Metrologia **47** (2010) 1–10 doi:10.1088/0026-1394/47/1/001

Long-term comparison of caesium fountain primary frequency standards

Thomas E Parker

National Institute of Standards and Technology, Time and Frequency Division, 325 Broadway, Boulder, CO 80305, USA

E-mail: tparker@boulder.nist.gov

Received 20 August 2009, in final form 21 October 2009
Published 7 December 2009
Online at stacks.iop.org/Met/47/1

Abstract
There are currently nine caesium fountain primary frequency standards regularly reporting calibrations of International Atomic Time to the Bureau International des Poids et Measures (BIPM). An investigation has been carried out using data from the BIPM publication *Circular T* to evaluate the frequency differences among these standards and to determine whether these offsets are consistent with the stated uncertainties. The fractional frequency uncertainties of some Cs fountains are now in the range of 4×10^{-16} to 5×10^{-16}. The results of this investigation show that the standards agree well with each other. An overall estimate of the caesium frequency is made using the weighted mean of all the fountains.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The first formal evaluation of a caesium fountain primary frequency standard (PFS) was reported to the Bureau International des Poids et Measures (BIPM) by the Observatoire de Paris in September 1995 [1]. However, only since November 1999 have laboratories with Cs fountain PFSs been regularly reporting evaluation results. As of June 2009 there are nine reporting fountains, and since 1999 there have been 187 formal reports published in *Circular T* (www.bipm.org/jsp/en/TimeFtp.jsp). Formal fountain reports to the BIPM are used to calibrate the rate (frequency) of International Atomic Time (TAI). The body of data is now large enough that a meaningful comparison can be made among the standards. There have been a few direct fountain comparisons reported in [2–5], and this is clearly the best way to compare fountains. In a direct comparison the operation times can be coordinated to minimize dead-time and the frequency-transfer uncertainty can also be reduced by optimizing the transfer processes. This minimizes the statistical uncertainties in the comparison and allows the best determination of the differences in the fountain frequencies. Any variations in the differences over time could then be observed. However, there have been only a few such comparisons. Currently the *Circular T* data provide the best overall comparison, although this is a long-term comparison covering many years. A comparison of fountain frequencies is a key test of whether the stated fountain uncertainties are consistent with the observed frequency differences.

In this study the comparison is made by using individual pairs of reports in *Circular T* that occur close together in time. Since the report periods generally do not exactly overlap (dead time is present) the stability of the reference flywheel must be taken into account in calculating the comparison uncertainty. Two different, independent, flywheel frequency references are used. One is TAI, and the other is the internal, post-processed, maser-based time scale, AT1E, at the National Institute of Standards and Technology (NIST). In addition to dead-time uncertainties, the uncertainty introduced by frequency transfer must also be included. For any two standards a number of data pairs are available over time, and these can be averaged to give an overall fractional frequency difference and a total uncertainty of comparison.

2. Fountains reporting to the BIPM

LPTF-FO1 (now SYRTE-FO1) was the first Cs fountain PFS to report to the BIPM in September 1995, and had a fountain uncertainty of 3×10^{-15} [1]. An equivalent amount of frequency-transfer uncertainty was also present. Twelve reports from LPTF-FO1 were submitted by the Laboratoire Primaire du Temps et des Frequences (now Laboratoire National de Métrologie et d'Essais, Systèmes de Référence

0026-1394/10/010001+10$30.00

Table 1. Typical recent Type A, u_A, Type B, u_B, and combined, u, uncertainties of fountains currently reporting to the BIPM.

#	Fountain	u_A (10^{-15})	u_B (10^{-15})	u (10^{-15})	Run length (days)	Number of formal reports (date started)
1	NIST-F1	0.15	0.33	0.36	25	36 (11/1999)
2	SYRTE-FO1	0.30	0.40	0.50	25	17[a] (9/1995)
3	SYRTE-FO2	0.35	0.45	0.57	30	36 (11/2002)
4	SYRTE-FOM	0.2	0.71	0.74	25	22 (11/2002)
5	PTB-CSF1	0.1	0.9	0.9	25	21 (8/2000)
6	IT-CsF1	0.9	0.5	1.0	20	20 (4/2003)
7	NICT-CsF1	1.0	0.8	1.3	10	7 (10/2006)
8	NPL-CsF1	0.5	1.8	1.9	35	8 (3/2004)
9	NMIJ-F1	0.7	3.9	4.0	30	20 (7/2005)

[a] Since year 2006.

Temps Espace (LNE-SYRTE)) in France from September 1995 to November 1997, then there were no additional reports from this standard until November 2006 [6, 7]. NIST-F1 from the National Institute of Standards and Technology in the USA started reporting in November 1999 [8] with a fountain uncertainty of 1.8×10^{-15} and a frequency-transfer uncertainty of 1.5×10^{-15}. NIST-F1 is still reporting regularly with improved uncertainty [9, 10]. In August 2000 PTB-CSF1 from Physikalisch Technische Bundesanstalt (PTB) in Germany made its first report [11, 12]. From this date on there have been at least two fountains reporting into *Circular T* several times per year. SYRTE-FO2 [6, 7], SYRTE-FOM [13] and IT-CsF1 [14] (from Istituto Nazionale di Ricerca Metrologica, INRIM, in Italy) all joined the fountain PFS community near the end of 2002. NPL-CsF1 [15], from the National Physical Laboratory (NPL) in the United Kingdom, started in 2004, NMIJ-F1 [16] from the National Metrology Institute of Japan (NMIJ) started in 2005 and NICT-CsF1 [17] from the National Institute of Information and Communication Technology (NICT) in Japan started in 2006. Currently these nine fountains from seven laboratories are reporting on a more or less regular basis. Every month now there is at least one fountain reporting into *Circular T*, and recently there have typically been three to four fountains. The uncertainty of the rate, or scale interval, of TAI is now sometimes as low as 4 to 5×10^{-15}. Data from August 2000 to June 2009 (June data published in July) are used in this study and all uncertainties presented in this paper are standard uncertainties (1 sigma).

Table 1 gives a list of the nine fountains, with typical uncertainties from recent evaluation reports sent to the BIPM (www.bipm.org/jsp/en/TimeFtp.jsp). The fountains are listed in the order of total combined uncertainty, u. The Type A and B uncertainties, u_A and u_B, are 'in laboratory' uncertainties and do not include frequency-transfer uncertainties. These fractional frequency uncertainties are in units of 10^{-15}. The corresponding run length in days is also given because the Type A (statistical) uncertainty is dependent on the run time. The Type B uncertainties [18] may vary somewhat from run to run, but in general they tend to decrease slowly with time as more is learned about each standard. See the appendix for a discussion of Type A and Type B uncertainties in PFSs. The last column in the table gives the number of reports from each standard that have been submitted to the BIPM as of June 2009 and the date of the first report. Note for SYRTE-FO1 that the number of reports is only since its return in 2006.

As shown in table 1, the combined uncertainties range over an order of magnitude among the fountains, but even the largest uncertainty is still lower than that of the best thermal beam standard. The three fountains with the lowest combined uncertainties are NIST-F1, SYRTE-FO1 and SYRTE-FO2, all with uncertainties in the mid-10^{-16} range. Given that these uncertainties vary a little from run to run, these three standards should be considered as nearly equivalent, although NIST-F1 has consistently had the lowest Type B uncertainty.

Figure 1(*a*) shows the fractional frequency (rate) offset of TAI as measured by each fountain PFS since November 1999. The reported uncertainty, including frequency-transfer uncertainty, is also shown for each fountain with the error bars. These data from *Circular T* are plotted as a function of Modified Julian Date (MJD) and cover a period of almost ten years. The long-term variations are in the rate of TAI, but the short-term fluctuations are from both the noise in TAI and variations in the fountain frequencies. There are a few apparent outliers near MJDs 52 800 and 53 600. Since MJD 54 000 there have been so many fountains reporting that it is difficult to resolve individual data points. Figure 1(*b*) shows an expanded view of these data after MJD 53 900. For this investigation we are interested in pairs of fountain measurements that occur within 100 days of each other. Using data with time interval offsets greater than that is not recommended since the dead-time uncertainty becomes quite large, and, more importantly, the long-term frequencies of TAI and AT1E are not independent of the fountain frequencies. The rate (frequency) of TAI is steered by the BIPM towards the fountain frequencies in small frequency steps once a month, although the effective steering time constant is well over one year. The frequency of AT1E is steered at NIST toward the fountain frequencies only through linear frequency ramps that are adjusted about once every 100 days.

A possible alternative flywheel to TAI could have been TT(BIPMXX), where XX represents the last two digits of the computation year. TT(BIPMXX) is a post-processed computation of Terrestrial Time (TT) performed on an annual basis at the BIPM and is a better overall estimate of TT than TAI, which is calculated monthly. However, since TAI has essentially the same stability over tens of days as

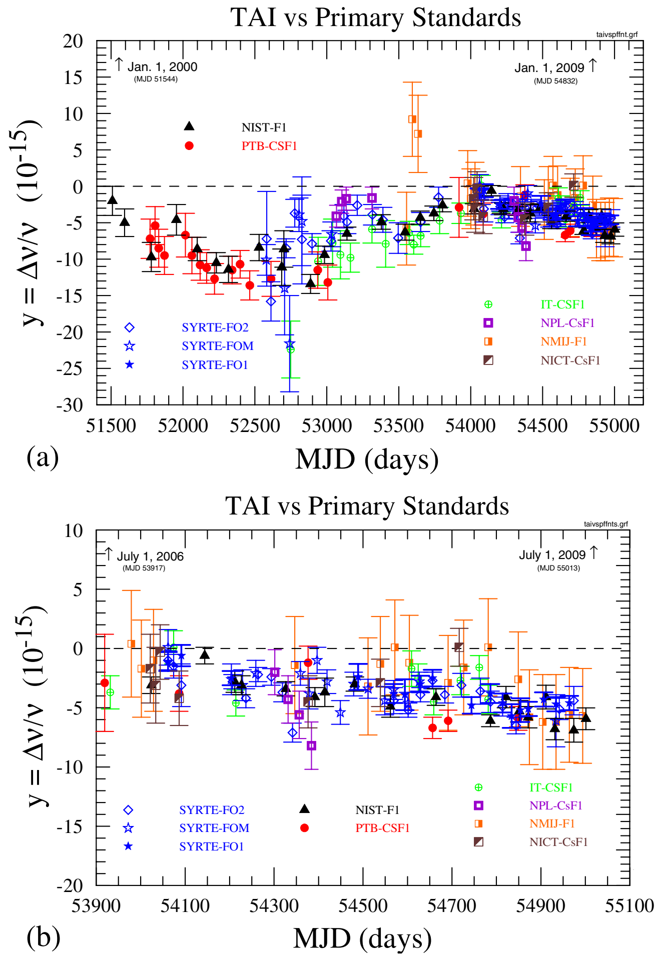

Figure 1. (*a*) Fractional frequency offset of TAI relative to Cs fountain PFSs as a function of MJD as reported in *Circular T* since November 1999. (*b*) Expanded view of the fractional frequency offset of TAI relative to Cs fountain PFSs from the data in figure 1(*a*) since June 2006.

TT(BIPMXX), and is available each month, it was chosen for this study.

3. Procedures for comparing fountains

A key test of fountain performance is to determine whether the scatter over time in a fountain frequency is consistent with the stated uncertainties. It is equally important to determine whether the various fountains agree with each other within the stated uncertainties. The study reported here is an attempt to answer these questions.

Clearly, it would be highly desirable to have direct fountain comparisons in which the start and stop times are coordinated and the frequency-transfer techniques are optimized. However, operating fountains on demand has been very difficult, and relatively few such comparisons have been accomplished [2–5]. Another approach is to use the data available in *Circular T*. Wolf *et al* in [19] used an approach in which variations of a TAI type time scale were calculated excluding one fountain at a time. This has the advantage of using all available data for the other fountains, but excluding a high weight fountain produces a time scale missing key

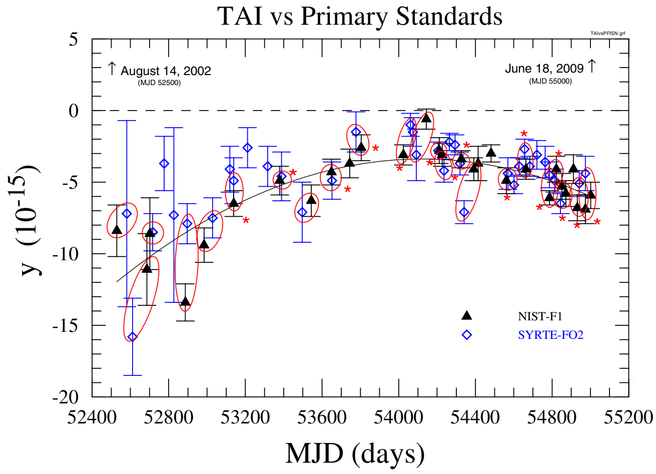

Figure 2. Fractional frequency offset of TAI versus NIST-F1 and SYRTE-FO2 as a function of MJD. The pairs of data points used for the frequency comparison are circled (in red). The 15 pairs marked with the (red) asterisks are used for a subset with a lower uncertainty.

information. Another significant difference in the Wolf *et al* study is that Type B uncertainties were treated as uncorrelated in time, and therefore were averaged down in a manner similar to Type A uncertainties (see section 5).

For the investigation reported in this paper a different approach is being used in which pairs of fountain data points from *Circular T* that are closely aligned in time are compared. This gives a fairly large number of data points, but they are generally not exactly aligned in time. Each data point from *Circular T* gives the rate of TAI [20] relative to a particular PFS. Differencing two data points that are closely aligned in time gives an estimate of the frequency difference between the two standards. However, this introduces a dead-time uncertainty that depends on how closely the two runs are aligned and on the frequency stability of the frequency reference [21, 22]. To help reduce this dead-time uncertainty, the fountain frequency values are also referenced to a post-processed, maser-based ensemble at NIST referred to as AT1E [23]. Although the clocks in AT1E are also in TAI, the two scales are essentially independent since the AT1E clocks have less than 5% of the total weight in TAI. The dead-time uncertainty is slightly reduced by averaging the results from both time scales. Typically, this results in a reduction of up to 15% in the total Type A uncertainty of the comparison.

Figure 2 shows only the set of data pairs for NIST-F1 and SYRTE-FO2 from figure 1. The (blue) diamonds show all the data points of TAI versus SYRTE-FO2 that have appeared in *Circular T* through June 2009. The (black) triangles are the data points for NIST-F1 that were reported over the same interval. There are 36 SYRTE-FO2 data points and 29 points for NIST-F1. Each pair of points used in this analysis is circled (in red). Only one data point from each standard was used for each comparison pair. There are 23 suitable pairs, with no pair having more than 75 days of centre-to-centre time offset. The mean centre-to-centre offset (using signs) is +3.8 days, with a mean magnitude of offset (ignoring signs) equal to 21.2 days. About 70% of the runs had at least some overlap in run time. Type B uncertainties tend

Table 2. Example of a fountain comparison data pair and the difference calculation for NIST-F1 and SYRTE-FO2. Fractional frequencies are in units of 10^{-15}.

NIST-F1

MJD	MJD	Duration	y(TAI-F1)	u_A	u_B	u_l	u_{TAI}
54314-54339	54326.5	25d	–3.4	0.3	0.3	0.1	0.4

u_{dead}	y(FO2-F1)	u_C	u_{CA}	u_{CB}	overlap
0.7	0.3	1.2	1.05	0.58	15d

SYRTE-FO2

MJD	MJD	Duration	y(TAI-FO2)	u_A	u_B	u_l	u_{TAI}
54309-54329	54319	20d	–3.7	0.3	0.5	0.1	0.5

to decrease over time, so a subset of 15 recent pairs has also been selected. These are indicated with (red) asterisks. These pairs also have a lower frequency-transfer uncertainty because transfer technology has been slowly improving [24, 25] and for a few points a smaller frequency-transfer uncertainty was used than that indicated in *Circular T*. (In September 2006 the BIPM started using an updated equation to calculate the frequency-transfer uncertainty in *Circular T*, but in fact the actual uncertainty was lower even before this change.) In addition, a tighter requirement on overlap was used such that the mean centre-to-centre offset is −1.0 days, with a mean magnitude of offset equal to 9.3 days. 93% of the runs had at least some overlap in run time. As a result this subset has smaller Type A and Type B comparison uncertainties. The same procedure as that shown in figure 2 was also used with AT1E as the frequency reference for exactly the same pairs.

The solid (black) line in figure 2 is a second order fit line to the NIST-F1 data. Its only function is to illustrate the long-term frequency drift of TAI. It does not represent the short-term stability of TAI.

The procedure used for combining uncertainties is as follows. All Type A uncertainties are treated as uncorrelated and are therefore added in quadrature when two standards are compared. Over time the uncertainty of the mean will average down. Type B uncertainties are treated as uncorrelated between standards, but correlated over time. Neither of these statements regarding Type B uncertainties is strictly true, but for the purpose of this study we will assume that they are a reasonable compromise. A rigorous process of combining the Type B uncertainties would require a detailed analysis of Type B biases and uncertainties for each standard, and how these biases vary as a function of time. Such an analysis is beyond the scope of this study. Therefore, Type B uncertainties of different standards will be combined in quadrature for a pair, but over time a weighted average of the Type B uncertainties will be used. Thus, Type B uncertainty will not average down and will never be smaller than the smallest individual Type B uncertainty of a single pair.

As an example, table 2 shows the details of how a pair of data points is handled for NIST-F1 and SYRTE-FO2. The first column shows the start and stop dates (in MJD) for the runs for each standard. The second column gives the midpoints of each run and the third column is the duration in days. The overlap in days is given in the last column of the middle (offset) row. As is typical, these two runs were made with no knowledge that the other standard was being operated. Columns 4 to 8 in the top and bottom rows are, respectively, (4) the fractional frequency difference between TAI and the particular fountain, (5) u_A, the reported Type A uncertainty of the fountain, (6) u_B, the Type B uncertainty of the fountain, (7) u_l, the uncertainty in the link between the PFS and the local clock contributing to TAI, sometimes dominated by fountain dead time (the fountain may not have operated continuously over the report interval) and (8) u_{TAI}, the frequency-transfer uncertainty in the link to TAI. All these uncertainties are obtained from *Circular T*.

The middle (offset) row contains the results of the frequency difference calculation. Here u_{dead} is a dead-time uncertainty [21, 22] introduced by the run misalignment and the noise of the frequency reference, in this case TAI. If the start and stop times of each fountain run were exactly the same this uncertainty would go to zero. Columns 2 to 5 in this middle row are, respectively, (2) the calculated fractional frequency difference between SYRTE-FO2 and NIST-F1, (3) the total combined uncertainty of the comparison, u_C, (4) the Type A uncertainty of the comparison, u_{CA}, and (5) the Type B uncertainty of the comparison, u_{CB}.

Note that the two u_{TAI}'s and u_{dead} are among the larger individual uncertainties. u_A, u_l, u_{TAI} (top and bottom rows) and u_{dead} are all Type A uncertainties and are combined in quadrature to give u_{CA} of the comparison in the middle row. The two individual Type B uncertainties in the top and bottom rows are also combined in quadrature to give the comparison Type B uncertainty, u_{CB}. u_{CA} and u_{CB} in the middle row are combined in quadrature to give the total uncertainty of the comparison, u_C.

For NIST-F1 and SYRTE-FO2 there are 23 data pairs as in table 2. The average fractional frequency offset, $y(\mathrm{FO2}-\mathrm{F1})_{avg}$, for all 23 is determined by calculating the weighted mean of $y(\mathrm{FO2}-\mathrm{F1})$ using $1/u_C^2$ as the weights. The Type A uncertainty for all 23 pairs averages down as

$$\frac{1}{U_{CA}^2}=\sum_{i=1}^{n}\frac{1}{u_{CAi}^2}, \tag{1}$$

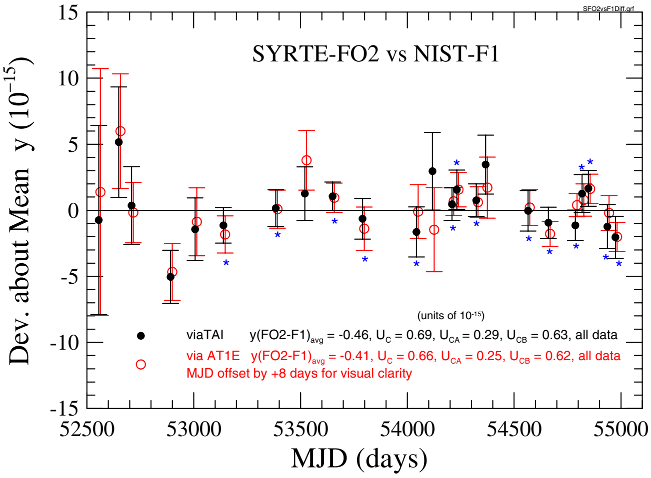

Figure 3. Comparison of SYRTE-FO2 and NIST-F1 on a point by point basis. Average values for all 23 points from TAI and AT1E are shown. The (blue) asterisks indicate the points used in the reduced data set.

where $n = 23$ in equation (1). The Type B uncertainty for the 23 data pair comparison is calculated as the weighted average of the individual Type B uncertainties, again using u_C to calculate the weights.

The value $y(\mathrm{TAI}-\mathrm{FO2})$ can be transformed into $y(\mathrm{AT1E}-\mathrm{FO2})$ by using data in *Circular T* and internal data at NIST. With these values the same procedures used for TAI can be applied to AT1E to give another set of data for $y(\mathrm{FO2}-\mathrm{F1})$, but using AT1E as a flywheel. The data sets from TAI and AT1E, having exactly the same pairs, are then averaged to obtain the final results, which are discussed in section 4. Averaging the TAI and AT1E results reduces the u_{dead} contribution by about a factor of $1/\sqrt{2}$, and this gives a modest reduction in comparison uncertainty if the other Type A and Type B uncertainties are not significantly larger than u_{dead}.

Figure 3 shows the point by point results for $y(\mathrm{FO2}-\mathrm{F1})$ using both TAI and AT1E. The error bars represent the total comparison uncertainty, u_C, for each point, as in table 2. The vertical axis in the figure is the deviation about the means, and the mean, $y(\mathrm{FO2}-\mathrm{F1})_{\mathrm{avg}}$, for all 23 points is shown in the figure for both TAI and AT1E, along with the resulting uncertainties of the means. Note that the results from TAI and AT1E are very similar. For all the fountains in this study the average difference between the TAI and AT1E results is -0.05×10^{-15}. The largest difference is -0.49×10^{-15}.

4. Comparison results

4.1. NIST-F1, SYRTE-FO2 and SYRTE-FO1

Table 3 shows the results of the comparisons of NIST-F1, SYRTE-FO1 and SYRTE-FO2, the three standards with the smallest total individual uncertainties (see table 1). Column 1 lists the two fountains being compared, and Column 8 shows the number of pairs averaged. Columns 2 to 5 are, respectively, (2) the average fractional frequency difference, y_{avg}, from TAI and AT1E, (3) the total comparison uncertainty, U_C, (4) Type A comparison uncertainty, U_{CA}, and (5) Type B comparison uncertainty, U_{CB}. All are in units of 10^{-15}. U_{CA} and U_{CB} are added in quadrature to get U_C. Note that a capital U is used to indicate the uncertainty of a mean calculated for a data set over time. Columns 6 and 7 are Birge ratios [26] as defined in equation (2).

The Birge ratio is equivalent to the square root of the reduced chi-square:

$$R_B = \sqrt{\frac{\sum_{i=1}^{n} \dfrac{(y_i - \bar{y})^2}{u_i^2}}{(n-1)}}. \qquad (2)$$

The term $(y_i - \bar{y})$ is the deviation of each data pair from the mean, and u_i is either u_{CAi} for R_{BA} or u_{Ci} for R_{BC}. n is the number of data points. The Birge ratio is used as a measure of whether the actual scatter about the mean in a data set is consistent with the uncertainty associated with each point. R_B will be close to 1 if the uncertainties are correct. R_B will be less than 1 if the uncertainties are overstated, and will be greater than 1 if they are understated. In a PFS the scatter in the data may come from just Type A frequency fluctuations or from both Type A and B fluctuations. If Type B bias correction errors are constant and do not change over time, the scatter about the mean in the data will be consistent with just Type A uncertainty. If both Type A and B fluctuations contribute to the scatter of the data, then the Birge ratio is best calculated using the total uncertainty. Therefore table 3 lists R_B calculated with either u_A, R_{BA} or u_C, R_{BC}. R_{BA} will always be larger than R_{BC}. A reasonable scenario is for R_{BA} to be a little larger than 1 and R_{BC} to be somewhat smaller than 1. This will occur when some of the frequency fluctuations over time are caused by variations in Type B bias errors. R_{BA} and R_{BC} will be relatively close in value if the Type B uncertainties are small compared with the Type A uncertainties. If Type B uncertainty is large, however, the difference between R_{BA} and R_{BC} will be larger. There is cause for concern if R_{BC} is found to be significantly larger than 1 or if R_{BA} is significantly smaller than 1. Note that the Birge ratio is biased below 1 when only a small number of degrees of freedom are present. For example, an entirely normal distribution with ten degrees of freedom will give a Birge ratio of 0.97. For four degrees of freedom R_B becomes 0.92. Also, the confidence limits are relatively poor for less than about ten data points.

The first row in table 3 shows the results for the set of 23 data points from $y(\mathrm{F1} - \mathrm{FO2})$. As can be seen, the two fountains are in very good agreement, with an average frequency difference of only -0.44×10^{-15} and a total comparison uncertainty of 0.67×10^{-15}. The average frequency offset is larger than the Type A uncertainty, $U_{CA} = 0.26 \times 10^{-15}$, indicating that the difference is likely not just a statistical fluctuation due to noise. However, the offset is well within the Type B uncertainty of $U_{CB} = 0.62 \times 10^{-15}$. Both Birge ratios are close to 1, indicating consistent data. Note that the data set is large enough that the overall Type A uncertainty is significantly smaller than the Type B uncertainty, even with the large frequency-transfer and dead-time uncertainties inherent in using data from *Circular T*. This result reflects the average performance of the two standards over a period of nearly seven years.

Table 3. Comparisons of NIST-F1, SYRTE-FO1 and SYRTE-FO2. y_{avg} is the average fractional frequency difference, U_C is the total comparison uncertainty, U_{CA} is the Type A comparison uncertainty, and U_{CB} is the Type B comparison uncertainty. R_{BA} and R_{BC} are, respectively, the Birge ratios calculated from the Type A or combined uncertainties. Fractional frequencies are in units of 10^{-15}.

Fountains	y_{avg}	U_C	U_{CA}	U_{CB}	R_{BA}	R_{BC}	# pairs
SYRTE-FO2 versus NIST-F1	−0.44	0.67	0.26	0.62	1.15	1.04	23
SYRTE-FO2[a] versus NIST-F1	−0.43	0.65	0.23	0.61	1.14	0.97	15
SYRTE-FO1 versus SYRTE-FO2	−0.32	0.64	0.16	0.62	0.76	0.53	13
SYRTE-FO1 versus NIST-F1	−0.80	0.60	0.30	0.52	0.86	0.77	13

[a] Selected data.

The second row (fountain names in red) shows the result for the selected set of 15 pairs discussed earlier. By using the smaller data set (blue asterisks in figure 3) a slightly smaller total uncertainty is obtained because of the smaller u_B and the reduced transfer uncertainty. Again NIST-F1 and SYRTE-FO2 show very good agreement, with an average frequency difference of only -0.43×10^{-15} and a comparison uncertainty of 0.65×10^{-15}. The Birge ratios are again close to 1. The larger and smaller data sets in rows 1 and 2 show essentially the same frequency offset between NIST-F1 and SYRTE-FO2. The uncertainty of 0.65×10^{-15} in the smaller data set is almost as good as a single, well coordinated 30-day direct comparison with no dead-time uncertainty and optimized frequency-transfer uncertainty, which would give a comparison uncertainty of about 0.63×10^{-15}. With the current fountains, even if the total Type A uncertainty could be reduced to a negligible level, the comparison uncertainty would be no smaller than about 0.5×10^{-15}, due to the Type B uncertainties.

The third row shows the results of a comparison of SYRTE-FO1 with SYRTE-FO2 using 13 data pairs from *Circular T*. Obviously such a comparison can be made in a better fashion within SYRTE, but six of the 13 pairs had exactly the same start and stop times. In this situation u_{dead} and u_{TAI} go to zero (u_{TAI} goes to zero since the two fountains are in the same location). Except for internal dead time in u_l, these six comparisons should be essentially in-house comparisons. The runs for the other seven pairs were not perfectly aligned and hence had larger uncertainties. In any case, the agreement between SYRTE-FO1 and SYRTE-FO2 is very good, with an average frequency offset of only -0.32×10^{-15} and a comparison uncertainty of 0.64×10^{-15}. Both Birge ratios are a little small, indicating that some of the uncertainties entering into the comparison may be overstated, or there may be some correlation in the bias correction errors since both standards are in the same laboratory. Although the frequency difference is within the total uncertainty, and even within U_{CB}, it is a factor of two larger than U_{CA}. Thus, it is likely that there is a systematic offset. Obviously, well conducted direct in-house comparisons of SYRTE-FO1 and SYRTE-FO2 should give more precise results than these data.

The fourth row shows a comparison of SYRTE-FO1 with NIST-F1. Here we see an offset of -0.80×10^{-15} that is slightly larger than the comparison uncertainty of 0.60×10^{-15}. This offset is consistent with the difference between SYRTE-FO1 and NIST-F1 obtained from rows 2 and 3, which is -0.75×10^{-15}. The fact that the frequency difference is slightly larger than the total comparison uncertainty is not as serious as it might seem at first. As shown in section 5, the weighted mean of the overall Cs frequency is approximately midway between the frequencies of NIST-F1 and SYRTE-FO1. Both standards are within 1 sigma of the overall mean (see figure 4 in section 5), although they are more than 1 sigma of the comparison uncertainty apart from each other. In fact, in a comparison between two standards, an offset of up to 1.4 sigma for the total comparison uncertainty should be considered reasonable. Nevertheless, these two standards have an offset that is near the limits of the stated uncertainties. Again the Birge ratios are a little small, indicating that some of the uncertainties entering into the comparison may be overstated.

4.2. All other standards versus NIST-F1 or SYRTE-FO2

Both NIST-F1 and SYRTE-FO2 were used as standards of comparison for the other fountains because they are in good agreement with each other, have low uncertainties, and have a large number of BIPM reports. Using both standards provides more usable pairs with smaller time interval offsets. Thus, table 4 shows the results of comparisons of SYRTE-FOM, IT-CsF1, PTB-CSF1, NPL-CsF1, NICT-CsF1 and NMIJ-F1 with either NIST-F1 or SYRTE-FO2. The standard which gave the smallest time offset was normally used, although an effort was made to employ both standards approximately equally. The procedures used to obtain these data were the same as those used to compare the standards in table 3. The only difference in table 4 is that the last column shows the number of pairs used along with the total number of possible data points for that particular fountain. There was not always a data point from either NIST-F1 or SYRTE-FO2 that was close enough to use, so some points were missed.

Overall, the fountains listed in table 4 show good agreement with NIST-F1 or SYRTE-FO2, with only one having a frequency offset of more than the 1 sigma. Since the frequencies of NIST-F1 and SYRTE-FO2 differ by about 4×10^{-16}, this contributes to some of the scatter and hence increases the Birge ratios. For fountain comparisons with uncertainties on the order of 1×10^{-15}, this difference increases the Birge ratios by about 20%. Four of the fountains have R_{BA} values equal to or greater than 1.5, and two of these also have R_{BC} values larger than 1.2. Even accounting for the frequency difference between NIST-F1 and SYRTE-FO2, this indicates that in some cases there is more scatter in the data than would be expected from the stated uncertainties.

The first row shows the results for IT-CsF1 using 18 of 20 possible data points, about equally divided between NIST-F1 and SYRTE-FO2. The average fountain frequency offset is quite small, smaller than both the Type A and Type B

Table 4. Comparisons of SYRTE-FOM, IT-CsF1, PTB-CSF1, NPL-CsF1, NICT-CsF1 and NMIJ-F1 with either NIST-F1 or SYRTE-FO2. y_{avg} is the average fractional frequency difference, U_C is the total comparison uncertainty, U_{CA} is the Type A comparison uncertainty, and U_{CB} is the Type B comparison uncertainty. R_{BA} and R_{BC} are, respectively, the Birge ratios calculated from the Type A or combined uncertainties. Fractional frequencies are in units of 10^{-15}.

Fountains	y_{avg}	U_C	U_{CA}	U_{CB}	R_{BA}	R_{BC}	# pairs
IT-CsF1 versus NIST-F1 or SYRTE-FO2	+0.01	0.88	0.42	0.77	1.50	1.34	18 of 20
SYRTE-FOM versus NIST-F1 or SYRTE-FO2	−0.89	0.98	0.22	0.95	1.22	0.94	19 of 22
SYRTE-FOM versus NIST-F1 or SYRTE-FO2[a]	−0.82	0.96	0.23	0.93	1.18	0.86	13 of 15
PTB-CSF1 versus NIST-F1 or SYRTE-FO2	+1.37	1.17	0.34	1.12	1.89	1.24	14 of 21
NICT-CsF1 versus NIST-F1 or SYRTE-FO2	−1.16	1.81	0.84	1.60	0.80	0.63	5 of 7
NPL-CsF1 versus NIST-F1 or SYRTE-FO2	−0.51	2.03	0.51	1.95	1.56	0.97	8 of 8
NMIJ-F1 versus NIST-F1 or SYRTE-FO2	−2.12	3.95	0.38	3.93	1.84	0.74	17 of 20

[a] After FOM rebuilt.

uncertainties. However, both the Birge ratios are significantly larger than 1, indicating more scatter than expected from the uncertainties. An important contributor to the large Birge ratios is an apparent outlier near MJD 52 740.

The second row in table 4 shows the frequency offset of SYRTE-FOM. More than 86% of the available runs could be used, with the comparison standards being about equally divided between NIST-F1 and SYRTE-FO2. The analysis shows that the average frequency offset of SYRTE-FOM is within the stated total uncertainty, but that it does have a bias about four times larger than the Type A comparison uncertainty. Thus the standard has a systematic offset, but it is still within the Type B uncertainty. Therefore, it is performing within its stated uncertainty. The Birge ratios are quite reasonable. SYRTE-FOM was rebuilt several years ago, so the third row (fountain names in red) shows the results for 13 of 15 data pairs acquired since the fountain came back on line in November 2006. Seven of the 13 points are compared against NIST-F1. The results are nearly the same as in row 1, but with a slightly smaller frequency offset. Note that the Type B uncertainty for SYRTE-FOM shown in table 1 is from its most recent report (April 2009), and that in many prior reports the Type B uncertainty was 0.9×10^{-15}. Thus, the data in table 4 reflect a larger Type B uncertainty for this standard than that shown in table 1.

The results for PTB-CSF1 are shown in row 4 with 14 out of 21 possible data points. Eight of the 14 data points are comparisons with NIST-F1. The average frequency offset is larger than the comparison uncertainty, and as shown in section 5, the standard is also a little more than 1 standard uncertainty from the weighted mean estimate of the Cs frequency. However, a few offsets outside the 1 sigma limits are to be expected. R_{BA} and R_{BC} are both larger than 1, indicating that the uncertainties are probably understated.

The results for NICT-CsF1, the newest PFS, are shown in row 5. It is behaving in a completely consistent manner although there are only a small number of points (three out of five are with SYRTE-FO2). Both the Birge ratios are low, indicating the possibility of overstated uncertainties, but with only five points the confidence in these values is low.

Row 6 shows the results for NPL-CsF1. All eight of the reports from this standard could be used, although only one aligned best with NIST-F1. These eight data points appear in two groups, with four data points occurring between MJD 53 049 and 53 329, and the remaining four between MJD 54 284 and 54 399. The frequency offset of this standard is equal to the Type A uncertainty, although it is well within its Type B uncertainty. The large R_{BA} value indicates that Type B bias correction errors are likely fluctuating.

Finally, row 7 shows the results for NMIJ-F1 with 17 out of 20 points, about equally divided between the two reference standards. This standard exhibits the largest offset, but it is within its uncertainty. R_{BA} is again relatively large, which indicates fluctuations in Type B bias correction errors. Given the large Type B uncertainty this is not surprising.

Of the 11 measurements on the nine standards in tables 3 and 4, two exhibit frequency offsets larger than the comparison uncertainty. For standard 1 sigma uncertainties this is entirely reasonable.

A weakness in using *Circular T* data is that there are significant contributions to the Type A uncertainties from dead time and frequency transfer (see table 2). Thus, many points need to be averaged to obtain a low uncertainty and, as a result, this study essentially becomes a long-term comparison. This makes it difficult to resolve changes in the frequency offset of a PFS as a function of time. Such information would be very useful and would best be observed by a series of well coordinated fountain comparisons with no dead time and optimized frequency transfer.

5. Overall estimate of the Cs frequency

Nine standards are reporting on a regular basis, and there are no significant indications from frequency offsets and Birge ratios that there are any excessively large errors in the stated uncertainties. In the few cases where the Birge ratios differed significantly from 1, some were larger than 1 and some were smaller. Therefore, there is no reason that the information from all nine fountains cannot be used to make an overall estimate of the Cs frequency.

A weighted mean of all the fountains can be calculated using NIST-F1 and SYRTE-FO2 as frequency references. For example, all the data in table 4 (excluding row 2) can be averaged to determine how these fountains compare as a group with SYRTE-FO2 and NIST-F1. The results are

$$y_{wtdavg} = -0.13 \times 10^{-15} \qquad U_{Cavg} = 0.52 \times 10^{-15}$$
$$R_B = 0.75,$$

where y_{wtdavg} is the weighted mean of the frequency differences relative to the approximate mean of NIST-F1 and SYRTE-FO2,

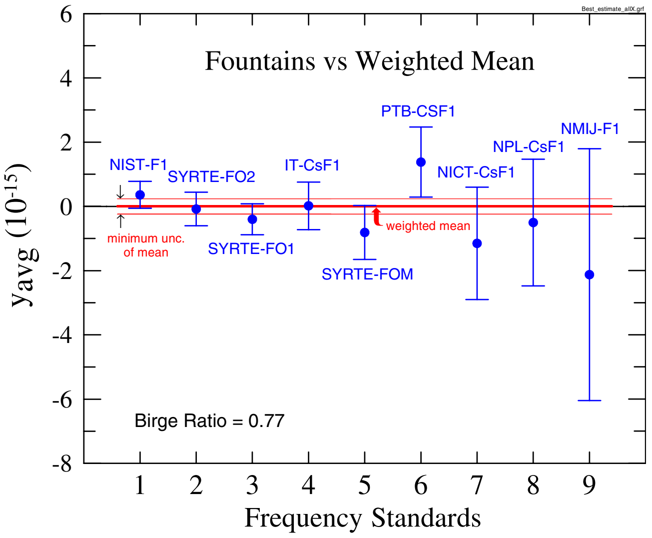

Figure 4. Average fractional frequency offsets of nine fountains relative to the combined estimate of the Cs frequency.

U_{Cavg} is the combined uncertainty and R_{B} is the Birge ratio (using U_{C} for the individual uncertainties). The combined uncertainties of the individual comparisons were treated as being uncorrelated in order to determine the value of U_{Cavg}. Note that the Birge ratio here represents the frequency scatter among a group of fountains, and not for particular fountains over time, as in tables 3 and 4. We see that these six fountains as a group agree with NIST-F1 and STRTE-FO2 at a level well within the comparison uncertainty. The small frequency offset is consistent with the assumption that the Type B bias correction errors among the standards are largely uncorrelated. The Birge ratio also shows the data to be reasonably consistent with the stated uncertainties, although they are a little on the low side (see discussion below).

An overall estimate for the Cs frequency using all nine fountains can be made by also using the results in table 3. Here we use the midpoint between NIST-F1 and SYRTE-FO2 as the working frequency reference, which is close to the reference for the data in table 4. Figure 4 shows the deviations of the individual fountain frequencies from the weighted mean obtained from the 117 comparisons of the nine standards. The overall estimate of the Cs frequency is 0.8×10^{-16} above the average frequency of SYRTE-FO2, 4.0×10^{-16} above the average frequency of SYRTE-FO1 and 3.6×10^{-16} below that of NIST-F1. The Birge ratio is equal to 0.77.

The average uncertainties for the individual fountains shown by the error bars in figure 4 were estimated from the information in tables 1, 3 and 4. For NIST-F1 and SYRTE-FO2 the comparison uncertainty of 0.67×10^{-15} in row 1 of table 3 was used as a starting point. If it were made up of equal contributions from the two standards, each standard individually would have an average uncertainty of $0.67 \times 10^{-15}/\sqrt{2} = 0.47 \times 10^{-15}$. However, NIST-F1 has a smaller Type B uncertainty, so estimated total uncertainties of 0.42×10^{-15} and 0.52×10^{-15} were used for NIST-F1 and SYRTE-FO2, respectively. The uncertainty for SYRTE-FO1 was estimated to be 0.48×10^{-15}, based on its lower Type B uncertainty in table 1 as compared with SYRTE-FO2. The uncertainties for the other standards were obtained by subtracting in quadrature the value 0.47×10^{-15} from the comparison uncertainties in table 4.

The uncertainty of the weighted mean of all the standards is difficult to calculate rigorously, but it falls in the range of 2×10^{-16} to 4×10^{-16}. The lower limit (used in figure 4) assumes that all the errors in corrected and uncorrected biases are independent (uncorrelated) between fountains, and therefore the uncertainty of the mean averages down. This assumption is undoubtedly not completely true, but determining how much correlation is present would be extremely difficult. The upper limit is that of the best single fountain.

In figure 4 it is seen that the frequency offsets of all but one fountain fall within their individual uncertainties. Given that the uncertainties are nominally 1 sigma, one might expect two or three to fall outside the uncertainties. This low number is consistent with the Birge ratio being only 0.77. There are at least two possible explanations for the low Birge ratio. One is that, on the average, the fountain uncertainties have been overstated by about 20%. This is not completely out of the realm of possibility (see the appendix). Another more likely possibility is that the assumption made in this study that the Type B uncertainties for the individual standards are correlated over time is not totally valid. There almost certainly are some random fluctuations in the Type B bias correction errors of the various fountains that would cause the Type B uncertainties of the mean to be somewhat smaller than the weighted mean of u_{B} that was used in this study. Thus, the U_{CB} values used here may be slightly overestimated. On the other hand, the Type B uncertainties are certainly not completely uncorrelated over time. If the Type B uncertainties were averaged down as if they were completely uncorrelated (as with the Type A uncertainties) the Birge ratio for the data in figure 4 would be nearly 2. This would indicate a factor of 2 understatement of the uncertainties, which is not likely. A definitive answer to the degree of correlation in Type B uncertainties over time and between standards would require a very detailed study of the Type B uncertainties of each standard and how they have changed with time. This is certainly beyond the scope of this study and may not even be possible for the data used here.

6. Summary

With nine fountains from seven laboratories reporting on a more or less regular basis, the uncertainty in TAI is now sometimes as low as 0.4×10^{-15}. The three Cs fountain PFSs with the lowest uncertainties, NIST-F1, SYRTE-FO1 and SYRTE-FO2, all agree with the overall mean at $\pm 0.4 \times 10^{-15}$ and are within their individual uncertainties of the overall Cs frequency. The comparison between NIST-F1 and SYRTE-FO2 shows a fractional frequency offset of -0.43×10^{-15}, with a comparison uncertainty of 0.65×10^{-15}. The six standards SYRTE-FOM, IT-CsF1, PTB-CSF1, NPL-CsF1, NICT-CsF1 and NMIJ-F1 have an average frequency offset of less than 0.2×10^{-15} when compared as a group with either NIST-F1 or SYRTE-FO2. Overall, the Birge ratios do not show significant inconsistencies in the stated uncertainties, although some individual standards may have modestly understated or overstated uncertainties. Only one of the nine standards has an average frequency offset from the mean that is larger than 1 sigma.

Based on the data presented in this paper it is clear that the community of Cs fountain PFSs is in a very healthy state, with nine fountains contributing to the calibration of the scale interval (rate) TAI, and the performance of the individual fountains being generally consistent with the stated uncertainties.

Acknowledgments

Very useful discussions with Steve Jefferts, Tom Heavner, David Smith and Mike Lombardi are gratefully acknowledged.

Appendix

Comments on the use of Type A and Type B uncertainties in PFSs.

Type A and B uncertainties are defined in the *Guide to the Expression of Uncertainty in Measurement* [18]. Type A uncertainties are determined by 'the statistical analysis of a series of observations'. Type B uncertainties are evaluated by 'means other than the statistical analysis of a series of observations'. In practice, this means that Type A uncertainties are determined by the evaluation of random fluctuations in the quantity being measured during the evaluation of the standard. The Type B label is applied to uncertainties that are determined by other methods, which generally means that the uncertainties are determined at some time other than during the formal evaluation. Type B uncertainties may be determined by theoretical considerations, from uncertainties in the literature or from parameters determined from measurements at some time prior to, or after, the evaluation of the standard.

For Cs fountain PFSs, Type A uncertainties inherent in the fountain are determined by a statistical analysis of the white FM noise present in the fountain during the formal evaluation. This gives the uncertainty of the average frequency of the fountain over the measurement interval. In some fountains, part, or all, of the uncertainty of the spin exchange bias correction may also be obtained from these measurements. Also, some dead-time uncertainty is introduced if a fountain does not operate 100% of the time during the evaluation interval [22]. In fountain comparisons significant Type A uncertainties may come from dead time and frequency transfer. A characteristic of Type A uncertainties, whether inherent in the PFS or originating in the comparison process, is that they average down with repeated measurements. The Type A uncertainties for fountains are always specified as 1 sigma.

The Type B uncertainties for a PFS are almost always applied to systematic bias corrections and generally come from a combination of measurements that were made outside of the time interval of the formal evaluation, and/or from theoretical considerations. For example, corrections for biases dependent on microwave power come from repeated measurements at different microwave power levels (usually at times other than during the evaluation interval) and from a theoretical understanding of the sources of these biases. Blackbody correction uncertainties are made up of a combination of temperature measurements during the evaluation and theoretical uncertainties. If an uncertainty for a particular bias is made up of a combination of measurements at the time of the evaluation (hence Type A), and other sources of information (hence Type B), it will usually be labelled as a Type B uncertainty if the Type B component is the largest.

Some biases that go into Type B uncertainties in fountains may vary over time as conditions change (intentionally or unintentionally) in the PFS, while others will be very stable. For example, the bias and uncertainty of the gravitational red shift is determined by the gravitational potential at the fountain laboratory and is a Type B bias that would not vary over time unless the PFS is moved. On the other hand, microwave leakage, which causes a bias with a power dependence, may change if physical changes to the fountain or microwave equipment are made. A consequence of this situation is that it cannot be assumed that Type B uncertainties will average down over repeated measurements, as with Type A uncertainties. On the other hand, there may be some fluctuations in Type B bias corrections. In this case the uncertainty of the mean for a Type B bias in a series of measurements may be slightly smaller than the individual uncertainties.

In the PFS community the uncertainties are conventionally stated as standard uncertainties (1 sigma). This is generally true for Type A uncertainties, although a few laboratories with only one fountain have occasionally taken a more conservative approach when the fountain stability at a few days could not be observed directly due to maser noise. The situation with Type B uncertainties is also open to question. The smaller, and less significant, uncertainties are sometimes determined theoretically in a manner such that the investigator determines that the uncertainty 'cannot be any larger than' some value. That is clearly not a 1 sigma uncertainty, but it has no practical impact for small uncertainties. The situation for the larger Type B uncertainties depends on the details of the evaluation process and the individuals investigating the uncertainties.

References

[1] Clairon A, Ghezali S, Santarelli G, Laurent Ph, Lea S, Bahoura M, Simon E, Weyers S and Szymaniec K 1996 Preliminary accuracy evaluation of a cesium fountain frequency standard *Proc. 5th Symp. on Frequency Standards and Metrology (Woods Hole, MA)* (Singapore: World Scientific) pp 49–59

[2] Parker T E, Hetzel P, Jefferts S R, Weyers S, Nelson L M, Bauch A and Levine J 2001 First comparison of remote cesium fountains *Proc. 2001 IEEE Int. Frequency Control Symp. (Seattle, WA)* pp 63–9

[3] Richard J Y *et al* 2004 Comparison of remote cesium fountains using GPS P3 and TWSTFT links *Proc. 18th European Frequency and Time Forum (Guildford, UK)*

[4] Bauch A *et al* 2006 Comparison between frequency standards in Europe and the USA at the 10^{-15} uncertainty level *Metrologia* **43** 109–20

[5] Fujieda M, Gotoh T, Piester D, Kumagai M, Weyers S, Bauch A, Wynands R and Hosokawa M 2007 First comparison of primary frequency standards between Europe and Asia *Proc. Joint Meeting of the 21st European

Frequency and Time Forum and the 2007 IEEE Int. Frequency Control Symp. (Geneva, Switzerland) pp 937–40
[6] Vian C *et al* 2005 BNM-SYRTE fountains: recent results *IEEE Trans. Instrum. Meas.* **54** 833–6
[7] Vian C *et al* 2005 Frequency comparison between two atomic fountain clocks at the 10^{-16} level *Proc. 19th European Frequency and Time Forum (Besançon, France)* pp 52–7
[8] Jefferts S R *et al* 2002 Accuracy evaluation of NIST-F1 *Metrologia* **39** 321–36
[9] Heavner T P, Jefferts S R, Donley E A, Shirley J H and Parker T E 2005 NIST-F1: recent improvements and accuracy evaluations *Metrologia* **42** 411–22
[10] Parker T E, Jefferts S R, Heavner T P and Donley E A 2005 Operation of the NIST-F1 caesium fountain primary frequency standard with a maser ensemble, including the impact of frequency transfer noise *Metrologia* **42** 423–30
[11] Weyers S, Hübner U, Schröder R, Tamm Chr and Bauch A 2001 Uncertainty evaluation of the atomic caesium fountain CSF1 of the PTB *Metrologia* **38** 343–52
[12] Weyers S, Bauch A, Schröder R and Tamm Chr 2002 The atomic caesium fountain CSF1 of PTB *Proc. 6th Symp. on Frequency Standards and Metrology (St Andrews, UK)* (Singapore: World Scientific) pp 64–71
[13] Marion H *et al* 2003 Search for variations of fundamental constants using atomic fountain clocks *Phys. Rev. Lett.* **90** 150801
[14] Levi F, Calonico D, Lorini L and Godone A 2006 IEN-CsF1 primary frequency standard at INRIM: accuracy evaluation and TAI calibrations *Metrologia* **43** 545–55
[15] Szymaniec K, Chalupczak W, Whibberley P B, Lea S N and Henderson D 2005 Evaluation of the primary frequency standard NPL-CsF1 *Metrologia* **42** 49–57
[16] Kurosu T, Fukuyama Y, Koga Y and Abe K 2004 Preliminary evaluation of the Cs atomic fountain frequency standard at NMIJ/AIST *IEEE Trans. Instrum. Meas.* **53** 466–71
[17] Kumagai M, Ito H, Kajita M and Hosokawa M 2008 Evaluation of caesium atomic fountain NICT-CsF1 *Metrologia* **45** 139–48
[18] 1993 *Guide to the Expression of Uncertainty in Measurement* 1st edn (Geneva: International Organization for Standards) ISBN 92-67-10188-9
[19] Wolf P *et al* 2006 Comparing high accuracy frequency standards via TAI *Proc. 20th European Frequency and Time Forum (Braunschweig, Germany)* pp 476–85
[20] Guinot B and Arias E F 2005 Atomic time-keeping from 1955 to the present *Metrologia* **42** S20–30
[21] Parker T E 2001 Comparing and evaluating the performance of primary frequency standards: Impact of dead time *Proc. 2001 IEEE Int. Frequency Control Symp. (Seattle, WA)* pp 57–62
[22] Yu D H, Weiss M and Parker T E 2007 Uncertainty of a frequency comparison with distributed dead time and measurement interval offset *Metrologia* **44** 91–6
[23] Parker T E 1999 Hydrogen maser ensemble performance and characterization of frequency standards *Proc. 1999 Joint Meeting of the IEEE Int. Frequency Control Symp. and the European Frequency and Time Forum (Besançon, France)* pp 173–6
[24] Parker T E and Panfilo G 2007 Experimental analysis of frequency transfer uncertainty *Proc. Joint Meeting of the 21st European Frequency and Time Forum and the 2007 IEEE Int. Frequency Control Symp. (Geneva, Switzerland)* pp 986–91
[25] Panfilo G and Parker T E 2009 A theoretical and experimental analysis of frequency transfer uncertainty, including frequency transfer into TAI *Metrologia* submitted
[26] Kacker R, Datla R and Parr A 2002 Combined results and associated uncertainty from interlaboratory evaluations based on the ISO Guide *Metrologia* **39** 279–93