Improved Scale Factors of the BKG Superconducting Gravimeters, Derived from Comparisons with Absolute Gravity Measurements.

Martina Harnisch *, Günter Harnisch *, Reinhard Falk **
* formerly Bundesamt für Kartographie und Geodäsie (BKG)
** Bundesamt für Kartographie und Geodäsie (BKG)
Richard-Strauss-Allee 11, D-60598 Frankfurt am Main

1. History

Calibration results of several superconducting gravimeters (SG) owned by BKG were presented at the Earth Tides Symposium 2000 at Mizusawa [1]. The calibrations were based mainly on comparisons with absolute gravimeters (AG). As far as possible, the scale factors were compared with results from the Frankfurt Calibration System (FCS), which uses an artificial calibration signal, generated by sinusoidal vertical movement of the SG. The scale factor of the CD030 at Bad Homburg derived from comparisons with AG is very reliable (9 campaigns with the absolute gravimeter FG5-101, error of the weighted mean ±0.88 nm s^{-2}/V for the lower and ±0.84 nm s^{-2}/V for the upper system). However, the correspondent FCS results deviate by 2.81 nm s^{-2}/Volt (lower system), or, 2.96 nm s^{-2}/Volt (upper system), respectively. Commonly the differences between the two independent calibration methods are near the significance level. It is striking that without exception the FCS results deviate in the same direction from those of the AG comparisons.

To find an explanation for these discrepancies, the evaluation procedures of both calibration methods were examined and new calibration experiments were performed. In the following, some critical points of the AG comparison method are considered. All of the data of AG comparisons concerning the CD030 were reprocessed under the new aspects. Finally, both calibration methods were reexamined again with respect to possible systematic deviations between the results of both methods. Detailed investigations of the FCS method shall be published in a separate contribution [2].

2. Basic principle of the calibration by comparisons with absolute gravity measurements

At the same site, SG and AG should measure the same gravity variations, e.g. the tidal signal. While the AG values are measured directly in cgs units, the output of the SG is given in Volt. Comparing the output of both instruments, a scale factor may be derived, which converts the data recorded by the SG into cgs units (µGal/V, better nm s^{-2}/V). Any environmental influences (e.g. air-pressure variations, hydrological influences, polar motion etc.) do not disturb because they affect both instruments in the same way. On the other hand, all disturbing influences acting on any one of the instruments alone, must be eliminated very carefully. This is valid e.g. for the instrumental drift of the SG.

A great advantage of the comparisons with AG is that they are made “in situ”, i.e. the calibration measurements do not influence the normal operation of the SG. This also implies that the whole recording system and its specific transfer characteristic (amplifier, digital voltmeter etc.) are the same during the calibrations as during the normal operation of the SG.

The very different accuracies of the of AG and SG are problematic. The result of gravity measurements with the AG is derived from a large number of individual fall experiments of a test mass (“drops”). In contrast, for calibration purposes the gravity values derived from each drop are used as basic data. These single values are much less accurate than the final result of the absolute measurements as well as also the gravity recorded by the SG.

The calibration of SG by comparison with synchronous measured absolute values works at the error limit of the modern AG. Commonly a calibration accuracy in the order of $1 \cdot 10^{-3}$ or better is expected. In consideration of the maximum tidal signal of about 3000 nm s^{-2}, it follows, that the error of the absolute measurements must not exceed 3 nm s^{-2}. It is clear, that such an accuracy may be reached only statistically by averaging over a large number of single free fall experiments (drops). The error of one drop from the fit to a parabolic trajectory is presently for FG5-101 in the range of 5 ... 10 nm s^{-2}.

For the comparison the data of AG and SG must be synchronized. For each drop-value the correspondent SG value is estimated by interpolation of the more accurate SG data, which nowadays are sampled at shorter time intervals than the AG data. It has been ensured, that no significant errors arise during this step of the data processing.

The modern AG are equipped with computers, which control the whole measuring process including the application of corrections for the different disturbing influences, data statistics and estimation of the final result. As a consequence, all data, especially the g-values of the single drops are corrected ones. However, in addition, the drop results as well as also the corresponding corrections can be read from the gravimeter, and, as a first step of the comparison procedure, the “original” drop values have to be reconstructed by undoing the corrections, which are not connected with the measuring process itself (tides, air pressure, polar motion).

The correlogram of the synchronized AG and SG data has the shape of an elongated cloud of points. Its slope is the scale factor of the SG, which can be estimated by linear adjustment.

At the stations equipped with SG owned by the BKG many absolute measurements took place, which commonly not were planned from the aspect of SG calibration. Nevertheless, all absolute measurements carried out in parallel with SG are used for calibration purposes. Therefore, the AG data are not always optimal from the viewpoint of calibration (small amplitudes of the tidal signal, short data series, data series broken into different subsets). But small errors of the scale factor may also be reached under unfavorable conditions as may be seen from fig. 7 (symbols with open circles, which mark measurements at low tidal signal).

To reach optimal accuracy for the scale factor it is necessary to check all steps of the calibration procedure and to exclude different error sources especially in the absolute measurements.

3. Tidal corrections

The basis for the estimation of the scale factor by linear regression are the original drop values, uncorrected for tides and other influences. Tidal corrections are only needed in an intermediate step when the AG data are prepared for the elimination of outliers. For that purpose however, the tidal correction must be very precise, above all in the short periodic range. Especially in older versions of the gravimeter software the accuracy of the tidal corrections was not sufficient. It is, however, no problem, to use any tidal corrections based on a series expansion of the tidal potential together with a set of tidal parameters valid for the station in consideration. These tidal corrections are without direct influence on the further evaluation process.

If the absolute measurements are extended over a large period, incompletely eliminated constituents of the tides or of the air pressure influence may feign a drift of the drop values. For the detection of outliers this “drift” has to be corrected by fitting a linear model. After the outliers have been eliminated the drift correction has to be cancelled again as it is done with other corrections. For example an apparent driftrate of about 0.128 nm s^{-2}/h was derived from the AG data of the campaign 7.-12.12.1999. After the drift correction has been applied the scale factor changed from (-735.26 ± 1.36) nm s^{-2}/V to (-735.79 ± 0.99) nm s^{-2}/V (moving window, asymptotic fit and extrapolation to zero size).

4. Instrumental drift

Commonly the instrumental drift of SG is very low. Therefore, and due to the short duration of the absolute measurements, in most cases the influence of the drift on the calibration results may be neglected. Only in few cases the drift is so large, that drift corrections must be applied. Especially in the period after initialization or re-initialization of the SG, anomalous drift behavior with non-linear constituents (“exponential drift”) may occur. However, in such cases, the drift during the period of the absolute measurements may be approximated by a straight line.

5. Offsets between data sets of the absolute measurements

Offsets are not characteristic for absolute gravity measurements. However, to avoid systematic errors caused by an incorrect adjustment of the apparatus, sometimes the adjustment is repeated after a certain time. In such cases the data set of the AG commonly is broken into different subsets. The same is valid if the gravimeter on the surface of the pillar is moved from one place to another.

There are several ways to determine the offsets between the subsequent subsets. In the simplest case the offsets are derived from the arithmetic mean values of the subsets. Another method is the use of a step algorithm as it is done during the preprocessing of tidal recordings. The best way is a third. It consists in a trial and error procedure, which minimizes the standard deviation m_0 and the error m_{SF} of the scale factor in dependence of different values of the offset. The left frame of fig. 1 shows, that the scale factor nearly linearly depends on the offset, which is applied to correct the data. This makes evident, that incorrectly estimated offsets necessarily must lead to falsified calibration results. On the other hand, the right frame shows that m_0 and m_{SF} go through a clear minimum. The optimum offset is that, at which the minimum of m_0 or m_{SF} is reached.

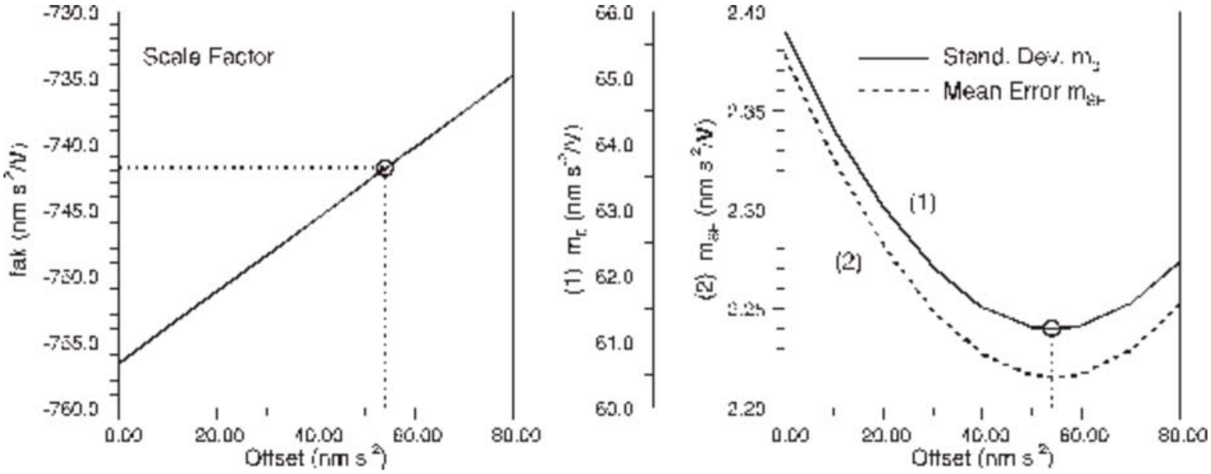

Fig. 1: Dependence of the scale factor of the SG and of its mean error on the offset, which is applied to correct the AG data (Example of the CD030, Bad Homburg, 13.-15.7.2001)

Fig. 2 gives an example of a data set, which is influenced by two offsets. If the offsets were not corrected, a scale factor of (-711.199 ± 2.401) nm s^{-2}/V would result (upper frame). Offset corrections change the result to (-739.446 ± 1.880) nm s^{-2}/V if they are derived from the difference of the mean values of neighbored sections of the data set and to (-737.787 ± 1.883) nm s^{-2}/V if they are derived by the m_0-criterion (lower frame). The influence of the step corrections is considerable. Leaving the uncorrected value out of consideration, the two other results with offset corrections derived in different ways also deviate from each other by 1.66 nm s^{-2}/V, i.e. by about 0.2 percent.

Fig. 2 also gives an impression of the different accuracies of the AG and the SG data. The drop values of the AG form an extended band-like cloud of data points with a vertical stripe pattern. The width of the band corresponds to the large scattering range of the absolute measurements. Although the most deviating values are deleted, a more or less large number of remaining outliers can be clearly recognized. The stripe pattern is caused by the succession of measurements ("sets") and breaks. The breaks between the sets are used to check the instrument. Due to the considerable less scattering of the SG data the correspondent cloud of data points looks like a smooth broken line, no deviating points are to be seen.

Offset correction and number of the deleted outliers are dependent from each other. The estimation of offsets needs a data set, which is freed from outliers. To this end in a first approximation the weak $2s_0$-criterion was used. This is valid for all examples summarized in table 1. In order to check the influence of outliers on the estimation of the offset, the data of calibration no. 8 were used. In a second step the skew criterion was applied with different threshold values. For each threshold the estimation of the offset was repeated. While the number of detected outliers clearly roses, the estimated offsets vary in a range of about ±1 nm s^{-2}. However, at high values of the threshold deviations from this general behavior are possible.

6. Elimination of outliers

Commonly in the data of absolute measurements several values occur, which deviate more or less from the majority of the others. Such outliers indicate that the individual drop is disturbed in any way. The reason of outliers can be mechanical perturbations of the drop sample. Single drop results can also be affected by a laser standard "out of lock" or a short disturbance of the rubidium frequency. The right selection of the outliers to be eliminated is the most crucial point in the calibration of SG by comparison with AG. - The error estimates of the parabolic fit of the single drops have not been used for the identification of outliers in this study.

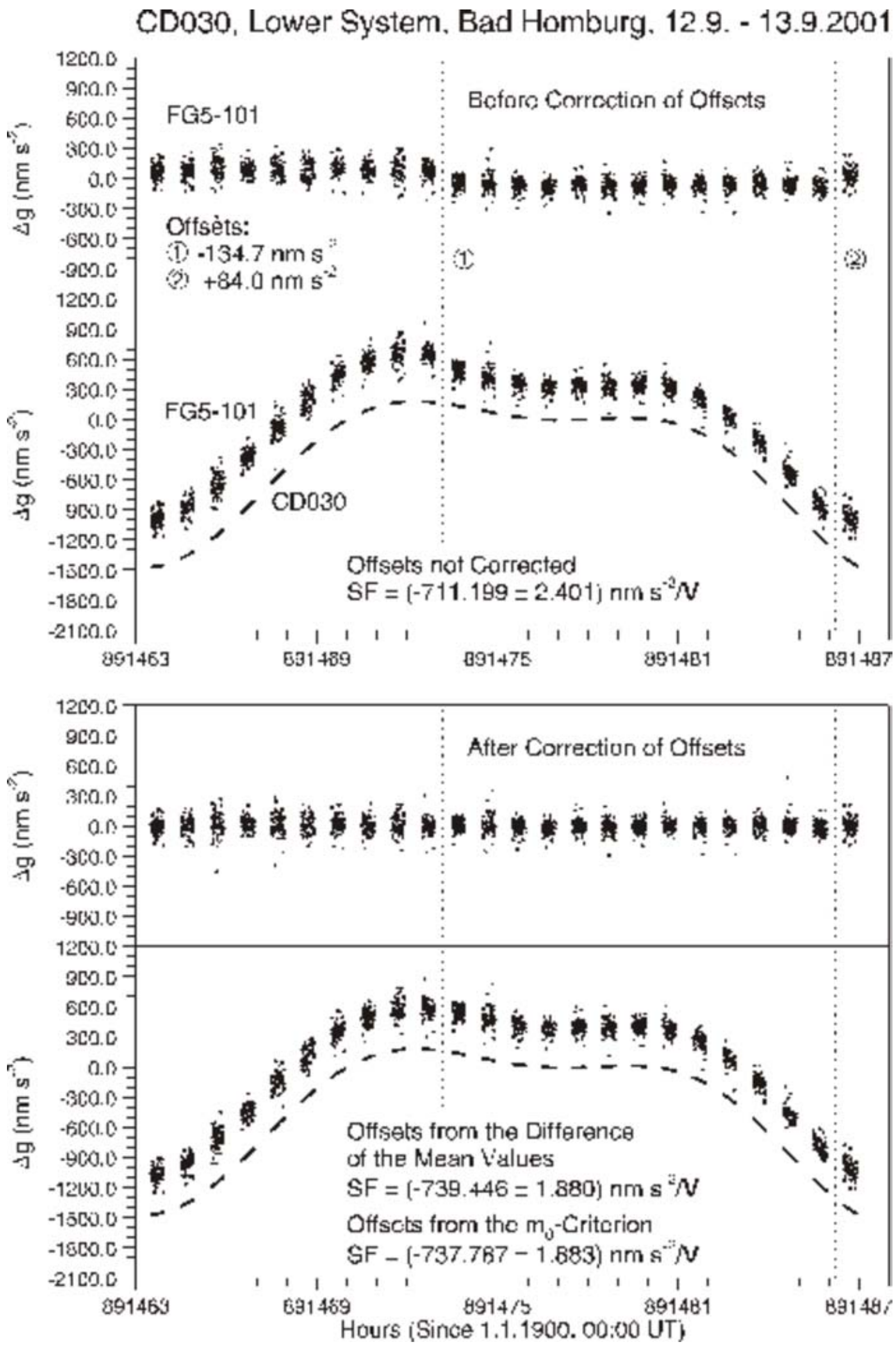

Fig. 2: Tidal signals measured by AG and SG. AG data influenced by offsets (Example of the CD030, Bad Homburg, 12.-13.9.2001)

To minimize the computational problems the absolute measurements should be arranged in such a way that generally from the start outliers are avoided as far as possible. From the viewpoint of the calibrations low repetition rates of the drops are to be preferred. Obviously the number of outliers reduces, if the drops follow at greater time intervals, e.g. every 20 s instead of 10 s as it is common use. However, up to now not enough data are available for a systematic investigation of this problem. On the other hand, if the assumption is correct, the absolute measurements themselves could benefit from a reduction of the repetition rates.

From the viewpoint of the accuracy the tidal variation during the calibration experiment should be as large as possible. Above all this depends from the right choice of the measuring period. Always, the data set of each measuring campaign has to be handled as a whole. This is especially valid for the elimination of outliers.

Statistical methods have to be used for the detection and elimination of outliers. It is clear that the rules for the application of the different statistical tests have correctly to be fulfilled. First of all the data must be homogeneous.

Therefore they have to be corrected with respect to several disturbing influences (e.g. offsets due to different installations or readjustment of the absolute gravimeter, earth tides, gravity effect of air pressure, polar motion). After the outliers have been eliminated, the original data are again used. The corrections are without influence on the further steps of the calibration procedure.

Due to the more or less large number of outliers, at least the raw data are not normally distributed. Therefore only such tests are allowed, which are not based on the assumption of a normal distribution. The tests are cyclically applied. If the test value points out that deviations from the normal distribution exist, the most distant value (with respect to the median value of the remaining data set) is deleted and the next test cycle starts. The procedure stops if the test value remains below a fixed threshold.

The following tests were tried.

- **Comparison of two variances (“2s_0-criterion”, modified F-test)**

The variance of the sample is computed twice. The first estimate s is derived in the common way as if the data would be normally distributed. The second estimate s_0 is derived from the quartiles q_{25} and q_{75} of the sample, taking into consideration that for a normally distributed sample the relation $s_0 = (q_{75} - q_{25})/1.349$ is valid. If s exceeds $2 \cdot s_0$ than deviations from the normal distribution have to be assumed. More correctly speaking the ratio s^2/s_0^2 is tested and the factor 2 stands in a simplified manner for the theoretical threshold defined by the F-distribution.

- **Deviation of the variance from a hypothetical value (χ^2-test)**

Using the same two estimates s and s_0 of the variance a test value $\chi^2 = s^2 \cdot (N-1)/s_0^2$ is derived. If χ^2 exceeds a certain tabulated value, deviations from the normal distribution and therefore the existence of outliers are to be assumed.

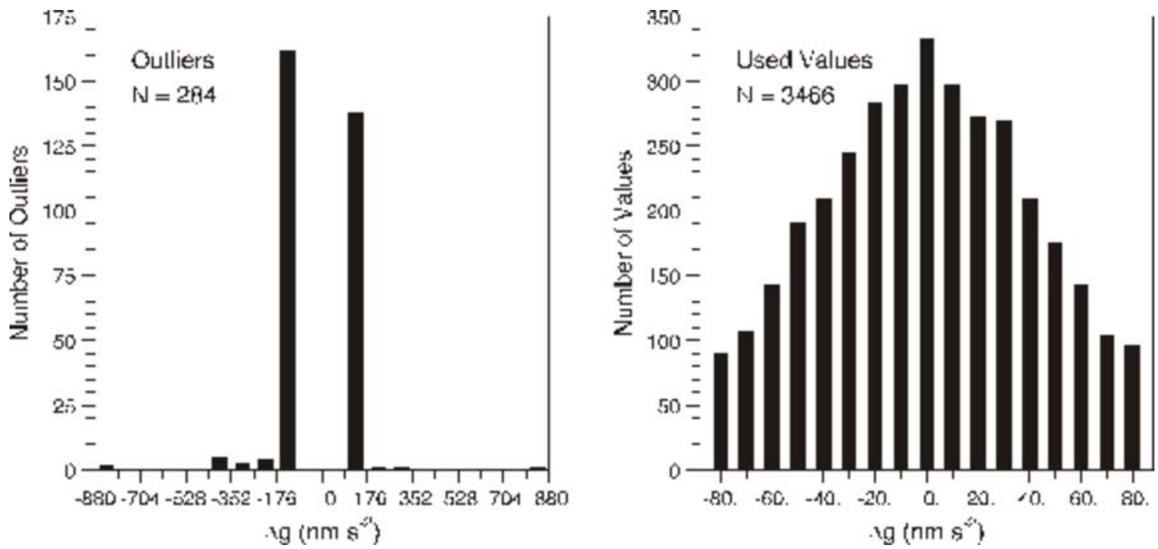

Fig. 3: Frequency distribution of the single drop values during a measuring campaign with the AG (Example FG5-103, Bad Homburg, 2.-3.12.2000). On the right: central part of the histogram. On the left: tails on both sides of the histogram, showing the skew of the frequency distribution

- **Deviation of the mean from a hypothetical value (u-test)**

The mean value of the data set, which is identical with the moment of first order m_1, is compared with the median. If the test value $u = |m_1 - med| \cdot SQRT(N)/s_0$ exceeds a certain tabulated threshold of the normal distribution (e.g. u = 1.960 for a error probability of 0.05), the mean value deviates significantly from the median, i.e. in this case a non-normal distribution and the existence of outliers have to be assumed.

- **Skew of the total data set (threshold criterion)**

A fundamental feature of the normal distribution is its symmetry. On the other hand, a skew unequal zero points out, that the distribution is non-normal. For instance fig. 3 shows the histograms of the outliers (left side) and of the remaining absolute values after the skew test (right side). The skew Sk depends on the statistic moment of third order m_3. It is defined by $Sk = m_3/s^3$, where $s^2 = [vv]/(N-1)$ and $m_3 = [vvv]/N$. The problem is, to find a suited threshold for the decision whether a sample is normally distributed or not. To this end the following variants were tested.

In the simplest case a single fixed threshold was used, e.g. 0.2. The greater the threshold the lower the number of detected outliers.

Normally the skew starts with large values (some tenth or more). After a sufficient number of outliers was deleted, the skew reaches the neighborhood of zero and the distribution of the remaining absolute values becomes nearly a normal one. In the following cycles if more and more outliers are deleted, the skew may move around the zero level. For some time an increase of the skew is also possible. Therefore the elimination process is stopped only when the skew remains at least for 5 cycles below the threshold.

More commonly it was tried, to vary the threshold from values of about 0.2 down to very low values when the number of outliers exceeds 2000 or about the half of the total number of data. The number of detected outliers rapidly rises with the diminution of the threshold. A clear tendency of the resulting scale factors is not to be seen. The variation of the scale factor of each calibration experiment follows an individual curve. Two examples are shown in fig. 4. The scattering range of all curves valid for the different calibration experiments is very large. The single results may be summarized, if an optimum scale factor is derived, which minimizes the differences to the minimally deviating result of each calibration experiment.

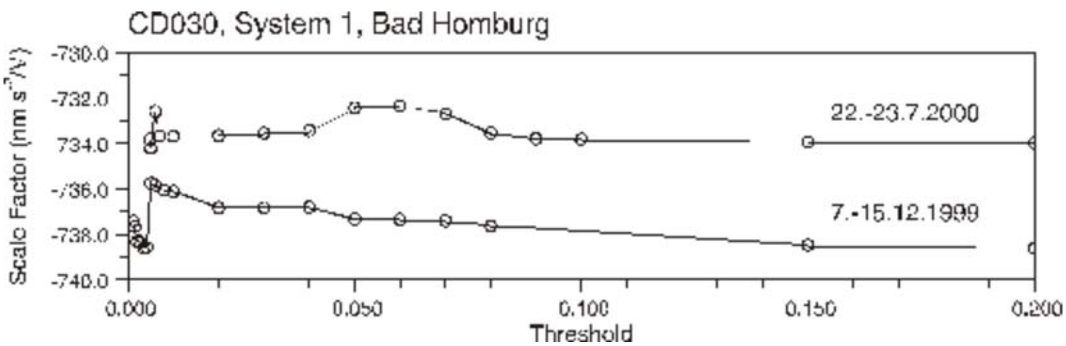

Fig. 4: Dependence of the scale factor on the applied threshold of the skew test. The threshold is lowered until the number of detected outliers exceeds about the half of the original number of AG data

- Minimum of the skew of different subsets (moving window)

No particular threshold is used. Instead of them the skew is estimated for consecutive subsets of given size (described by a certain percentage of the total data set), which are moved step by step over the sorted set of AG data. In a certain position of each of these windows the skew reaches a minimum. The correspondent data set is used for the estimation of the scale factor while the data outside the window automatically are rejected as outliers. The procedure is repeated with windows of different size. In this way for each calibration experiment a series of scale factors results, each of them valid for a certain size of the data window. Two of such curves with clear different behavior are given in fig. 5. Additionally for each calibration experiment an asymptotic value may be derived by fitting a polynomial of third degree with a horizontal tangent in the inflection point (“extrapolation to zero size”).

The influence of the number of eliminated outliers on the difference between mean value and median is shown in fig. 6. It may be seen, that there is no asymptotic approach to the zero level. Due to the outliers, which are included in the data set, the difference starts with large values. If more and more outliers are eliminated the difference decreases, passes the zero level, rises again and finally it varies in a narrow stripe around zero. This behavior is similar to that of the skew as it was discussed above. If the elimination of outliers is stopped too early, the remaining asymmetry of the data set may falsify the estimation of the scale factors.

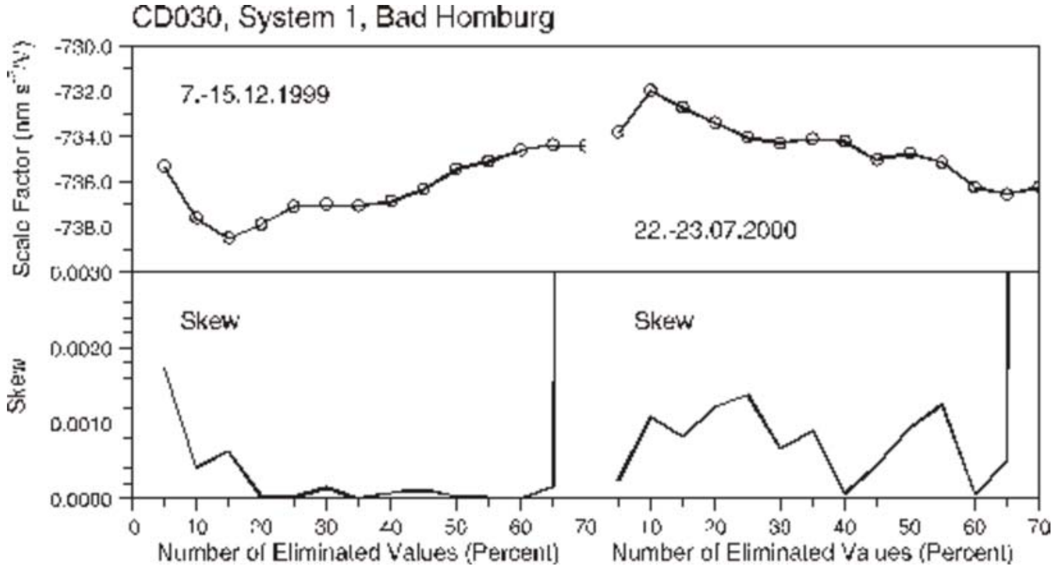

Fig. 5: Influence of the size of the moving window on the estimated scale factor. The size of the window is diminished step by step until the skew increases abruptly.

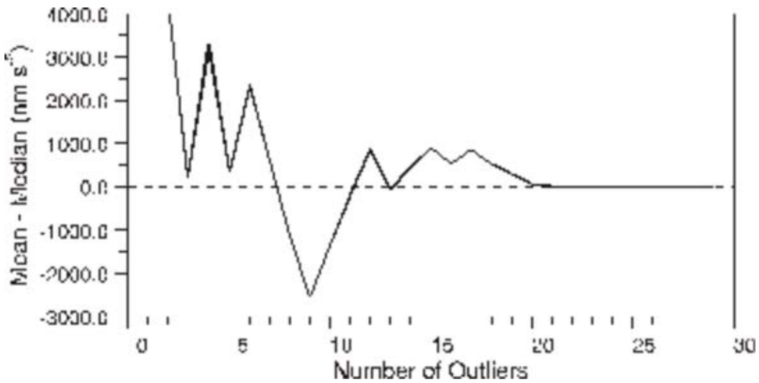

Fig. 6: Influence of the number of eliminated outliers on the difference between mean value and median of the AG data set (Example FG5-101, Bad Homburg, 12.-13.9.2001)

7. Results from the Dual Sphere Gravimeter CD030 at Bad Homburg

All earlier calibrations of the CD030 based on comparisons with AG were reprocessed with regard to the new viewpoints described above (m_0-criterion for the estimation of offsets, skew-criterion for the detection of outliers). In doing this, not all calibration results changed. Several AG comparisons remained unchanged because there neither were offsets nor was the frequency distribution significantly asymmetric. If the skew is near zero, the $2s_0$-criterion is sharp enough for the elimination of outliers. - Additionally the results of some new calibration experiments were included (nos. 10 - 20).

In all cases where the AG data are split into different subsets, the minimum-criterion of m_0 was used for a new estimation of the offsets (calibrations nos. 1, 2, 3, 4, 8, 15, 16, 18, 19). The values of the revised offsets deviate more or less from the previous results (tab. 1, columns 2 and 5). In all examples the mean square errors of the scale factors decrease (columns 4 and 6).

Due to the more stringent skew criterion a greater number of outliers is detected than with the $2s_0$-criterion used before (columns 3 and 7). In some cases the number of values to be deleted changes only slightly and therefore the influence on the scale factor is low. In the other cases the number of outliers grows stronger and as a consequence the scale factor and its mean square error are influenced considerably (columns 4 and 8). However, the results change upward as well as downward as may be seen from the up- and down-arrows in column 9. Though the majority of the results tends to lower values the weighted overall mean increases by about 0.6 nm s^{-2}/V [1].

Tab. 1: Change of the calibration results of the lower system of CD030 due to the new estimation of offsets (minimum of

m_0) and the stronger test for outliers (test for the skew of the frequency distribution). V1 = $2s_0$-criterion, V4 = skew-criterion). Calibration no. 19 was not included in the mean values given in the last line

	Period	Offset old	Out-liers old	Scale Factor old	Offset new	Scale Factor new, V1	Out-liers new	Scale Factor new, V4	D.
		$nm\,s^{-2}$		$nm\,s^{-2}/V$	$nm\,s^{-2}$	$nm\,s^{-2}/V$		$nm\,s^{-2}/V$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	7.-15.12.1999	- 51 +372 - 6 -365	82	-735.86 ± 1.79	- 58 +366 - 12 -370	-738.51 ± 1.78	83	-738.66 ± 1.77	↑
2	23.-24.1.2000	+ 5 -16	4	-734.30 ± 2.29	0 - 9	-732.17 ± 2.29	16	-734.20 ± 1.47	↓
3	7.-9.3.2000	+11 +21	5	-742.01 ± 2.90	+ 9 + 9	-741.98 ± 2.88	13	-739.76 ± 2.10	↓
4	28.-29.3.2000	- 3 +19	4	-739.00 ± 5.00	+ 2 +16	-741.08 ± 5.00	14	-738.68 ± 3.26	↓
5	17.-18.4.2000	–	13	-744.42 ± 8.20	–		22	-741.51 ± 7.59	↓
6	20.-22.5.2000	–	32	-740.74 ± 2.69	–		44	-738.84 ± 1.81	↓
7	8.-9.6.2000	–	26	-734.50 ± 4.05	–		111	-743.23 ± 2.54	↑
8	29.-30.6.2000	+ 4	2	-737.02 ± 1.44	+13	-740.38 ± 1.44	10	-740.40 ± 0.78	↑
9	22.-23.7.2000	–	3	-734.09 ± 2.48	–		10	-733.98 ± 2.44	↓
10	24.-25.8.2000	–	6	-738.35 ± 2.12	–		54	-736.75 ± 1.18	↓
11	28.-29.9.2000	–	5	-739.45 ± 2.58	–		11	-738.38 ± 1.86	↓
12	31.10.-1.11.00	–	1	-739.76 ± 2.57	–		8	-738.36 ± 1.49	↓
13	2.-3.12.2000	–	1	-737.81 ± 1.99	–		40	-739.45 ± 1.44	↑
14	17.-18.5.2001	–	15	-736.67 ± 4.33	–		20	-737.08 ± 3.50	↑
15	13.-15.7.2001		7		+54	-741.84 ± 2.22	17	-742.59 ± 1.43	↑
16	12.-13.9.2001	-142.8 + 89.6	23		-134.7 + 84.0	-737.79 ± 1.88	29	-737.63 ± 1.62	↓
17	20.-21.9.2001	–	18	-741.39 ± 2.88	–		300	-738.15 ± 1.38	↓
18	13.-14.1.2002		3		- 9.6	-736.00 ± 2.10	14	-736.32 ± 1.28	↑
[19]	12.-13.2.2002		0		- 5.8	-730.52 ± 2.36	19	-729.58 ± 1.48	↓
20	26.-29.3.2002		6		-23.0 +3.5 +19.6	-742.57 ± 3.25	97	-736.82 ± 1.59	↓
Mean				-737.95 ± 0.58 19 Measurem.				-738.51 ± 0.52 19 Measurem.	↑

Correspondent conclusions with concern to the reprocessing of the comparisons with absolute measurements are valid also for the gravimeters C023 at Medicina and CD029 at Wettzell.

Considering the fact, that deviations from the normal distribution have a great influence on the estimated scale factors, some additional experiments have been made with different strategies for the detection and elimination of outliers. This corresponds to the processing of the absolute measurements themselves, where commonly a greater number of outliers is eliminated.

At first the skew criterion was applied with different thresholds. Two examples of the influence on the estimated scale factors are given in fig. 4. The curves of the different calibration experiments scatter over a range of more than 10 nm s^{-2}/V. As a common tendency for each calibration experiment the estimated scale factors decrease with decreasing threshold. However, in detail there are many deviations. If the scale factors, resulting for a certain threshold are averaged over all the experiments, the same tendency results. If the threshold 0.2 is used, between 8 and 111 outliers are deleted and a scale factor of (-737.93 ± 0.67) nm s^{-2}/V results. On the other hand, if the threshold is lowered to 0.05 between 20 and 1561 outliers are deleted and the scale factor changes to (-737.72 ± 0.53) nm s^{-2}/V. In a similar way from the first minimum of the skew the value (-736.93 ± 0.42) nm s^{-2}/V results, while the smallest useful threshold leads to (-736.54 ± 0.52) nm s^{-2}/V. The optimum scale factor, which summarizes the results of all the different calibration experiments results to -737.17 nm s^{-2}/V.

Finally the moving window technique was applied with window sizes between 30% and 95%, i.e. between 5% and 70%

of the data were excluded (fig. 5). If windows of 60% and 95% are used, scale factors (-737.10 ± 0.50) nm s^{-2}/V and (-737.67± 0.61) nm s^{-2}/V result. This confirms again that the greater the window, i.e. the smaller the number of values, which are not considered ("outliers"), the greater the resulting scale factor. If the asymptotic values of the single calibration experiments are averaged (extrapolation to windows with zero size), a value of (-736.72 ± 0.50) nm s^{-2}/V results, which corresponds to the optimum value given above.

Compared with the FG5-101 (BKG, Frankfurt a.M., Germany) the data of FG5-103 (POL, Bidston, U.K., calibration no. 13) show a deviating behavior. At low numbers of outliers both instruments agree well. If the number of eliminated outliers rises, the scale factor derived from the data of FG5-103 tends to significantly greater values (about –744 nm s^{-2}/V), i.e. the difference between the results of both AG increases. However, this statement is only based on one calibration experiment and it is dangerous to rush to conclusions.

Tab. 2: Calibration of the dual sphere gravimeter CD030 by comparisons with absolute gravimeters and by the Frankfurt Calibration System (FCS)

CD030, Bad Homburg	Lower System			Upper System		
	Scale Factor nm s^{-2}/V	Phase L. s	N [4])	Scale Factor nm s^{-2}/V	Phase L. s	N [4])
Comparison with Absolute Gravimeters						
Dec. 1999 - July 2000 [1]) Dec. 1999 - Mar. 2002 [2])	-736.90 ± 0.88 -738.51 ± 0.52		9 19	-676.26 ± 0.84 -677.91 ± 0.60		9 16
Frankfurt Calibration System (FCS)						
February 2000 February 2000 [3])	-739.71 ± 0.23 -739.75 ± 0.25	40.18 40.18		-679.22 ± 0.41 -678.68 ± 0.72	41.58 41.60	
December 2000	-739.58 ± 0.19	40.37	2/12			2/15
Mean [3])	-739.66 ± 0.16	40.28		-678.68 ± 0.72	41.60	
Difference FCS – Absolute Comparison						
August 2000 [1]) April 2002 [2])	2.81 1.15			2.96 0.77		

[1]) State August 2000 (ETS2000, Mizusawa [1])

[2]) State April 2002

[3]) Revised result, March 2002

[4]) N means
- number of absolute measuring campaigns for the comparison method
- number of calibration sequences during one calibration experiment for the FCS method. The number of different periods during each calibration sequence follows the slash.

From the different experiments with more stringent criteria for the elimination of outliers scale factors in the range between about -736.5 and -737.0 nm s^{-2}/V were derived. This moment the single results shall not be discussed in detail. Generally, it may be stated that with increasing number of outliers the scale factors tend to smaller values. At the same time the results seem to be stabilized. Obviously the results are influenced by systematic errors, depending on the number of outliers. Therefore only groups of scale factors may be averaged, which are derived with similar criteria for the elimination of outliers. It has to be assumed that the asymmetry in the frequency distribution is fundamentally (at least to a certain, very low extent) and not only caused by a more or less large number of single outliers. A decision, which are the most reliable results would be made easier if the results could be compared with reliable results from the Frankfurt Calibration System (FCS), which is based on a fundamentally different principle. At present it must be stated, that all scale factors derived from comparisons with AG are less than the results of the FCS, which are available up to now. The smallest deviations occur, if the skew test with a threshold of 0.2 is applied. All attempts to get a better agreement by more stringent criteria for the elimination of outliers result in smaller scale factors, i.e. in an increasing difference between both calibration methods. Therefore the following considerations exclusively refer to the skew test with threshold 0.2.

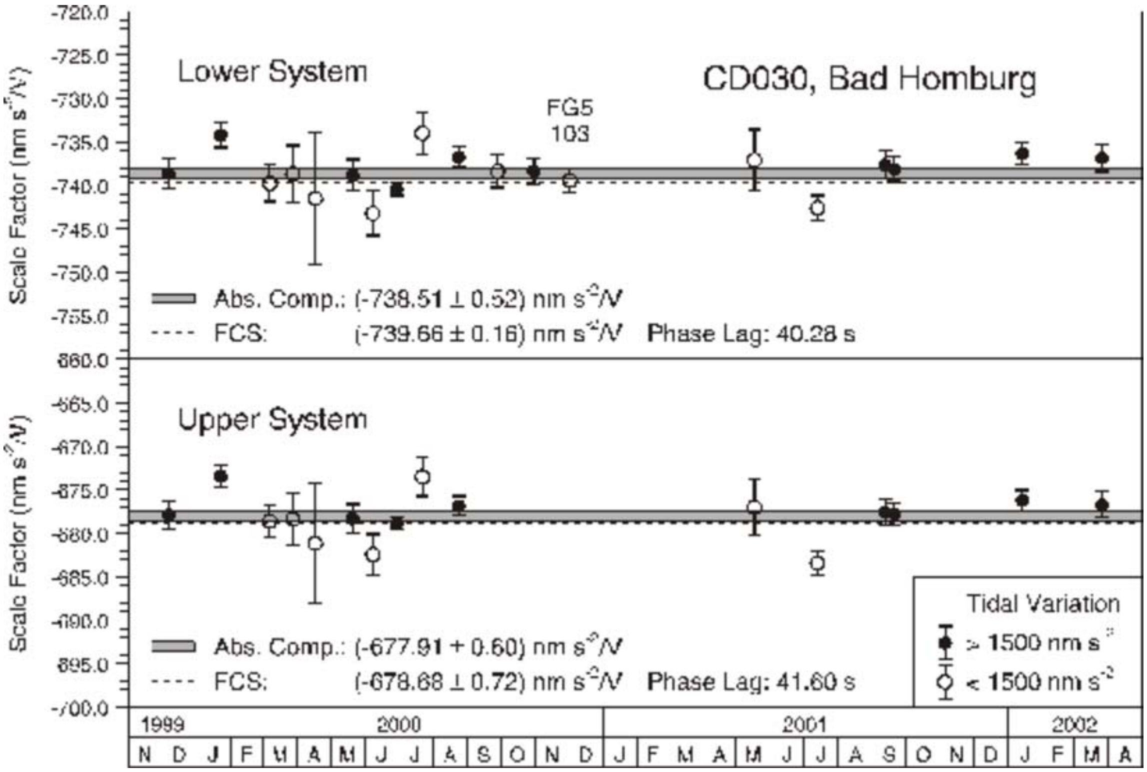

Fig. 7: Scale factors of the CD030, derived by comparisons with AG. Grey stripe: ±1σ-range of the weighted mean of the 19 single values (The strongly deviating result of calibration no. 19 was not included). Broken lines: results of the Frankfurt Calibration System (FCS), state of March 2002

8. Comparison with older results and with the FCS

A summary of the results of different calibrations of the dual sphere gravimeter CD030 is given in tab. 2 and fig. 7. In the upper frame of fig. 7 the values of the lower system are compared with the FCS result. The results of the AG comparisons are given in detail in table 1, column 8. The lower frame demonstrates the correspondent results concerning the upper system.

Generally since the ETS2000 the discrepancy between the comparisons with absolute measurements and the FCS results diminished. Both calibration methods contribute to this improvement.

The comparisons with AG were reprocessed with respect to the improved estimation of offsets and more stringent criteria for the elimination of outliers. Moreover 11 new absolute measurements at Bad Homburg could be additionally taken into the considerations. In this way, the scale factor of the lower system changed from (-736.902 ± 0.882) to (-738.512 ± 0.515) nm s^{-2}/V and that of the upper system from (-676.263 ± 0.839) to (-677.914 ± 0.604) nm s^{-2}/V. As a general tendency the revised results lower the discrepancy between the FCS and the comparisons with AG.

The calibration results on the basis of the FCS were also revised and the results of a new FCS experiment could be included. Very encouraging was a separate experimental determination of the frequency transfer function of the FCS. A detailed description of some new aspects of the FCS method shall be given in a separate contribution [2].

As may be seen from tab. 2 for the lower system two FCS calibrations are available (February and December 2000). The second calibration deviates by only 0.171 nm s^{-2}/V from the revised result of the first one. Roughly speaking the scale factor remains nearly unchanged with respect to the state of August 2000 [1] (with a very slight tendency to decrease the discrepancy between the both calibration methods). During the FCS calibration in December 2000 the upper system of the CD030 was out of order. Therefore, only results from the calibration in February 2000 are available for the upper system. In contrast to the lower system the revised result changed by 0.547 nm s^{-2}/V with a clear tendency to reduction in the discrepancy with respect to the results of comparisons with absolute measurements.

Altogether, the discrepancy between the FCS method and the comparisons with AG reduces from 2.810 nm s^{-2}/V (state

August 2000) to 1.062 nm s^{-2}/V (state March 2002) for the lower system or from 2.962 nm s^{-2}/V to 0.684 nm s^{-2}/V for the upper system, respectively. Also, if the remaining discrepancy is not significant in the strong sense of statistics, it needs to be explained. It is striking, that without any exception the absolute value of the FCS calibration is always greater than the mean value of the scale factor derived from comparisons with absolute gravity measurements.

9. Conclusions

From the critical revision of the evaluation procedures, the results of different comparisons with absolute gravimeters and some applications of the Frankfurt Calibration System (FCS) the following conclusions may be drawn.

1. The example of the dual sphere gravimeter CD030 encourages use of comparisons with absolute gravimeters as a reliable method for the calibration of superconducting gravimeters. While the accuracy derived from a single measuring campaign may reach the order of 1.5 nm s^{-2}/V, the mean of several comparisons, spread over a longer period, gives stable results with an accuracy of better than $1 \cdot 10^{-3}$.

2. Great care has to be taken in the elimination of outliers and the estimation of offsets. The absolute gravity data taken in this analysis could not be seen as a priori normally distributed. Therefore, common criteria based on the variance of the normal distribution are not suitable. For the elimination of outliers the skew of the data distribution has proved to be a reliable test value. Offsets may be estimated by a trial and error procedure, which searches for a minimum of the mean error of the calibration result.

3. The calibration of SG by comparison with AG has the great advantage, that the normal operation of the SG is not disturbed. Especially the entire recording electronics need not be changed. This implies, that the transfer characteristic (frequency dependence) remains unchanged and the scale factor cannot be influenced by the calibration procedure.

4. Calibrations using the Frankfurt Calibration System (FCS) are more accurate by at least one order of magnitude in comparison with calibrations on the basis of absolute measurements. The high accuracy of the FCS may be reached during a single calibration experiment, which however needs a large expenditure of work and a time of several days. As a consequence the operation of the SG is interrupted during the period of the calibration experiment.

5. The experiments with different skew thresholds for the detection of outliers show, that the calibration results systematically depend on the number of eliminated outliers. Though no objective criterion for a certain value of the threshold may be given, the results related to the threshold 0.2 are preferred.

6. Since ETS2000 [1] the discrepancy of about 3 nm s^{-2}/V between AG comparisons and FCS partly could be cleared up. About two parts of the reduced discrepancy are due to a reprocessing of the comparisons with absolute gravimeters and about one part is contributed by a revision of the FCS calibrations. In both cases new calibration experiments could also be included in the investigations. However, the remaining discrepancy in the order of 1 nm s^{-2}/V (not significant in the strong sense of statistics) also requires further explanation. It is remarkable that the absolute value of the scale factor based on the FCS is always greater than the mean derived from the comparisons with absolute measurements.

References

1. Falk, R., Harnisch, M., Harnisch, G., Nowak, I., Richter, B., Wolf, P.
Calibration of the Superconducting Gravimeters SG103, C023, CD029 and CD030
J. Geod. Soc. Jap., Japan 47(2001)1, pp. 22 – 27.

2. Richter, B., Harnisch, G., Nowak, I.:
Experimental and Computational Contributions to Estimate the Accuracy and Reliability of the Frankfurt Calibration System (FCS).
Proc. Third Workshop of GGP, Jena, March 2002. In Preparation.

Figures

Fig. 1: Dependence of the scale factor of the SG and its mean error on different amounts of an offset in the AG data (Example of the CD030, Bad Homburg, 13.-15.7.2001)

Fig. 2: Tidal signals measured by AG and SG. AG data influenced by offsets (Example of the CD030, Bad Homburg, 12.-13.9.2001)

Fig. 3: Frequency distribution of the single drop values during a measuring campaign with the AG (Example FG5-103, Bad Homburg, 2.-3.12.2000). On the right: central part of the histogram. On the left: tails on both sides of the histogram, showing the skew of the frequency distribution

Fig. 4: Dependence of the scale factor on the applied threshold of the skew test. The threshold is lowered until the number of detected outliers exceeds about the half of the original number of AG data

Fig. 5: Influence of the size of the moving window on the estimated scale factor. The size of the window is diminished step by step until the skew increases abruptly

Fig. 6: Influence of the number of eliminated outliers on the difference between mean value and median of the AG data set (Example FG5-101, Bad Homburg, 12.-13.9.2001)

Fig. 7: Scale factors of the CD030, derived by comparisons with AG. Grey stripe: ±1σ-range of the weighted mean of the 19 single values. Broken lines: results of the Frankfurt Calibration System (FCS), state of March 2002

---

[1] If scale factors are compared, in the following text always the absolute values are considered.