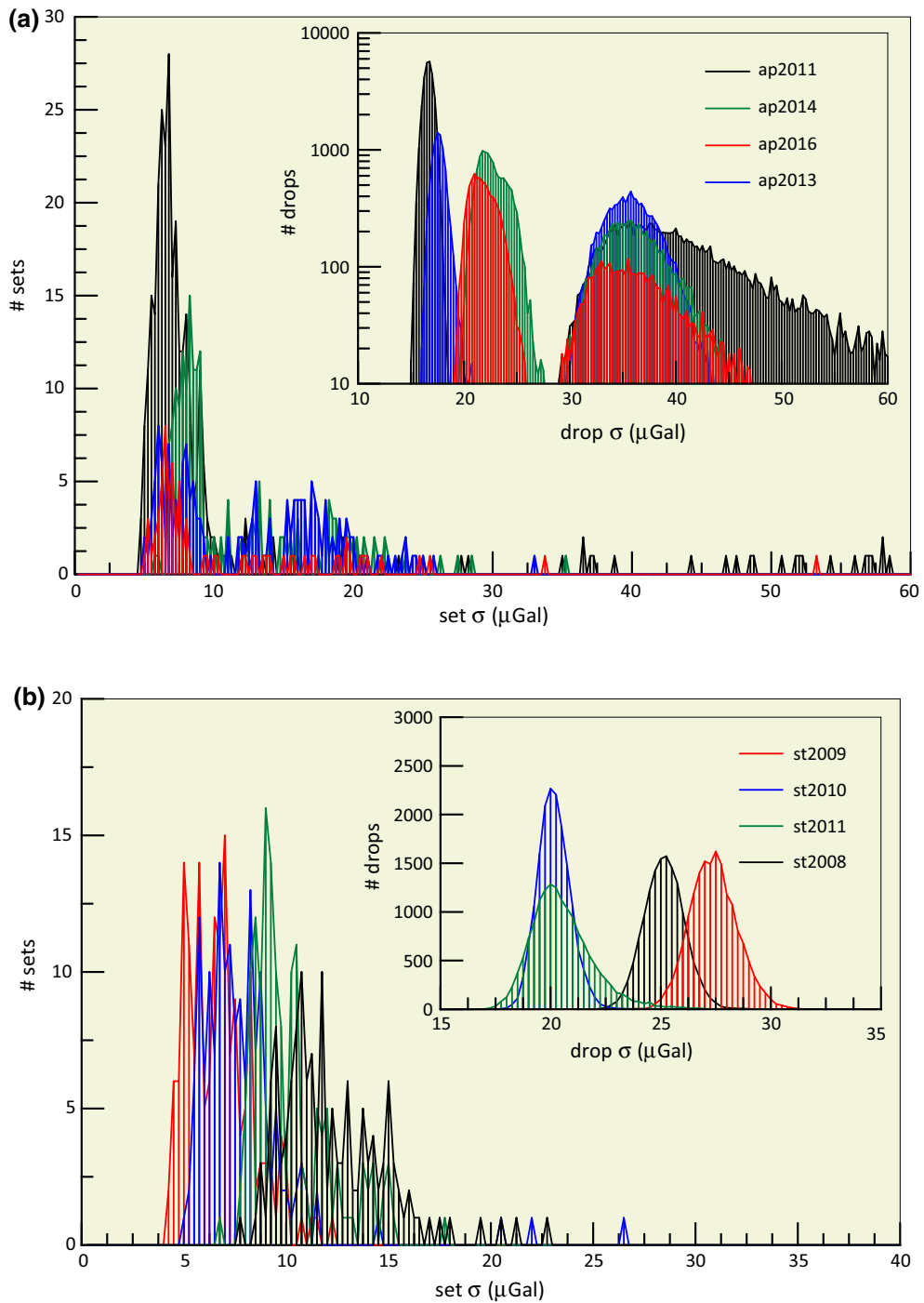


Figure 3

Histograms of standard deviations of drop and set values for 4 datasets from **a** Apache Point and **b** J9 Strasbourg. The AP drop σ 's are divided into two groups separated just below $30 \mu\text{Gal}$ for the better data and $> 30 \mu\text{Gal}$ for the poor nighttime data. Note that the range of the σ 's for data $< 30 \mu\text{Gal}$ is similar for both drop and set errors at both stations. One could also plot the set histograms based on the 'Error' column in the set files, in which case the set σ 's are about a factor 10 smaller when there are ~ 100 acceptable drops per set

Table 3
Set versus drop scale factors β ($\mu\text{Gal}/V$) for AP experiments; drop and set $\delta g_m = 30 \mu\text{Gal}$

Start date	2011 Jul 28	2013 Nov 18	2014 Sept 8	2016 Jun 2
Sets/drops recorded	422/42,200	229/22,900	207/20,700	117/11,700
Sets used/% data	283/67%	64/28%	113/55%	55/47%
Drops used/% data	27,493/65%	6083/27%	11,018/53%	5092/44%
(a) DROP SF's				
JS0	-79.044 ± 0.092	-112.13 ± 0.525	-94.732 ± 0.270	-96.503 ± 0.245
JS1 pass 1	-79.342 ± 0.093	-94.148 ± 0.437	-94.686 ± 0.261	-94.930 ± 0.311
JS1 pass 2	-79.384 ± 0.094	-94.357 ± 0.445	-94.541 ± 0.263	-94.830 ± 0.315
JS3 pass 1	-79.341 ± 0.093	-94.451 ± 0.626	-94.262 ± 0.283	-94.948 ± 0.313
JS3 ^a pass 2	-79.389 ± 0.094	-94.420 ± 0.628	-94.317 ± 0.283	-94.847 ± 0.315
(b) SET SF using σ_s				
JS0	-79.328 ± 0.408	-94.039 ± 1.240	-94.123 ± 0.847	-94.692 ± 1.125
JS1 pass 1	-79.325 ± 0.405	-94.001 ± 1.731	-94.439 ± 0.976	-94.823 ± 1.149
JS1 pass 2	-79.348 ± 0.406	-94.013 ± 1.735	^b	^b
JS3 pass 1	-79.373 ± 0.405	-94.399 ± 2.433	-94.331 ± 1.016	-94.911 ± 1.147
JS3 pass 2	-79.395 ± 0.406	^b	^b	-94.958 ± 1.149
(c) SET SF using SEM				
JS0	-79.325 ± 0.041	-94.029 ± 0.125	-94.121 ± 0.085	-94.674 ± 0.114
JS1 pass 1	-79.329 ± 0.041	-93.994 ± 0.175	-94.431 ± 0.099	-94.821 ± 0.117
JS1 pass 2	-79.351 ± 0.041	-94.006 ± 0.176	^b	^b
JS3 pass 1	-79.378 ± 0.041	-94.401 ± 0.248	-94.287 ± 0.162	-94.957 ± 0.118
JS3 pass 2	-79.398 ± 0.041	^b	^b	^b

^aJS3 (bold) for drops is the preferred solution. Sections (a) and (b), drop and set errors from 'Sigma' column of respective text files, (c) set errors from 'Error' column of set files

^bNo pass 2 solution required, same value as pass 1

approach contained in Meurers (2012). We mention again the goal here was to process the AG and SG data using only the information in the FG5 text files, and further to do this on files that come directly from the FG5 without any pre-processing of the AG data using the g-software. This was the case for the AP text files received by mail from NGA, whereas the ST data were already carefully pre-processed to reject bad drops before the text files were written—a significant difference (and improvement).

The following discussion addresses a number of options that we tried in rejecting bad drops, which is the key to getting SFs that are consistent when using both drops and sets. The options are summarized in Table 8 in the "Appendix" with the same abbreviations as here:

1. w_1-w_0 : weighting the data when computing means. When the g-software records each drop it is compared to an evolving mean value that eventually becomes the set mean, and there is no chance to weight the drops. But after the drops and

sets are recorded one can re-compute the set means by weighting the drops with their σ_d 's, in principle this being a better approach. So we decided to base all subsequent processing on just the drop data files, and used the set file data only as a check. We could recover the set file data by computing the unweighted mean of the set mean, and also by adopting the 3- σ rejection of drop outliers. We verified that exactly the same drops were rejected as reported in the set files (columns 'accept', 'reject') and with exactly the same set means. The 3- σ rejection is iterated successively until the number of rejected drops does not change (a maximum of 5 iterations proved adequate). The unweighted solutions are designated ' w_0 ' and the weighted solutions are ' w_1 '. The weighting (or not) is applied to every instance in the program where set means are required; our default preference is to use the weighted option.

2. s_3-s_1 : using 3- σ or 1- σ criterion for rejecting outliers. Amalvict et al. (2002) reported on the

Table 4
As Table 3 but for 4 Strasbourg J9 calibrations

Start Date	2008 Dec 17	2009 Jun 16	2010 Jun 10	2011 Oct 11
Sets/drops recorded	143/14,300	167/16,700	168/16,800	164/16,400
Sets used/% data	139/97%	164/98%	161/96%	159/97%
Drops used/% data	13,791/96%	16,478/99%	15,678/93%	15,093/92%
(a) DROP SF's				
JS0	– 79.133 ± 0.390	– 79.154 ± 0.230	– 79.008 ± 0.162	– 78.989 ± 0.248
JS1 pass 1	– 78.951 ± 0.396	– 79.216 ± 0.231	– 78.998 ± 0.162	– 78.930 ± 0.249
JS1 pass 2	– 78.875 ± 0.397	– 79.183 ± 0.232	– 78.999 ± 0.164	– 78.965 ± 0.249
JS3 pass 1	– 78.956 ± 0.396	– 79.212 ± 0.231	– 79.012 ± 0.162	– 78.931 ± 0.248
JS3 ^a pass 2	– 78.877 ± 0.397	– 79.183 ± 0.232	– 79.003 ± 0.164	– 78.969 ± 0.249
ST processed	– 78.842 ± 0.389	– 79.231 ± 0.230	– 78.549 ± 0.161	– 79.020 ± 0.246
(b) SET SF's using σ_s				
JS0	– 78.841 ± 1.725	– 79.160 ± 0.557	– 78.999 ± 0.574	– 78.892 ± 1.144
JS1 pass 1	– 78.824 ± 1.722	– 79.162 ± 0.556	– 78.986 ± 0.573	– 78.998 ± 1.111
JS1 pass 2	– 78.832 ± 1.722	– 79.156 ± 0.557	– 78.975 ± 0.579	– 78.974 ± 1.112
JS3 pass 1	– 78.856 ± 1.723	– 79.189 ± 0.556	– 79.001 ± 0.573	– 79.037 ± 1.111
JS3 pass 2	– 78.863 ± 1.723	– 79.183 ± 0.556	– 78.991 ± 0.579	– 78.930 ± 1.123
ST processed	– 78.879 ± 1.707	– 79.148 ± 0.554	– 78.975 ± 0.572	– 78.981 ± 1.139
(c) SET SF's using SEM				
JS0	– 78.837 ± 0.173	– 79.160 ± 0.056	– 78.997 ± 0.058	– 78.894 ± 0.115
JS1 pass 1	– 78.825 ± 0.173	– 79.162 ± 0.056	– 78.985 ± 0.058	– 78.928 ± 0.115
JS1 pass 2	– 78.833 ± 0.173	– 79.156 ± 0.056	– 78.978 ± 0.058	– 78.917 ± 0.115
JS3 pass 1	– 78.857 ± 0.173	– 79.188 ± 0.056	– 79.001 ± 0.058	– 78.963 ± 0.115
JS3 pass 2	– 78.864 ± 0.173	– 79.183 ± 0.056	– 78.993 ± 0.058	– 78.954 ± 0.115
ST processed	– 78.879 ± 0.341	– 79.148 ± 0.111	– 78.975 ± 0.114	– 78.981 ± 0.228

^aJS3 (bold) for drops is the preferred solution. Sections (a) and (b), drop and set errors from 'Sigma' column of respective text files, (c) set errors from 'Error' column of set files

choice of ' n ' in using an ' n - σ ' selection. They chose 3- σ for drops and 1- σ for sets, so we decided to test both options when rejecting outliers. It is expected that this choice depends on the noise in the experiment; for good data it should make little difference.

3. co–nc data: should we do drop selection on the uncorrected or corrected data? We use 'TLBP' to refer to the geophysical corrections (tide, ocean load, barometric pressure loading, polar motion) that are generally removed to get a regular AG site measurement 'co', as opposed to the 'nc' option that does not remove TLBP when rejecting outliers. This point has been emphasized in the papers by Meurers (2012), Calvo et al. (2014), and Van Camp et al. (2016). As discussed previously, if corrections are not made, the drop rejection is compromised by the inclusion of tides (predominantly) which vary throughout the experiment. Once the drops are rejected, the TLBP corrections can then be reapplied to the accepted AG data for

use in the calibrations. At AP we requested the NGA operator to re-run the four calibration experiments with corrections applied, and based our rejections on those files rather than the uncorrected data. In this case it is necessary to have exactly consistent drop and set files to transfer the accept/reject criteria between corrected and uncorrected versions of the same data. This worked well for the AP data, but unfortunately for the Strasbourg experiments the loss of a computer disk in 2014 with all the processed data meant that we did not have ready access to the corrected text files, although the raw AG files are archived. Up to the present we have not tested the 'co–nc' option for the ST data.

4. p_1 – p_2 processing: adding a drop/set rejection of outliers after the first fit was obtained. This turned out to be a significant point. Once pass 1 is made, one has access to the residuals—the deviations between the SG curves and the AG drop or set values. These residuals have a more or less normal

distribution about the SG curve, so it is easy to reject deviations on a $3\text{-}\sigma$ (or $1\text{-}\sigma$) reject criterion; this is equivalent to refining the pass 1 solution based on rejecting outliers, or equivalently choosing drops or sets deviations that are close to the SG curve. For most datasets, even those that had been carefully screened manually using the g-software in Strasbourg, pass 2 found sufficient drops or sets to discard on all the datasets that the solution was noticeably improved. For some of the poorer AP sets up to 900 new drops were rejected and up to 5 more sets. It should be said that all the processing for drop or set SFs up to pass 1 use exactly the same accept/reject drops or sets, but at the last step, the pass 2 solution for drops rejects only drops, and the pass 2 solution for sets rejects only sets, so a small difference appears in the data used for the two types of SF.

An illustration of the effectiveness of this procedure is shown in Fig. 4 for the AP2011 experiment. As seen in Table 3, most calibrations are improved using pass 2 in the sense that the drop and set SFs are brought closer.

In addition to the above options (1)–(4) that could be invoked, we chose 3 additional methods to discard drops/sets. Again these are summarized in Table A1, and described below as a series of processing steps. Within each step we can choose any of the above options.

Step 1: From the drop files with TLBP corrections, compute the set means and implement rejection of drop outliers (with iteration), flag all drops as accept/reject, and save these flags as method ‘JS1’ (rejection based on set means). For the drop SFs only accepted drops are used, and for the set SF the accepted drops are gathered into sets and the set means are used as the data (as usual). Complete sets are rejected if their $\sigma_s > \delta g_m$ and such sets even with good drops are discarded when doing set SFs. Step 2: From the same drop files, reject drops based on the drop $\sigma > \delta g_m$ (the cut-off σ shown in Fig. 3), flag all drops as accept/reject, and save these flags as method ‘JS2’ (rejection based on drop σ). The procedure then follows JS1.

Step 3: Combine the accept/reject flags from the previous 2 steps, so drops are rejected as method

‘JS3’ (drop rejection based on both set means and drop σ 's).

Step 4: From the drop text files for the calibration with no corrections ‘nc’, we apply the reject flags on all drops identified from the previous three methods (JS1, JS2, and JS3). This completes the preselection of AG drops.

Step 5: Prepare the SG data in two ways. First we use the old method (JS0) for interpolating the 1 min data to either AG drop times, or AG set times, depending on whether we are using drop or set data; this method is combined with the JS1 and JS2 AG selection above. Second we use the SG 1 s files, select all data spanning the AG drop times, then filter the data to reduce the effect of the microseismic noise; this was suggested explicitly by Van Camp et al. (2016) and turns out to be a valuable suggestion for improving the SFs.

Figure 5 shows a suite of filters, based on the Parzen data window, with lengths from 11 to 501 points and with frequency cutoffs between 3 and 50 s. They are designated $sf_1\text{--}sf_9$, where the last filter sf_9 is the original 1 s to 1 min GGP filter. The effect of applying these filters to the four ST datasets on the calibrations is shown in Fig. 6.

It is clear that there can be a bias in the SF, depending on the data quality, if the SG data are inadequately smoothed. For one dataset, ST2010, the SF using no filter (sf_0) shows a noticeable shift in SF that can be corrected using a filter such as sf_6 (or longer). For this experiment there was a large earthquake that had to be removed from the data, and it was assumed the rest of the data was OK, but the effect of the earthquake carried over unseen in the data. All scale factors using the Calvo et al. (2014) method are obtained from the raw 1 s data without filtering. Here, filtering using sf_6 is used in all the calibration solutions with the ‘JS3’ designation. It should be pointed out that Meurers (2012) uses the SG 1 min data resampled to the AG drop times in his solution, which is equivalent to the original JS0 processing here. Even though it might be technically better to use filtered raw 1 s data, the SF difference between sf_6 -filtered 1 s data and the standard 1 min GGP data is negligible and there appears to be no bias introduced using the 1 min SG data directly. One other advantage of filtering the SG data at AP is a

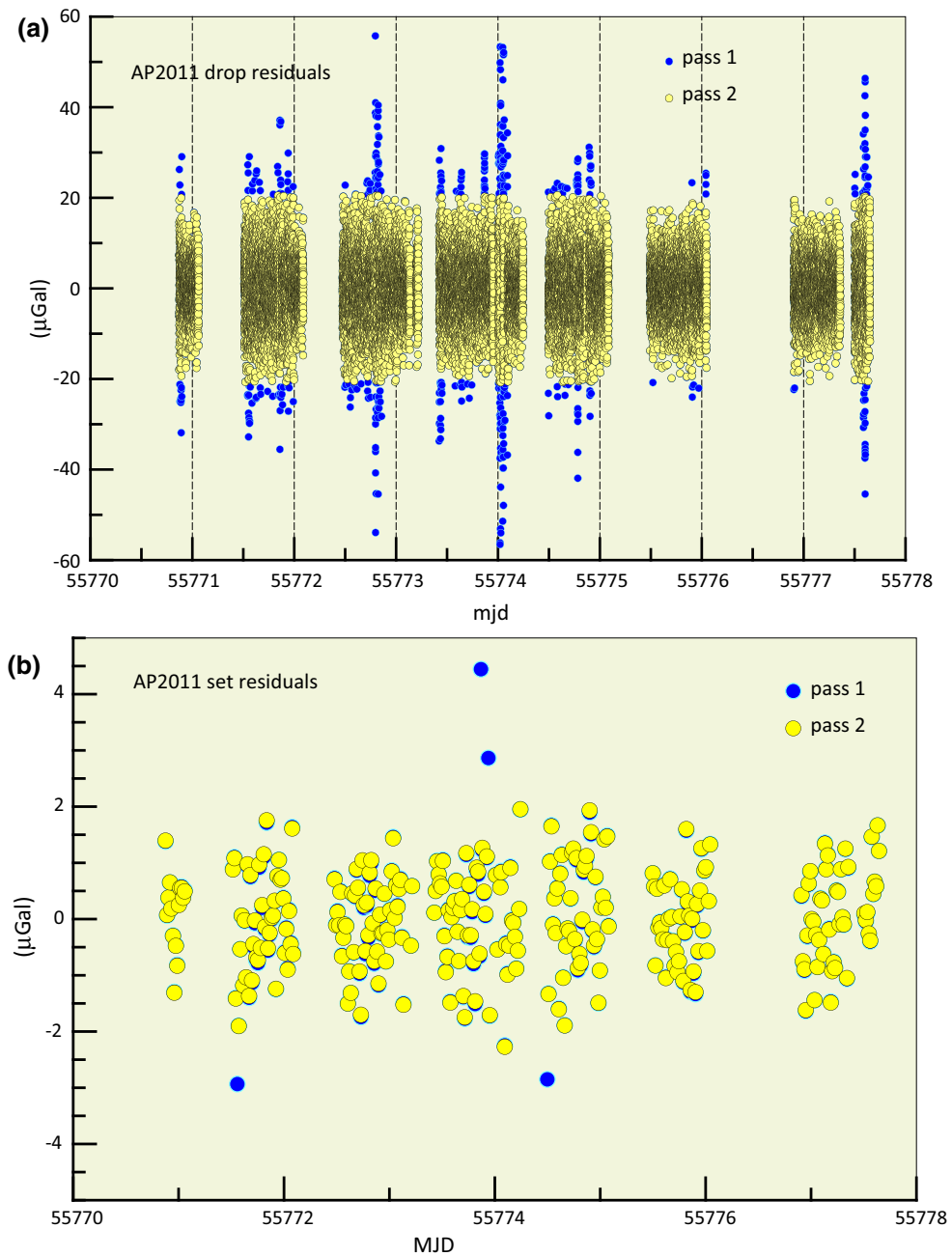


Figure 4

AG residuals **a** drop, and **b** set, after pass 1 and pass 2 of the AP 2011 calibration processed using the JS3 option (see text). The residuals show the departure of the AG values from the SG curve, but these are not related to the drop and set σ 's shown in Fig. 3a and b. The second pass successfully removes about 400 drops and 4 sets to improve the solution

noticeable reduction in the amplitude of the telescope glitches that are at the level of 0.5 μ Gal or less.

The JS1 and JS2 methods combine the different preselection of drops with the SG data interpolated from 1 min data. The JS3 method uses the combined

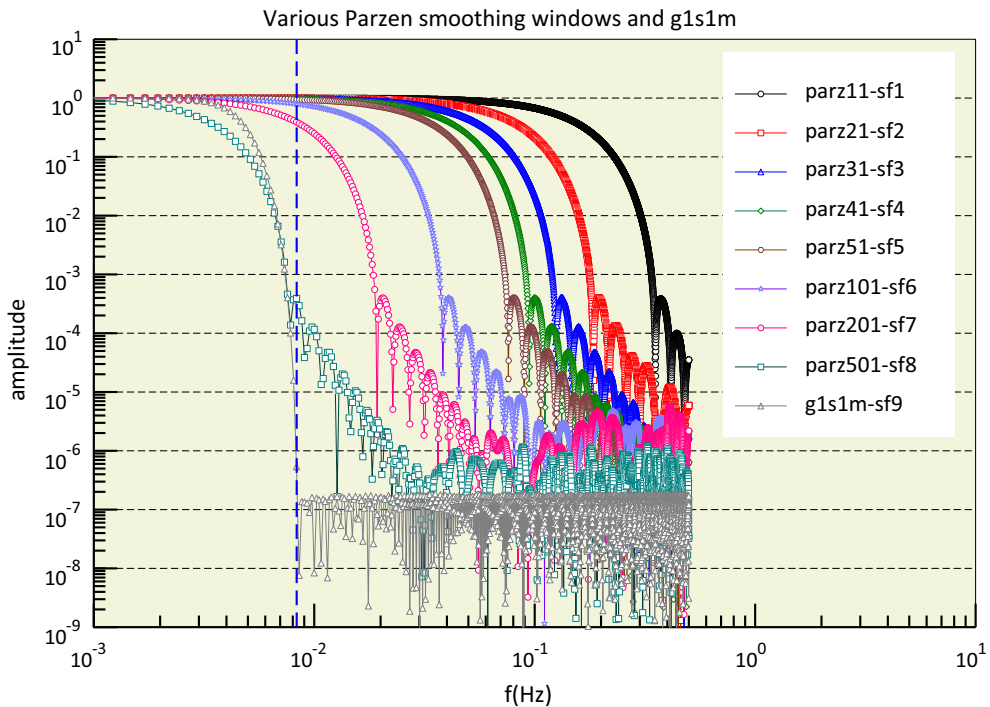


Figure 5

Amplitude transfer function of filters applied to the SG 1 s data prior to sampling at AG drop times. The Parzen filter windows of lengths 11–500 points are simple smoothing filters with cut-offs between 3 and 50 s. The standard GGP filter g_{1s1m} , length 1009 and with a Nyquist cutoff for 1 min sampling at 8.3×10^{-3} Hz (vertical dashed line) is used to decimate the 1 s data to 1 min

preselection of drops, indicated above, but the SG data are preselected at AG drop times and either used directly for the drop SF, or gathered into sets exactly as the AG drops are done, for the set SF. JS3 is thus the only combination that follows the procedure of Meurers (2012) and Calvo et al. (2014), and so we consider it to be the ‘best’ method of getting the SFs, especially with a pass 2 refinement. This is borne out by the results presented below.

2.4. Results of Tests for the Scale Factors on AP Data

From the previous section we may, therefore, summarize the various tests as follows. For the AP data we have these multiple solutions:

Method JS0: 1 solution for drops, 1 for sets; uses fixed options [w_1 , s_3 , co (AP) or nc (ST), pass 1] for 2 types of solution

Methods JS1–JS3: There are 2 solutions each for (w_0/w_1 , s_3/s_1 , co/nc, and p_1/p_2); the number of solutions computed is, therefore (16 options \times 3 methods \times 2 types = 96 calculations of the scale factors). Together with JS0, we therefore computed the SF for each experiment 98 times.

We then group all the solutions according to which factor we want to isolate (e.g., w_1/w_0) and compute the mean of the absolute difference between the solutions for w_0 and the solutions for w_1 (in units of $\mu\text{Gal/V}$). This provides a metric to judge which of the various options have the most effect on the final scale factors. Some of the more important results are shown in Table 9 (“Appendix”) under the 4 methods and 2 types (drops or sets). In this table we have also graded the datasets according to whether they are of high, medium or low quality, based on both the amount of data discarded in Table 3, and on the results in Table 9 themselves. It is important to also evaluate our options on the higher quality datasets

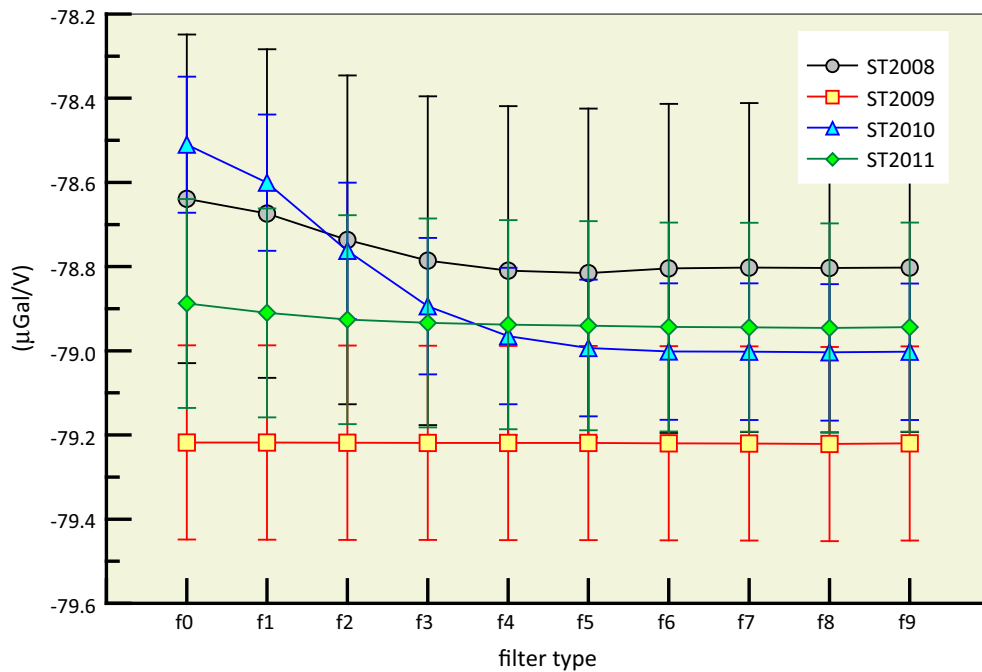


Figure 6

Effect of filtering the 1 s SG data on the scale factor for the Strasbourg datasets. The error bars come from the Sigma column of the set data. The 2010 experiment in particular shows a pronounced bias if the data are not filtered, whereas using filters between f_6 and f_9 suppresses noise at periods < 25 s. Other datasets ST2008 and ST2011 are affected to a lesser extent

that most users will encounter, so we combine all the results with a weighting on a scale of (1 = low, 2 = med, 3 = high) for both the AP and ST datasets combined. The result is shown schematically in Fig. 7 where the two options drops/sets are given with the combined score for the 4 options.

For the w_0/w_1 choice, it depends on whether we are doing drops or sets. For drops the effect is small, but for sets the choice is moderately important, and it seems better to weight the set means as a matter of principle. For the co/nc option, the means differences are surprisingly small, despite the seeming theoretical advantage in rejecting drops on the corrected data (previous discussion), as opposed to the uncorrected data used for the calibration. It is unfortunate that we did not have the corrected ST data at hand to test this result, but it appears that selecting drops directly from the uncorrected files may not in practice give a large SF error. For the s_3/s_1 option the effect is modest for either drops or sets, therefore, retaining the 3- σ outlier is OK for the better data. It is clear from Table 3, however, for the lower quality data there is a

difference (improvement) in choosing a tighter control of outliers using the 1- σ criterion. The final option p_1/p_2 , whether or not to have a second pass, makes the biggest difference in determining the SF, especially for the drop solutions. The advantage for sets is much less obvious, as might be expected because the set solutions are weighted more robustly. The number of tests is halved for the ST datasets because there is no corrected data ('co'); thus there are only 49 computations of the various SFs for ST.

2.5. Strasbourg Processing and Calibration Results

We turn from the problematic AG data of AP to station J9 in Strasbourg, which has a very long series of AG–SG calibrations, beginning in the early 1990's (Hinderer et al. 1991; Amalvict et al. 2001, 2002; Rosat et al. 2009; Calvo et al. 2014). We repeat the same calculations as for AP on data for SG CO26 in Strasbourg, using the 4 datasets from 2008, 2009, 2010, and 2011 as shown in Table 4.

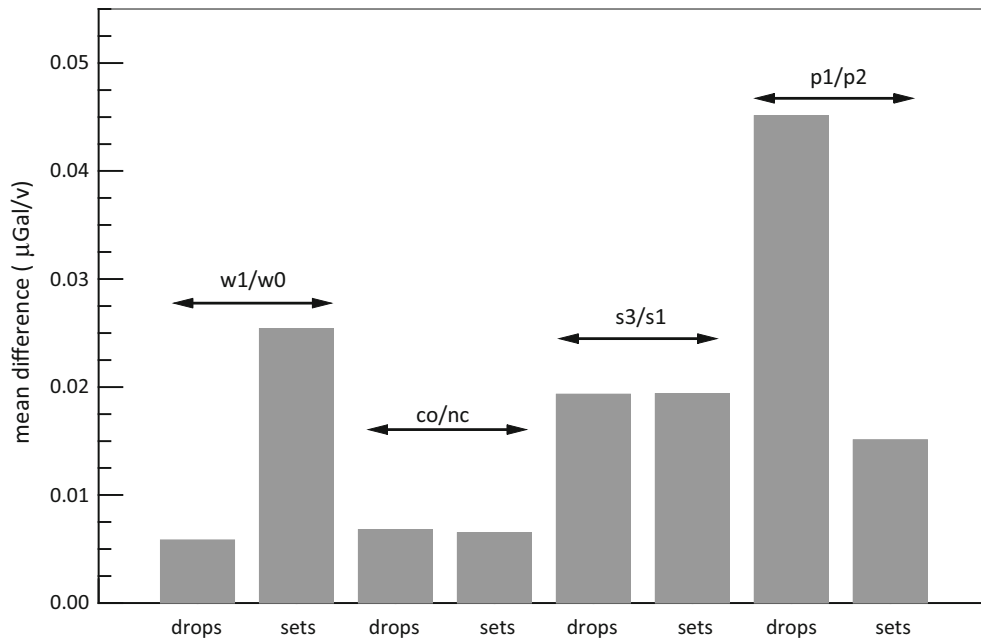


Figure 7

Summary of drop and set tests on options for computing the scale factor, based on Table 9 (see text). The bars represent the mean amplitude of changes in the scale factor over all calculations for both stations AP and ST related to: w_1/w_0 —option to weight the set averages, co/nc —option to use corrected or uncorrected AG data, s_3/s_1 —option to use 3- σ versus 1- σ rejection of outliers, and p_1/p_2 —option to choose a second pass rejecting residual fit outliers. A difference of 0.025 $\mu\text{Gal}/V$ is equivalent to a change of 0.03% in the SF

Because of the much better site conditions, the AG data are much cleaner than AP, and with the previously determined cutoff of 35 μGal for both the sets and drops, only a small percentage of the data are excluded. Comparing Tables 4 with 3, we note that the ST scale factors are more consistently determined than at AP, but the error in the SFs is not uniformly better. For example, experiment AP2011 in Table 3 has smaller drop and set errors than any of the data for ST, due probably to the larger number of sets used (283) compared to 139–164 sets for ST. Even so, with careful rejection of the bad data, a satisfactory SF can be obtained, even at AP. As in Table 3, we compare the JS1 and JS3 methods for passes 1 and 2 and note that pass 2 is always required, even though the ST data is much better overall than that at AP. This is a good time to point out that based on set standard deviations, the set SFs in Tables 3(b) and 4(b), have much larger uncertainties than the drop SFs in sections (a) of the Tables, especially for ST2008 and ST2011. The only way the drop and set errors can be comparable, as Amalvict et al. (2002) indicated, is

if we use standard error of the mean (that is to say the standard deviation divided by $\sqrt{(N)}$) as the uncertainty in the set SFs. To show this we recomputed the set SF's using the Error column of the set files, shown in sections (c) of Tables 3 and 4 for the set errors. It is clear that the differences in SF's are very small, and the SF errors are almost exactly a factor 10 smaller than when using the Sigma column. Thus, one can get almost the same result by finding the standard deviation of the SF using the Sigma column (sections (b) of Tables 3 and 4) and dividing by $\sqrt{(N)}$ to get the SEM. The SEM of a weighted mean has the more precise form indicated in the "Appendix" following Eq. (A2).

We see in Fig. 8 a comparison of 7 solutions (JS0, JS1, JS2, and JS3 for drops and sets) for 2 experiments at AP and two from ST, based on sections (a) and (b) of Tables 3 and 4.

The x-axis gives the different solutions comparing pass 1 and 2. Note that the SFs alternate high and low depending on which pass, but with the most important solutions (JS3 pass 2) there is a good

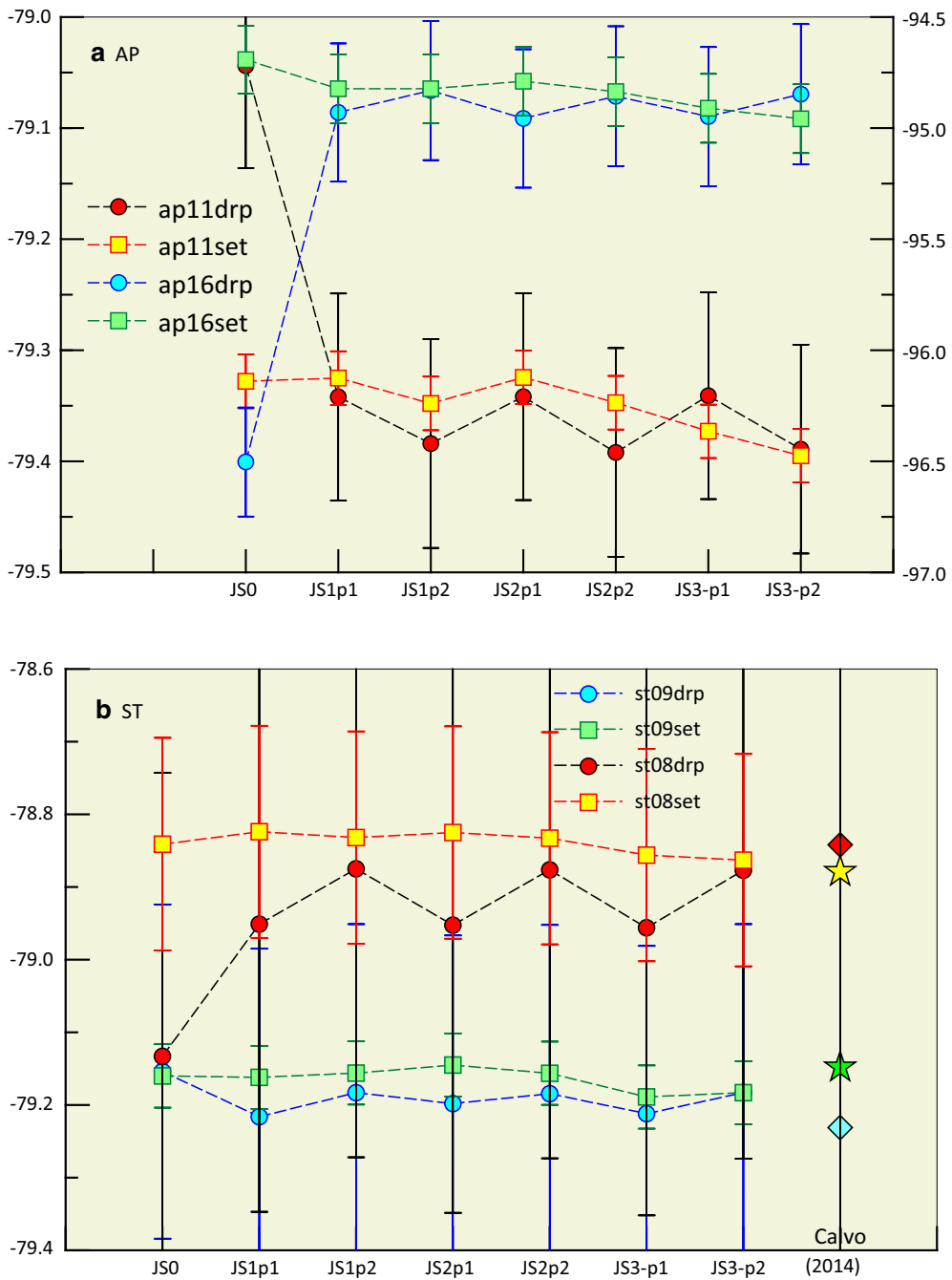


Figure 8

Comparison of scale factors for **a** two AP datasets (2011 left axis, 2016 right axis) and **b** two ST datasets (2008, 2009). All vertical axes are in $\mu\text{Gal/V}$. For each, the drop and set SFs are given for the various methods on the x axis. Abbreviations JS0, JS1, etc., are given in Table 9. Note that the pass 2 solutions are often quite distinct from the pass 1 solutions, and there is a good final convergence between drop and set values for the best datasets (all but AP 2016). The final column in **(b)** shows the drop and set factors obtained by Calvo et al. (2014), diamonds for the drops and stars for the sets (with matching colors). All error bars arise from using the Sigma column in the set files

convergence of the SFs from drops and sets. The original solutions (JS0) are quite different for the AP data, but with the improved processing the results for AP become almost as good as at ST. In Fig. 8b, we also show for comparison the solutions obtained by Calvo et al. (2014) where the drop and set values are more separated. This improvement is due probably to our filtering of the SG 1 s data, as well as using pass 2 to clean up the residuals.

3. Combining Different Scale Factors

Figure 9 shows the result of 51 AG set calibrations from instrument CO26 at J9 in Strasbourg, from 1997 to 2012 (Calvo et al. 2014). The SG sensor did not change over this period, but the data acquisition system electronics changed in 1997/12 (time lag reduced from 36.0 s using the TIDE filter to 17.18 s using the GGP2 filter) and again in 2010/04 when a new GGP filter board (GGP1 filter) was installed with a time lag of 8.16 s. The evolution of the measurements indicates a reduction in scatter of the calibrations with time, but no clear convergence to a

unique value. All these drop scale factors were computed using the drop and set methods in Calvo et al. (2014), similar to our Pass 1 determination.

What is the best way to combine such different estimates of the SG scale factors? The answer partially depends on whether the scale factor should be treated as a constant from one calibration experiment to another. Physicists have long been faced with this problem in the determination of fundamental quantities, for example, the Newtonian gravitational constant G , and in such cases the proper procedure is *conflation*, see “Appendix” Eq. (A1). In the case of an SG, it is assumed (e.g., Hinderer et al. 2015) that the scale factor is determined by the factory magnetic field configuration, or, as stated by Goodkind (1999): “The calibration constant is fixed by the geometry of the coils and suspended mass so that it remains the same if the instrument is turned off and on again no matter how long the time between”.

Assuming this is true, the scatter in Fig. 9 must then be attributed to random (and probably also systematic) factors in the experimental setup and environmental noise rather than in the instrument, and this would suggest that conflation is appropriate.

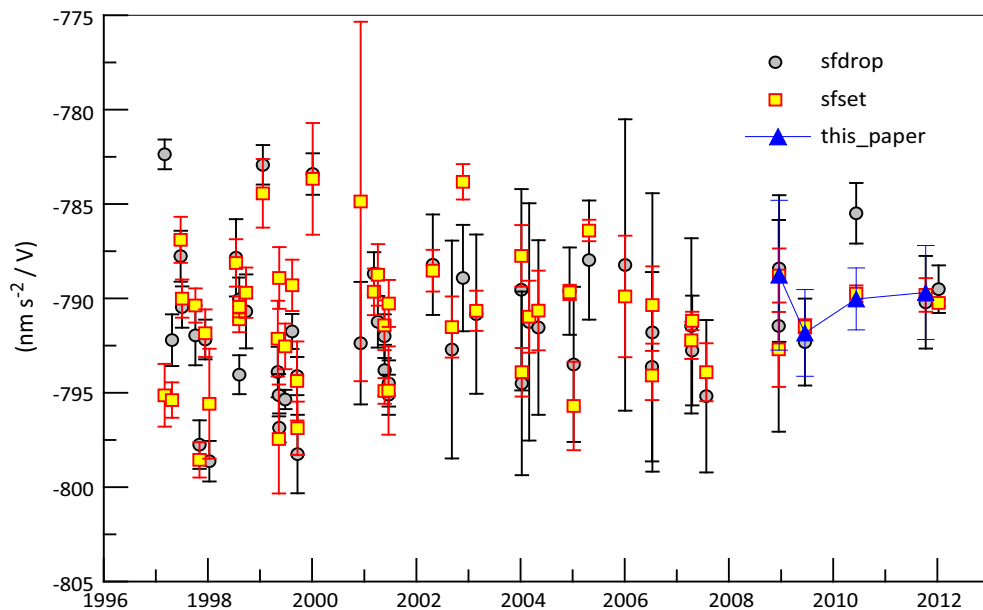


Figure 9

Strasbourg scale factors from sets and drops, 1997–2012. We show set SF’s (‘sfset’) based on the Error column of the set data. Four values denoted by blue triangles (‘this_paper’) are added from the drop SF’s in Table 4(a) for 2008–2011

As discussed in the “[Appendix](#)”, the calculation of a weighted mean of a series of measurements is unique, but there are two ways to compute its variance, depending on the purpose. One can use the weighted sample variance Eq. (A2), which measures the spread of estimates about the weighted mean. This is appropriate to indicate the scatter of the measurements, but in the case of the SG scale factor it is assumed that the repeated measurements should converge to a unique value, which is the actual calibration. This is indeed the case for SGs, where numerous studies indicate that β can be quite stable. Even the relocation of an SG between two quite different sites does not change the scale factor, as documented by Meurers (2012) for the transition between Vienna and the Conrad Observatory. The SF errors are then appropriately combined by conflation, Eq. (A1).

In principle set and drop scale factors are not independent, but they arise from different procedures and we cannot strictly average them, or conflate them, and they should be treated separately. We

apply (A1) to set and drop data independently, and assume the scale factor is not influenced by the changes in electronics, to get the result in Fig. 10.

The initial scale factors prior to 2000 are quite divergent, but each scale factor has more or less stabilized between 2005 and the end of 2012, and the set values are somewhat higher than that used in the Strasbourg data files which are in agreement with the drop value at about $-792.00 \text{ nm s}^{-2}/V$. The conflated values are much more revealing of the evolution of the calibration experiments than the scatter plot in Fig. 9, and are a useful way to give a unique value to an SG in a database.

It is to be noted that Eq. (A1) ensures that the eventual error of a long series of scale factor measurements will eventually approach zero, and thus may seem ‘unrealistic’. To test this we artificially extended the calibrations at J9 by repeating the same data as shown in Fig. 9 between 2001 and 2012 and adding this series as being qualitatively representative of future (yet to be done) measurements. From the total of 107 calibration experiments (those

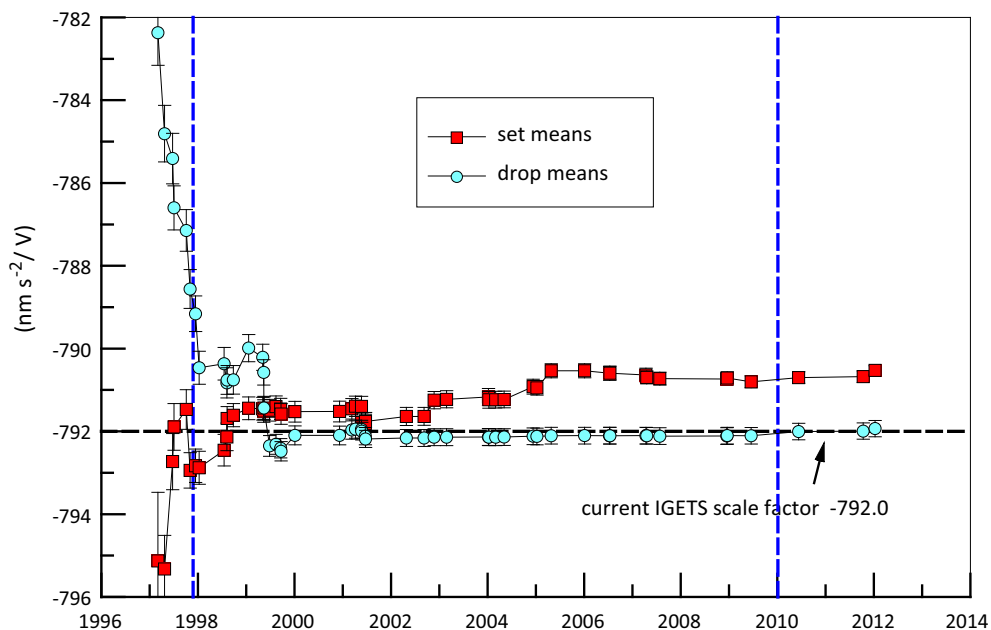


Figure 10

Conflation of the CO26 set and drop scale factors assuming no effect of the time delay (changes at vertical dashed lines). Error bars on the set SF's arise from using the set file Error column, as in Table 4(c), but the drop SF's use the Sigma column of the drop files, as in Table 4(a). The current SF value in the IGETS database is $-792.0 \pm 1.0 \text{ nm s}^{-2}/V$ (0.1%), close to the 2012 conflated drop mean of -791.934 ± 0.195 .

For sets the SF is slightly reduced to -790.527 ± 0.113

beyond 2012 being repeats) the conflated set error would have dropped slowly from $0.11 \text{ nm s}^{-2}/V$ (Fig. 10) to $0.07 \text{ nm s}^{-2}/V$, so the decrease for even a long series of measurements is quite slow.

To be complete, we compare the evolution of the two variances from Eqs. (A1) and (A2) in Fig. 11. The error in the weighted mean (A2) varies somewhat from the scatter of individual scale factor estimates, but does show the same overall downward trend as from conflation (A1). Assuming the actual SF is constant over long time periods, it seems plausible to expect an eventual convergence of the mean and error estimate from Eq. (A1). We also note that conflation can be used for SG scale factors determined using different AG instruments.

4. The AG Mean Value

Returning to the basic Eqs. (1)–(3), we note that the AG mean value y_0 includes certain AG static corrections such as the transfer height and gradient effects that are applied to standardize the absolute site

level. During the processing of AG data, the operator sets the transfer height and gradient for the experiment. The former is simply a height at which the gravity value is desired for the particular site (frequently ground level), which should be constant from experiment to experiment for consistency. This can be computed from a combination of the actual height for the dropping chamber (in fact the sum of the setup height and a specific height close to 1.2 m given by the manufacturer for each instrument where g is computed from the trajectory over a distance of about 20 cm inside the dropping chamber) and a gradient to be used for the transfer. Ideally, the observed gravity gradient should replace the default $-3.0 \mu\text{Gal}/\text{cm}$ (standard free-air gradient).

For various reasons these static corrections (transfer height and gradient) were not always kept constant at Apache Point. In one experiment, the transfer height was set to 0, and the gradient set to $-2.79 \mu\text{Gal}/\text{cm}$, whereas normally we had used 100 or 130 cm for the transfer height and the default gradient. For a long time we did not know the actual gradient below the telescope where the AG

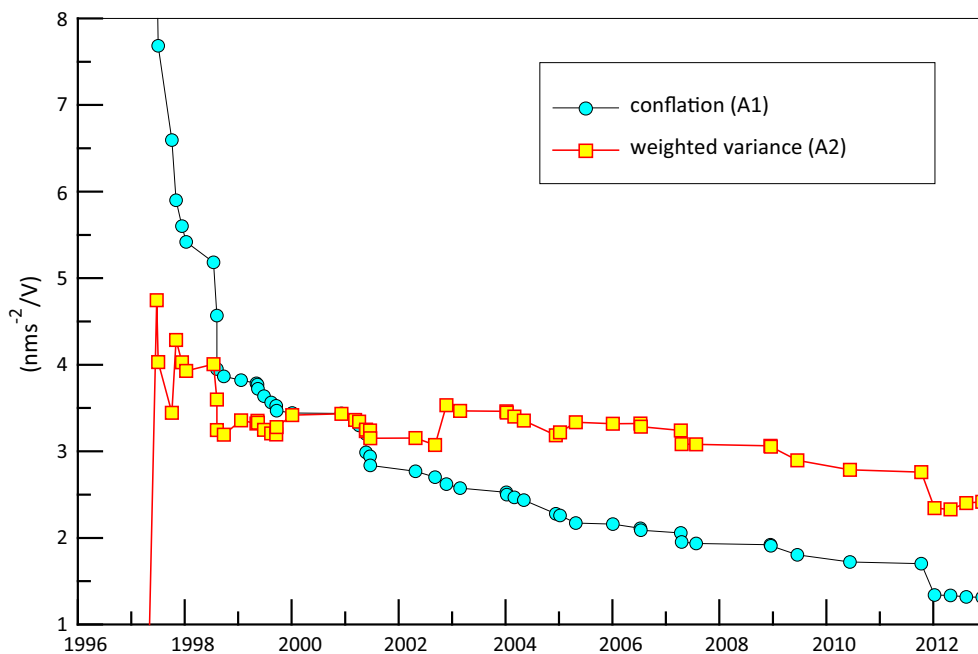


Figure 11

Evolution of the error in the J9 scale factor (Fig. 10) from set data for conflation, Eq. (A1), vs weighted variance, Eq. (A2). Errors are defined in the same way as for Fig. 10

measurements were made. This was eventually measured in 2015 with a Scintrex CG5 both in the cone room ($-4.42 \mu\text{Gal}/\text{cm}$) and outside the telescope building at ground level ($-3.87 \mu\text{Gal}/\text{cm}$). The gradient resulting from the potential of a homogenous ellipsoidal model at AP is $-3.08 \mu\text{Gal}/\text{cm}$, but cannot be applied to a station at an elevation of 2788 m (referred to WGS84); the difference with the measured value is most likely due to the assumption of a radial Earth model; the discrepancy being due to local topography (including the building) and lateral crustal density anomalies. It then became necessary to adjust not only the transfer height but also the gradient from the values used in the experiment to consistent values. We refer to discussion in the “[Appendix](#)” showing how this can be done without access to the *g*-software.

During a regular AG site measurement, assume $y(t)$ is measured as previously. But this time the geophysical corrections are applied for tides from local gravimetric factors, barometric pressure, and polar motion (TLBP), therefore we write these as:

$$g_s(t) = [g_{\text{tide}} + g_{\text{press}} + g_{\text{polar}}] = g_{s0} + g_{s1}(t) \quad (5)$$

so that during a site measurement, the corrected AG measurements are

$$y_c(t) = y(t) - g_s(t). \quad (6)$$

Introducing the mean and time-varying parts from (3) and (5), and recognizing that the corrected gravity $y_c(t)$ should ideally be the site AG value g_0 , free of time-varying effects, we find for the constant and time-varying parts

$$g_0 = y_0 - g_{s0} = \alpha + x_0 \times \beta - g_{s0} \quad (7a)$$

$$0 = y_1(t) - g_{s1}(t) \quad (7b)$$

Equation (7b) ensures that all time-varying parts of the measured gravity field are accounted for by the time-varying part of $g_s(t)$, provided we ignore the errors in the model and other factors such as non-tidal ocean effects and hydrology (but see below).

Equation (7a) shows how to use the mean value y_0 to get g_0 , i.e., by subtracting the mean value (or zero level) of the geophysical corrections applied by the *g*-software. These are the tidal amplitudes with a non-zero mean level, the applied nominal pressure corrections with a specific reference pressure p_0 (calculated for each AG site) and admittance ($-0.3 \mu\text{Gal}/\text{hPa}$), and the mean level of polar motion. It was for this reason that we kept track of the constants α , β , and x_0 in the solution of (1).

Normally no other time-varying effects are explicitly involved in the site AG measurements such as local hydrology attraction, non-local hydrology loading, non-tidal ocean effects, and tectonics (see e.g., Pálinkáš et al. 2010); indeed some of these are the target being measured. But many SG users routinely consider such further corrections to their data, so we could assume another model for these: $gh(t) = gh_0 + gh_1(t)$, dominated by hydrology, similar to (6). Adding $gh(t)$ to $g_s(t)$ then changes (7a) and (7b) to:

$$g_0 = y_0 - g_{s0} - gh_0 \quad (8a)$$

$$y_1(t) = g_{s1}(t) + gh_1(t) \quad (8b)$$

Again, the time-varying part of the AG measurements can be accounted for in (8b), for which many models exist. It is Eq. (8a) that poses a problem for AG measurements because it is not easy to define the mean level of hydrology gh_0 . Unlike the other mean levels, there is no obvious reference level for hydrology; the relative level used for SG studies (e.g., supplied as a loading correction by the EOST/IGETS loading service) is arbitrary and may have no relevance for a particular site. One might use the hydrology levels expected from local environmental parameters (rain, snow, evapotranspiration ...) which could define a mean hydrology based, for example, on decades-long averaging, or a reference hydrology level established after a prolonged drought, which might be empirically estimated. This is an interesting unresolved problem that may arise when considering further corrections to the AG site measurements.

Aside from the problem of hydrology, we can still find the site gravity g_0 from Eq. (7a), and compare it

with the AG site measurements when re-running the g-software with the geophysical corrections turned on. To estimate g_{s0} , we turn to the SG-derived version of $g_s(t)$ that is readily available at all SG stations due to the need for such modeling, noting this may differ from the FG5 corrections. For example, the SG may provide a superior tidal model (using local gravimetric factors), and we can apply the polar motion as published by IERS instead of predicting the polar motion from a model done in the g-software.

We applied the above method to finding the mean value y_0 and the estimate AG gravity g_0 for the 4 calibrations at AP based on set-derived solutions from the calibration—to be consistent with set means used in the FG5 processing. Table 5 shows the AG mean values from set and drop estimates, estimating g_0 from (5) using the procedure above, and g_0 from a usual FG site measurement.

For AP, the standard transfer height and gradient are (130 cm, $-4.42 \mu\text{Gal}/\text{cm}$). We note that the largest component of g_{s0} is not necessarily the tide, and TLBP is quite variable over the 4 years. The mean value y_0 comes directly from the fitting (1), with x_0 being found before the fit, after which we read in the corrections $g_s(t)$ and find the mean level of the tides, pressure, and polar motion from all the 1 min values coincident with the SG data values. Then $g_s(t)$ are splined to the AG set times, and subtracted from the AG values $y(t)$ as in Eq. (6). The corrected AG values $y_c(t)$ allow a weighted mean of the set

values that leads to a ‘simulated’ FG5 site value denoted by g_1 in Table 5. Alternatively, the mean level of the components of g_{s0} are added together and subtracted from y_0 using (7a), giving an AG site value g_0 from y_0 , as advertised. Finally the FG5 operator can reprocess the AG data using the g-software with corrections applied to get the actual FG5 site gravity. Some subtleties exist. For example, the SG mean value x_0 and the AG mean value y_0 , which can also be obtained directly from the input AG data, must use the highest sampling of the data, either 1 s or 1 min for the SG, and drop data for the AG even though the drop data may be noisy. But the mean TLBP level should be based on the set times, which is based only on good data, and g_0 and g_1 should match the set data as recorded in a site measurement. All the solutions in Tables 5 and 6 are based on the JS3, pass 2 method.

In Table 5, we note that some errors, e.g., the site value g_0 from y_0 , seem large, especially for the AP data where there are relatively few TLBP data at set times used in finding g_{s0} . Thus, quantities derived from g_{s0} tend to have larger errors than one might expect. The final two lines of Table 5 show that the difference $g_1 - g_0$ is consistently less than $1 \mu\text{Gal}$, clearly indicating that Eqs. (7a) and (7b) work in practice for all datasets. The discrepancy between the g_1 and FG5 site values are more variable but still generally within the error bars. The smallest error is on the value g_1 which uses the external corrections directly on the AG data, but other contributions to the

Table 5

Processing of Apache Point AG mean values y_0 from SG to AG calibrations as AG site measurements; all units μGal

	2011 Aug	2013 Nov	2014 Sept	2016 Jun
Means levels				
Tide	-8.2077	18.8580	4.253	-32.4812
Press	-3.9618	-1.7205	-3.030	-3.3393
Polar motion	6.9382	4.5114	4.411	7.8291
g_{s0}	-5.231 ± 5.003	21.649 ± 4.762	5.635 ± 6.923	-27.991 ± 12.801
AG—9788050000.0				
Mean value y_0	5158.34 ± 0.87	5189.66 ± 1.69	5161.53 ± 1.40	5136.14 ± 2.35
g_0 from y_0 Eq. (7a)	5163.57 ± 5.08	5168.01 ± 5.05	5155.90 ± 7.063	5164.13 ± 13.02
Simulated site value g_1	5164.36 ± 0.88	5167.59 ± 0.90	5155.05 ± 2.71	5164.07 ± 1.676
FG5 site value	5164.11 ± 1.81	5167.45 ± 1.82	5155.29 ± 1.83	5164.39 ± 1.84
$g_1 - g_0$	0.782 ± 5.079	-0.420 ± 5.131	-0.844 ± 7.562	-0.062 ± 13.122
$g_1 - \text{FG5}$	0.246 ± 1.811	0.143 ± 2.029	-0.234 ± 3.271	-0.322 ± 2.489

Table 6

Processing of Strasbourg AG mean values y_0 from SG to AG calibrations as AG site measurements; all units μGal

	2008 Dec 17	2009 Jun 16	2010 Jun 10	2011 Oct 11
Means levels				
Tide	23.8074	23.9927	19.084	22.0829
Press	– 4.1595	– 2.3362	0.194	– 3.4177
Polar motion	0.1435	0.4463	– 1.064	2.5759
gs_0	19.791 \pm 3.673	22.103 \pm 5.670	18.214 \pm 6.070	21.24 \pm 4.19
AG—980870000.0				
Mean value y_0	7828.45 \pm 1.98	7810.51 \pm 0.98	7804.63 \pm 1.15	7819.30 \pm 1.59
g_0 from y_0 Eq. (7a)	7808.66 \pm 4.17	7788.41 \pm 5.75	7786.42 \pm 6.18	7798.06 \pm 4.48
Simulated site value g_1	7809.07 \pm 1.27	7789.37 \pm 0.36	7797.88 \pm 1.28	7797.88 \pm 1.28
FG5 site value	7811.62 \pm ^a	7793.15 \pm 1.46	7789.24 \pm 1.51	7800.49 \pm 1.30
$g_1 - g_0$	0.412 \pm 4.362	0.960 \pm 5.766	– 0.547 \pm 6.235	– 0.186 \pm 4.657
$g_1 - \text{FG5}$	– 2.546 \pm 1.265*	– 3.784 \pm 1.504	– 3.370 \pm 1.729	– 2.612 \pm 2.273

^aFG5 error unavailable for this experiment

total error, i.e., the ‘uncertainty’, are not added as done within the g-software.

Confirmation of this procedure is provided from the 4 Strasbourg experiments, Table 6. We see that the tide mean value is significantly higher than at AP and does dominate gs_0 —it is a significant effect. The overall agreement $g_1 - g_0$ is very close and more consistent than for AP due to the better AG data, but there are discrepancies between g_1 and the standard FG5 set measurement, whose origin may be due to the corrections being either FG5 or ‘SG-derived’ values.

5. Summary and Conclusions

We show in Table 7, a summary of the amplitudes of the various effects that influence the SG scale factor, according to our estimates. These are taken from the various tables and figures, with additional estimates for the difference between JS0, JS1, and JS3. Notice that the effects are in units of the scale factor ($\mu\text{Gal/V}$), but when translated into percentages the values are close to % errors, e.g., an error of 0.025 in scale factor is equivalent to 0.03%. Certain effects are more important than others, i.e., moving from JS0 to JS3 (using departures from set means which is standard in FG5 processing) when processing drops,

Table 7

Summary of all factors influencing the SG scale factor

Factor	Amplitudes	Effect or decision to be made
Time lag	0.01 for 30 s	Influence of instrumental time lag on SG selection
AG trend	0.01 ($\mu\text{Gal/day}$)	Highly dependent on data
w_1/w_0	0.025 (sets)	Weighting the drops when finding set means
	0.006 (drops)	
co/nc	0.0067	Correcting the AG values for TLBP prior to drop and set selection
s_3/s_1	0.0194	Using a 1- σ versus 3- σ rejection of drops and sets
p_2/p_1	0.045 (drops)	Second pass to eliminate residual outliers far from the SG curve
	0.015 (sets)	
SG 1 s	\leq 0.5	Filtering of 1 s SG data prior to selection, highly data dependent
JS1/JS0	0.529 (drops)	Rejecting outliers from set means vs rejection based on drop δg_m
	0.038 (sets)	
JS3/JS2	0.005 (drops)	For drops, spline-interpolate SG values vs sampling SG value at drop time; for sets, gathering SG drops at AG set times vs interpolation of 1 min data
	0.059 (sets)	

See Table 8 for abbreviations. Values in ($\mu\text{Gal/V}$) except where indicated; a single entry applies to both drops and sets

but this is less important for sets. Two factors have been improved over the processing of Calvo et al. (2014), namely smoothing the SG 1 s data, and doing a second pass—especially in the case of drop SFs. For set SFs the dominant effects are to weight the drops when finding set means, and JS3–JS2—gathering the SG data at AG times into sets rather than interpolating SG 1 min data to AG set times. Any difference below $0.01 \mu\text{Gal/V}$ (such as having to make corrections to AG data before selecting drops), is considered a minimal effect, but we still recommend processing using JS3.

In addition, we have shown that the SG data should be selected at beginning and ending at the AG drop times, and especially for the AG mean value (if used) it can be important to include a trend to account for a drift in one of the instruments but not the other. The SG electronics time delay, which ideally should also be included, has almost negligible effect. Another feature of our study is that we use only the drop text files, because we can compute everything from them, including all the set processing required for a set SF. We do not need special pre-processing of the recorded data using the g-software to reject drops, although this can of course be done by groups that have the facilities and manpower. In terms of choosing whether to report drop or set SFs, both should be computed. Where there is a discrepancy it is likely that the set value may be less affected by bad data. On the other hand if the values are close, the drop SF is preferred as its error is statistically better defined in the sense one does not have to choose between using the ‘Sigma’ and ‘Error’ columns of the set data as errors.

We also recommend the use of conflation to combine different estimations of the SF for a particular SG, as this is the best way to characterize the SF for stations in a database. Finally, for users who do not have the g-software, or are reluctant to spend the effort to use it for their calibrations, we have shown it is possible to turn an SG calibration experiment into an AG site measurement at a site by subtracting the geophysical TLBP corrections from the AG mean value. Also it might be useful on occasion to determine the internal distance parameter D , through Eq. (A5), for an FG5 to enable a precise conversion of an AG mean value from one gradient to another.

Acknowledgements

Two anonymous reviewers provided many insightful comments and criticisms that allowed us to rethink some of our initial results and ideas, from which we trust the reader will benefit. We also thank Bruno Meurers as editor for excellent suggestions throughout the reviewing process. We are deeply indebted to R. David Wheeler, from the National Geospatial-Intelligence Agency stationed at Holloman AFB, New Mexico, for making the difficult AG measurements in the cone room at Apache Point Observatory. DC benefitted from very useful discussions with Hartmut Wziontek (BKG, Leipzig) and Derek Van Westrum (National Geodetic Survey, Boulder). The work was done as a subcontract originating from the pioneering work of Tom Murphy (UCSD, California) on the APOLLO LLR system, and his effort to install an SG to improve the LLR; funding came from Grants 10-APRA10-0045 (NASA), PHY-1068879 and 10322410-SUB (NSF). The Strasbourg SG data is available at <https://doi.org/10.5880/igets.st.11.001>.

Appendix

Weighted Mean and Variance

There are two approaches to finding the variance of weighted samples, assuming N samples of a quantity x_i , each assumed to have a Gaussian probability distribution with a standard deviation σ_i . Interpreting the weight of each sample as its inverse variance, $w_i = 1/\sigma_i^2$, the mean (x_m) and variance σ^2 of the resulting combination are given by

$$x_m = \Sigma_i(w_i x_i) / V_1; \sigma^2 = 1/V_1 \quad (\text{A1})$$

where $V_1 = \Sigma_i(w_i)$, assuming the data are uncorrelated. Equation (A1) gives the *variance of the weighted mean*, used in combining different quantities, derived for example from the LSQ inversion or fitting of parameters to data. It is also the error of compound quantities, so that when all weights are equal, $\sigma^2 = (1/N^2) \Sigma_i(\sigma_i^2)$. A second approach is to compute the *unbiased weighted sample variance*

$$\sigma^2 = \Sigma_i \left[w_i (x_i - x_m)^2 \right] / [V_1 - (V_2/V_1)] \quad (\text{A2})$$

where $V_2 = \sum_i (w_i^2)$, as given for example in the GNU Fortran library (function `gsl_stats_wvariance` at <https://www.gnu.org/software/gsl/manual/gsl-ref.html#Weighted-Samples>). The mean x_m remains as in (A1) and the weighted SEM is defined as $\sigma_{\sqrt{(V_2)/V_1}}$. Equation (A2) is used to assess the variance of the data about its mean, and is the weighted version of the usual formula for Gaussian mean and variance. In his useful little book, Topping (1979, Section 43) refers to (A1) as measuring *internal consistency of the data*, (using the errors associated only with each experiment) as opposed the *external consistency* of the data where the errors are determined by the spread of each experiment about a common mean.

Hill and Miller (2011) give the name *conflation* to (A1) and argue this is the correct way to combine different experiments to determine the best value of an unchanging physical quantity. They show that conflation is (a) commutative and associative, (b) iterative, and the (c) conflations of normal distributions are normal. Property (b) is useful as new data can be easily combined by adding to the conflations of the previous datasets. A more mathematically oriented justification can be found in Hill (2011).

AG Transfer Height

AG transfer height and gradient effects were discussed by Niebauer (1989) and then extended by Nagorny (1995) and Timmen (2003) who provided more instrumental details. From a user's point of view it is not possible to deal with changing the gradient using only the transfer height adjustment provided in the FG5 manual. Denote this transfer height correction by

$$\begin{aligned} \delta g_1 &= -(\text{actual height} - \text{transfer height}) \\ &\quad \times \text{gradient} \\ &= -(\text{AH} - h) \times \Delta g \end{aligned} \quad (\text{A3})$$

where AH is the actual height and h the transfer height. The actual height is the sum of the factory height and the setup height, and these quantities are given in the FG5 project files. But (A3) is insufficient to recover the transfer height correction if the gradient Δg changes. A series of experiments was performed by NGA (National Geospatial-Intelligence Agency) by varying the transfer heights and gradients in the FG5 processing and recording the calculated values from the FG5 merged project files. Starting with $g_c(h, \Delta g)$ as the calculated value, we first

Table 8

Terminology and abbreviations used in the paper

Options	Comment
TLBP corrections	Tides and ocean loading from local gravimetric factors, barometric pressure, and polar motion; for AG data
AG mean value	y_0 The quantity y_0 in Eqs. (2) and (3)
AG site value	g_0 The result of a normal AG determination with TLBP corrections
Sigma rejection	δg_m Choice of drop and set sigma rejection level for disturbed data, effective for AP data in rejecting nighttime noise
Type of FG5 file	co/nc co—corrected for TLBP, as in a normal AG measurement nc—not corrected, as in a calibration experiment
Weighting used for set means	w_1/w_0 w_1 —set means are computed with drop errors as weights w_0 —no weighting used for set means, as in FG5 files
Reject criterion	s_3/s_1 s_3 —3- σ rejection of outliers for set means s_1 —1- σ rejection
Treatment of residuals	p_1/p_2 p_1 —pass 1: all residuals accepted after pass 1 p_2 —pass 2: 3- σ rejection of residual outliers and revised fit
SG smoothing	$s_f/0-9$ Application of various smoothing filters on SG 1 s data prior to selection of SG value at AG drop times, see Fig. 5
Mechanism for drop rejection	JS0 JS1 JS2 JS3 JS0: drop selection based only on drop sigma, set rejection based on δg_m ; SG 1 min data interpolation JS1: drop selection based on 3 σ rejection of drops when computing set means; SG 1 min data interpolated JS2: drop selection based on rejecting drops with errors $> \delta g_m$ SG data 1 min interpolated JS3: drop selection based on both JS1 and JS2; SG data from smoothed 1 s data at closest 1 s AG drop times, averaged as for AG sets from gathered drop times

Table 9

Test results for AP and ST scale factors. Shown are mean differences between equivalent calculations (keeping other options constant); values in ($\mu\text{Gal/V}$)

Station AP	2011	2013	2014	2016
AG data quality	High	Low	Med	Low
Drops w_1-w_0	0.00167	0.0217	0.00584	0.0218
Sets w_1-w_0	0.00507	0.0210	0.00348	0.0358
Drops $co-nc$	0.00809	0.0294	0.00716	0.0610
Sets $co-nc$	0.00646	0.0378	0.00287	0.0608
Drops s_3-s_1	0.00403	0.0375	0.0337	0.0408
Sets s_3-s_1	0.00637	0.0658	0.0337	0.0531
Drops p_2-p_1	0.0507	0.0838	0.0507	0.0727
Sets p_2-p_1	0.0218	0.00330	0.0218	0.0290
Station ST	2008	2009	2010	2011
AG data quality	High	High	High	High
Drops w_1-w_0	0.00359	0.00130	0.00024	0.0118
Sets w_1-w_0	0.0169	0.0102	0.00233	0.105
Drops s_3-s_1	0.0334	0.0222	0.00784	0.00621
Sets s_3-s_1	0.0252	0.0208	0.00723	0.00114
Drops p_2-p_1	0.0766	0.0336	0.00068	0.0383
Sets p_2-p_1	0.00802	0.00452	0.00999	0.0261

subtract the standard correction (A1), and also the calculated gravity for zero gradient at this transfer height $g_c(h, 0)$, thus

$$g'_1(h, \Delta g) = g_c(h, \Delta g) - \delta g_1 - g_c(h, 0) \quad (\text{A4})$$

We established that the left hand side (LHS) is a linear function of the gradient, i.e.,

$$g'_1(h, \Delta g) = D \times \Delta g \quad (\text{A5})$$

where for NGA's FG5-107 used at AP, $D = 8.0258$ cm. Thus

$$g_c(h, \Delta g) - g_c(h, 0) = (h - h_e) \times \Delta g \quad (\text{A6})$$

introducing $h_e = (AH - D) = 122.694$ cm as an effective height, such that when $h = h_e$ the gradient has the least effect on the gravity value. Because it is not possible to find h_e or D from the project files, Pálinkáš et al. (2010) observed "...Some users of the gravity data have no access to the FG5g-software; they cannot accurately correct for the new gradient without knowledge of the effective position. There is even a risk that they will compute the new transfer

correction with respect to the top of the drop, because this is often presented as the instrument's reference height."

This was indeed our earlier experience, and we found h_e by adding one additional step, i.e., re-processing the same FG5 data with zero gradient at the same transfer height, $g'_1(h, 0)$, and subtract this to get $g'_1(h, \Delta g)$ using (A5) to find D . We can then correct for both transfer height and gradient from one AG setup to another using

$$g_c(h_2, \Delta g_2) - g_c(h_1, \Delta g_1) = (\delta g_1 - \delta g_2) + D \times (\Delta g_2 - \Delta g_1) \quad (\text{A7})$$

Note the distance D in (A5) remains the same when AH changes, and therefore, needs to be determined only once, whereas the effective height h_e depends on AH . In one of our experiments we start with $AH = 130.72$ cm, a transfer height of 100 cm, and a gradient of -3.0 $\mu\text{Gal/cm}$ but want gravity at a height of 130 cm and a gradient of -4.42 $\mu\text{Gal/cm}$; using Eq. (A7) gives -100.37 μGal , but Eq. (A3) gives a value of -90.0 μGal , a difference of more than 10 μGal . Equation (A5) in fact is entirely

consistent with Eq. (3) in Pálinská et al. (2010) to which the reader should refer for complete details (Tables 8, 9).

REFERENCES

- Amalvict, M., Hinderer, J., Boy, J.-P., & Gegout, P. (2001). A three year comparison between a superconducting gravimeter (GWR C026) and an absolute gravimeter (FG5#206) in Strasbourg (France). *Journal of the Geodetic Society of Japan*, 47, 410–416.
- Amalvict, M., Hinderer, J., Gegout, P., Rosat, S., & Crossley, D. (2002). On the use of AG data to calibrate SG instruments in the GGP network. *Bull d'Inf Marees Terr*, 135, 10621–10626.
- Baker, T. F., & Bos, M. S. (2003). Validating Earth and ocean tide models using tidal gravity measurements. *Geophysical Journal International*, 152, 468–485.
- Boy, J.-P., Llubes, M., Hinderer, J., & Florsch, N. (2003). A comparison of tidal ocean loading models using superconducting gravimeter data. *Journal of Geophysical Research*, 108(B4), 2193. <https://doi.org/10.1029/2002JB002050>.
- Calvo, M., Hinderer, J., Rosat, S., Legros, H., Boy, J.-P., Ducarme, B., et al. (2014). Time stability of spring and superconducting gravimeters through the analysis of very long gravity records. *Journal of Geodynamics*, 80, 20–33. <https://doi.org/10.1016/j.jog.2014.04.009>.
- Crossley, D., & Hinderer, J. (2010). GGP (global geodynamics project): An international network of superconducting gravimeters to study time-variable gravity. *IAG Symposia, Gravity, Geoid, and Earth Observation*, 135, 627–635. https://doi.org/10.1007/978-3-642-10634-7_83.
- Francis, O. (1997). Calibration of the C021 superconducting gravimeter in Membach (Belgium) using 47 days of absolute gravity measurements. In: *Gravity, Geoid and Marine Geodesy, Tokyo, Japan, IAG Symposium* (Vol. 117, pp. 212–219). Berlin: Springer.
- Goodkind, J. (1999). The superconducting gravimeter. *Review of Scientific Instruments*, 70(11), 4131–4152.
- Hill, T. (2011). Conflations of probability distributions. *Transactions of the American Mathematical Society*, 363(6), 3351–3372.
- Hill, T., & Miller, J., (2011). An optimal method to combine results from different experiments. [arXiv:1005.4978v3](https://arxiv.org/abs/1005.4978v3) [physics.data-an].
- Hinderer, J., Crossley, D., & Warburton, R. J. (2015). Superconducting gravimetry. In Gerald Schubert (Ed.), *Treatise on geophysics* (2nd ed., Vol. 3, pp. 59–115). Oxford: Elsevier.
- Hinderer, J., Florsch, N., Mäkinen, J., Legros, H., & Faller, J. (1991). On the calibration of a superconducting gravimeter using absolute gravity measurements. *Geophysical Journal International*, 106, 491–497.
- Imanishi, Y., Higashi, T., & Fukuda, Y. (2002). Calibration of the superconducting gravimeter T011 by parallel observation with the absolute gravimeter FG5 210—A Bayesian approach. *Geophysical Journal International*, 151, 867–878.
- Imanishi, Y., Nawa, K., Tamura, Y., & Ikeda, H. (2018). Effects of horizontal acceleration on the superconducting gravimeter CT #036 at Ishigakijima, Japan. *Earth, Planets and Space*, 70, 9. <https://doi.org/10.1186/s40623-018-0777-9>.
- Kren, P., Palinkas, V., & Masika, P. (2016). On the effect of distortion and dispersion in fringe signal of the FG5 absolute gravimeters. *Metrologia*, 53(1), 27–40.
- Mäkinen, J., Virtanen, H., Bilker-Koivula M, Ruotsalainen, H., Näränen, J., and Raja-Halli, A. (2015). The effect of helium emissions by a superconducting gravimeter on the rubidium frequency standards of absolute gravimeters In: *International Association of Geodesy Symposia*. New York: Springer. https://doi.org/10.1007/1345_2015_205.
- Meurers, B. (2002). Aspects of gravimeter calibration by time domain comparison of gravity records. *Bull. d'Inf Marées Terr*, 135, 10643–10650.
- Meurers, B., (2012). Superconducting gravimeter calibration by colocated gravity observations: Results from GWRC025. *International Journal of Geophysics*. <https://doi.org/10.1155/2012/954271>. <https://www.hindawi.com/journals/ijge/2012/954271/>.
- Nagornyi, V. (1995). A new approach to absolute gravimeter analysis. *Metrologia*, 32(3), 201–208.
- Niebauer, T. M. (1989). The effective measurement height of freefall absolute gravimeters. *Metrologia*, 26, 115–118.
- Pálinská, V., Kostecký, J., & Simek, J. (2010). A feasibility of absolute gravity measurements in geodynamics. *Acta Geodynamica Geomaterial*, 7(1), 61–69.
- Rosat, S., Boy, J.-P., Ferhat, G., et al. (2009). Analysis of a ten-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. *Journal of Geodynamics*, 48(3–5), 360–365.
- Tamura, Y., Sato, T., Fukuda, Y., & Higashi, T. (2005). Scale factor calibration of a superconducting gravimeter at Esashi Station, Japan, using absolute gravity measurements. *Journal of Geodesy*, 78, 481–488.
- Timmen, L. (2003). Precise definition of the effective measurement height of free-fall absolute gravimeters. *Metrologia*, 40, 62–65.
- Topping, J. (1979). *Errors of observation and their treatment*. New York: Halstead Press.
- Van Camp, M., Meurers, B., de Viron, O., & Forbriger, T. (2016). Optimized strategy for the calibration of superconducting gravimeters at the one per mille level. *Journal of Geodesy*, 90, 91–99.
- Voigt, C., Förste, C., Wziontek, H., Crossley, D., Meurers, B., Pálinská, V., Hinderer, J., Boy, J.-P., Barriot, J.-P., Sun, H. (2016). Report on the data base of the international geodynamics and earth tide service (IGETS). *Scientific Technical Report STR—Data*. Potsdam: GFZ German Research Centre for Geosciences. <http://doi.org/10.2312/GFZ.b103-16087>.
- Warburton, R., & Goodkind, J. (1976). Search for evidence of a preferred reference frame. *Astrophysical Journal*, 208, 881–886.
- Wziontek, H., Falk, R., Wilmes, H., & Wolf, P. (2006). Rigorous combination of superconducting and absolute gravity measurements with respect to instrumental properties. *Bull d'Inf Marées Terr*, 142, 11417–11422.