New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Validating frequency transfer via interferometric fiber links for optical clock comparisons

To cite this article: Sebastian Koke *et al* 2021 *New J. Phys.* **23** 093024

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
1 April 2021

REVISED
20 August 2021

ACCEPTED FOR PUBLICATION
26 August 2021

PUBLISHED
20 September 2021

PAPER

Validating frequency transfer via interferometric fiber links for optical clock comparisons

Sebastian Koke*, Erik Benkler, Alexander Kuhl and Gesine Grosche

Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany
* Author to whom any correspondence should be addressed.

E-mail: sebastian.koke@ptb.de

Keywords: optical frequency dissemination, optical fiber links, optical clocks, optical clock comparisons, ultra-stable lasers

Abstract

We investigate the validation of fiber-based optical frequency transfer for frequency comparison applications. We specifically consider the frequency transfer validation for remote optical clock comparisons and want to ensure interferometric fiber link uncertainty contributions below the combined uncertainty of the clocks under test. The validation is based on signals obtained via looping back from the remote end to the sender site and comparing the input with the output of the loop. These loop-back data need to be averaged over intervals for reaching the validation goal, as the short-term instability of long-distance interferometric fiber links is typically higher than that of optical clocks. We introduce a two-step validation approach and address the finding of a compromise between opposing aspects of averaging: reaching low uncertainties versus achieving a high data coverage of the validated data set via a high temporal resolution of the fault identification. We discuss the impact of different averaging types and of the tolerance of filtering criteria on the achievable estimated uncertainty and on the coverage of the validated data set. Data from four multiple-week-long measurement campaigns on the fiber link between Physikalisch-Technische Bundesansanstalt and University of Strasbourg are used for this assessment.

1. Introduction

Optical frequency dissemination via interferometric fiber links (IFLs) [1, 2] has proven to support frequency transfer uncertainties $u_{\mathrm{link}} \ll 10^{-18}$ between two locations separated by > 600 km over selected time spans [3–7]. At present, transportable optical clocks [8–10] and microwave satellite connections [11–13] do not reach these uncertainty levels. Hence, IFLs are currently the only means capable of remotely comparing [14] the best optical clocks now reaching systematic uncertainties u_{clock} close to or below 10^{-18} [10, 15–20].

This outstanding frequency transfer performance can be exploited for applications such as relativistic geodesy [10, 14, 19, 21–25] or tests of fundamental physics [26, 27]. Despite the proven performance, however, comparing remote optical clocks at the 10^{-18} level separated by continental scales requires careful validation of the IFL frequency transfer performance for achieving an IFL uncertainty u_{link} below the combined systematic uncertainty of the clocks u_{clocks} in a specific clock comparison.

The following treatment is related to the fiber-network-based comparison formalism for multiple (optical) clocks that has been published recently [28]. In that formalism, the measurement chain is divided into separated entities, which reflects the spatial separation and organizational boundaries between the measurements along the chain. For an optical clock comparison, each of these entities reports a comparison value when its operation was valid. This work deals with the determination of the validity of a long-distance IFL. This task is different from assessing the performance of an IFL: for determining the IFL's resolution limit, the statistical uncertainty is minimized to the extent possible by analyzing long, phase-coherent data segments. This allows to reach low uncertainty levels by applying metrics with a strong discrimination for white phase noise, i.e. phase-averaging (also known as Λ averaging) and modified Allan deviations

(modADEVs) [6, 29, 30]. Using these metrics, the IFL-induced instability (modADEV) can initially fall off with averaging time τ as white-phase noise $\propto \tau^{-3/2}$ or even faster with up to τ^{-2} behavior, if phase noise decreases towards lower Fourier frequencies [31, 32].

The IFL uncertainty contribution to (optical) clock comparisons, however, is impacted by the mutual overlap of operation of the measurement chain. The unavailability of optical clocks, frequency combs and ultra-stable lasers alter which IFL data enter the comparison result and, hence, the IFL uncertainty contribution. An underlying question in this paper is the assessment of the IFL uncertainty contribution in the presence of gaps. Here, we treat the gapped frequency transfer data set as if it was contiguous and apply offset and instability calculations without considering the gaps. This matches the way IFL data are processed in optical clock comparisons. However, as elucidated below, the interruptions cause a higher statistical uncertainty due to an additional uncertainty contribution resulting from the loss of phase coherence. The latter is observable as an instability contribution falling off as white frequency noise $\propto \tau^{-1/2}$. In the appendix, we will briefly analyze the mitigation of the impact of gaps on the statistical uncertainty estimation using adapted Allan deviation calculations routines [33] or autocovariance-based routines [34].

In addition, optical clock comparison campaign data may be collected over time spans longer than the contiguous data segments employed for IFL characterization. Hence, there might be an impact of external influences not observed in IFL characterizations. Frequency offsets of the transferred optical signal are monitored out-of-loop (OOL) by returning the transferred optical frequency back to the sender over a second stabilized fiber connection. In [7] we showed that, on the considered IFL, the fiber phase noise is correlated on two fibers from the same installed underground cable using data samples Λ-averaged with a gate time of 1s (in the following Λ_{1s}). Hence, a true upper bound for the $\tau = 1$s offset of the transferred frequency is obtained measuring the OOL frequency offset [7].

Here we elucidate our approach to validate the one-way frequency transfer over the IFL between Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, Germany, and University of Strasbourg (UoS), France, used for optical clock comparison campaigns with international partners [14, 26, 27, 35]. A central question guiding the analysis below is how to ensure $u_{\text{link}} \ll u_{\text{clocks}}$ while keeping the largest possible data coverage (DC). For the validation, we analyze the OOL signal that has traveled the 1400 km long fiber loop PTB–UoS–PTB. Our approach consists of two steps, which we will call *integrity check (IC)* and *performance filtering (PF)* in the following. We presented this core idea already in a conference contribution [36] and meanwhile our French partners have also detailed their alternative but similar validation strategy [37, 38]. While our approach is applicable regardless of the IFL technology used, details of the checks to be included in the two steps and how results of the two steps are merged may differ. The procedure presented here reflects to some extent the characteristics of the testbed PTB–UoS: this IFL is the first to combine [7] fiber Brillouin amplification (FBA) [4, 6, 39] with a repeater laser station (RLS) [5, 40].

Under optimal conditions, FBA-based IFLs can operate with a cycle-slip (CS) rate of less than 1 per 5 days [6, 7]. In case of an unlock of the FBA pump laser, however, reestablishing the pump laser lock requires some time, especially if FBA is cascaded several times for mitigation of the fiber attenuation along the IFL. This is different in long-haul IFLs based on cascaded erbium-doped fiber amplifiers. These IFLs are known to suffer from more frequent but short signal drop-outs [3, 5, 41] resulting from destructive multipath interference due to double reflection or double Rayleigh scattering [42, 43] in combination with a time varying path length between the reflection points. Hence, for these IFLs the procedure may require adaption to yield optimal performance.

2. The interferometric fiber link PTB–UoS

The frequency transfer performance over the IFL loop PTB–UoS–PTB has been extensively investigated in [7], where the experimental setup is also described in more detail. Of importance for the study presented here is that the RLS cascades the individually stabilized, FBA-based uplink and downlink to form the IFL loop PTB–UoS–PTB. In the RLS, a laser is phase-locked to the signal received on the uplink [5, 40]. For (optical) clock comparisons, a fraction of the repeater laser output is tapped off into the so-called *user output*. The impact of the radio-frequency reference used to phase lock the repeater laser cancels for the OOL signal (PTB–UoS–PTB) and can be compensated on the user output using an appropriate frequency shift [5]. The endpoint of the uplink (*B* in figure 1), the starting point of the downlink (*C*), and the reference point for the two-way comparison (*E*) reside in a temperature stabilized box [5]. Differential optical path-length variation between *B*–*C* and *B*–*E* cannot be characterized using the OOL signal PTB–UoS–PTB (*A*–*B*–*C*–*D*–*A*). These have been addressed and minimized in the design of the RLS [44, 45].

Similarly, optical path-length variations *D*–*A* inside the PTB laboratory set a resolution limit for the OOL characterization of the one-way frequency transfer PTB–UoS (*A*–*E*). As in previous publications

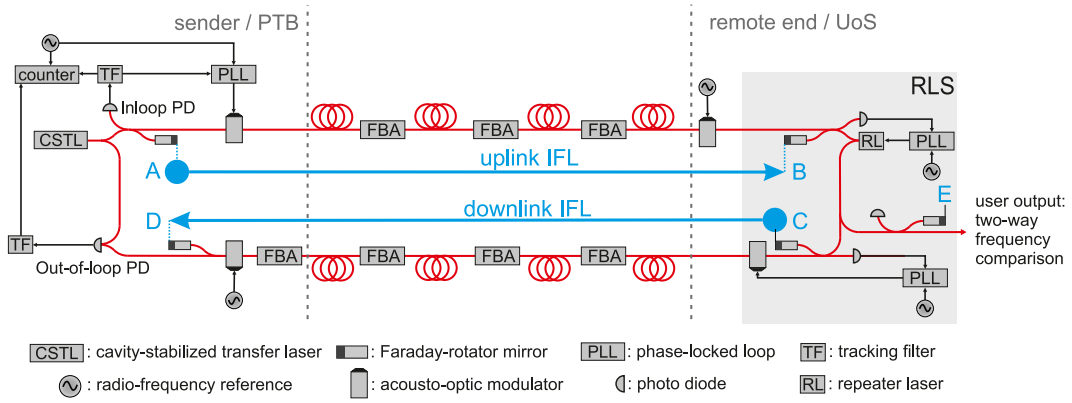

Figure 1. Setup of IFL loop PTB–UoS–PTB. For the compensation of transmission fiber phase noise [1, 2], part of the light is retroreflected at the remote end and brought to interference with a copy of the input light. The resulting beat is stabilized with respect to a radio-frequency reference using a phase-locked loop (PLL). An acousto-optic modulator at the remote end serves to discriminate light from the wanted backreflection from spurious backreflections and backscattering along the IFL. Within the locking bandwidth, the PLL establishes phase coherence between the two reference points given by the retroreflection in the short reference arm *A* and at the remote end *B*. This phase-coherent frequency dissemination is schematically denoted by the blue arrows. The frequency of a cavity-stabilized laser is disseminated from *A* to *B* via the phase stabilized uplink. Inside the RLS, the optical frequency is distributed from point *B* to the starting point *C* of the downlink and the reference point *E* of the user output. The differential phase noise is passively minimized by a temperature controlled environment and length matched fibers [5, 37, 44]. The phase-stabilized downlink transfers light back to point *D* at PTB. Mediated by a passive fiber connections, a phase comparison between light fields at *A* and *D* is performed for the OOL characterization.

[6, 7], we kept the two interferometers separate to simulate the frequency transfer to a remote place for the data analyzed here. The fiber interferometers are placed in separate aluminum housings being placed under separate wooden boxes for passive thermal isolation. These wooden boxes reside adjacent to each other and the fiber connection *A*–*D* mostly runs underneath these boxes. Potential (anti)correlations between these error source and other error sources on the round-trip (*A*–*B*–*C*–*D*) or one-way (*A*–*E*) are unknown. Hence, at this stage we assume that the OOL signal gives an upper bound of the frequency transfer error PTB–UoS (*A*–*E*) even at time spans where the spurious path-length variations *A*–*D* are dominating. For the purpose of improving the one-way frequency transfer validation, one could employ a passive two-way comparison on *A*–*D* [37, 46] or a more compact interferometer layout [45, 47, 48].

In the following, we will analyze the signal used for the uplink path-length stabilization (in-loop (IL), *A*–*B*–*A*) and the OOL heterodyne beat between the light entering the uplink and the light returning to PTB on the downlink (*A*–*B*–*C*–*D*–*A*). The beat signals are phase-tracked using phase-locked voltage controlled oscillators (locking bandwidth ≈ 100 kHz) for removal of high-frequency noise contributions. For short, we will call these tracking filters in the following. For each signal, we employ two tracking filters with slightly different settings in order to check for the consistency of the tracking process. The output frequencies of these tracking filters are recorded using a dead-time-free frequency counter [49] (K & K FXE80) in phase-averaging mode with a report interval $\Delta t = 1\,\mathrm{s}$ ($\Lambda_{1\,\mathrm{s}}$).

The analyzed data set has been made available [50] and stems from four multiple-week-long measurement campaigns: the first one from 2015-06-01 to 2015-06-24; the second from 2016-03-14 to 2016-04-30; the third from 2018-03-27 to 2018-05-02; and the fourth from 2018-12-03 to 2018-12-31. Total number of $\Lambda_{1\,\mathrm{s}}$ data samples within these campaigns amounts to ≈ 11.9 million.

3. Typical PTB–UoS–PTB IFL performance for long continuous data segments

For long continuous data segments, we achieve instabilities of the IL signal and the OOL beat (*A*–*B*–*C*–*D*–*A*) as shown in figure 2(a) estimated by the overlapping Allan deviation (OADEV), which is based on Π-averaging metrics [29, 30]. Figure 2(b) shows the results for the Λ-averaging-based modADEV. The instabilities start off in the high 10^{-16} range [7] and initially fall off proportional to τ^{-1} (OADEV) and $\approx \tau^{-1.75}$ (modADEV). The modADEV's double-logarithmic slope below $-3/2$ indicates sub-white phase noise in stabilized operation [6, 31, 32]. The capability of the modADEV to differentiate sub-white frequency noise sources and its steeper averaging behavior reveal noise contributions not observable in the OADEV. For instance, a bump is visible around $\tau = 15\,\mathrm{s}$, which we attribute to a temperature stabilization system [7]. For averaging times larger than $\tau \approx 100$ s, optical path length variation between *D* and *A* start to dominate [6, 7] leading to a significantly slower averaging behavior of the modADEV with τ. The dashed horizontal lines show the fractional frequency offsets over this data segment. These offsets are typically at

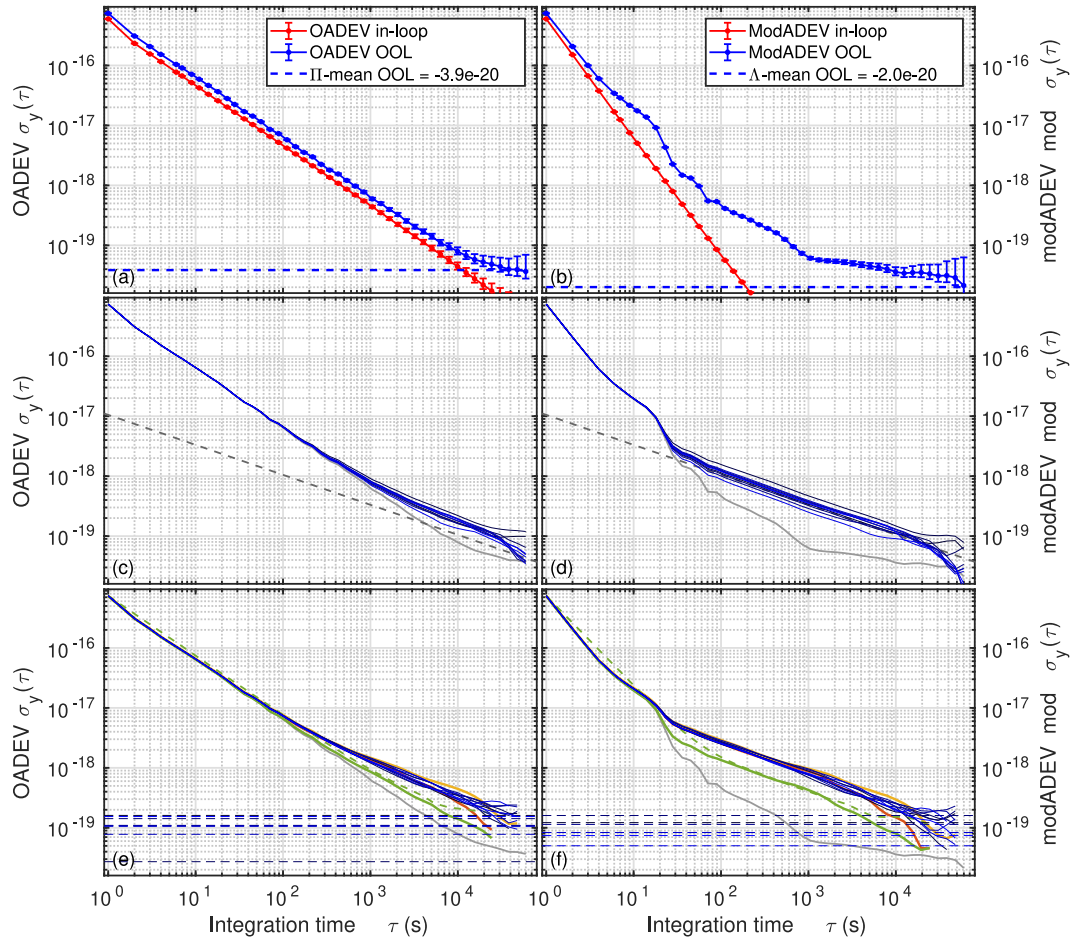

Figure 2. (a) Typical OADEV of IL stabilization beat (red) and of the OOL characterization signal (blue solid) and the Π-averaged fractional frequency offset (blue dashed) for a typical long continuous data segment. Error bars indicate the 1σ confidence interval estimated by the method of Greenhall and Riley [51]. (b) Same as (a) but for modADEV and Λ-averaging. (c) The solid blue curves of different saturation show the same data as the blue line in (a), but a CS of 2π has been added to the data at different positions. The gray solid line shows the original OADEV for comparison. The gray dashed line indicates the curves asymptote $a(\tau)$ (see text). (d) Same as (c) but for modADEV. (e) The solid blue curves of different saturation show the same data as the blue line in (a), but with artificially chopping the time series and regarding the remaining one as a new continuous data set. The blue curves show the variance for normal distributed extracted segment lengths ($\bar{\mu}_s = 200$ s, $\sigma_s = 50$ s) and gaps introduced ($\bar{\mu}_g = 50$ s, $\sigma_g = 10$ s). The dashed blue horizontal lines show the corresponding Π-averaged fractional frequency offsets. The dark orange and the green solid curve in the background show the instability when a fixed interval and gap length of 200 s and 1000 s is used, respectively. The dashed green curve shows the instability of a purely white-phase noise surrogate data set after applying a data selection with an interval and gap length 1000 s. The light orange curve shows the instability for a fixed interval of 200 s and a gap length of 1s. The gray curve shows the original OADEV for comparison. (f) Same as (e) but for modADEV and Λ-averaging.

the level of or below the last point of the corresponding instability, which can be regarded as an estimate for the statistical uncertainty [30].

The subfigures figures 2(c)–(f) exemplify the impact of non-ideal operation conditions. The blue solid curves of differing color saturation in figures 2(c) and (d) show the OADEV and modADEV estimates, respectively, when a CS with 2π radians magnitude is artificially added to the original long continuous data set at different instances of time (see similar considerations in [4]). We heuristically find that with the cycle slips the asymptote $a(\tau) = 1\ \mathrm{Hz}/(\nu_{\mathrm{laser}} \times \sqrt{T\tau})$ is an approximate lower bound for the instabilities (T being the time span covered by the data set and ν_{laser} being the frequency of the cavity-stabilized transfer laser in figure 1). The absence of such a bound in the instabilities determined from the original data set is an indirect verification of CS-free operation.

Of greater importance for the discussion below is the impact of chopping the data set into disconnected segments of varying length and with variable gap length. This is the condition for the IFL data entering clock comparisons after taking into account the overlapping uptime of the complete measurement chain. As an example, figures 2(e) and (f) show the OADEV and modADEV estimates, respectively, assuming that the interval and gap length follow a normal distribution. For this we used standard normal distributed random

numbers r_k to generate random extracted segment lengths ($l_{s,k}$) and the gaps ($l_{g,k}$) using

$$l_{X,k} = \bar{\mu}_X + \sigma_X \cdot r_k, \tag{1}$$

where $\bar{\mu}_X$ and σ_X denote the expectation value and the standard deviation of the distributions, respectively. We use $\bar{\mu}_s = 200$ s, $\sigma_s = 50$ s for the segments and $\bar{\mu}_g = 50$ s, $\sigma_g = 10$ s for the gaps. For the calculation of the Allan deviations, the remaining data set has been treated as if it was contiguous without applying a gap mitigation technique. In the resulting instability curves, a kink is observable again with a smooth transition. Deviations from the original data set are already perceivable for averaging times shorter than the characteristic timescales given by the Gaussian distribution parameters (shortest segment and gap length is on the order of $\bar{\mu}_s - \sigma_s$ and $\bar{\mu}_g - \sigma_g$, respectively). The existence of deviations is also observed by the dark orange and solid green curves in the background showing the instability of the chopped data set for a fixed segment and gap lengths of 200 s and 1000 s, respectively. The dashed green curve shows the same 1000 s filtering for a white-phase noise data set. Also for this curve, the deviation from the $\tau^{-3/2}$-scaling sets in at averaging times considerably shorter than the phase-coherent segment length. Hence, this finding is independent from the underlying noise processes. The dark orange curve for 200 s chopping essentially follows the curves when Gaussian distributions are employed, and also the course of the light orange curve, which shows the data for a fixed segment length of 200 s and a gap length of 1s. This shows that the gaps only have a minor contribution to the kink characteristics. For averaging times τ beyond the inset of the kink, the double-logarithmic slope is close to the one of white frequency noise ($-1/2$), which is an indication of the loss of phase coherence. Keeping in mind the larger uncertainty of the instability values at the longest averaging times τ (cf to e.g. figures 2(a) and (b)) [51], the fractional frequency offsets shown as the dashed horizontal lines are still at or below the corresponding instabilities. This indicates stationary noise processes on our IFL in such well-behaved data segments over time scales of $\approx$ 200 s.

4. Parameter study for IFL validation on PTB–UoS

4.1. Integrity check

As mentioned above, we follow a two-step approach to IFL validation. In the first step, the IC, we assess the prerequisites of reasonable IFL operation: we create a validity flag time series and mark those points as invalid where no data was collected, the self-assessment of the transfer laser lock to the ultra-stable cavity indicates unstabilized operation, or where the in-lab FBA-pump laser amplifying downlink light before reaching PTB is out of lock. In future, this self-assessment can be enhanced by also including the monitoring signals of the remote FBA-modules or of the RLS. To fill this missing piece of self-assessment, we complement the IC by a coarse check for large frequency deviations on the IL and OOL beat. Figure 3 shows the distributions of the Λ_{1s} frequency offsets observed on our IFL. The IL (figure 3(a)) and the OOL (figure 3(b)) distributions show near-Gaussian shapes with widths $\sigma \approx 0.13$ Hz, which is compatible with the 1s instability values of $\approx 6.5 \times 10^{-16}$ in figure 2(a). The width of the distributions of the double tracking filter frequency difference are almost three orders of magnitude narrower. Given these values, we will consider Λ_{1s} data points as valid if frequency offsets from the nominal value and the frequency difference of the two tracking filters are below $\approx 10\sigma$: $\Delta f_{IL} \leqslant 1.5$ Hz and $\Delta f_{OOL} \leqslant 1.5$ Hz, $\Delta f_{TF,IL} \leqslant 7.5$ mHz and $\Delta f_{TF,OOL} \leqslant 7.5$ mHz.

We will show below in figure 7 that the number of points removed by IC filtering is >13% and differs within the campaigns. This mainly results from the different distributions of longer interruptions in the IFL operation.

4.2. Performance filtering

The goal of the validation is to constrain the IFL uncertainty in the presence of intermittent disturbances due to, e.g. a perturbed lock or time spans of high phase noise levels due to human activity. The second PF step is where we select time spans that fulfill $u_{link} \ll u_{clocks}$. The short-term uncertainty of the IFL is dictated by the free-running fiber noise and the delay-limited phase noise suppression [2]. For IFLs over buried fiber connections, typical short-term uncertainties are above demonstrated short-term optical clock uncertainties u_{clock} already for IFL lengths of $\approx$ 100 km. Hence, achieving the goal of $u_{link} \ll u_{clocks}$ for long-haul IFLs requires to average over data intervals.

Due to this averaging, the PF step validates or invalidates data intervals of length $T = N\Delta t$, with $\Delta t = t_{i+1} - t_i$ denoting the sampling interval (here $\Delta t = 1$s). Figure 4 illustrates the strategy applied in the following to combine the results of the N samples long PF with the data-point-wise IC. We also tested out a strategy consecutively applying IC and PF. This strategy is detailed in appendix B together with the corresponding results.

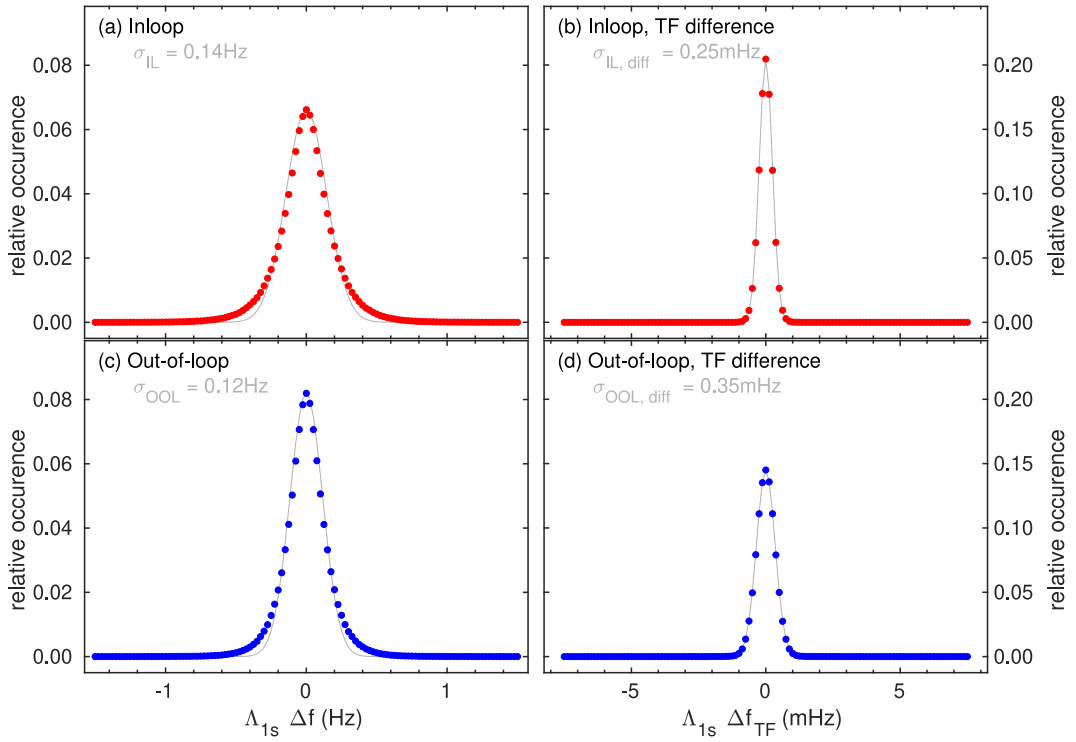

Figure 3. Typical distributions of Λ_{1s} frequency offsets of (a) the IL beat, (b) the difference of the two IL tracking filters, (c) the OOL beat, and (d) the difference of the two OOL tracking filters. The analyzed data stems from the 2018-12 campaign. The distributions are normalized to the total number of points considered and depend on the experimental settings: for instance, the IL distribution can be squeezed by increasing the fiber link stabilization servo loop gain. The OOL distribution, however, is fundamentally limited by the delay limit [2].

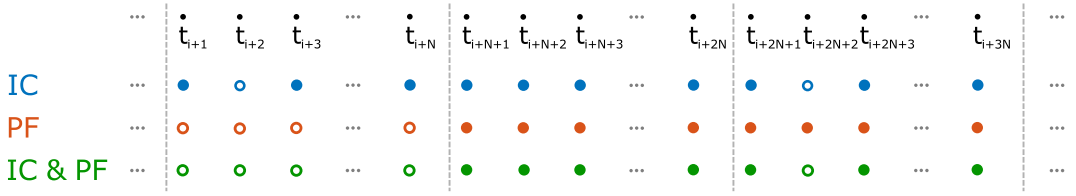

Figure 4. Merging of the IC and PF results to an overall validation flag (valid = solid symbols, invalid = open symbols): IC and PF are separately applied to the original data set. The results of the PF step, validating or invalidating intervals of N data points, are combined with the IC results by a data-point-wise logical AND operation. This may result in an exclusion of N samples due to only a few disturbed data points. An improved strategy in this respect is discussed in appendix B.

Natural quantities to assess the IFL uncertainty contribution u_{link} over these intervals are the fractional frequency offset $\langle y(t_i)\rangle$ and the Allan deviation, which is a measure for the statistical uncertainty contribution [30]. Constraining the IFL uncertainty contribution always requires balancing the contradicting aims of achieving low uncertainty levels, obtained by averaging over longer intervals, and the aim of achieving a high DC, accomplished by minimizing the amount of data removal using short interval lengths. In the following, we will characterize the variability of performance measures for different interval lengths using Π- and Λ-weighted averaging metrics [29, 30] and we will discuss this balance.

Figures 5 and 6 show the observed distributions of fractional frequency offsets and instabilities versus the interval length T in Π-weighted averaging metrics ($\langle y\rangle_{T,\Pi}$, $\sigma_y(T/4)$) and in Λ-weighted averaging metrics ($\langle y\rangle_{T,\Lambda}$, mod $\sigma_y(T/4)$), respectively. The different subfigures show the same type of quantities for the different campaigns. Hence, in the discussion to follow, it is most instructive for the reader to first concentrate on the data of one of the campaigns and analyze the differences between the campaigns afterwards.

Comparing the distributions for Π-weighted and Λ-weighted averaging metrics for the same length T reveals that the distributions of Π-weighted averaging metrics are located at values above the Λ-weighted distributions. This is the behavior expected from the different scaling already observable in figures 2(a) and (b). Analyzing the distributions in closer detail one notes:

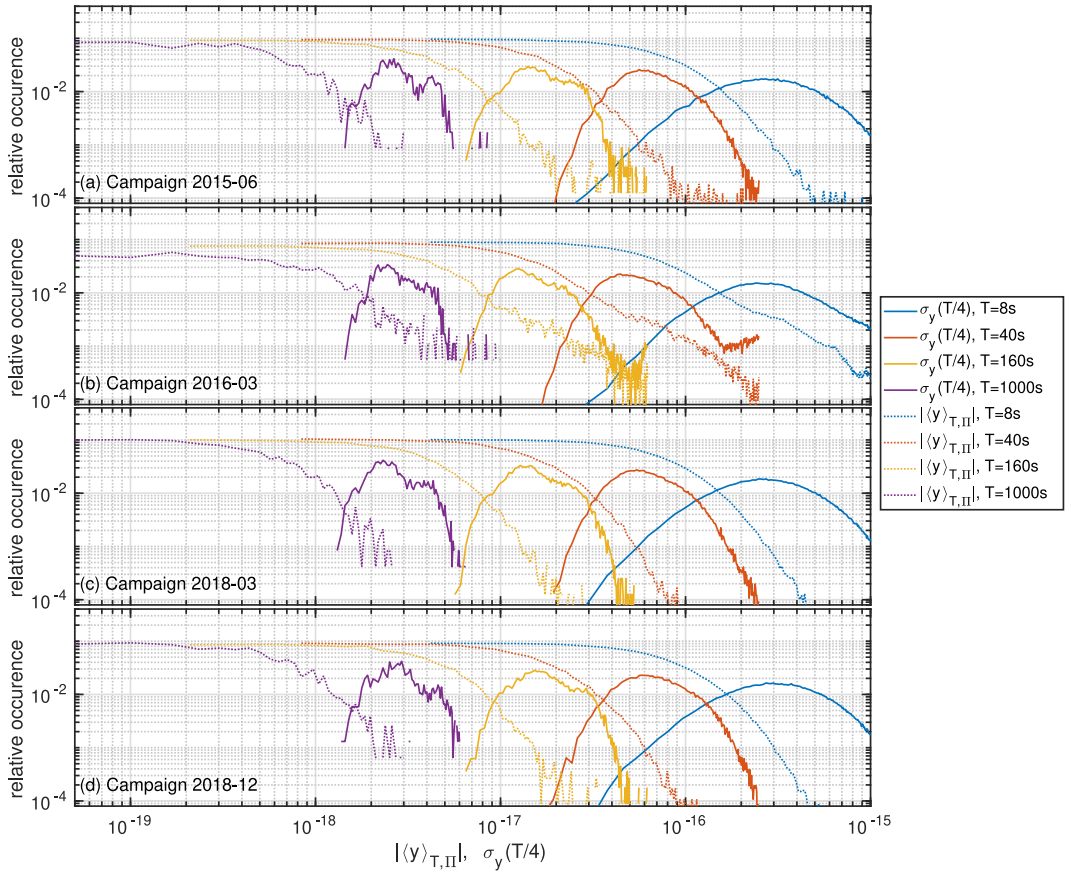

Figure 5. Distribution of the Π-averaged fractional frequency offset $\langle y\rangle_{T,\Pi}$ and the OADEV $\sigma_y(T/4)$ over data intervals of length T. The histograms are calculated over bins equally spaced on a linear grid and then plotted on a logarithmic scale. The data basis are all intervals which passed the IC without any point being marked as invalid.

(a) For Λ-weighted averaging metrics, the position of the distributions' peaks D_p largely match the progression of the modADEV for the continuous data set in figure 2(b). For short interval lengths $T \leqslant 160$ s, however, the peaks appear slightly below the corresponding modADEV $\mathrm{mod}\,\sigma_y(\tau = T/4)$ values, with ratios $D_\mathrm{p}/\mathrm{mod}\,\sigma_y(T/4) \approx 0.7$–$0.9$.

(b) The central parts of the distributions of the Π- and Λ-averaged offsets $\langle y\rangle_{T,\Pi}$, $\langle y\rangle_{T,\Lambda}$ are located left of the OADEV $\sigma_y(T/4)$ and modADEV $\mathrm{mod}\,\sigma_y(T/4)$ distributions, respectively. Hence, averaged over all data intervals $\langle y\rangle_{T,\Pi} < \sigma_y(T/4)$ and $\langle y\rangle_{T,\Lambda} < \mathrm{mod}\,\sigma_y(T/4)$ holds. Partly, this results from the fact that the effective averaging length is different for both quantities for a data set of given length, which can be compensated by extrapolating the Allan deviation from $T/4$ to T. For instance, in regions where white phase noise dominates (i.e. at short integration times), one can extrapolate the modADEV via $\mathrm{mod}\,\sigma_y(T) = \mathrm{mod}\,\sigma_y(T/4)/8$. This leads to a better match of the location of the distributions. The decrease of the slope of the OADEV and the modADEV with increasing averaging time τ already observable in figure 2(b) leads to a better match of the distributions of the Allan deviation and of the fractional frequency offsets for larger interval lengths T even without applying the extrapolation.

The meaning of these distributions become clear by comparison with demonstrated optical clock performance. Several optical atomic clocks have reached an estimated systematical uncertainty u_clock close to the 1×10^{-18} level or below [10, 15–20]. The instability of optical atomic clocks is subject to white frequency noise scaling at averaging times for which the lock to the atomic transition is active. The best demonstrated instabilities reach down to $\sigma_{y,\mathrm{clock}}(\tau) \approx 5 \times 10^{-17}/\sqrt{\tau/1\ \mathrm{s}}$ [19, 52, 53]. As already noted earlier [6, 7, 30], the distributions show again that a combination of Λ- and Π-averaging is required for a meaningful execution of optical clock comparisons. For a hypothetic comparison of these low short-term instability clocks over the considered IFL, for instance, figure 6 shows that $T = 160$ s is a good choice for Λ-averging length as then the observable offsets and modADEVs are mostly smaller than $\sigma_{y,\mathrm{clock}}(T)$.

In the following, we perform PF by threshold selection of the fractional frequency offsets and instabilities. For Π-weighted averaging metrics, we consider data intervals of averaging length T as valid

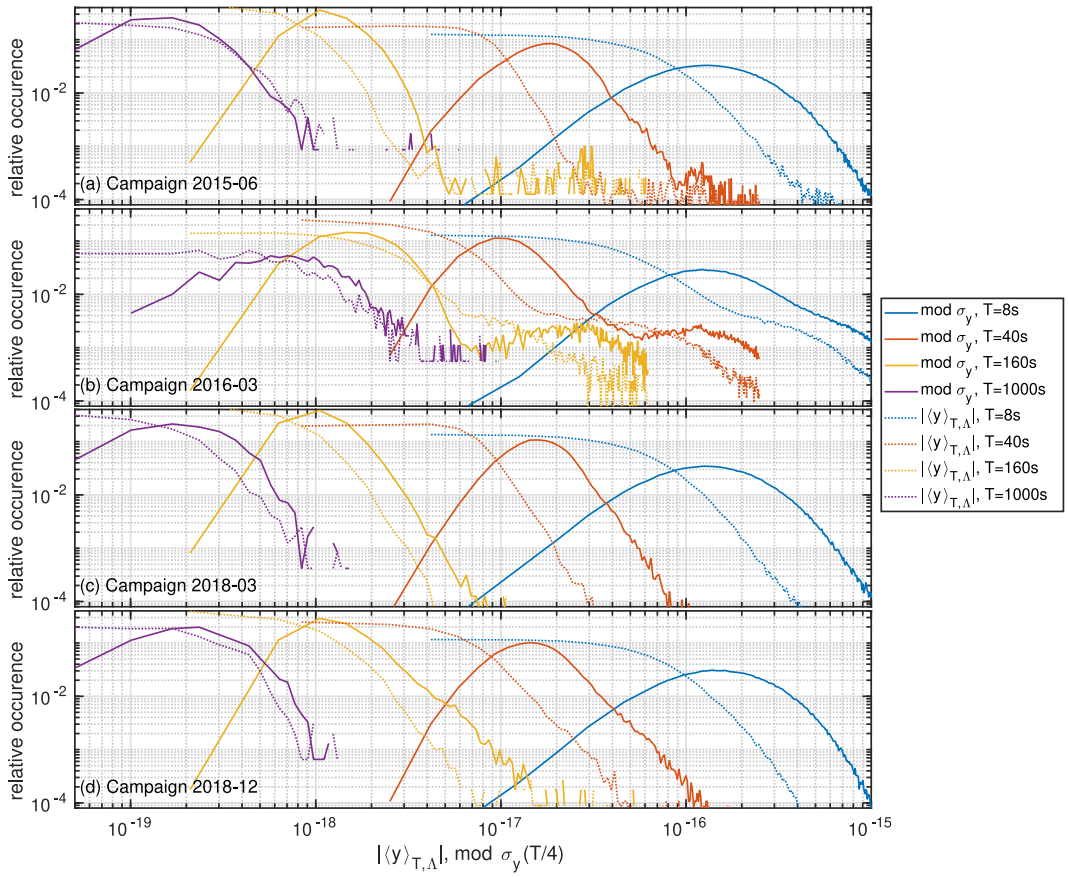

Figure 6. Distribution of the Λ-averaged fractional frequency offset $\langle y\rangle_{T,\Lambda}$ and the modADEV $\mathrm{mod}\,\sigma_y(T/4)$ determined from data intervals of length T. Except for the campaign 2016-03, where the IFL operation performance has been sub-optimal, the distribution of the performance measures are quite reproducible.

when the following criteria are fulfilled

$$|\langle y\rangle_{T,\Pi}| \leqslant y_{\mathrm{th},\Pi}(T,\xi) = \xi \times 1.5 \times 10^{-15}/(T/1\ \mathrm{s}), \tag{2}$$

$$\sigma_y(T/4) \leqslant \sigma_{y,\mathrm{thresh}}(T,\xi) = \xi \times 7.0 \times 10^{-15}/(T/1\ \mathrm{s}). \tag{3}$$

Similarly, the thresholds applied for Λ-weighted averaging metrics are:

$$|\langle y\rangle_{T,\Lambda}| \leqslant y_{\mathrm{th},\Lambda}(T,\mu) = \xi\sqrt{\left[2.5 \times 10^{-15}/(T/1\ \mathrm{s})\right]^2 + \left[1.0 \times 10^{-17}/\sqrt{T/1\ \mathrm{s}}\right]^2}, \tag{4}$$

$$\mathrm{mod}\,\sigma_y(T/4) \leqslant \mathrm{mod}\ \sigma_{y,\mathrm{thresh}}(T,\xi) = \xi\sqrt{\left[2.0 \times 10^{-14}/(T/1\ \mathrm{s})^{1.75}\right]^2 + \left[1.5 \times 10^{-17}/\sqrt{T/1\ \mathrm{s}}\right]^2}. \tag{5}$$

These threshold functions are chosen to approximate the position where the upper edge of the corresponding distributions have fallen to 1/10 of their maximum value for $\xi = 1$ (in campaigns 2015-06, 2018-03, and 2018-12). In this way, the multiplicative parameter ξ determines the tolerance of the PF step globally for all considered interval lengths T.

5. Properties of validated data sets

5.1. Data coverage of the validated data set

Figures 7(a) and (b) visualize the uncertainty constraints resulting from the PF threshold functions defined in equations (2)–(5). The percentage of data points marked valid after PF with these criteria is depicted in

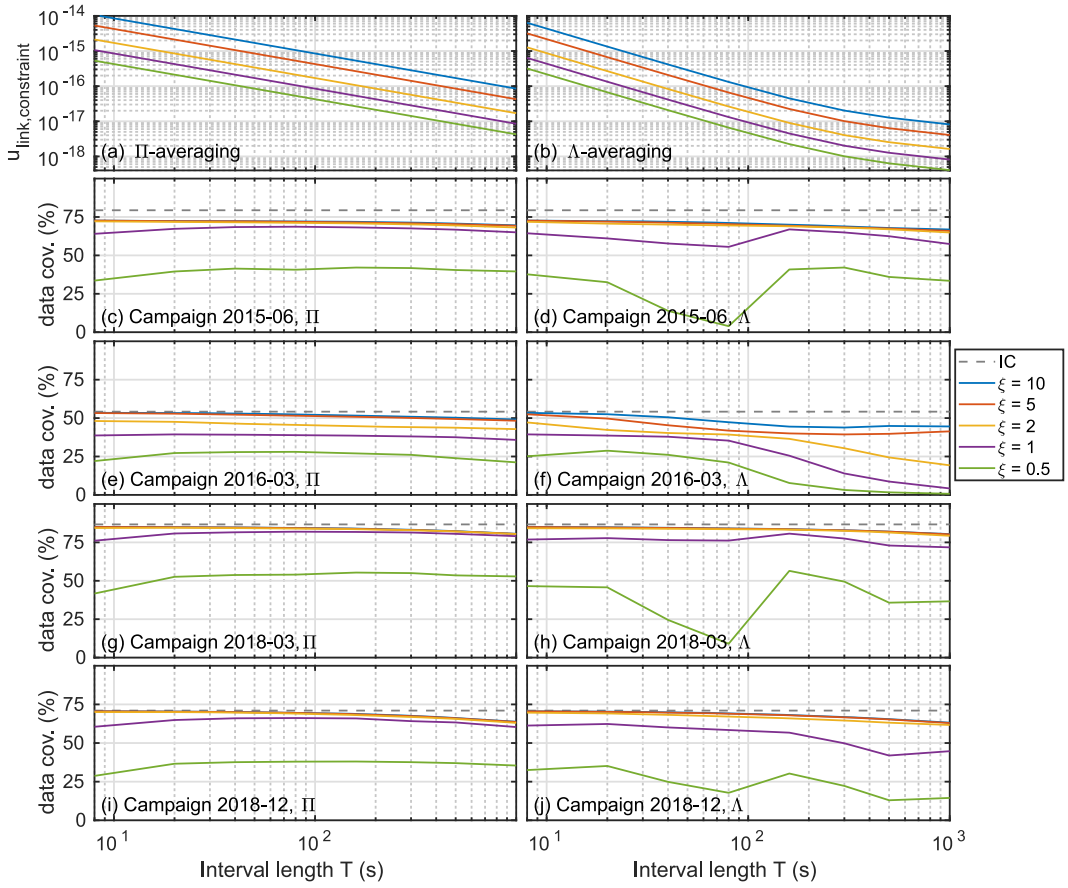

Figure 7. Visualization of the IFL uncertainty constraint imposed in PF for Π-weighted averaging ((a), $u^{\Pi}_{\text{link,constraint}}(T,\xi) = y_{\text{th},\Pi}(T,\xi) + \sigma_{y,\text{thresh}}(T,\xi)$) and for Λ-weighted averaging ((b), $u^{\Lambda}_{\text{link,constraint}}(T,\mu) = y_{\text{th},\Lambda}(T,\mu) + \mod \sigma_{y,\text{thresh}}(T,\xi)$) for different values of ξ. Subfigures (c), (e), (g), and (i) show the percentage of data points marked as valid achieved in the four campaigns using constraints shown in (a). The dashed gray line indicates the DC after IC only. Subfigures (d), (f), (h), and (j) show the DC achieved in the four campaigns using constraints shown in (b).

figures 7(c)–(j). Except for the data from the campaign 2016-03, for which deviating performance has already been visible in figure 6, the DC is only weakly dependent on the tolerance parameter ξ for $\xi \geqslant 2$. For these values of ξ, the majority of data points is already rejected by the IC. The DC after solely the IC is between 71.0% and 86.7% for campaigns 2015-06, 2018-03, and 2018-12.

For these campaigns and ξ values, the DC tends to decrease with interval length T for a fixed tolerance parameter ξ. For $T = 1000$ s, Λ-averaging, and $\xi = 2$, the PF only rejects 7.4% to 14.48% additional data points for constraining the IFL frequency transfer uncertainty to $u^{\Lambda}_{\text{link,constraint}}(1000\text{ s}, 2) \leqslant 2 \times 10^{-18}$. For $\xi = 1$ the amount of data rejected by PF starts to increase significantly (14.9%–26.3% for $T = 1000$ s) and values of $> 30\%$ are reached already for $\xi = 0.5$.

As a next step, we will consider the IFL frequency transfer uncertainty after validation. This consists of the systematic uncertainty contribution $u_{vl,\text{sys}}$ and the statistical uncertainty contribution. The former is determined from the validated data set by calculating the mean OOL frequency offset and the latter by calculating its instability at maximum averaging time.

5.2. Instability of the validated IFL data

In order to take a closer look at the statistical uncertainty contribution, we will now analyze the validate data set instabilities versus integration time τ. Figure 8 compares the instabilities for Π- and Λ-averaging metrics for different interval lengths T and $\xi = 2$ (blue lines) to the PF set instability constraints (orange symbols). As expected, we observe lower long-term instabilities when applying the lower constraints corresponding to longer interval lengths T. For instance, for $T \geqslant 160$ s and Π-averaging the achieved validated data set instability is typically sub-10^{-18}. However, the results for the campaign 2016-03 in figure 8(c) show that in some cases the instabilities stagnate at a higher level (e.g. $> 3 \times 10^{-18}$ for $T = 160$ s). This is different when applying Λ-averaging metrics and corresponding constraints. In this case, we achieve a validated data set uncertainty of $\leqslant 10^{-19}$ in all campaigns.

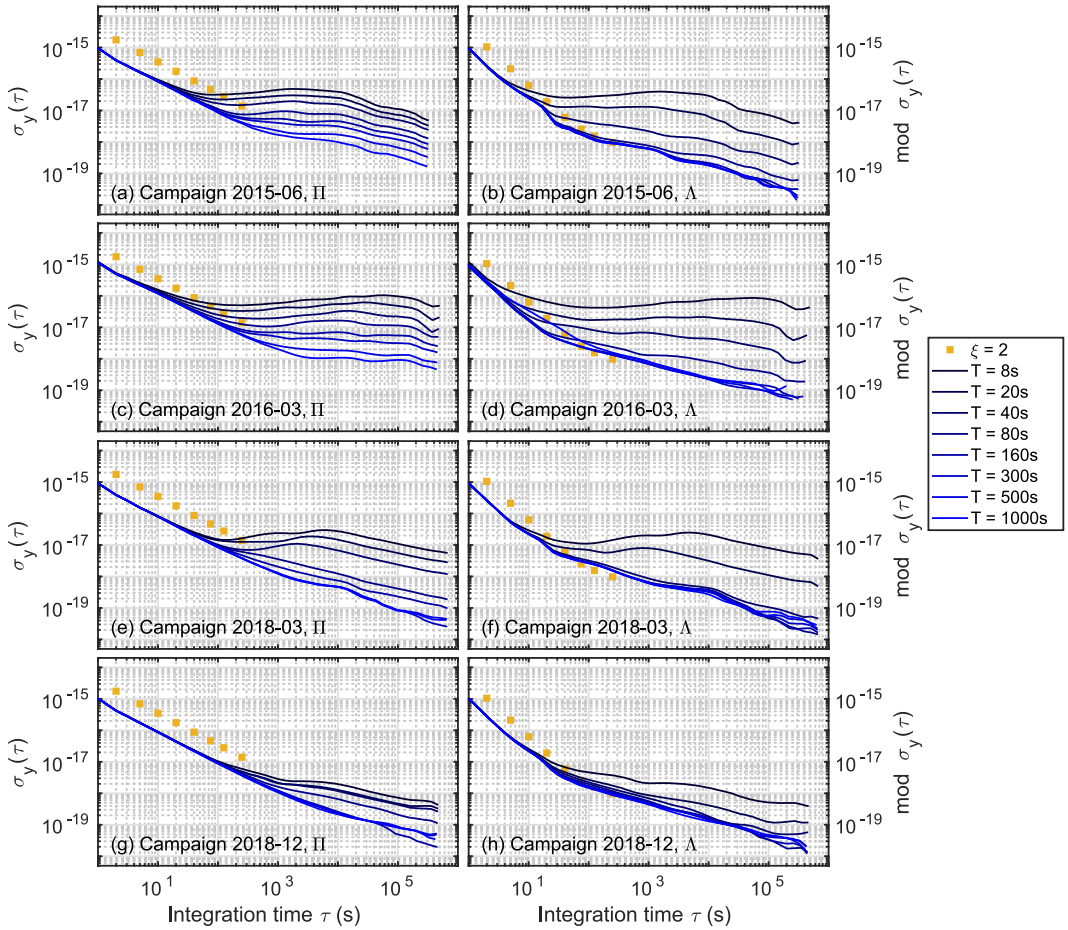

Figure 8. Instability of all points marked as valid after applying IC and PF versus interval length T for $\xi = 2$ (blue lines). The orange symbols show the instability constraint set by the PF.

For the two most recent campaigns including IFL improvements, the instabilities average down significantly beyond the PF constraints and reach values below 2×10^{-19} already for interval lengths $T \geqslant 80$ s irrespective of the averaging type. In these campaigns, the achieved instabilities for $T = 80$ s fall below the PF constraints by more than two orders of magnitude. This documents the effectiveness of the PF step.

In figures 8(d) and (f) for $T \geqslant 500$ s, the modADEVs of the validated data set are above the constraint mod $\sigma_{y,\mathrm{thresh}}(T, 2)$. This behavior is also observable for the other campaigns for $\xi \leqslant 1$. As a kink starting before $\tau = T/4$ is visible for the respective modADEVs, this phenomenon is related to the behavior of interrupted data sets already discussed earlier in conjunction with figures 2(e) and (f). The $\tau^{-1/2}$-scaling observable in figure 8 for large T and for averaging times $\tau = 10^2$ s to 10^5 s is also reminiscent of the behavior in figures 2(e) and (f).

5.3. Achievable IFL uncertainty contribution after validation

In the following, we will now combine the statistical uncertainty with the systematic uncertainty $u_{\mathrm{vl,sys}}$ for assessing the minimal achievable IFL uncertainty contribution $u_{\mathrm{vl}}(T, \xi)$ to a clock comparison. We conservatively assess the IFL uncertainty by adding the two contributions as if they are correlated:

$$u_{\mathrm{vl}}^{\Pi}(T, \mu) = |\langle y \rangle_{\tau_{\max}(T,\xi),\Pi}| + \sigma_y(\tau_{\max}(T, \xi)/4)\,, \tag{6}$$

$$u_{\mathrm{vl}}^{\Lambda}(T, \xi) = |\langle y \rangle_{\tau_{\max}(T,\xi),\Lambda}| + \mathrm{mod}\,\sigma_y(\tau_{\max}(T, \xi)/4), \tag{7}$$

where $\tau_{\max}(T, \xi)$ denotes the length of the respective validated data set.

These quantities are plotted as lines in figure 9 versus the interval length T and the tolerance parameter ξ. These curves show that in each of the four campaigns we are able to find a set of PF parameters yielding an overall uncertainty of the validated IFL data of less than $\approx 5 \times 10^{-19}$. Figure 9 confirms that applying Λ-averaging yields low uncertainties more reliably. Using Λ-averaging, uncertainties below 2×10^{-19} have been achieved for a wide range of PF parameters in three out of the four campaigns. In these campaigns, a

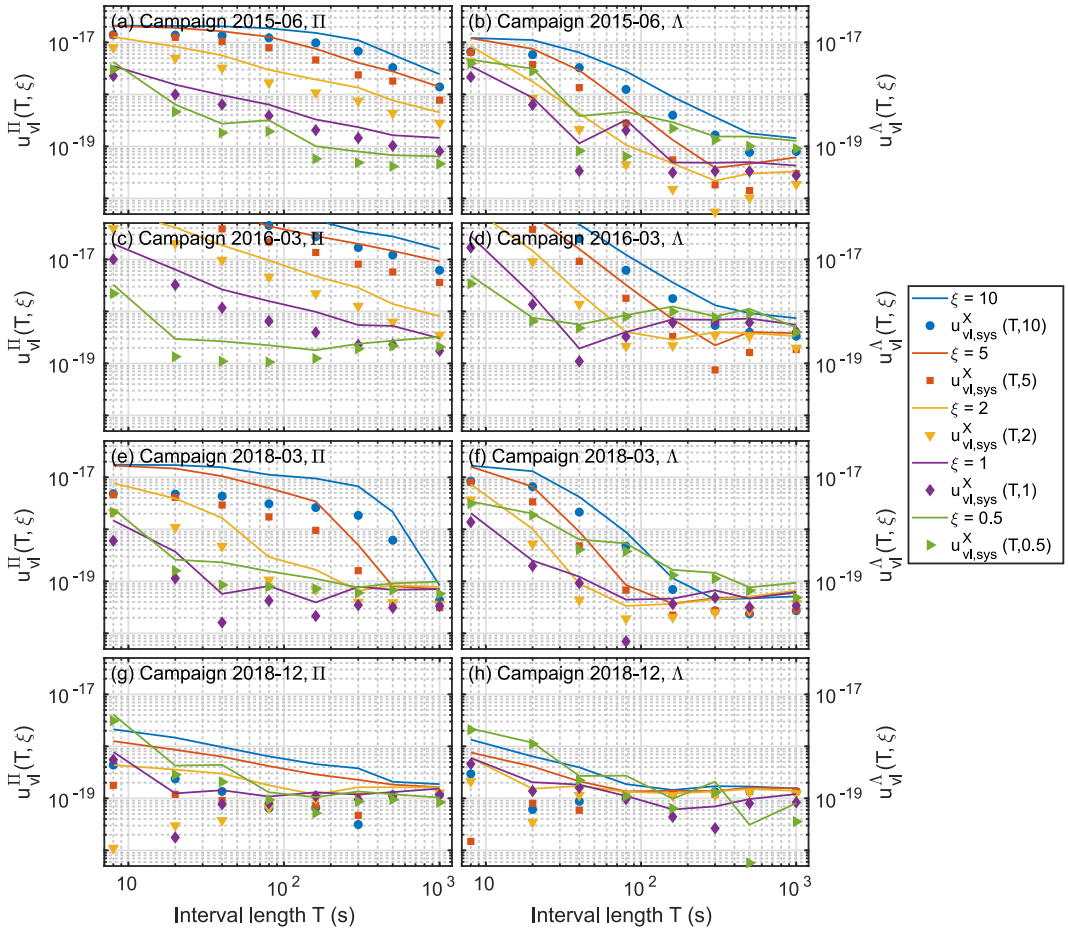

Figure 9. Systematic uncertainty (symbols) and total uncertainty (lines) of all points marked as valid after applying IC and PF versus interval length T and tolerance factor ξ.

good compromise between a validated data set uncertainty $u_{\mathrm{vl}}(T,\xi) \ll u_{\mathrm{clocks}}$ and high DC is achieved for interval lengths around $T \geqslant 100$ s. While for $T \geqslant 100$ s the achieved uncertainties for $\xi \geqslant 5$ still show a larger deviation in some cases, the achieved uncertainties for $\xi = 2$ are often in the range of the more stringent filtering criteria $\xi \leqslant 1$. Given the distinct decrease in DC observable in figure 7 for $\xi \leqslant 1$, tolerance parameter values of $\xi \approx 2$ appear optimal in terms of the DC versus overall uncertainty compromise.

The symbols in figure 9 show the corresponding systematic uncertainty contributions $u^{X}_{\mathrm{vl,sys}}(T,\mu) = |\langle y \rangle_{\tau_{\max}(T,\xi),X}|$ $(X = \Pi, \Lambda)$. In the early campaigns 2015-06 and 2016-03, the systematic and total uncertainty show the same progression with the systematic uncertainty typically being a constant factor below the total uncertainty. For the later campaigns 2018-03 and 2018-12, this relation between systematic and total uncertainty is not as distinctive as in the earlier campaigns. This might result from the achievement of slightly lower uncertainties for which other noise sources may dominate.

In appendix A, we present and discuss the results obtained using a gap-tolerant ADEV calculation routine for the statistical uncertainty estimation and for Π-averaging the data set validated employing Λ metrics. Both differences only lead to minor changes and do not change the conclusions drawn here.

Interestingly, both systematic and overall uncertainties show that applying the lowest tolerance parameter values ξ often does not result in achieving the lowest uncertainties for the complete validated data set. This shows that a careful analysis of the PF is mandatory for obtaining best filtering results.

6. Conclusion

Using the introduced two-step approach to interferometric fiber link validation, we are able to find validation parameters yielding a frequency transfer uncertainty of the validated data set below 5×10^{-19} for our ≈ 700 km long interferometric fiber link between Physikalisch-Technische Bundesanstalt and University of Strasbourg in all of the four considered campaigns. This has also been achieved for one campaign having suboptimal performance, which shows that our approach can also be applied in optical comparisons with a

more unfavorable relation between clock and interferometric fiber link uncertainty. In future, this might become reality when, e.g. comparing the current clock generation over even longer delay-limited IFLs or when comparing significantly improved clocks over existing IFLs. In these cases, more stringent data selection may be worthwhile compared to upgrading the IFLs in order to achieve better short-term performance. Using Λ-averaging based metrics, we are typically able to constrain the validated interferometric fiber link uncertainty to $\leqslant 2 \times 10^{-19}$ over a wide range of the PF parameters. In the corresponding PF parameter range, the coverage of the validated data set varies between 61% and 80%.

The interferometric fiber link uncertainty values discussed in this paper only quantify the lowest achievable uncertainty of the considered interferometric fiber link to a specific optical clock comparison in the sense that the mutual operation overlap with the remaining measurement chain will exclude further data points. This data selection will depend on the temporal statistics of the overlap of operation times and will alter the interferometric fiber link uncertainty contribution to the specific clock comparison.

In view of the steady progress in lowering optical clock uncertainties, it is highly desirable to improve the capability of interferometric fiber link validation performance in future. One direction in this respect may be to lower the uncertainty of the interferometric OOL characterization setup [45, 47, 48], which was found to be the uncertainty limitation that is encountered at the shortest integration times. A second worthwhile interferometric fiber link improvement already under way [46] is the enhancement of the uptime of IFLs. Minimizing the gaps between time spans of continuous operation will lower the data-selection-induced uncertainty limitations (cf figure 2). The mitigation of these data-selection-induced uncertainty contributions using estimators improved for interrupted data sets [54, 55] may enhance the conservative interferometric fiber link uncertainty estimation presented here.

We believe that the approach presented here will help to reliably perform optical clock comparisons to pursue the road-map towards redefinition of the SI second [56, 57] and for applications in relativistic geodesy [10, 19, 24, 25] or fundamental physics [26, 27].

Acknowledgments

We acknowledge the fruitful collaboration with Olivier Lopez, Nicolas Quintin, Anne Amy-Klein (Laboratoire de Physique des Lasers) and Etienne Cantin, Paul-Eric Pottie (LNE-SYRTE) that has led to the data set [50] analysed in this publication. We thank Thomas Waterholter for jointly operating the IFL in the campaigns and for his critical proof-reading of the manuscript. We are grateful to Thomas Legero and Uwe Sterr for operating the cavity-stabilized laser. Furthermore, we thank Paul-Eric Pottie and Jochen Kronjäger for general discussion on the task of IFL validation and acknowledge fruitful discussions with Sebastian MF Raupach, Christian Grebing, Christian Lisdat, Uwe Sterr, and Andreas Koczwara. This work has been partially funded by the European Metrology Programme for Innovation and Research (EMPIR) in project 15SIB05 (OFTEN). The EMPIR is jointly funded by the EMPIR participating countries within EURAMET and the European Union. This work has been partially funded by EMPIR TiFOON. This project 18SIB06 TiFOON has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme. This work has been partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the collaborative research centers SFB 1464—Project-ID 434617780—and SFB 1128 geo-Q (subproject A04).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.4046057.

Appendix A. Statistical uncertainty contribution evaluated by gap-tolerant estimates

In the main part, we analyzed the validated data set as if it was a contiguous data set and applied Π- and Λ-averaging metrics to assess the overall frequency transfer uncertainty. In contrast to this procedure, figure A1 shows the overall uncertainty using the gap-tolerant OADEV [33] as an estimate for the statistical uncertainty. This method does not rely on phase noise modelling for achieving a phase coherent filling of the gaps as it is done in more sophisticated methods [54, 55]. Furthermore, Π-averaging is now applied for determining the systematic uncertainty of the Λ-validated data sets in figures A1(b), (d), (f), and (h). This procedure matches the data processing in recent optical clock comparisons [14] where Π-averaging of all Λ_{1s} measurements along the chain has been employed.

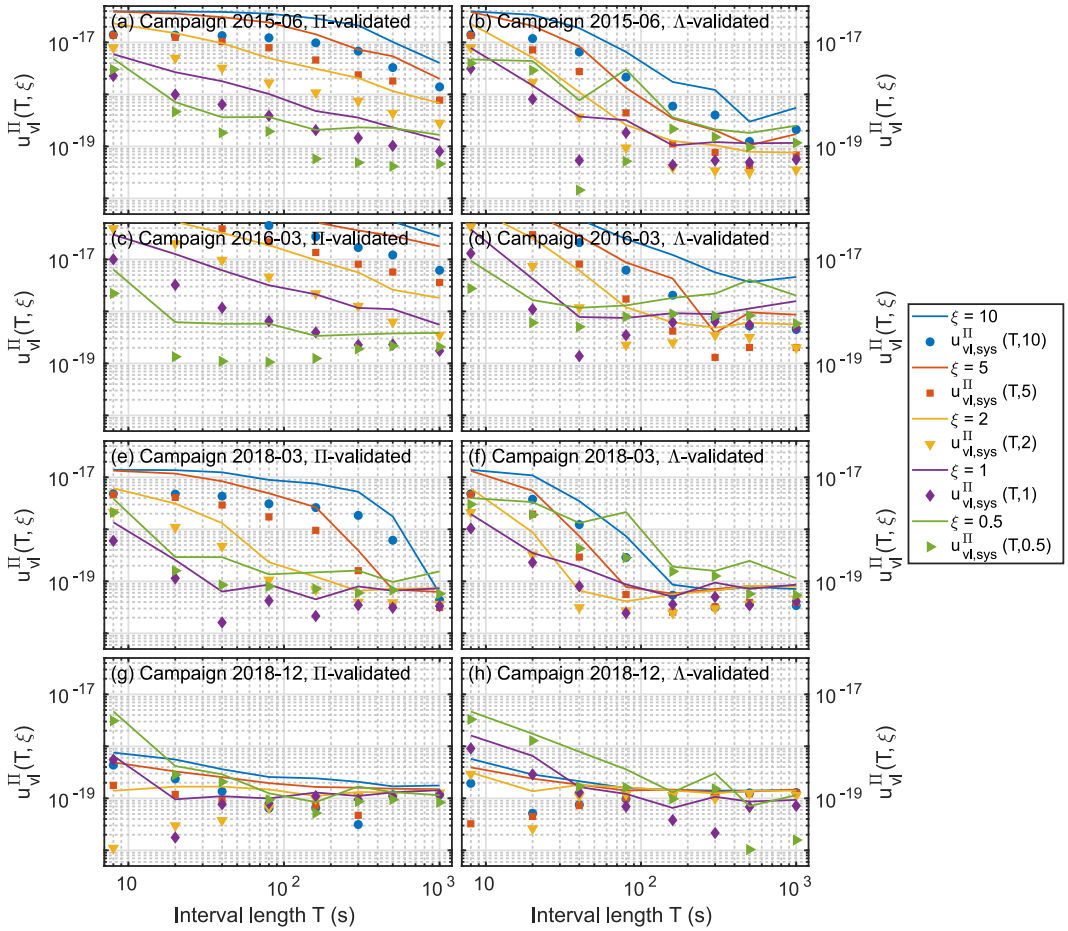

Figure A1. Π-weighted systematic uncertainty (symbols) and total uncertainty (lines) of all points marked as valid after Π-averaging-based (left column) and Λ-averaging-based (right column) validation.

Comparing the systematic uncertainties (symbols) in the right columns of figures 9 and A1 does not reveal pronounced changes between using Λ- or Π-averaging for systematic uncertainty determination of Λ-validated data sets. Also for the overall uncertainties (lines) including the statistical contribution only minor differences without any clear tendency are observable. The major findings of the main part remain unchanged.

We also checked for the results when estimating the statistical uncertainty of the Π-weighted mean of the validated data set using the auto-covariance-based treatment developed recently [34]. It turned out, however, that our IFL data show long-range correlations which results in a statistical uncertainty estimate depending on the cut-off parameter l_{cut} up to the maximum tested $l_{\mathrm{cut}} = 40\,000$ s (for a definition see discussion around equation (8) of [34]). Hence, even larger values of l_{cut} are required for a truthful statistical uncertainty estimation. Due to the computational costs associated with the auto-covariance calculation for large lags and the large data sets we analyze here, we did not pursue this path further.

Appendix B. A consecutive two-step strategy for improved data coverage

In the strategy for merging IC and PF validation results shown in figure 4, a few corrupted data points can lead to exceedance of the PF criteria in the corresponding interval and, hence, to an invalidation of N data points. We tested out whether consecutive application of the two validation steps as illustrated in figure B1 minimizes the impact of corrupted data points. The idea of this approach is to perform PF on a surrogate data set consisting of all data points passing the IC. The advantage of this approach is the mitigation of the impact of far off-lying outliers. However, the downside is that the phase-coherence is altered in the data set noted 'IC valid' in figure B1 due to the data selection (see figure 3). In the following, we analyze if one of these two effects dominates over the other.

Using the processing strategy from figure B1, the distributions of the fractional frequency offsets and the instabilities show the formation of tails at higher values with relative occurrences up to the low 10^{-3} range

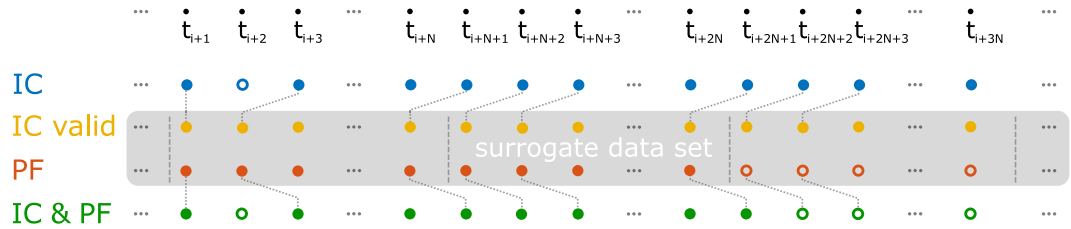

Figure B1. A strategy consecutively applying IC and PF: all points passing the IC are transferred to a new surrogate data set. The PF is then applied to the surrogate data set validating or invalidating data intervals of length N. The combined validations result of IC and PF is obtained in a third step where the PF results are mapped back to the corresponding data indices in the original data set using the inverse of the mapping IC $\rightarrow$ IC valid.

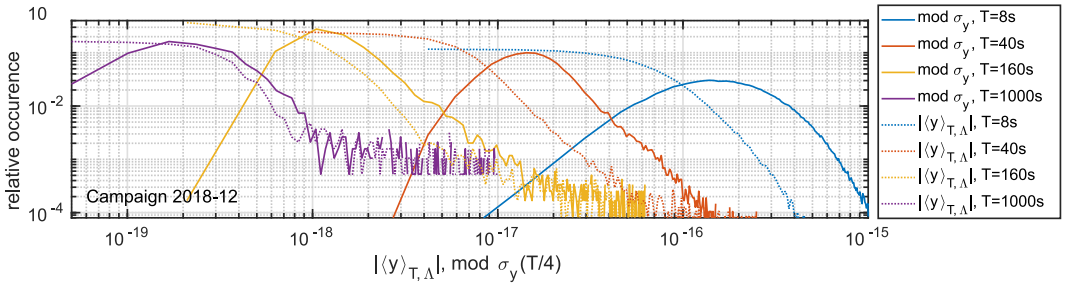

Figure B2. Distribution of the Λ-averaged fractional frequency offset $\langle y\rangle_{T,\Lambda}$ and the modADEV mod $\sigma_y(T/4)$ determined from data denoted 'IC valid' in figure B1.

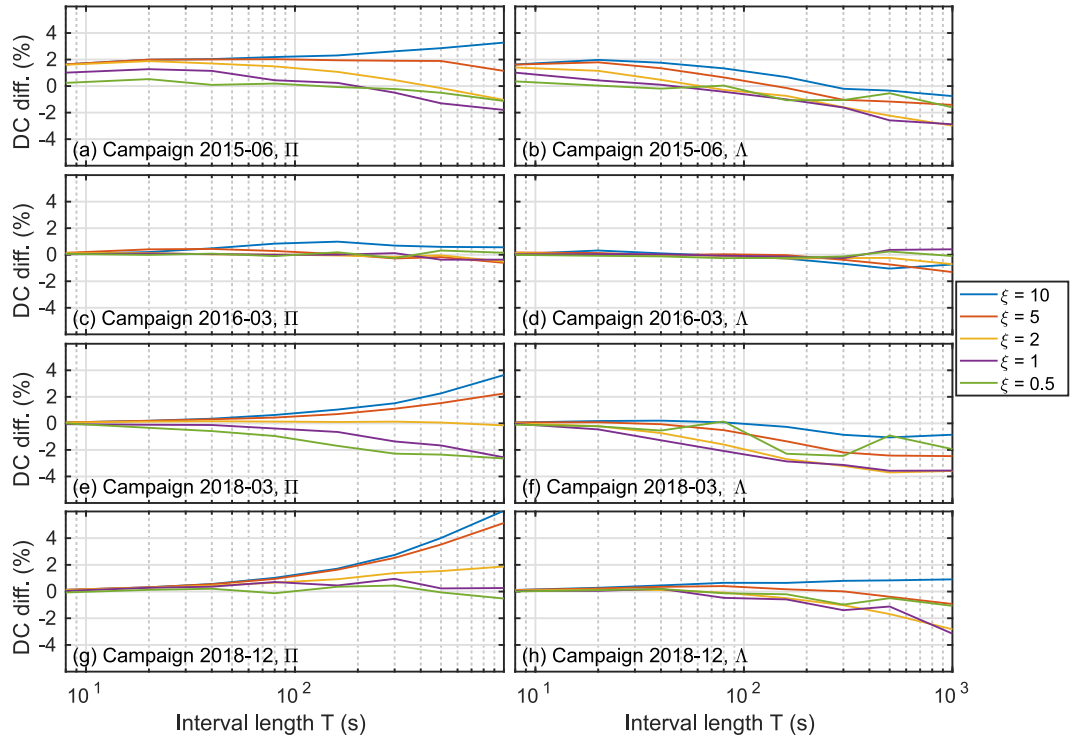

Figure B3. Difference in the DC between the two processing strategies. Positive values indicate a higher DC of the consecutive approach detailed in figure B1.

for $T > 8$ s. These are displayed in figure B2 for the 2018-12 campaign, as an example. The tails are more noticeable when applying Λ-averaging based metrics and are more prominent the larger the interval length T is. As these tails are significantly reduced in figure 6, these new features seem to result from the removal of data points before applying the PF.

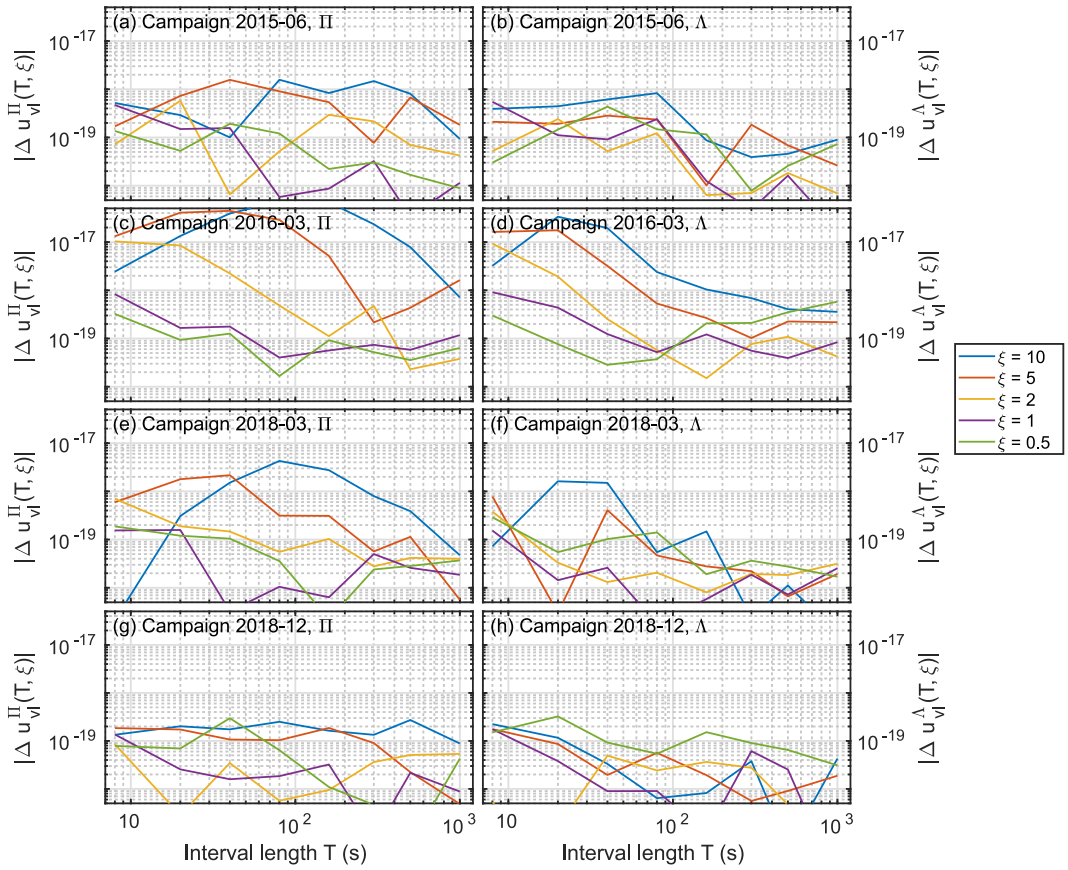

Figure B4. Difference of the validated data set uncertainties between the two processing strategies.

Figure B3 compares the DC after validation using the two different processing strategies. We observe DC difference between the two processing strategies within the range of ±6%. The sign of the difference changes and, hence, which of the two strategies is advantageous changes depending the PF criteria. One observable tendency is, however, that for Λ-averaging, interval lengths $T \geqslant 100$s, and $\xi = 2$—which we identified as the optimal PF criteria region in the main part—the non-consecutive processing yields higher DCs. This is probably related to the formation of the tails, i.e. to the loss of phase coherence impacting the Λ-averaging metrics stronger than Π-averaging metrics.

The difference of the achieved overall uncertainty of the validated data set between the two processing strategies is shown in figure B4. Comparison with figure 7 reveals that the differences are mostly at or below the level of the observed individual validated data set uncertainties. Hence, the uncertainty differences between the two processing strategies are not significant.

In conclusion, for the data of the IFL PTB–UoS studied here, there are only minor differences between the two processing strategies with a slight advantage for the non-consecutive strategy in the case of Λ-averaging metrics.

ORCID iDs

Sebastian Koke https://orcid.org/0000-0002-2393-5701
Erik Benkler https://orcid.org/0000-0001-7907-849X
Alexander Kuhl https://orcid.org/0000-0003-4306-8075

References

[1] Foreman S M, Holman K W, Hudson D D, Jones D J and Ye J 2007 *Rev. Sci. Instrum.* **78** 021101
[2] Williams P A, Swann W C and Newbury N R 2008 *J. Opt. Soc. Am.* B **25** 1284–93
[3] Calonico D *et al* 2014 *Appl. Phys.* B **117** 979–86
[4] Raupach S M F, Koczwara A and Grosche G 2014 *Opt. Express* **22** 26537–47
[5] Chiodo N *et al* 2015 *Opt. Express* **23** 33927–37
[6] Raupach S M F, Koczwara A and Grosche G 2015 *Phys. Rev.* A **92** 021801

[7] Koke S *et al* 2019 *New J. Phys.* **21** 123017
[8] Koller S, Grotti J, Vogt S, Al-Masoudi A, Dörscher S, Häfner S, Sterr U and Lisdat C 2017 *Phys. Rev. Lett.* **118** 073601
[9] Cao J, Zhang P, Shang J, Cui K, Yuan J, Chao S, Wang S, Shu H and Huang X 2017 *Appl. Phys.* B **123** 112
[10] Takamoto M, Ushijima I, Ohmae N, Yahagi T, Kokado K, Shinkai H and Katori H 2020 *Nat. Photon.* **14** 411–5
[11] Fujieda M *et al* 2018 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65** 973–8
[12] Jiang Z, Zhang V, Parker T E, Petit G, Huang Y-J, Piester D and Achkar J 2019 *Metrologia* **56** 025005
[13] Droste S, Grebing C, Leute J, Raupach S M F, Matveev A, Hänsch T W, Bauch A, Holzwarth R and Grosche G 2015 *New J. Phys.* **17** 083044
[14] Lisdat C *et al* 2016 *Nat. Commun.* **7** 12443
[15] Chou C W, Hume D B, Rosenband T and Wineland D J 2010 *Science* **329** 1630–3
[16] Ushijima I, Takamoto M, Das M, Ohkubo T and Katori H 2015 *Nat. Photon.* **9** 185–9
[17] Nicholson T L *et al* 2015 *Nat. Commun.* **6** 6896
[18] Huntemann N, Sanner C, Lipphardt B, Tamm C and Peik E 2016 *Phys. Rev. Lett.* **116** 063001
[19] McGrew W F *et al* 2018 *Nature* **564** 87–90
[20] Brewer S, Chen J S, Hankin A, Clements E, Chou C, Wineland D, Hume D and Leibrandt D 2019 *Phys. Rev. Lett.* **123** 033201
[21] Vermeer M 1983 *Chronometric Levelling* (*Reports of the Finnish Geodetic Institute No. 83:2*) (Geodeettinen Laitos: Geodetiska Institutet)
[22] Bjerhammar A 1985 *Bull. Geod.* **59** 207–20
[23] Takano T *et al* 2016 *Nat. Photon.* **10** 662–6
[24] Grotti J *et al* 2018 *Nat. Phys.* **14** 437
[25] Mehlstäubler T E, Grosche G, Lisdat C, Schmidt P O and Denker H 2018 *Rep. Prog. Phys.* **81** 064401
[26] Delva P *et al* 2017 *Phys. Rev. Lett.* **118** 221102
[27] Roberts B M *et al* 2020 *New J. Phys.* **22** 093010
[28] Lodewyck J, Le Targat R, Pottie P E, Benkler E, Koke S and Kronjäger J 2020 *Phys. Rev. Res.* **2** 043269
[29] Dawkins S T, McFerran J J and Luiten A N 2007 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **54** 918–25
[30] Benkler E, Lisdat C and Sterr U 2015 *Metrologia* **52** 565
[31] Droste S, Ozimek F, Udem T, Predehl K, Hänsch T W, Schnatz H, Grosche G and Holzwarth R 2013 *Phys. Rev. Lett.* **111** 110801
[32] Calosso C E, Clivati C and Micalizio S 2016 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **63** 646–55
[33] Sesia I and Tavella P 2008 *Metrologia* **45** S134–42
[34] Riedel F *et al* 2020 *Metrologia* **57** 045005
[35] Guéna J *et al* 2017 *Metrologia* **54** 348–54
[36] Koke S, Grebing C, Kuhl A and Grosche G 2016 Validating frequency transfer via a stabilised fibre link for optical clock comparisons (York) *European Frequency and Time Forum 2016*
[37] Xu D, Cantin E, Frank F, Quintin N, Meynadier F, Tuckey P, Amy-Klein A, Lopez O and Pottie P 2019 *IEEE Trans. Instrum. Meas.* **68** 1–6
[38] Tønnes M, Cantin E, Xu D, Frank F, Lopez O, Amy-Klein A and Pottie P E 2020 Scientific data processing of a fiber network for optical frequency transfer: methods and studies *IEEE Int. Frequency Control Symp. & IEEE Int. Symp. on Applications of Ferroelectrics 2020*
[39] Terra O, Grosche G and Schnatz H 2010 *Opt. Express* **18** 16102–11
[40] Lopez O, Haboucha A, Kéfélian F, Jiang H, Chanteau B, Roncin V, Chardonnet C, Amy-Klein A and Santarelli G 2010 *Opt. Express* **18** 16849–57
[41] Predehl K *et al* 2012 *Science* **336** 441–4
[42] Cahill J P, Zhou W and Menyuk C R 2017 *Appl. Opt.* **56** B18–25
[43] Bromage J, Winzer P J and Essiambre R J 2004 Multiple path interference and its impact on system design *Raman Amplifiers for Telecommunications 2: Sub-systems and Systems* (*Springer Series in Optical Sciences*) ed M N Islam (New York: Springer) pp 491–568
[44] Stefani F, Lopez O, Bercy A, Lee W-K, Chardonnet C, Santarelli G, Pottie P-E and Amy-Klein A 2015 *J. Opt. Soc. Am.* B **32** 787–97
[45] Akatsuka T *et al* 2020 *Opt. Express* **28** 9186–97
[46] Guillou-Camargo F *et al* 2018 *Appl. Opt.* **57** 7203–10
[47] Jürß T, Koke S and Grosche G 2020 Monolithic interferometer for accurate optical frequency dissemination *IEEE Int. Frequency Control Symp. & IEEE Int. Symp. on Applications of Ferroelectrics 2020*
[48] Cantin E, Tønnes M, Targat R L, Amy-Klein A, Lopez O and Pottie P E 2021 *New J. Phys.* 053027
[49] Kramer G and Klische W 2001 Multi-channel synchronous digital phase recorder *Frequency Control Symp. and PDA Exhibition, 2001. Proc. of the 2001 IEEE Int.* pp 144–51
[50] Koke S *et al* 2020 Data set from 'Combining fiber Brillouin amplification with a repeater laser station for fiber-based optical frequency dissemination over 1400 km' https://doi.org/10.5281/zenodo.4046057
[51] Greenhall C and Riley W 2003 Uncertainty of stability variances based on finite differences *Proc. 2003 PTTI Meeting* pp 267–80
[52] Oelker E *et al* 2019 *Nat. Photon.* **13** 714–9
[53] Schwarz R *et al* 2020 *Phys. Rev. Res.* **2** 033242
[54] Grebing C *et al* 2016 *Optica* **3** 563–9
[55] Galleani L and Sesia I 2019 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66** 1667–83
[56] Gill P 2016 *J. Phys.: Conf. Ser.* **723** 012053
[57] Riehle F 2017 *Nat. Photon.* **11** 25–31