# Direct comparisons of European primary and secondary frequency standards via satellite techniques

(Article begins on next page)

17 April 2024

arXiv:1910.06736v1 [physics.ins-det] 9 Oct 2019

# Direct comparisons of European primary and secondary frequency standards via satellite techniques

**F. Riedel[1], A. Al-Masoudi[1], E. Benkler[1], S. Dörscher[1], V. Gerginov[1], C. Grebing[1], S. Häfner[1], N. Huntemann[1], B. Lipphardt[1], C. Lisdat[1], E. Peik[1], D. Piester[1], C. Sanner[1], C. Tamm[1], S. Weyers[1], H. Denker[2], L. Timmen[2], C. Voigt[2], D. Calonico[3], G. Cerretto[3], G. A. Costanzo[3,4], F. Levi[3], I. Sesia[3], J. Achkar[5], J. Guéna[5], M. Abgrall[5], D. Rovera[5], B. Chupin[5], C. Shi[5], S. Bilicki[5], E. Bookjans[5], J. Lodewyck[5], R. Le Targat[5], P. Delva[5], S. Bize[5], F. N. Baynes[6], C. F. A. Baynham[6], W. Bowden[6], P. Gill[6], R. M. Godun[6], I. R. Hill[6], R. Hobson[6], J. M. Jones[6], S. A. King[6], P. B. R. Nisbet-Jones[6], A. Rolland[6], S. L. Shemar[6], P. B. Whibberley[6] and H. S. Margolis[6]**

[1]Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany
[2]Institut für Erdmessung, Leibniz Universität Hannover, 30167 Hannover, Germany
[3]Istituto Nazionale di Ricerca Metrologica INRIM, Torino, Italy
[4]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Torino, Italy
[5]LNE-SYRTE, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, 61 avenue de l'Observatoire, 75014 Paris, France
[6]National Physical Laboratory, Teddington, UK

E-mail: `erik.benkler@ptb.de`

16 October 2019

**Abstract.** We carried out a 26-day comparison of five simultaneously operated optical clocks and six atomic fountain clocks located at INRIM, LNE-SYRTE, NPL and PTB by using two satellite-based frequency comparison techniques: broadband Two-Way Satellite Time and Frequency Transfer (TWSTFT) and Global Positioning System Precise Point Positioning (GPS PPP). With an enhanced statistical analysis procedure taking into account correlations and gaps in the measurement data, combined overall uncertainties in the range of $1.8 \times 10^{-16}$ to $3.5 \times 10^{-16}$ for the optical clock comparisons were found. The comparison of the fountain clocks yields results with a maximum relative frequency difference of $6.9 \times 10^{-16}$, and combined overall uncertainties in the range of $4.8 \times 10^{-16}$ to $7.7 \times 10^{-16}$.

## 1. Introduction

Primary and secondary frequency standards find broad application in metrology, e.g. for steering time scales, and in fundamental research. Currently, the definition of the SI second is based on the ground-state hyperfine transition in the caesium-133 atom, with fountain clocks providing the best realizations with respect to accuracy and frequency instability, reaching relative uncertainties in the low $10^{-16}$ range [1–4]. Optical frequency standards, on the other hand, with relative frequency uncertainties that range down below $10^{-18}$ [5–8], are promising candidates for a new definition of the SI second [9–11]. Comparisons of frequency standards in different laboratories, both by absolute frequency measurements and by frequency ratio measurements, are necessary to provide consistency checks of the performance and accuracy of the various clocks. This is of utmost importance to ensure the continuity of time scales in the case of a redefinition. Some direct fountain clock comparisons have taken place in the past [12–16], but only a few bilateral remote comparisons of optical clocks have been performed so far [17–22].

In this work we present remote comparisons of primary and secondary frequency standards located at four European metrology laboratories: the Italian Istituto Nazionale di Ricerca Metrologica (INRIM), the French Laboratoire National de métrologie et d'Essais - SYstème de Références Temps-Espace (LNE-SYRTE), the UK National Physical Laboratory (NPL) and the German Physikalisch-Technische Bundesantalt (PTB).

For this purpose we employed different satellite-based comparison techniques: Two-Way Satellite Time and Frequency Transfer (TWSTFT), which uses geostationary satellites with transmission frequencies in the Ku-band, and time and frequency transfer via satellites of Global Navigation Satellite Systems (GNSS), the Global Positioning System (GPS) in our case. In contrast to optical-fibre-based techniques, these are applicable at the intercontinental scale and it is easier to extend the link network, because only local installations are required to add an additional station. Since these techniques are used on a daily basis for remote comparisons of clocks currently contributing to the computation of TAI/UTC (International Atomic Time and Coordinated Universal Time), all involved laboratories already had the basic required measurement equipment and infrastructure. However, temporary setup changes and refinements were needed.

The instability of TWSTFT is limited by the modulation rate of the pseudo-random code. Using a higher bandwidth for the satellite transmission compared to the routine TWSTFT measurements allows us to increase the modulation rate. As a consequence, the short-term instability at averaging times up to about 100 s is reduced, similar to an increase of the signal-to noise-ratio, and lower instabilities can also be expected for longer averaging times [23].

Here we present the results of a 26-day measurement campaign, where a simultaneous comparison of five optical clocks and six caesium or rubidium fountain clocks, located at INRIM, LNE-SYRTE, NPL and PTB was carried out by using both broadband TW-STFT and GPS PPP. This was the first time that the full available modulation rate of commercially available TWSTFT modems (20 Mchip/s) was used for optical clock comparisons. The campaign also represented the most comprehensive coordinated simultaneous operation of optical clocks to date.

Section 2 describes the technical setups of all building blocks of the experiment such as the clocks, the local hydrogen masers serving as flywheel oscillators for connecting the satellite equipment to the clocks, and the satellite links. Data analysis follows in section 3, and was a major challenge of this work, since time-deviation data from the microwave links, which was dominated by white phase noise, had to be combined with relative frequency deviaton data from the measurements between the optical clocks and the hydrogen masers, in the presence of clock and link downtimes. The non-standard analysis procedure for the comparison of the optical clocks is explained in more detail in the same section, together with a description of the data processing for the fountain clock comparisons. The results are presented and discussed in section 4, and a conclusion is given in section 5.

## 2. Experimental realization

The setup for the remote comparison of the atomic clocks via satellite is depicted in fig. 1. At each institute, a hydrogen maser (HM) was used as a continuously operating flywheel oscillator, simultaneously serving as the reference for the satellite time transfer equipment and being measured against the atomic clocks. As a consequence, fluctuations of the flywheel oscillator are cancelled out in the clock comparison to the extent given by common mode rejection as detailed below. While the frequency comparisons between the fountain clocks and the HMs were directly carried out in the radio frequency domain, the comparisons between the optical clocks and the HMs used optical frequency combs for the transfer between the optical and radio frequency domains. The satellite microwave links provided information on the phase and thus frequency difference between the HMs in the different institutes. In this constellation, each HM contributed to both the remote HM comparisons and the local measurements of clocks versus the HMs. In the ideal case these HM
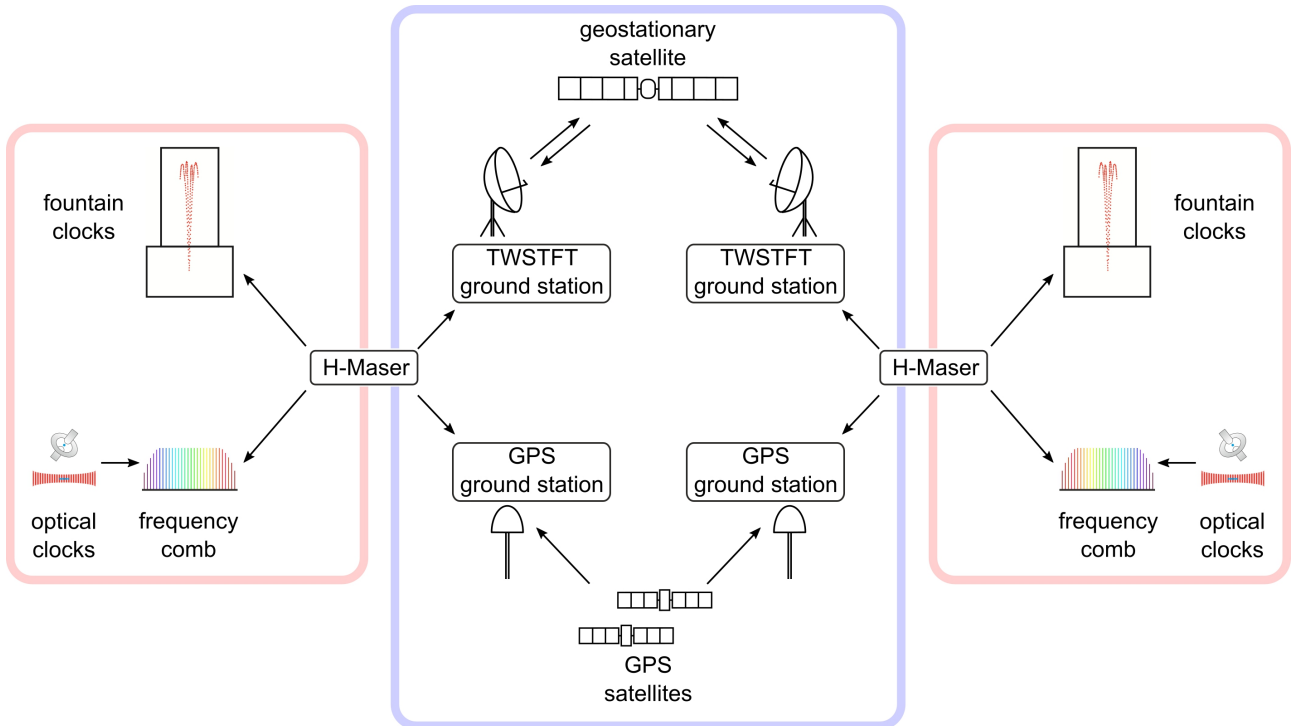
**Figure 1.** Basic setup of two remote stations. At each side, a continuously operating hydrogen maser was used as reference for both the satellite time transfer equipment and the atomic clocks. In this way, the clocks could be measured locally against the hydrogen masers (red boxes), which were remotely compared via the satellite link (blue box).

contributions would be identical and thus would cancel out completely in the remote clock comparison. In practice, the HM data were exploited over slightly different time periods in the local clock measurements and in the link comparisons in order to enable suppression of white phase noise on the satellite link data in the data processing. This will be discussed in more detail in section 3.

*2.1. Clocks operated during the campaign*

An overview of all the primary and secondary frequency standards involved is given in table 1, with typical values of their systematic uncertainties and with their uptimes during the campaign. The uptime coverage of the optical clocks refers to the whole 26-day interval, and duty cycles up to 77 % were achieved. For the combination of two optical clocks at one institute, i.e. when at least one clock out of two was operated, duty cycles of 88 % were reached. The uptime coverage of the fountain clocks refers to intervals of 12 and 16 days in length that were used for the analysis (see sect. 3).

The short- and long-term frequency instability characteristics of the hydrogen masers can be found in table 2. At LNE-SYRTE, the HM is filtered by a microwave oscillator to improve the short term stability of the local clock measurements against the HM.

The clock comparisons must account for the relativistic redshifts of the clock frequencies, which depend on the gravity (gravitational plus centrifugal) potential experienced by each clock. In this context, the spatial variations of the gravity potential are most important, while temporal variation effects (mainly due to solid Earth and ocean tides) for the clocks are below a few parts in $10^{17}$ [29] and have been neglected in this work. Two classical geodetic methods exist to determine the static (spatially varying) potential. These are the geometric levelling approach (levelling together with gravity measurements along the levelling path) and the GNSS/geoid approach, using GNSS positions (ellipsoidal heights) and a high-resolution gravimetric (quasi)geoid model. The GNSS/geoid approach was chosen because it is not affected by systematic levelling errors over long distances. An improved European gravimetric (quasi)geoid model (EGG2015), incorporating new gravity measurements around all clock sites, was utilized in the first instance for the computation of the geopotential numbers for specific existing reference markers near the clock sites ($C_{\mathrm{RefMk}}$). In addition to this, local levelling measurements were carried out to transfer the geopotential numbers from the reference markers to the actual clock locations, giving $C_{\mathrm{clock}} = C_{\mathrm{RefMk}} + g\Delta H$, where $g$ is the local gravity acceleration, and $\Delta H$ is the height difference between the clock point and

**Table 1.** Overview of the clocks that were compared and their estimated systematic uncertainties $u_{\mathrm{B}}$. In addition, the uptimes during the campaign are listed. For the optical clocks, the uptime refers to the whole duration of the campaign, whereas for the fountain clocks two different measurement intervals were analysed (see sect. 3). This is the reason why in three cases an uptime span is indicated.

| Institute | Clock | $u_{\mathrm{B}}$ | Uptime | References |
|---|---|---|---|---|
| INRIM | ITCsF2 | $2.3 \times 10^{-16}$ | 97% | [2] |
| LNE-SYRTE | Sr2 $^{87}$Sr lattice | $4.1 \times 10^{-17}$ | 68% | [20, 24] |
| | FO1 | $3.6 \times 10^{-16}$ | 71% − 75% | [3] |
| | FO2 | $2.5 \times 10^{-16}$ | 77% − 79% | [3, 25] |
| | FO2-Rb | $2.7 \times 10^{-16}$ | 80% | [3, 26] |
| NPL | $^{87}$Sr lattice | $6.8 \times 10^{-17}$ | 77% | |
| | $^{171}$Yb$^{+}$ single-ion (E3) | $1.1 \times 10^{-16}$ | 74% | [27] |
| PTB | $^{87}$Sr lattice | $1.9 \times 10^{-17}$ | 49% | [20] |
| | $^{171}$Yb$^{+}$ single-ion (E3) | $3.2 \times 10^{-18}$ | 33% | [6] |
| | CSF1 | $3.0 \times 10^{-16}$ | 98% | [4, 28] |
| | CSF2 | $3.0 \times 10^{-16}$ | 84% − 87% | [4, 28] |

**Table 2.** Overview of the hydrogen masers used in the campaign and their typical instabilities at short (1 s) and long (1 d) averaging times. The short-term instability of HM-889 at LNE-SYRTE is superior due to filtering with a microwave oscillator.

| Institute | HM | $\sigma_y(\tau = 1\,\mathrm{s})$ | $\sigma_y(\tau = 1\,\mathrm{d})$ |
|---|---|---|---|
| INRIM | HM3 | $3 \times 10^{-13}$ | $4 \times 10^{-16}$ |
| LNE-SYRTE | HM-889 | $3 \times 10^{-15}$ | $5 \times 10^{-16}$ |
| NPL | HM2 | $5 \times 10^{-13}$ | $7 \times 10^{-16}$ |
| PTB | H9 | $1 \times 10^{-13}$ | $5 \times 10^{-16}$ |

the reference marker (with $\Delta H = H_{\mathrm{clock}} - H_{\mathrm{RefMk}}$). Finally, the fractional relativistic redshift is computed according to

$$\frac{\Delta \nu}{\nu} = \frac{C_{\mathrm{clock}}}{c^2}, \tag{1}$$

where $c$ is the speed of light in vacuum. All relevant values are provided in table 3, where the underlying coordinate reference frame is ITRF2008 with the epoch 2005.0, and the zero reference potential (consistent with the IAU recommendations from the year 2000 and with the resolution 2 of the 26$^{\mathrm{th}}$ CGPM 2018) for the geopotential numbers is $W_0 = 62\,636\,856.00\,\mathrm{m}^2\mathrm{s}^{-2}$. The uncertainties of the differential redshift determinations for the distant clocks are less than $4 \times 10^{-18}$, corresponding to less than 4 cm uncertainty in height [30–32].

## 2.2. TWSTFT Setup

For the broadband TWSTFT experiment, the ground station equipment at each site comprised a very small aperture terminal (VSAT) roof station and a satellite time and ranging equipment modem (SATRE modem, TimeTech GmbH). The HM signal was directly used as the reference for the modem. As the SATRE modems are usually equipped with a maximum of two receive (RX) channels, two modems were daisy-chained to allow all three remote stations to be tracked whilst simultaneously performing a satellite ranging measurement. The modems were operating at the maximum modulation chip rate of 20 Mchip/s, requiring a total bandwidth of 36 MHz, i.e. one full Ku-band satellite transponder. For this experiment, transponder capacity on satellite ASTRA 3B at 23.5°E, operated by SES (Société Européenne des Satellites), was leased.

The principle of a TWSTFT measurement [33], i.e. the bidirectional exchange of pseudo-random code signals modulated onto a microwave carrier between two remote ground stations over a satellite, results in a compensation of the overall transit times originating from the signal paths between the stations and the satellite. However, when performing TWSTFT measurements at high resolution, various influences that cause small violations of the reciprocity of the signal paths will become significant and have to be taken into account. Since frequencies were compared and no absolute phase measurement was carried out, only the temporal variations of these effects need to be considered here.

In two-way frequency transfer, geometric terms need to be taken into account, i.e. the variation of

**Table 3.** Data used to derive the relativistic redshifts of the clock frequencies. $C_{\mathrm{RefMk}}$ is the geopotential number for a specific nearby reference marker based on the geodetic GNSS/geoid approach [30,32], $\Delta H$ is the local height difference between the clock and a specific reference marker with $\Delta H = H_{\mathrm{clock}} - H_{\mathrm{RefMk}}$, $g$ is the local acceleration due to gravity derived from an FG5-X absolute gravimeter measurement (with an uncertainty far below the significant digits quoted here), and $C_{\mathrm{clock}}$ is the final geopotential number for the clock point. The underlying coordinate reference frame is ITRF2008 with the epoch 2005.0, and the zero reference potential for the geopotential numbers is $W_0 = 62\,636\,856.00\,\mathrm{m^2s^{-2}}$. The uncertainties for $\Delta H$ and for the geopotential numbers $C_{\mathrm{clock}}$ are given in parentheses, the uncertainty of $C_{\mathrm{RefMk}}$ is 0.22 $\mathrm{m^2s^{-2}}$.

| Institute | Clock | Reference marker ID | Geopotential number for reference marker $C_{\mathrm{RefMk}}$ [$\mathrm{m^2s^{-2}}$] | Height difference $\Delta H$ [m] | Local gravity acceleration $g$ [$\mathrm{ms^{-2}}$] | Geopotential number for clock $C_{\mathrm{clock}}$ [$\mathrm{m^2s^{-2}}$] | Redshift correction $\frac{\Delta\nu}{\nu}$[$10^{-15}$] |
|---|---|---|---|---|---|---|---|
| INRIM | ITCsF2 | CS104 | 2323.32 | 1.020(10) | 9.8053 | 2333.32(24) | -25.9617(27) |
| LNE-SYRTE | Sr2 | SR2 | 545.06 | 0.201(5) | 9.8093 | 547.03(23) | -6.0865(25) |
| | FO1 | FO1 | 613.42 | 0.761(10) | | 620.89(24) | -6.9083(27) |
| | FO2-Cs | FO2 | 579.63 | 0.962(10) | | 589.06(24) | -6.5542(27) |
| | FO2-Rb | FO2 | 579.63 | 0.886(10) | | 588.32(24) | -6.5459(27) |
| NPL | $^{87}$Sr | G4L10 | 96.54 | 1.290(10) | 9.8118 | 109.20(24) | -1.2150(27) |
| | $^{171}$Yb$^+$ E3 | G4L16 | 96.58 | 1.081(5) | | 107.19(23) | -1.1926(25) |
| PTB | $^{87}$Sr | PB02 | 763.84 | 0.538(3) | 9.8125 | 769.12(22) | -8.5576(25) |
| | $^{171}$Yb$^+$ E3 | KB02 | 753.04 | 1.001(3) | | 762.86(22) | -8.4880(25) |
| | CSF1 | KB02 | 753.04 | 1.632(5) | | 769.05(23) | -8.5569(25) |
| | CSF2 | KB02 | 753.04 | 1.526(5) | | 768.01(23) | -8.5453(25) |

the Sagnac effect and the variation of the path delay difference, while the gravitational terms cancel [34]. Both geometric terms arise from the residual motion of the satellite around its nominal geostationary position, a daily oscillation with a maximum amplitude of about 30 km. In order to compensate for the variations of the path delay difference, delays with respect to the local UTC($k$) time scales were introduced to the reference 1 pulse per second (PPS) signals to ensure all signals arrive at about the same time at the satellite. The variations of the Sagnac effect and the residual variations of the path delay difference were then calculated using satellite position and velocity data provided by SES. This results in peak-to-peak relative frequency corrections between $8 \times 10^{-16}$ (LNE-SYRTE – INRIM) and $2 \times 10^{-15}$ (PTB – NPL), with the Sagnac term being dominant over the residual path delay difference contribution.

Another influence on the signal path delay is the dispersion of the atmosphere. The tropospheric contribution can be neglected because of its low frequency dependence and the short signal path, due to the high elevation of the satellite [33, 35]. For the ionosphere, the frequency dependence of the signal transit time is larger, and a correction can be calculated based on the total electron content (TEC) of the ionosphere [33]. For the calculation, the values provided in [36] were used, and the projection mapping function as described in [37] was applied to estimate the extension of the signal path due to the slant propagation path through the ionosphere. The peak-

to-peak relative frequency corrections due to variation of the difference of ionospheric delays were about $1.3 \times 10^{-15}$ for the link between PTB and INRIM (maximum influence of the ionosphere) and $4 \times 10^{-16}$ for the link LNE-SYRTE – NPL (minimum influence of the ionosphere).

### 2.3. GPS Setup

As for the TWSTFT ground stations, the GPS ground station equipment regularly employed for contributions to TAI was used, consisting of a GNSS receiver and a GNSS choke ring antenna connected by a coaxial cable with a length between 30 m and 50 m. The receivers were Javad Legacy at INRIM, Septentrio PolaRx4-TR at LNE-SYRTE, DICOM GTR50 at NPL, and Septentrio PolaRx4-TR at PTB. Like the TWSTFT modems, the GPS receivers were directly referenced to the hydrogen masers. All receivers provided data in the RINEX format (versions 2.10 and 2.11), enabling PPP processing. The PPP algorithm by the National Resources Canada (NRCan) [38] was used for the computation of GPS link phase data. It takes into account various effects such as the propagation through the troposphere (based on models and meteorological data) and ionosphere, antenna phase centre variations [39], carrier-phase windup [40], relativistic effects [41], and site displacements, e.g. due to solid Earth tides.

## 3. Clock comparison data analysis

As usual in time and frequency metrology, we denote the fractional frequency offset from the nominal frequency as "relative frequency deviation" (symbol $y$) and use the symbol $x$ for the time deviation associated with the phase [42]. The calculation of the relative frequency deviation difference between two remote atomic clocks located at laboratories 1 and 2, respectively, involves two frequency data sets (hydrogen maser (HM) against atomic clock (AC) at each institute, called "clock data" in this paper) and one phase data set for the link (remote comparison of hydrogen masers, "link data" in this paper):

$$y^{c1}(t) = y^{AC1}(t) - y^{HM1}(t) \qquad (2)$$
$$y^{c2}(t) = y^{AC2}(t) - y^{HM2}(t) \qquad (3)$$
$$x^{link}(t) = x^{HM1}(t) - x^{HM2}(t). \qquad (4)$$

If the relative frequency deviation data sets $y(t)$ are represented by a uniform time series $y_i = y(t_i)$ with sampling times $t_i$ on an equispaced time grid with steps $\Delta t$, and the phase data set $x^{link}(t)$ is represented by a corresponding time series $x_i^{link} = x^{link}(t_i)$, it can be converted into a relative frequency deviation time series:

$$
\begin{aligned}
y_i^{link} &= \frac{x_{i+1}^{link} - x_i^{link}}{\Delta t} \\
&= y_i^{HM1} - y_i^{HM2}.
\end{aligned} \qquad (5)
$$

The time series of relative frequency deviation differences between the remote clocks can be calculated as:

$$
\begin{aligned}
y_i^{AC1-AC2} &= y_i^{AC1} - y_i^{AC2} \\
&= y_i^{c1} - y_i^{c2} + y_i^{link},
\end{aligned} \qquad (6)
$$

from which we could directly compute the mean value and its statistical uncertainty.

However, the measurement data in practice have different start and end times, and have gaps at different times and of different durations (as seen in fig. 2 for the gaps in the optical clock and the TWSTFT data during the campaign). They are thus not straightforwardly representable by uniform time series as assumed above. In addition they have different noise properties: the clock data are dominated by the noise of the HMs, whereas the link data are dominated by the satellite link noise up to averaging times of about 0.5 d, as can be seen in the modified Allan deviations of the TWSTFT link data (see fig. 3). Only at longer averaging times does the HM noise prevail together with other coloured noise contributions on the link, most probably caused by environmental influences such as temperature variations.

Different boundary conditions are given in the cases of optical and caesium fountain clocks. For the optical clocks, the data are available on a 1 s grid, and they contain a relatively large number of gaps. The uncertainties of the optical clocks are known from local comparisons to range below $10^{-16}$, so in this case, it is necessary to suppress all disturbances which can deteriorate the remote clock comparison uncertainty.

The fountain clock comparisons were operated without any major interruptions, but the concept of operation provides frequency values at intervals which do not in general correspond to integer seconds. Furthermore, the statistical uncertainty of the fountain clocks for averaging times of 1 d is in the few $10^{-16}$ range.

For these reasons we applied different analysis procedures to the optical clock comparisons and to the fountain clock comparisons, although both methods utilize some kind of pre-averaging of the link phase data in order to suppress the link noise.

### 3.1. Determination of the relative frequency deviation difference between pairs of optical clocks (OC)

If we were to naively apply eqs. 2 – 6, the instability and the resulting statistical uncertainty would be limited by the unsuppressed white phase noise of the link and possibly by a residual contribution from HM fluctuations in the presence of gaps in the phase and frequency datasets, to a level much larger than $10^{-16}$. In fact, due to the white phase noise, the instability would not be limited to approximately $5 \times 10^{-10}(\tau/\text{s})^{-1}$ as in fig. 3, but rather to $5 \times 10^{-10}(\tau/\text{s})^{-1/2}$, which would render the remote optical clock comparisons useless. Hence, the link data should be pre-averaged to suppress the link noise, and the different data sets should be combined in a way that suppresses the HM fluctuations. Such an analysis procedure is described in the following.

According to the *Guide to the expression of Uncertainty in Measurement* (GUM) [43], the standard deviation $\sigma_{\bar{y}}$ of the mean describes the type-A uncertainty of the mean. The choice of the estimator for $\sigma_{\bar{y}}$ depends on knowledge of the noise types involved, or in other words of serial correlations in the relative frequency deviation data. The commonly used estimator

$$\hat{\sigma}_{\bar{y}} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N(N-1)}} \qquad (7)$$

ignores correlations and is therefore only unbiased in the case of white frequency noise. If other noise contributions are present it can significantly underestimate the standard deviation of the mean.
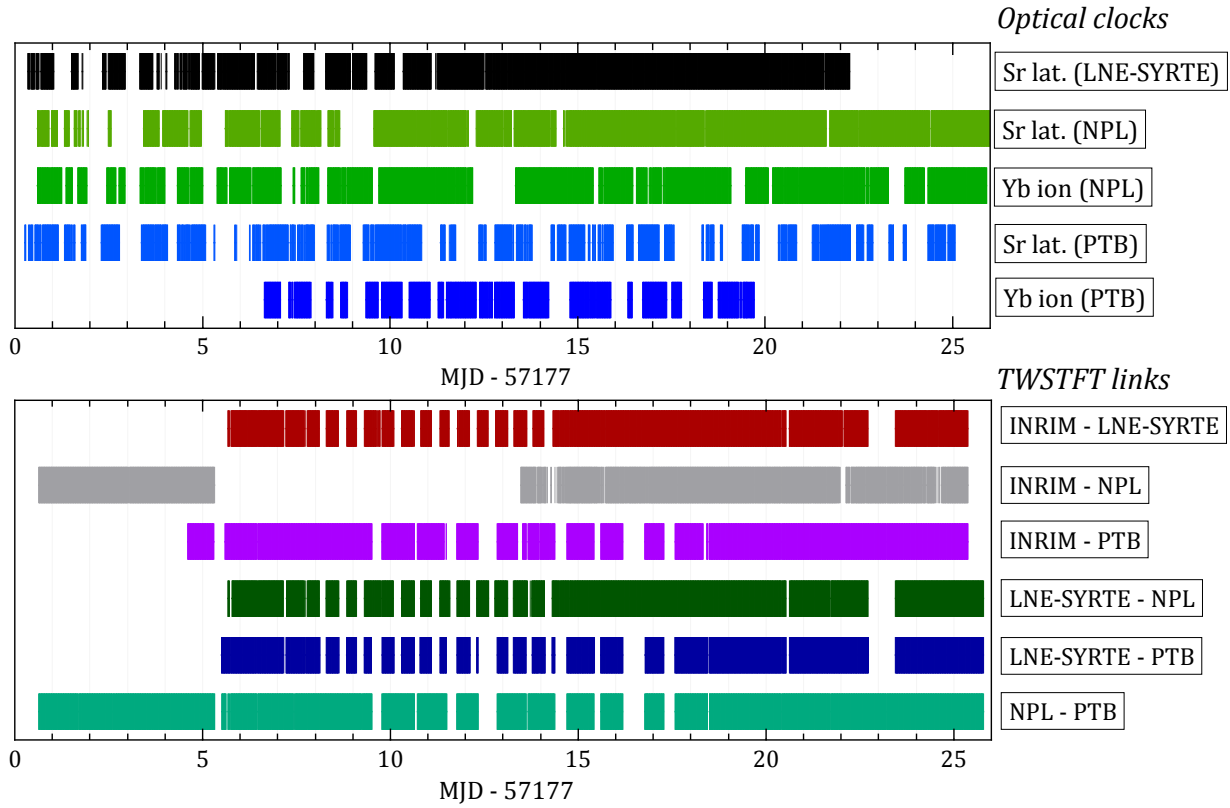
**Figure 2.** Measurement times during the campaign of all optical clocks involved (above) and the TWSTFT links (below). Periods with technical disturbances resulting in outliers have been excluded. The GPS links were continuously operated, but the intervals used for the clock comparisons were limited by discontinuities arising from technical disturbances. Hence, data from the GPS receivers at INRIM and LNE-SYRTE were only available for the intervals 57177.0 − 57198.0 and 57183.0 − 57197.0. The numbers refer to to Modified Julian Date (MJD), which is based on a continuous count of days, in which 57177.0 corresponds to 4th June 2015, 00:00:00 UTC.
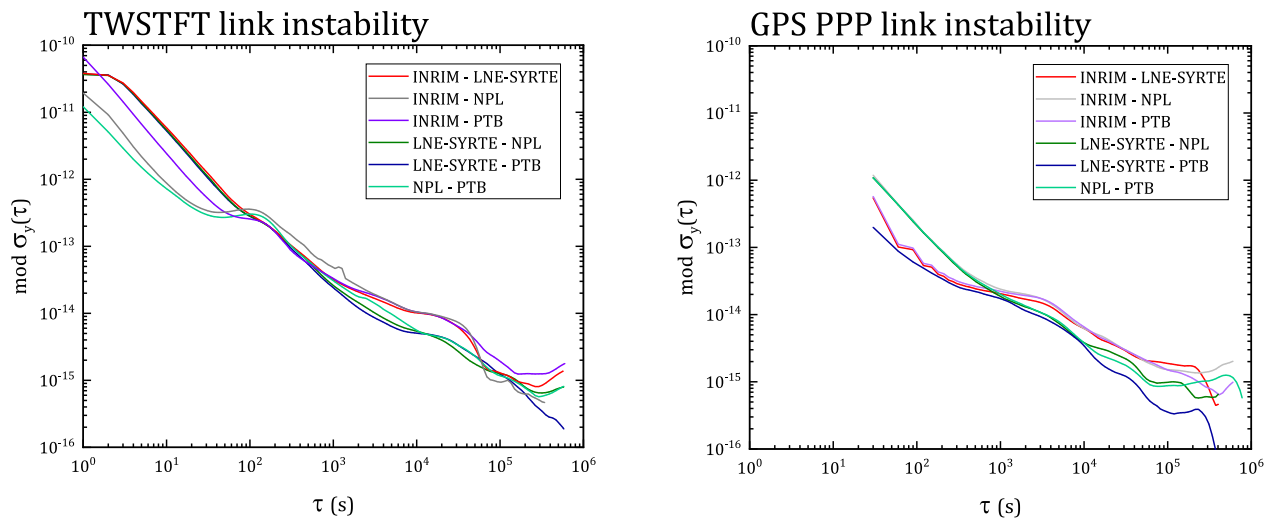


**Figure 3.** Modified Allan deviation of all link data, calculated over the whole measurement period. For averaging times below 1 day, the modified Allan deviation reflects the frequency instability of the links. Above 1 day, the modified Allan deviations are dominated by the HM fluctuations.

In the optical clock comparisons via satellite link we have a mixture of coloured noise processes, i.e. rather intricate correlations. Furthermore, the relative frequency deviation data contain gaps. Therefore, we derive an alternative approach for an estimator of $\sigma_{\bar{y}}$ (see appendix), which is based on ideas from [44–46]. This takes into account both the correlations and the data gaps. The correlations are included via autocovariances, which can themselves be estimated from the relative frequency deviation data. The gaps are handled by allowing for weighting of the relative frequency deviation data and setting the weighting factors to zero at gaps. For this purpose, the relative frequency deviation data are represented by a time series $y_i$ on a uniform time grid and with weighting factors $w_i$, with the index $i = 1 \ldots N$, where $N$ is the total number of elements of the time series. As derived in the appendix, an estimator for the standard deviation of the weighted mean is:

$$\hat{\sigma}_{\bar{y}}^2 = \sum_{i=1}^{N} w_i (y_i - \bar{y})^2 + 2 \sum_{l=1}^{l_{\text{cut}}} \hat{R}_l \sum_{j=1}^{N-l} \sqrt{w_j w_{l+j}}, \qquad (8)$$

where $\hat{R}_l$ is the estimator of the lag-$l$ autocovariance given by equation 17 in the appendix, and $l_{\text{cut}}$ is a cutoff index. The latter is introduced because the autocovariance can only be calculated with sufficient accuracy for lower lags. At higher lags the coefficients become erroneous [46].

With this tool we can now develop an analysis strategy for the optical clock comparisons. The determination of the relative frequency deviation difference from the three time series $x_i^{\text{link}}$, $y_i^{\text{OC1}}$ and $y_i^{\text{OC2}}$ for the example optical clock comparison $^{171}$Yb$^+$(NPL) versus $^{171}$Yb$^+$(PTB) is illustrated by figure 4.

There are two opposing requirements which should be satisfied, but which can both degrade the statistical uncertainty. Hence, we have to find a compromise which minimizes the statistical uncertainty. On the one hand, we should combine clock- and link-data only at times at which they overlap, in order to achieve the best common mode rejection of the HM fluctuations. On the other hand, the phase data should be pre-averaged, and with regard to the uncertainty determination, a time series with a small time grid is desirable in order to estimate the autocovariance reliably.

We apply a sliding arithmetic mean to the phase data and convert the pre-averaged phase data into frequency data by using eq. 5 with $\Delta t = 1\,$s. As a negative consequence, HM fluctuations during gaps in the data are partially folded to adjacent points of time, at which the link frequency data is subsequently combined with the clock data. Hence the common mode rejection of the HM fluctuations is not perfect

any more. For the sliding average of the link phase data, an averaging window length of 1 day is chosen as a compromise between the two opposing aspects: diurnal oscillations as well as other disturbances present on the link phase data at averaging times between 0.5 and 1 d are suppressed along with the white and flicker phase noise, without introducing systematic errors, while the HM fluctuations at up to 1 d are still tolerable.

Three details should be noted in connection with the conversion of the pre-averaged phase data to frequency data, especially when there are gaps in the phase data:

First, we use a time step of $\Delta t = 1\,$s for the conversion with eq. 5. Larger intervals would introduce artificial correlations at larger lags $l > l_{\text{cut}}$, which could ultimately yield a biased statistical uncertainty. At gaps, however, the introduction of such correlations cannot be totally avoided. In order to minimize their influence, we cut the original gaps of the raw link phase data into the pre-averaged data, which also reduces the negative influence on the common mode rejection of the HM fluctuations discussed above.

Second, if the data are asymmetrically distributed within the phase data averaging window due to gaps, the effective point of time at which the phase average is valid is not in the centre of the averaging window, but instead at the temporal centre of gravity. For this reason, the temporal centres of gravity rounded to values on the time series grid are used instead of the original times for the conversion to frequency values.

Third, the number $n_i$ of phase values $x_i$ falling within the averaging window varies near gaps, and larger weights should be assigned to frequency values derived from a larger number of phase values. Hence, we derive semi-heuristic weights for the link frequency data from the time series of the $n_i$:

$$w_i^{\text{link}} = \frac{1}{\frac{1}{n_i} + \frac{1}{n_{i+1}}} = \frac{n_i n_{i+1}}{n_i + n_{i+1}}. \qquad (9)$$

These weights are not yet normalized to unity, but can be interpreted as an effective number of frequency values in the 1 day phase averaging window. This is motivated by the fact that each frequency value $y_i^{\text{link}}$ on the 1 s grid is determined from the difference between the 1 day averages $\bar{x}_i$ and $\bar{x}_{i+1}$ over $n_i$ and $n_{i+1}$ phase values, respectively, divided by the difference between the time centres of gravity. In fig. 4 (a), the original phase data and the averaged phase data are shown for a two-day interval. The resulting relative frequency deviation and the associated weights $w_i^{\text{link}}$ are visualized in fig. 4 (b).

We should also avoid the introduction of artificial correlations on the clock data, so we do not pre-average $y^{\text{c1}}$ and $y^{\text{c2}}$. However, when taking only the data for which $y^{\text{link}}$, $y^{\text{c1}}$ and $y^{\text{c2}}$ overlap, we would discard
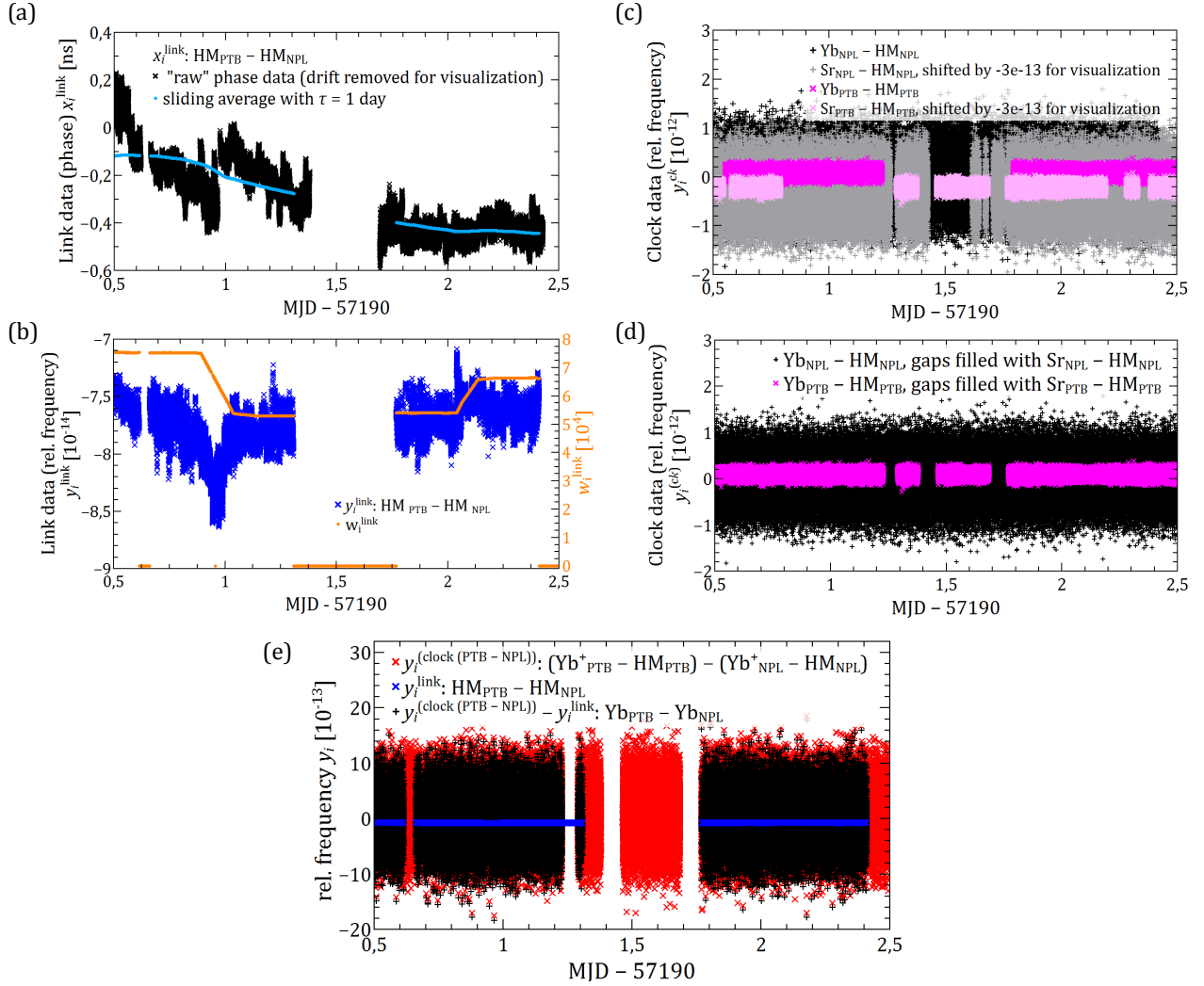
**Figure 4.** Scheme of the data analysis procedure, shown for the comparison of Yb$^+$(PTB) and Yb$^+$(NPL) via TWSTFT for two days of the campaign. (a) Phase data of the link between the hydrogen masers at PTB and NPL, with the phase drift (frequency offset) over the whole campaign interval removed for better visualization. Black symbols: Measured data (outliers removed), blue dots: Moving average with a window of 1 day. The time points of the pre-averaged phase data are different from those of the measured phase data (centres of gravity as described in the text). For this reason, the gap edges are different for the black and blue curve. (b) Blue symbols: Relative frequency deviation calculated from the averaged phase data. Orange: Weights $w_i^{\text{link}}$. (c) Locally measured differences between the hydrogen maser and the optical clocks (pink: PTB, black/grey: NPL). Dark symbols: Yb$^+$ clocks, light symbols: Sr clocks. The NPL data show increased noise in comparison to the PTB data because of the different short-term stability of the hydrogen masers (see table 2). The Sr clock data are used to fill the gaps in the Yb$^+$ data at the same institute, which results in the graphs shown in (d). (e) The final step is depicted. The two time series of (d) are merged into one $y_i^{\text{clock(PTB−NPL)}}$ (red), and in blue the frequency difference of the hydrogen masers (blue graph from (b)) is shown. Both time series are subtracted in order to get the frequency difference of the two Yb$^+$ clocks (black symbols).

a lot of information. Hence, if an institute operates two clocks (say A and B), we fill the gaps in the data of clock A, whenever clock B is running while clock A is not and vice versa. For this purpose, we use the ratio between clock A and B measured with small uncertainty during the periods when A and B are operative during the campaign ( 4 (c) and (d)). This is justified, because the uncertainty of the local ratio is negligible with respect to the uncertainty of the remote comparison. This approach has been applied before in

an optical clock comparison between NICT in Japan and PTB in Germany [18]. Just as for the link data, weights, not yet normalised, are assigned to each clock data frequency value:

$$w_i^{y^c} = \begin{cases} 0 & \text{at gaps} \\ 1 & \text{elsewhere.} \end{cases} \tag{10}$$

In the next step, the three resulting frequency data time series are merged according to eq. 6. An overall weight was assigned to each frequency value, given by the product of the individual weighting factors and

normalized to unity:

$$w_i = \frac{w_i^{y_1} w_i^{y_2} w_i^{\text{link}}}{\sum_{i=1}^{N} w_i^{y_1} w_i^{y_2} w_i^{\text{link}}}. \qquad (11)$$

The time series $y_i^{\text{link}}$, $(y_i^{\text{c1}} - y_i^{\text{c2}})$ and $y_i^{\text{OC1}-\text{OC2}}$ are depicted in fig. 4 (e). From this, a weighted mean of the relative frequency deviation difference between the two optical clocks located at laboratories 1 and 2 $\bar{y}^{\text{OC1}-\text{OC2}}$ was calculated for the overall campaign.

In the case of GPS PPP data, the time series of the link data is on a $\Delta t = 30\,\text{s}$ grid. The procedure was adapted accordingly.

*Determination of uncertainties.* The statistical uncertainty of the relative frequency deviation difference between two optical clocks is calculated as described above with eq. 8. We chose a cutoff index $l_{\text{cut}} = 4000$. Above this lag, no significantly non-zero values of the autocovariance estimator were observed.

As known from local optical clock comparisons, the contribution of the optical clocks to the statistical uncertainty is well below $10^{-16}$ and thus negligible in the remote comparison.

For the systematic uncertainty, the considered contributions originate from both the clocks and the link. To determine the systematic uncertainty of the satellite links, various influences were investigated, with the temperature variations both inside the laboratory and outdoor near the antennas being the dominant contributions. During the campaign the relevant temperatures were monitored at all stations. To estimate possible frequency errors due to the observed temperature changes, phase deviations $x_{\text{T,i}}$ were deduced from the recorded temperature data and temperature sensitivities $c_{\text{T,i}}$ for individual components $i$ of the link equipment. With these phase data, the same analysis process was performed as for the raw link phase data, resulting in the component-related temperature induced frequency shift $\bar{y}_{\text{T,i}}$. Because of the uncertainty in the temperatures and sensitivities of the various components, we quadratically add all individually inferred frequency shifts and use the result as estimate for the systematic uncertainty of the link. For the TWSTFT links, measurements of the temperature coefficients for both indoor and outdoor equipment were performed before the campaign. Since not all components of every base station could be tested, the largest temperature sensitivities observed for the class of components were used for a conservative estimation: $c_{\text{T}} = 5\,\text{ps/K}$ for the SATRE modems and $c_{\text{T}} = 5\,\text{ps/K}$ for the combination of frequency converters and amplifier.

For the GPS measurement analysis, temperature coefficients for the main components (receiver, antenna and antenna cable) were taken from the literature for the analysis [47–49]: $9\,\text{ps/K}$ for the PolaRX4 receivers, $13\,\text{ps/K}$ for the GTR50, $0.03\,\text{ps/K·m}$ for the antenna cables and $10\,\text{ps/K}$ for the antennas.

Details about the systematic uncertainties of the individual optical and fountain clocks can be found in the references in table 1.

### 3.2. Determination of the relative frequency deviation difference between pairs of fountain clocks

As mentioned above, the situation is different for the fountain clock comparisons. The high duty cycles of the fountain clocks allow for the approach of determining average frequency values for each time series separately (eq. 2 to 4). So, we first identified two appropriate measurement intervals common for fountain clocks and links of at least two of the partners INRIM, LNE-SYRTE and PTB. These intervals chosen for analysis are:

- MJD 57183.0 – MJD 57199.0 (16 d) for broadband TWSTFT for all three links between INRIM, LNE-SYRTE, and PTB;
- MJD 57184.0 – MJD 57196.0 (12 d) for GPS PPP on all three links.

Subsequently, the individually measured frequency differences between a fountain and the local hydrogen maser were averaged for such intervals. To evaluate the mean frequency difference between the remote hydrogen masers $\bar{y}^{\text{link}}$, we calculated the average phase values for the TWSTFT or GPS link over 24 hours at the beginning and at the end of the respective interval, centred around 0:00 UTC of the first/last day of the respective comparison interval. We then obtained the mean frequency difference $\bar{y}^{\text{link}}$ by dividing the difference between the two phase averages by the duration of the interval ($\Delta t_{\text{FC}}$, with FC = fountain clock comparison). By this means, the influence of white phase noise and of the other disturbances like diurnals on the link was reduced. For a frequency averaging interval length of 1 day, this processing is equivalent to $\Lambda$-averaging, but for longer frequency averaging intervals, the frequency weighting function has a trapezoidal shape instead of the $\Lambda$-shape.

The averaged frequency values $\bar{y}^{\text{link}}$, $\bar{y}^{\text{c1}}$ and $\bar{y}^{\text{c2}}$ of the three time series were merged into one relative frequency deviation difference between the remote fountain clocks in analogy to eq. 6.

*Determination of uncertainties* The determination of the statistical uncertainties of the fountain clocks is based on the measured short-term instability and the known averaging behaviour (white frequency noise) for longer measurement times. The systematic uncertainties have already been presented in section 2.1, table 1.

In addition, a dead-time uncertainty was taken into account, as generally found in Circular T [50].

The uncertainty evaluation for the link during the frequency averaging interval required several steps. The fluctuations of the phase data during the first and last day of the frequency averaging interval carry information about the short term ($\leq 1\,\mathrm{d}$) link instability. Together with knowledge about the relation between the instability and statistical uncertainty $u_\mathrm{A}$ [51], we can thus derive the uncertainty for a frequency averaging interval length of 1 day. As mentioned before, for a 1 day frequency averaging interval the frequencies are $\Lambda$-weighted, so the uncertainty at 1 day is given by the modified Allan deviation at 1 day, multiplied by a form factor of $\sqrt{2/3}$ for the link data dominated by white phase noise:

$$u_\mathrm{A}(1\,\mathrm{d}) = \sqrt{2/3}\,\mathrm{mod}\,\sigma_y(\tau = 1\,\mathrm{d}). \tag{12}$$

However, at the longer averaging intervals used for the fountain comparisons, we cannot derive such information about the pure link instability solely from the modified Allan deviation estimated from the link data: First, the modified Allan deviation is dominated by the HM fluctuations at $\tau > 1\,\mathrm{d}$. Second, the modified Allan deviation can usually be estimated only at averaging times shorter than the total length of the measurement interval. However, we can instead make use of the statistical uncertainties $u_{\mathrm{A,\,OC}}$ determined in the optical clock comparisons on intervals of length $\Delta t_\mathrm{OC}$ longer than $\Delta t_\mathrm{FC}$, because there the influence of the optical clocks is negligible and $u_{\mathrm{A,\,OC}}$ is marginally affected by the HMs. Thus, we can interpolate linearly between $u_\mathrm{A}(1\,\mathrm{d})$ derived using eq. 12 and $u_{\mathrm{A,\,OC}}(\Delta t_\mathrm{OC})$, in order to obtain the link uncertainty $u_{\mathrm{A,\,FC}}(\Delta t_\mathrm{FC})$ at the frequency averaging interval length used for the fountain clock comparison. For the values of $u_{\mathrm{A,\,OC}}$ and $\Delta t_\mathrm{OC}$, see tables 4 and 5 in the next section.

This procedure was applied directly for the link LNE-SYRTE – PTB. Without optical clock data at INRIM, however, for the INRIM-related links we used the typical uncertainties $u_{\mathrm{A,\,OC}}(\Delta t_\mathrm{OC})$ observed with the other links.

The systematic uncertainty of the link was determined in the same way and with the same parameters as for the optical clock comparisons. To complete the information on the temperature coefficients of the GPS equipment, the values used for INRIM were $13\,\mathrm{ps/K}$ for the receiver, $0.5\,\mathrm{ps/K\cdot m}$ for the antenna cable and $10\,\mathrm{ps/K}$ for the antenna.

## 4. Results and discussion

### 4.1. Comparison of optical clocks

In table 4, the results for the relative frequency deviation differences between the optical clocks of the same type are listed, with the statistical and systematic uncertainty contributions. For a graphical visualization of clock comparison results with clocks located at three institutes, we introduce triangle plots as shown in fig. 5 for clock pairs of the same species. The intention behind such triangle plots is that some important features of a triangle clock comparison might be grasped more intuitively than in the form of a mere table or of a Cartesian scatter plot. For example, it is easier to identify a clock whose frequency deviates from the frequency of the other two clocks. For this purpose, each corner of the triangle corresponds to one institute. Between the corners, there are relative frequency deviation axes, with the relative frequency deviation difference $y = 0$ in the centre, which for clock pairs of the same species corresponds to a frequency ratio of 1 between the clocks located at the institutes indicated on the respective corners. If the relative frequency deviation difference differs from 0 (or the frequency ratio deviates from 1), we use the following convention: For $y^{(\mathrm{c1-c2})} > 0$, the data point is drawn on the axis in the direction of the corner where clock c1 is located. The axes range from zero in the centre to $1 \times 10^{-15}$ at each of the two corners. If we consider for example the TWSTFT result for the clock pair $\mathrm{Sr(NPL)} - \mathrm{Sr(PTB)}$, with $\bar{y}^{(\mathrm{Sr(NPL)-Sr(PTB)})} = -2.9 \times 10^{-16}$, the according data point is drawn at $2.9 \times 10^{-16}$ on the NPL – PTB axis in the direction of the "PTB" corner.

For clock pairs of different species, similar considerations hold true, but zero relative frequency deviation difference means that the frequency ratio is referenced to the ratio resulting from CIPM recommended values.

First, we discuss the results with respect to the differences between TWSTFT and GPS PPP satellite links.

All combined comparison uncertainties are equal or smaller than $3.5 \times 10^{-16}$. The GPS PPP links yield smaller statistical uncertainties than broadband TWSTFT. This is also true for the combined uncertainties, which are dominated by the statistical uncertainties. The smallest combined uncertainty in the case of the GPS PPP link NPL – PTB is $1.8 \times 10^{-16}$, which is smaller than for the according TWSTFT link by almost a factor of 2. The TWSTFT comparisons, however, yield slightly smaller systematic uncertainties. Both satellite link techniques lead to results that are compatible within a confidence level of 68 %.

**Table 4.** Results for the comparisons of optical clocks of the same type. The third column gives the average optical frequency difference, $u_{\mathrm{B,c}}$ is the combined systematic clock uncertainty, $u_{\mathrm{A,l}}$ and $u_{\mathrm{B,l}}$ correspond to the statistical and systematic link uncertainties, $u$ is the combined uncertainty, and $\Delta t_{\mathrm{OC}}$ corresponds to the effective length of the optical clock comparison. All values except for the last column $\Delta t_{\mathrm{OC}}$ are in $10^{-16}$.

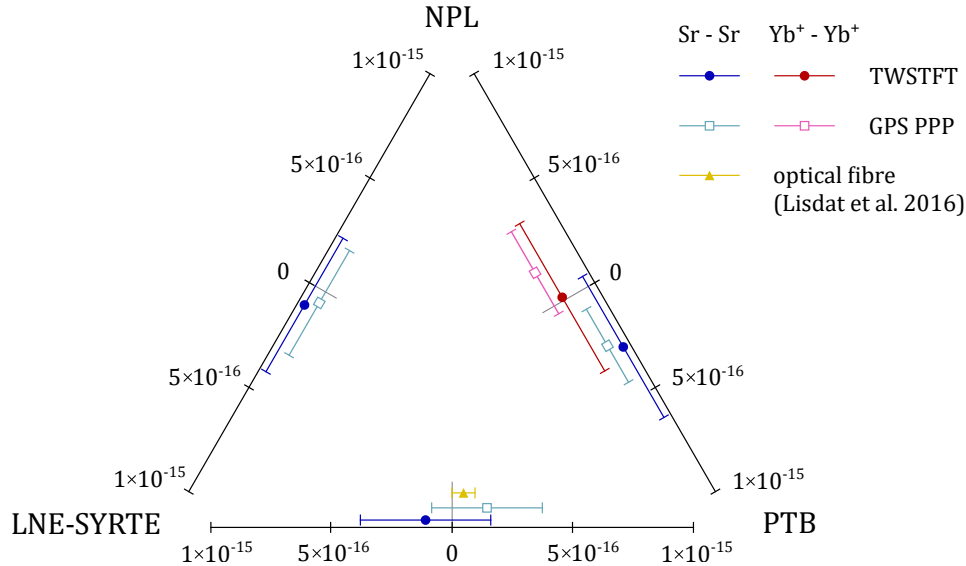| Clock pair | link | difference | $u_{\mathrm{B,c}}$ | $u_{\mathrm{A,l}}$ | $u_{\mathrm{B,l}}$ | $u$ | $\Delta t_{\mathrm{OC}}$ [d] |
|---|---|---|---|---|---|---|---|
| Sr(LNE-SYRTE) | TWSTFT | 0.9 | | 3.0 | 0.7 | 3.2 | 16.6 |
| – Sr(NPL) | GPS PPP | 0.5 | 0.8 | 2.3 | 0.8 | 2.5 | 13.5 |
| Sr(LNE-SYRTE) | TWSTFT | 1.1 | | 2.5 | 0.9 | 2.7 | 16.4 |
| – Sr(PTB) | GPS PPP | -1.4 | 0.5 | 1.9 | 1.2 | 2.3 | 13.5 |
| Sr(NPL) | TWSTFT | -2.9 | | 3.3 | 0.5 | 3.4 | 24.4 |
| – Sr(PTB) | GPS PPP | -2.5 | 0.7 | 1.5 | 0.6 | 1.8 | 24.5 |
| Yb$^+$(NPL) | TWSTFT | 0.2 | | 3.3 | 0.5 | 3.5 | 24.4 |
| – Yb$^+$(PTB) | GPS PPP | 1.6 | 1.1 | 1.5 | 0.6 | 2.0 | 24.5 |



**Figure 5.** The results of the relative frequency deviation differences for the optical clock comparisons between the three institutes LNE-SYRTE, NPL and PTB, for clocks of the same type. The triangle plot is explained in the text at the beginning of section 4. The result for the fibre comparison is taken from [20].

Concerning the results in view of the optical clocks, the relative frequency offsets for the clock pairs Yb$^+$(NPL)–Yb$^+$(PTB) and Sr(LNE-SYRTE)–Sr(PTB) are compatible with zero, independent of the link technique. Furthermore, the results for Sr(LNE-SYRTE) – Sr(PTB) agree with the result of a comparison between these clocks via optical fibres $(0.47 \pm 0.5 \times 10^{-16})$, carried out simultaneously with our measurement campaign [20].

The Sr(LNE-SYRTE)–Sr(NPL) and Sr(PTB)–Sr(NPL) comparisons both show a positive offset, however with the Sr(LNE-SYRTE)–Sr(NPL) comparison

still compatible with zero within the $1\sigma$-uncertainty, while the Sr(LNE-SYRTE)–Sr(PTB) results do not indicate a significant offset. Thus, the observed deviations from zero of the Sr(NPL)-related comparisons may indicate that the Sr(NPL) optical frequency was lower by a few $10^{-16}$ compared to the other Sr clocks.

This observed discrepancy illustrates the value of coordinated clock comparison experiments. At the time of this campaign, accounting for all known systematic effects, the total estimated systematic uncertainty of NPL(Sr) was $6.8 \times 10^{-17}$, as shown in table 2. However, later investigations performed by

means of interleaved self-comparison data indicated another uncharacterised systematic frequency shift: depending on the delay between spin-preparation and clock interrogation, the clock frequency could vary by up to $4 \times 10^{-16}$. The working theory is that this could have been a Doppler shift due to radial motion in the lattice, but unfortunately its magnitude at the time of the campaign cannot be quantified in retrospect as the experimental setup has subsequently been modified. The shift has now been eliminated by replacing the lattice laser delivery optics, which has greatly improved the spatial beam quality.

The relative frequency deviation differences for the clocks of different types (Sr lattice against Yb ion) are shown in table 5. They are presented relative to the 2017 CIPM recommended values for the absolute frequencies of the $Yb^+$ octupole transition, $\nu_{Yb^+_{oct}} =$ 642 121 496 772 645.0 Hz and for the Sr transition $\nu_{Sr} =$ 429 228 004 229 873.0 Hz [52]. The results are visualized in fig. 6, in a similar way as in fig. 5. All results deviate from zero in a symmetric way, i.e. $y(Sr - Yb^+) <$ 0 for all comparisons, which indicates that the true frequency ratio $\nu_{Yb^+}/\nu_{Sr}$ is different by a few $10^{-16}$ from the one resulting from the CIPM recommended frequency values. Nevertheless, it remains compatible with the CIPM recommended values within their combined uncertainty of $7.2 \times 10^{-16}$ [52].

### 4.2. Fountain clock comparisons

The results of the comparisons between the remote fountain clocks are compiled in table 6 and depicted in fig. 7 and 8. We find good agreement between all six participating fountain clocks at the low $10^{-16}$ level, compatible with the overall comparison uncertainties. We also observe a good agreement between TWSTFT and GPS PPP satellite link techniques. The results for the comparisons between LNE-SYRTE and PTB agree with the frequency differences measured via optical fibre [16].

### 5. Conclusion

We have carried out simultaneous frequency comparisons between remote optical and fountain clocks via satellite-based techniques and reached uncertainties between $1.8 \times 10^{-16}$ and $3.5 \times 10^{-16}$ for the optical clocks and between $4.8 \times 10^{-16}$ and $7.7 \times 10^{-16}$ for the fountain clocks. For this campaign, five optical clocks and six fountain clocks located at four European metrology institutes, INRIM, LNE-SYRTE, NPL and PTB, were compared over a measurement time of several weeks with duty cycles of up to 77 % for the optical clocks and 98 % for the fountain clocks. The operated optical clocks were the Sr lattice clocks at NPL, PTB

and LNE-SYRTE, and the $Yb^+$ (E3) clocks at NPL and PTB.

The results obtained with the different satellite link techniques are compatible with each other within the combined standard uncertainty. The relative frequency deviation differences between the fountains are consistent with zero within the combined standard uncertainty. The results for comparisons of Sr lattice and Cs/Rb fountain clocks between LNE-SYRTE and PTB agree with comparisons carried out via optical fibre simultaneously with this campaign.

The uncertainties obtained mark a significant improvement for clock comparisons via satellite–based techniques [13,18,21]. The statistical uncertainty of the link dominates the overall link uncertainty for all clock comparisons, with GPS PPP yielding lower statistical and thus combined uncertainties than broadband TWSTFT for all optical clock and most fountain clock comparisons. The TWSTFT measurements suffered from unforeseen technical disturbances causing gaps and an increased instability for averaging times larger than 1 day. GPS measurements on the other hand were uninterrupted, and apart from a few limitations for the receivers of LNE-SYRTE and INRIM, data over the whole campaign could be used.

To achieve even lower uncertainties, further characterization and improvement of the robustness of stations and links is needed, especially in the TWSTFT case. Particularly, software defined radio (SDR) receivers as a replacement for the analog delay locked loops of the SATRE modemsâĂŹ Rx modules demonstrated a significant reduction of instabilities in TWSTFT links contributing to the production of UTC [53]. SDR receivers are significantly less sensitive to interferences between several signals transmitted through a single transponder as used in this comparison [54].

In summary, the two exploited satellite link techniques are complementary as they may be considered as completely independent. This offers the possibility to find undiscovered or study obscured systematic effects in both techniques limiting the significance of frequency comparisons especially for those long-baseline links were optical fiber links do not exist.

We have shown an example of how data sets with different noise types, gaps and correlations can be treated while avoiding unnecessarily discarding data and being limited by short-term noise processes.

This first large-scale coordinated international clock comparison campaign indicated a possible unexpected systematic frequency shift in NPL's Sr clock. Further investigations motivated by this result have led to an understanding and significant suppression of a systematic frequency shift related to

**Table 5.** Results for the comparisons of optical clocks of different type. The third column gives the optical frequency ratio as an offset from the exact ratio $r_0 = 1.495\,991\,618\,544\,900$, which is approximately $1.60 \times 10^{-16}$ smaller than the ratio determined from the 2017 CIPM recommended values. The fourth column contains the average relative frequency deviation differences referenced to the 2017 CIPM recommended frequency values. $u_{B,c}$ is the combined systematic clock uncertainty, $u_{A,l}$ and $u_{B,l}$ correspond to the statistical and systematic link uncertainties, $u$ is the combined uncertainty, and $\Delta t_{OC}$ corresponds to the effective length of the optical clock comparison. All values except for the last column $\Delta t_{OC}$ are in $10^{-16}$.

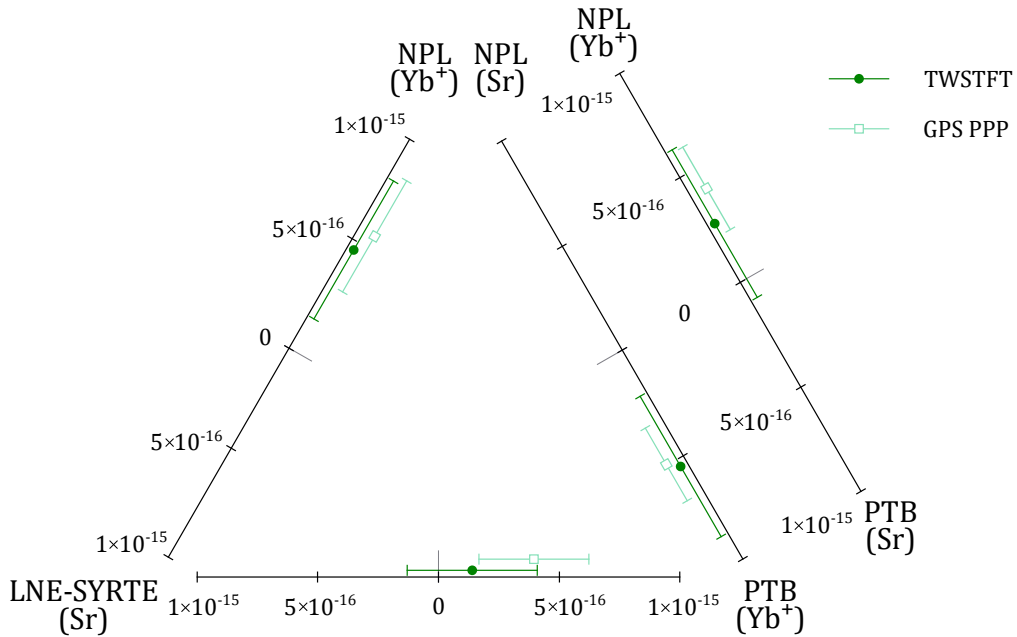| Clock pair | link | $\frac{\nu_{Yb+}}{\nu_{Sr}} - r_0$ | $y_{Sr} - y_{Yb+}$ | $u_{B,c}$ | $u_{A,l}$ | $u_{B,l}$ | $u$ | $\Delta t_{OC}$ [d] |
|---|---|---|---|---|---|---|---|---|
| Sr(LNE-SYRTE) | TWSTFT | 9.76 | -4.9 | 1.2 | 3.0 | 0.7 | 3.3 | 16.6 |
| – Yb$^+$(NPL) | GPS PPP | 11.13 | -5.8 | | 2.3 | 0.8 | 2.7 | 13.5 |
| Sr(LNE-SYRTE) | TWSTFT | 4.59 | -1.4 | 0.4 | 2.5 | 0.9 | 2.7 | 16.4 |
| – Yb$^+$(PTB) | GPS PPP | 8.40 | -3.9 | | 1.9 | 1.2 | 2.3 | 13.5 |
| Sr(NPL) | TWSTFT | 10.54 | -5.4 | 0.7 | 3.3 | 0.5 | 3.4 | 24.4 |
| – Yb$^+$(PTB) | GPS PPP | 9.99 | -5.0 | | 1.5 | 0.6 | 1.8 | 24.5 |
| Sr(PTB) | TWSTFT | 6.44 | -2.6 | 1.1 | 3.3 | 0.5 | 3.5 | 24.4 |
| –Yb$^+$(NPL) | GPS PPP | 8.58 | -4.1 | | 1.5 | 0.6 | 2.0 | 24.5 |



**Figure 6.** The results of the relative frequency deviation differences for the optical clock comparisons between the three institutes LNE-SYRTE, NPL and PTB, for clocks of different type. The triangle plot is explained in the text at the beginning of section 4. All results refer to the 2017 CIPM recommended values for the absolute frequencies of the clock transitions for Yb$^+$ and Sr(lattice).
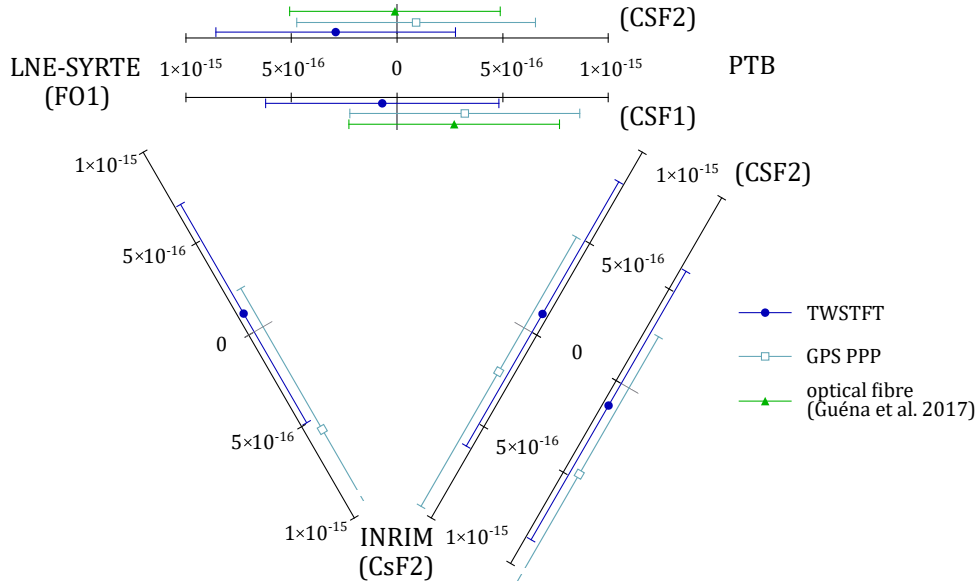
**Figure 7.** The results of the relative frequency deviation differences for all fountain clock comparisons between INRIM and PTB, and all comparisons to FO1 of LNE-SYRTE. The triangle plot is explained in the text at the beginning of section 4. The results for the fibre comparison are taken from [16].
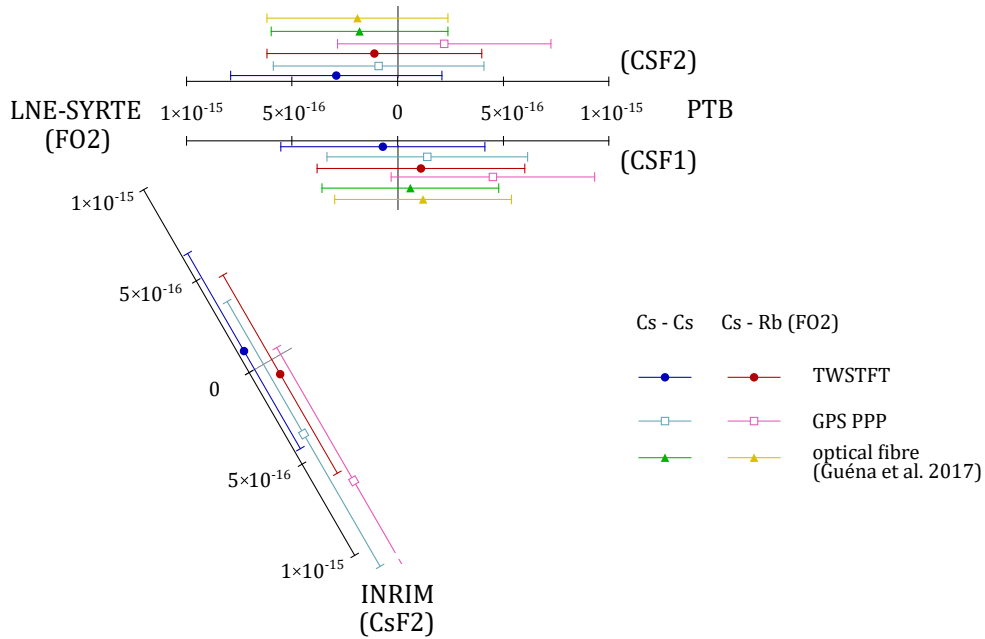


**Figure 8.** The results of the relative frequency deviation differences for all fountain clock comparisons between FO2 (both Cs and Rb) of LNE-SYRTE and the Cs fountains of INRIM and PTB. The results for the fibre comparison are taken from [16]. The triangle plot is explained in the text at the beginning of section 4. The third side of the triangle, which corresponds to the comparisons between INRIM and PTB, is left out, the results can be found in fig. 7.

**Table 6.** Summary of the remote fountain clock comparisons. The third column gives the average fountain frequency difference, $u_{A,c}$ and $u_{B,c}$ are the combined statistical and systematic clock uncertainties (with the dead-time uncertainty as part of the statistical uncertainty), $u_{A,l}$ and $u_{B,l}$ correspond to the statistical and systematic link uncertainties, and $u$ is the combined uncertainty. The value of the absolute frequency of the $^{87}$Rb ground state hyperfine transition used in these comparisons is the CIPM recommended value 6 834 682 610.904 312 6 Hz of 2017 [52]. All values in $10^{-16}$.

| Clock pair | link | difference | $u_{A,c}$ | $u_{B,c}$ | $u_{A,l}$ | $u_{B,l}$ | $u$ |
|---|---|---|---|---|---|---|---|
| ITCsF2 (INRIM) | TWSTFT | -1.0 | 2.8 | | 3.2 | 0.4 | 6.0 |
| - FO1 (LNE-SYRTE) | GPS PPP | 5.6 | 3.2 | 4.3 | 5.6 | 0.6 | 7.7 |
| ITCsF2 (INRIM) | TWSTFT | -1.0 | 2.8 | | 3.2 | 0.4 | 5.4 |
| - FO2 (LNE-SYRTE) | GPS PPP | 3.8 | 3.2 | 3.4 | 5.6 | 0.6 | 7.3 |
| ITCsF2 (INRIM) | TWSTFT | 0.8 | 2.7 | | 3.2 | 0.4 | 5.4 |
| - FO2Rb (LNE-SYRTE) | GPS PPP | 6.9 | 3.1 | 3.5 | 5.6 | 0.6 | 7.3 |
| ITCsF2 (INRIM) | TWSTFT | -1.0 | 2.9 | | 5.5 | 0.3 | 7.2 |
| - CSF1 (PTB) | GPS PPP | 2.4 | 3.3 | 3.8 | 5.4 | 0.9 | 7.4 |
| ITCsF2 (INRIM) | TWSTFT | 1.2 | 3.3 | | 5.5 | 0.3 | 7.3 |
| - CSF2 (PTB) | GPS PPP | 4.7 | 3.8 | 3.8 | 5.4 | 0.9 | 7.5 |
| FO1 (LNE-SYRTE) | TWSTFT | 0.7 | 1.4 | | 2.6 | 0.3 | 5.5 |
| - CSF1 (PTB) | GPS PPP | -3.2 | 1.7 | 4.7 | 2.0 | 1.0 | 5.5 |
| FO1 (LNE-SYRTE) | TWSTFT | 2.9 | 2.1 | | 2.6 | 0.3 | 5.8 |
| - CSF2 (PTB) | GPS PPP | -0.9 | 2.5 | 4.7 | 2.0 | 1.0 | 5.7 |
| FO2 (LNE-SYRTE) | TWSTFT | 0.7 | 1.3 | | 2.6 | 0.3 | 4.9 |
| - CSF1 (PTB) | GPS PPP | -1.4 | 1.5 | 3.9 | 2.0 | 1.0 | 4.8 |
| FO2 (LNE-SYRTE) | TWSTFT | 2.9 | 2.0 | | 2.6 | 0.3 | 5.1 |
| - CSF2 (PTB) | GPS PPP | 0.9 | 2.4 | 3.9 | 2.0 | 1.0 | 5.0 |
| FO2Rb (LNE-SYRTE) | TWSTFT | -1.1 | 1.2 | | 2.6 | 0.3 | 4.9 |
| - CSF1 (PTB) | GPS PPP | -4.5 | 1.4 | 4.0 | 2.0 | 1.0 | 4.8 |
| FO2Rb (LNE-SYRTE) | TWSTFT | 1.1 | 2.0 | | 2.6 | 0.3 | 5.2 |
| - CSF2 (PTB) | GPS PPP | -2.2 | 2.3 | 4.0 | 2.0 | 1.0 | 5.1 |

radial motion in the lattice.

This work provides satisfactory tests of reliability between clocks of the same species, with agreement matching the uncertainty of the current definition of the SI second: it is an essential step towards a possible redefinition of the second. The clock comparisons pushed towards the ascertainting the ratio between different species, which is useful in the perspective of standards that will be secondary representations after a redefinition of the second. Finally, the enhanced microwave satellite techniques implemented, despite the fact that they remain behind the optical fiber links for the frequency comparisons, allow these comparisons to the very low $10^{-16}$ level.

## Acknowledgment

discussions and support, and the Institute of National Resources Canada for providing the software license of their PPP calculation software. The authors thank Sébastien Merlet for his support to gravimetric and levelling measurements at and around the LNE-SYRTE site.

## Appendix

*Estimation of autocovariances and of the standard deviation of the mean*

In this appendix we describe how the mean frequency value and its statistical uncertainty, i.e. the standard deviation of the mean, are determined from the frequency data. The frequency values are serially correlated and are given on a time grid with regular spacing (e.g. 1 s), however with some of the frequency values missing or invalid at gaps. Text-book statistical procedures for the determination of the standard deviation of the mean either do not take into account serial correlations or do not allow for gaps. They can only be used to determine conservative estimates for the standard deviation of the mean, which are not useful in our case. For this reason, we have derived a procedure which takes all these aspects into account. As the fundamental prerequisite for the concept of representing a measurement result by a mean value and its statistical uncertainty, ergodicity of the statistical processes is assumed as usual.

The frequency data given on a regular time grid can be represented by a uniform discrete time series of frequency values $y_i$, $i = 1 \ldots N$. For the treatment of the gaps, a time series of the same length $N$ comprising associated weights $w_i$ is introduced, which are set to zero at times of missing or invalid frequency data. The weights $w_i$ used are not all zero and not fully random, sometimes referred to as reliability weights. This allows us to derive an estimate of the mean and its standard deviation from the frequency data.

Independently of the correlations, the unbiased estimate for the weighted mean is given by the weighted sample mean

$$\bar{y}_w = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}. \tag{13}$$

In general, the variance of a random variable $X$ is defined as

$$\sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[(X - \mu)^2\right]$$
$$= \mathbb{E}\left[X^2\right] - \mu^2, \tag{14}$$

with $\mathbb{E}[\ldots]$ denoting the expected value of the random variable in the square brackets, and with $\mu = \mathbb{E}[X]$. In order to estimate the population variance $\sigma^2$ and the standard deviation of the weighted sample mean $\sigma_{\bar{y}_w}$,

we must take both the correlations and the weighting into account.

It is conducive to start with the estimation of the autocovariances. They carry the information about the serial correlations and thus play an essential role in the derivation of the estimators for the population variance and for the variance of the mean. For a stationary process such as any ergodic process, the lag-$l$ autocovariance is defined as

$$R_l = \mathbb{E}\left[(y_i - \mu)(y_{i+l} - \mu)\right]$$
$$= \mathbb{E}\left[(y_i y_{i+l})\right] - \mu^2. \tag{15}$$

For lag $l = 0$, this is the population variance

$$R_0 = \sigma^2. \tag{16}$$

An estimator for the lag-$l$ autocovariance is the weighted sample autocovariance

$$\hat{R}_l = \frac{\sum_{i=1}^{N-l} \sqrt{w_i w_{i+l}}\,(y_i - \bar{y}_w)(y_{i+l} - \bar{y}_w)}{\sum_{i=1}^{N-l} \sqrt{w_i w_{i+l}}}. \tag{17}$$

We will now derive its expected value. First, note the relation

$$\mathbb{E}\left[\bar{y}_w^2\right] = \sum_{i=1}^{N} w_i \sum_{j=1}^{N} w_j \mathbb{E}\left[y_i y_j\right]$$
$$= \mathbb{E}\left[y_i \sum_{j=1}^{N} w_j y_j\right]$$
$$= \mathbb{E}\left[y_i \bar{y}_w\right]. \tag{18}$$

With this at hand, we can show for the expected value of the weighted sample autocovariance:

$$\mathbb{E}\left[\hat{R}_l\right] = \frac{\sum_{i=1}^{N-l} \sqrt{w_i w_{i+l}}\,\mathbb{E}\left[(y_i - \bar{y}_w)(y_{i+l} - \bar{y}_w)\right]}{\sum_{i=1}^{N-l} \sqrt{w_i w_{i+l}}}$$
$$= \mathbb{E}\left[(y_i - \bar{y}_w)(y_{i+l} - \bar{y}_w)\right]$$
$$= \mathbb{E}\left[y_i y_{i+l}\right] - \mu^2 - \mathbb{E}\left[y_i \bar{y}_w\right] - \mathbb{E}\left[y_{i+l}\bar{y}_w\right]$$
$$\quad + \mathbb{E}\left[\bar{y}_w^2\right] + \mu^2$$
$$= \left(\mathbb{E}\left[y_i y_{i+l}\right] - \mu^2\right) - \left(\mathbb{E}\left[\bar{y}_w^2\right] - \mu^2\right)$$
$$= R_l - \sigma_{\bar{y}_w}^2. \tag{19}$$

Since the variance of the mean is always larger than zero, the estimator $\hat{R}_l$ is negatively biased. To the best of our knowledge no unbiased estimator of the autocovariance has been proposed yet and it probably does not exist, but the sample covariance estimator is sometimes denoted as "unbiased estimator" in statistics textbooks even though it is actually biased [55]. For zero lag we get the relation

$$\mathbb{E}\left[\hat{R}_0\right] = R_0 - \sigma_{\bar{y}_w}^2. \tag{20}$$

For the derivation of the variance of the weighted mean first note that

$$\bar{y}_w^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j y_i y_j$$

$$= \sum_{i=1}^{N} w_i^2 y_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} w_j w_{i+j} y_j y_{i+j},$$

$$(21)$$

and its expected value

$$\mathbb{E}\left[\bar{y}_w^2\right] = \sum_{i=1}^{N} w_i^2 \mathbb{E}\left[y_i^2\right] + 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} w_j w_{i+j} \mathbb{E}\left[y_j y_{i+j}\right]$$

$$= \sum_{i=1}^{N} w_i^2 \left(R_0 + \mu^2\right)$$

$$+ 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} w_j w_{i+j} \left(R_i + \mu^2\right)$$

$$= \sum_{i=1}^{N} w_i^2 R_0 + 2 \sum_{i=1}^{N-1} R_i \sum_{j=1}^{N-i} w_j w_{i+j}$$

$$+ \left(\sum_{i=1}^{N} w_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} w_j w_{i+j}\right) \mu^2$$

$$= \sum_{i=1}^{N} w_i^2 R_0 + 2 \sum_{i=1}^{N-1} R_i \sum_{j=1}^{N-i} w_j w_{i+j}$$

$$+ \left(\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j\right) \mu^2$$

$$= \sum_{i=1}^{N} w_i^2 R_0 + 2 \sum_{i=1}^{N-1} R_i \sum_{j=1}^{N-i} w_j w_{i+j} + \mu^2. \quad (22)$$

Hence, we get for the variance of the weighted mean

$$\sigma_{\bar{y}_w}^2 = \mathbb{E}\left[\bar{y}_w^2\right] - \mu^2$$

$$= R_0 s_0 + 2 \sum_{i=1}^{N-1} R_i s_i$$

$$= r R_0 \qquad (23)$$

with

$$s_i = \sum_{j=1}^{N-i} w_j w_{i+j} \quad \text{for} \quad i = 0 \dots N - 1 \qquad (24)$$

and

$$r = s_0 + 2 \sum_{i=1}^{N-1} \frac{R_i}{R_0} s_i. \qquad (25)$$

If all weighting factors are equal, $w_i = 1/N$, it follows that $s_i = (N - i)/N$ in accordance with the formulae in [44, 45]. The factor $r$ itself depends on all autocovariances $R_l$ and has been reported previously [44], though without a derivation. Equation 23 is the generalization of the well-known relation

$$\sigma_{\bar{y}_w}^2 = \sigma^2/N, \qquad (26)$$

which is only applicable to uncorrelated and unweighted data. Due to this fact, the inverse of the

factor $r$ is often interpreted as an effective number of independent observations ($N_{\text{eff}}$, or similar nomenclature). Although this is undoubtedly a very descriptive interpretation in the case of an equally weighted mean [45], it at least partially loses its sense and can actually be misleading in the case of nonuniform weights, especially if some of the weighting factors are zero or very small.

Since we do not know the true autocovariances $R_l$, and unbiased estimates of the autocovariances do not exist, the best that can be done in this situation is to follow the approach of references [44] and [46] and plug the biased estimators $\hat{R}_l$ from eq. 17 into eq. 23:

$$\hat{\sigma}_{\bar{y}_w}^2 = \hat{R}_0 s_0 + 2 \sum_{i=1}^{N-1} \hat{R}_i s_i. \qquad (27)$$

This results in a biased estimator $\hat{\sigma}_{\bar{y}_w}^2$ for the variance of the mean. However, the bias is reduced with respect to naively using eq. 26, because at least approximate information on the correlations and weightings is used.

When doing so, we have to make sure that the estimators for the autocorrelation coefficient are statistically representative, i.e. that the sums in equation 17 include enough elements. For small lags $l$, this is usually the case, but not for large lags. In many practical cases, we know that the envelope of the autocorrelation asymptotically decreases to zero for large lags. For this reason, the usual approach is to set $\hat{R}_l = 0$ for lags $l$ larger than a cutoff-lag $l_{cut}$. Several criteria to determine $l_{cut}$ have been suggested, such as a "first transit through zero" (FTZ)-criterion [46], an $l_{cut} \approx N/4$ rule of thumb, or more sophisticated criteria for specific statistical processes [44]. It can also be noticed from eq. 27 that setting negative $\hat{R}_l < 0$ to zero yields a more conservative estimator for the variance of the mean.

## References

[1] R. Wynands and S. Weyers, "Atomic fountain clocks," *Metrologia*, vol. 42, no. 3, pp. S64–S79, 2005.

[2] F. Levi, D. Calonico, C. E. Calosso, A. Godone, S. Micalizio, and G. A. Costanzo, "Accuracy evaluation of ITCsF2: a nitrogen cooled caesium fountain," *Metrologia*, vol. 51, no. 3, pp. 270–284, 2014.

[3] J. Guéna, M. Abgrall, D. Rovera, P. Laurent, B. Chupin, M. Lours, G. Santarelli, P. Rosenbusch, M. E. Tobar, R. Li, K. Gibble, A. Clairon, and S. Bize, "Progress in Atomic Fountains at LNE-SYRTE," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 3, pp. 391–410, 2012.

[4] S. Weyers, V. Gerginov, M. Kazda, J. Rahm, B. Lipphardt, G. Dobrev, and K. Gibble, "Advances in the accuracy, stability, and reliability of the PTB primary fountain clocks," *Metrologia*, vol. 55, no. 6, pp. 789–805, 2018.

[5] T. L. Nicholson, S. L. Campbell, R. B. Hutson, G. E. Marti, B. J. Bloom, R. L. McNally, W. Zhang, M. D. Barrett, M. S. Safronova, G. F. Strouse, W. L. Tew, and J. Ye, "Systematic evaluation of an atomic clock at

$2 \times 10^{-18}$ total uncertainty," *Nature Communications*, vol. 6, p. 6896, 2015.

[6] N. Huntemann, C. Sanner, B. Lipphardt, C. Tamm, and E. Peik, "Single-ion atomic clock with $3 \times 10^{-18}$ systematic uncertainty," *Physical Review Letters*, vol. 116, no. 6, p. 063001, 2016.

[7] W. F. McGrew, X. Zhang, R. J. Fasano, S. A. Schäffer, K. Beloy, D. Nicolodi, R. C. Brown, N. Hinkley, G. Milani, M. Schioppa, T. H. Yoon, and A. D. Ludlow, "Atomic clock performance enabling geodesy below the centimetre level," *Nature*, vol. 564, no. 7734, pp. 87–90, 2018.

[8] S. M. Brewer, J.-S. Chen, A. M. Hankin, E. R. Clements, C. W. Chou, D. J. Wineland, D. B. Hume, and D. R. Leibrandt, "$^{27}$Al$^+$ quantum-logic clock with a systematic uncertainty below $10^{-18}$," *Phys. Rev. Lett.*, vol. 123, p. 033201, 2019.

[9] F. Riehle, "Towards a redefinition of the second based on optical atomic clocks," *Comptes Rendus Physique*, vol. 16, no. 5, pp. 506–515, 2015.

[10] P. Gill, "Is the time right for a redefinition of the second by optical atomic clocks?," in *Journal of Physics: Conference Series*, vol. 723, p. 012053, IOP Publishing, 2016.

[11] S. Bize, "The unit of time: Present and future directions," *Comptes Rendus Physique*, vol. 20, pp. 153–168, 2019.

[12] T. Parker, P. Hetzel, S. Jefferts, S. Weyers, L. Nelson, A. Bauch, and J. Levine, "First comparison of remote cesium fountains," in *Proceedings of the 2001 IEEE International Frequency Control Symposium and PDA Exhibition.*, pp. 63–68, IEEE, 2001.

[13] A. Bauch, J. Achkar, S. Bize, D. Calonico, R. Dach, R. Hlaváč, L. Lorini, T. Parker, G. Petit, and D. Piester, "Comparison between frequency standards in Europe and the USA at the $10^{-15}$ uncertainty level," *Metrologia*, vol. 43, no. 1, pp. 109–120, 2005.

[14] M. Fujieda, T. Gotoh, D. Piester, M. Kumagai, S. Weyers, A. Bauch, R. Wynands, and M. Hosokawa, "First comparison of primary frequency standards between Europe and Asia," in *IEEE International Frequency Control Symposium, Joint with the 21st European Frequency and Time Forum.*, pp. 937–941, IEEE, 2007.

[15] A. Zhang, K. Liang, Z. Yang, F. Fang, T. Li, D. Piester, V. Gerginov, S. Weyers, A. Bauch, M. Fujieda, I. Blinov, A. Boiko, Y. Domnin, A. Naumov, Y. Smirnov, A. S. Gupta, P. Arora, A. Acharya, and A. Agarwal, "Comparison of caesium fountain clocks in Europe and Asia," in *European Frequency and Time Forum (EFTF)*, pp. 447–450, 2014.

[16] J. Guéna, S. Weyers, M. Abgrall, C. Grebing, V. Gerginov, P. Rosenbusch, S. Bize, B. Lipphardt, H. Denker, N. Quintin, S. M. F. Raupach, D. Nicolodi, F. Stefani, N. Chiodo, S. Koke, A. Kuhl, F. Wiotte, F. Meynadier, E. Camisard, C. Chardonnet, Y. Le Coq, M. Lours, G. Santarelli, A. Amy-Klein, R. Le Targat, O. Lopez, P. E. Pottie, and G. Grosche, "First international comparison of fountain primary frequency standards via a long distance optical fiber link," *Metrologia*, vol. 54, no. 3, pp. 348–354, 2017.

[17] A. Yamaguchi, M. Fujieda, M. Kumagai, H. Hachisu, S. Nagano, Y. Li, T. Ido, T. Takano, M. Takamoto, and H. Katori, "Direct comparison of distant optical lattice clocks at the $10^{-16}$ uncertainty," *Appl. Phys. Express*, vol. 4, no. 8, p. 082203, 2011.

[18] H. Hachisu, M. Fujieda, S. Nagano, T. Gotoh, A. Nogami, T. Ido, S. Falke, N. Huntemann, C. Grebing, B. Lipphardt, C. Lisdat, and D. Piester, "Direct comparison of optical lattice clocks with an intercontinental baseline of 9 000 km," *Optics Letters*, vol. 39, pp. 4072–4075, 2014.

[19] T. Takano, M. Takamoto, I. Ushijima, N. Ohmae, T. Akat-

suka, A. Yamaguchi, Y. Kuroishi, H. Munekane, B. Miyahara, and H. Katori, "Geopotential measurements with synchronously linked optical lattice clocks," *Nature Photonics*, vol. 10, pp. 662–666, 2016.

[20] C. Lisdat, G. Grosche, N. Quintin, C. Shi, S. Raupach, C. Grebing, D. Nicolodi, F. Stefani, A. Al-Masoudi, S. Dörscher, S. Häfner, J.-L. Robyr, N. Chiodo, S. Bilicki, E. Bookjans, A. Koczwara, S. Koke, A. Kuhl, F. Wiotte, F. Meynadier, E. Camisard, M. Abgrall, M. Lours, T. Legero, H. Schnatz, U. Sterr, H. Denker, C. Chardonnet, Y. Le Coq, G. Santarelli, A. Amy-Klein, R. Le Targat, J. Lodewyck, O. Lopez, and P.-E. Pottie, "A clock network for geodesy and fundamental science," *Nature Communications*, vol. 7, p. 12443, 2016.

[21] J. Leute, N. Huntemann, B. Lipphardt, C. Tamm, P. B. R. Nisbet-Jones, S. A. King, R. M. Godun, J. M. Jones, H. S. Margolis, P. B. Whibberley, A. Wallin, M. Merimaa, P. Gill, and E. Peik, "Frequency comparison of $^{171}$Yb$^+$ ion optical clocks at PTB and NPL via GPS PPP," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 7, pp. 981–985, 2016.

[22] W. F. McGrew, X. Zhang, H. Leopardi, R. J. Fasano, D. Nicolodi, K. Beloy, J. Yao, J. A. Sherman, S. A. Schäffer, J. Savory, R. C. Brown, S. Römisch, C. W. Oates, T. E. Parker, T. M. Fortier, and A. D. Ludlow, "Towards the optical second: verifying optical clocks at the SI limit," *Optica*, vol. 6, no. 4, pp. 448–454, 2019.

[23] D. Piester, A. Bauch, J. Becker, E. Staliuniene, and C. Schlunegger, "On measurement noise in the European TWSTFT network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 9, pp. 1906–1912, 2008.

[24] J. Lodewyck, S. Bilicki, E. Bookjans, J.-L. Robyr, C. Shi, G. Vallet, R. Le Targat, D. Nicolodi, Y. Le Coq, J. Guéna, M. Abgrall, P. Rosenbusch, and S. Bize, "Optical to microwave clock frequency ratios with a nearly continuous strontium optical lattice clock," *Metrologia*, vol. 53, no. 4, pp. 1123–1130, 2016.

[25] J. Guéna, R. Li, K. Gibble, S. Bize, and A. Clairon, "Evaluation of Doppler shifts to improve the accuracy of primary atomic fountain clocks," *Phys. Rev. Lett.*, vol. 106, no. 13, p. 130801, 2011.

[26] J. Guéna, M. Abgrall, A. Clairon, and S. Bize, "Contributing to TAI with a secondary representation of the SI second," *Metrologia*, vol. 51, no. 1, pp. 108–120, 2014.

[27] C. F. A. Baynham, R. M. Godun, J. M. Jones, S. A. King, P. B. R. Nisbet-Jones, F. Baynes, A. Rolland, P. E. G. Baird, K. Bongs, P. Gill, and H. S. Margolis, "Absolute frequency measurement of the optical clock transition in with an uncertainty of using a frequency link to international atomic time," *Journal of Modern Optics*, vol. 65, no. 5-6, pp. 585–591, 2018.

[28] The relevant fountain systematic uncertainty budgets of CSF1 and CSF2 of PTB are close to those reported to the BIPM for the June 2015 evaluations (Circular T330 [50]). In the case of CSF1, a reevaluation of the distributed cavity phase shift was performed after the comparison campaign, which retrospectively leads to a significantly reduced overall systematic uncertainty.

[29] C. Voigt, H. Denker, and L. Timmen, "Time-variable gravity potential components for optical clock comparisons and the definition of international time scales," *Metrologia*, vol. 53, no. 6, p. 1365, 2016.

[30] H. Denker, L. Timmen, C. Voigt, S. Weyers, E. Peik, H. S. Margolis, P. Delva, P. Wolf, and G. Petit, "Geodetic methods to determine the relativistic redshift at the level of $10^{-18}$ in the context of international timescales: a review and practical results," *Journal of Geodesy*, vol. 92,

no. 5, pp. 487–516, 2018.

[31] T. Mehlstäubler, G. Grosche, C. Lisdat, P. Schmidt, and H. Denker, "Atomic clocks for geodesy," *Rep. Prog. Phys.*, vol. 81, p. 064401, 2018.

[32] P. Delva, H. Denker, and G. Lion, "Chronometric Geodesy: Methods andApplications," in *Relativistic Geodesy: Foundations andApplications* (D. Puetzfeld and C. Lämmerzahl, eds.), Fundamental Theories of Physics, pp. 25–85, Cham: Springer International Publishing, 2019.

[33] "The operational use of two-way satellite time and frequency transfer employing pseudorandom noise codes." Recommendation ITU-R TF. 1153-4, ITU, Geneva, Switzerland, 2015.

[34] G. Petit and P. Wolf, "Relativistic theory for picosecond time transfer in the vicinity of the Earth," *Astronomy and Astrophysics*, vol. 286, pp. 971–977, 1994.

[35] D. Piester, A. Bauch, M. Fujieda, T. Gotoh, M. Aida, H. Maeno, M. Hosokawa, and S. Yang, "Studies on instabilities in long-baseline two-way satellite time and frequency transfer (TWSTFT) including a troposphere delay model," in *Proc. 39th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 27-29 Nov 2007, Long Beach, California, USA*, pp. 211–222, 2007.

[36] "Ionosphere GNSS products." `ftp://gnss.oma.be/gnss/products/IONEX/`. Retrieved in July 2017.

[37] G. Xu, *GPS - Theory, Algorithms and Applications*, ch. Physical Influences of GPS Surveying: Ionospheric effects, pp. 39–50. Springer-Verlag Berlin Heidelberg New York, 2003.

[38] J. Kouba and P. Héroux, "Precise point positioning using igs orbit and clock products," *GPS Solutions*, vol. 5, pp. 12–28, 2001.

[39] R. Schmid, P. Steigenberger, G. Gendt, M. Ge, and M. Rothacher, "Generation of a consistent absolute phase-center correction model for GPS receiver and satellite antennas," *Journal of Geodesy*, vol. 81, pp. 781–798, 2007.

[40] J. T. Wu, S. C. Wu, G. A. Hajj, W. I. Bertiger, and S. M. Lichten, "Effects of antenna orientation on GPS carrier phase," in *Astrodynamics 1991* (P. A. Penzo and D. F. Bender, eds.), pp. 1647–1660, 1992.

[41] N. Ashby, "Relativity in the Global Positioning System," *Living Reviews in Relativity*, vol. 6, no. 1, p. 1, 2003.

[42] D. Sullivan, D. Allan, D. Howe, and F. Walls, "Characterization of clocks and oscillators," NIST tech. note 1337, NIST, U.S Department of Commerce, National Institute of Standards and Technology, 1990. Online available at `https://nvlpubs.nist.gov/nistpubs/Legacy/TN/nbstechnicalnote1337.pdf`.

[43] "Guide to the Expression of Uncertainty in Measurement." ISO/TAG 4. Published by ISO, 1993 (corrected and reprinted, 1995) in the name of the BIPM, IEC, IFCC, ISO, UPAC, IUPAP and OIML, 1995. ISBN number: 92-67-10188-9, 1995.

[44] N. F. Zhang, "Calculation of the uncertainty of the mean of autocorrelated measurements," *Metrologia*, vol. 43, no. 4, pp. S276–S281, 2006.

[45] A. Zięba, "Effective number of observations and unbiased estimators of variance for autocorrelated data-an overview," *Metrology and Measurement Systems*, vol. 17, no. 1, pp. 3–16, 2010.

[46] A. Zięba and P. Ramza, "Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation function estimated from the data," *Metrology and Measurement Systems*, vol. 18, no. 4, pp. 529–542, 2011.

[47] J. Prillaman, E. Powers, B. Fonville, S. Mitchell, and E. Goldberg, "Continued Evaluation of Carrier-Phase GNSS Timing Receivers for UTC/TAI Applications," tech. rep., NAVAL OBSERVATORY WASHINGTON DC, 2010.

[48] J. Ray and K. Senior, "IGS/BIPM pilot project: GPS carrier phase for time/frequency transfer and timescale formation," *Metrologia*, vol. 40, no. 3, pp. S270–S288, 2003.

[49] U. Weinbach, *Feasibility and impact of receiver clock modeling in precise GPS data analysis*. PhD thesis, Leibniz University Hannover, 2013.

[50] BIPM Circular T. `http://www.bipm.org/jsp/en/TimeFtp.jsp`.

[51] E. Benkler, C. Lisdat, and U. Sterr, "On the relation between uncertainties of weighted frequency averages and the various types of Allan deviations," *Metrologia*, vol. 52, no. 4, pp. 565–574, 2015.

[52] F. Riehle, P. Gill, F. Arias, and L. Robertsson, "The CIPM list of recommended frequency standard values: guidelines and procedures," *Metrologia*, vol. 55, no. 2, pp. 188–200, 2018.

[53] Z. Jiang, V. Zhang, Y.-J. Huang, J. Achkar, D. Piester, S.-Y. Lin, W. Wu, A. Naumov, S. hoon Yang, J. Nawrocki, I. Sesia, C. Schlunegger, Z. Yang, M. Fujieda, A. Czubla, H. Esteban, C. Rieck, and P. Whibberley, "Use of software-defined radio receivers in two-way satellite time and frequency transfers for UTC computation," *Metrologia*, vol. 55, no. 5, pp. 685–698, 2018.

[54] Z. Jiang, V. Zhang, T. E. Parker, G. Petit, Y.-J. Huang, D. Piester, and J. Achkar, "Improving two-way satellite time and frequency transfer with redundant links for UTC generation," *Metrologia*, vol. 56, no. 2, p. 025005, 2019.

[55] D. B. Percival, "Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean," *The American Statistician*, vol. 47, no. 4, pp. 274–276, 1993.