Robustness tests for an optical time scale

Metrologia

Robustness tests for an optical time scale

OPEN ACCESS
IOP Publishing | Bureau International des Poids et Mesures Metrologia
Metrologia **59** (2022) 015002 (11pp) https://doi.org/10.1088/1681-7575/ac3801

Robustness tests for an optical time scale

V Formichella[1], L Galleani[2], G Signorile[1] and I Sesia[1,*]

[1] Quantum Metrology and Nano Technologies Division, INRiM, Turin, Italy
[2] Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy

E-mail: i.sesia@inrim.it

Received 7 July 2021, revised 18 October 2021
Accepted for publication 9 November 2021
Published 7 December 2021

Abstract

Optical clocks have reached such an impressive accuracy and stability that the future redefinition of the second will be probably based on an optical transition. Consequently, building time scales based on optical clocks has become a key problem. Unfortunately, optical clocks are still laboratory prototypes and are not yet capable of long times of autonomous operation. It is hence critical to understand the impact of this limited optical clock availability on the generated time scale. In this work, after describing a simple and effective optical time scale algorithm, based on the steering of a flywheel oscillator towards the optical clock, we investigate in detail the impact of the limited availability of the optical clock on the performances of the steering algorithm and of the generated time scale through numerical simulations. In particular, we simulate a time scale generated by a hydrogen maser (with a flicker floor of 5.5×10^{-16}) steered towards an optical clock, by considering six different scenarios for the availability of the latter, spanning from the ideal one, i.e. continuous operation of the optical clock, to the worst one, i.e. non-uniformly distributed frequency measurements with long unavailability periods. The results prove that the steering algorithm is robust and effective despite its very simple implementation, and it is capable of very good performances in all the considered scenarios, provided that the hydrogen maser behaves nominally. Specifically, they show that a time scale with an accuracy of a few hundreds of picoseconds can be easily realized in the ideal scenario, whereas in a more realistic scenario, with one measurement per week only, the time accuracy is nonetheless of a few nanoseconds, competing with the best time scales currently realized worldwide. The performances degradation due to a non-nominal maser behaviour is also discussed.

Keywords: time scale, steering algorithm, optical clocks, optical time scale, flywheel oscillator, robustness

(Some figures may appear in colour only in the online journal)

1. Introduction

Optical clocks have reached an impressive accuracy, surpassing by orders of magnitude that of microwave clocks [1], so that the optical transitions on which they are based can be considered as ideal candidates for a possible future redefinition of the second [2]. For these reasons, it is fundamental to develop an effective way to generate time scales based on optical clocks. However, optical clocks are still far from matching the robustness, reliability and long times of autonomous operation typical of microwave clocks: despite the impressive results achieved by some groups [3–5], the majority of the optical clocks under development in time laboratories all over the world, is still at the prototype level and only intermittently operated, with a relatively small total uptime. Therefore, an optical time scale is commonly built by steering the frequency of a master clock towards the optical clock [6]. The master clock, typically an active hydrogen maser (AHM), is usually referred to as a flywheel oscillator, since it allows the generation of the time scale also when the optical clock is unavailable.

It is crucial to understand the impact of the unavailability of the optical clock on the performances of the steering

* Author to whom any correspondence should be addressed.

algorithm and of the generated time scale. For example, it is fundamental to understand the minimum availability of the optical clock data needed to guarantee a given level of performances. To this aim, we simulated a time scale steered with an optical clock by considering several possible scenarios for the availability of the latter, as well as different scenarios for the stability of the flywheel oscillator. After modifying the steering algorithm described in [7], we simulated the measurement data of the frequency offset of an AHM with respect to an optical clock, and we used them to steer the AHM for the time scale generation. Then, we tested the algorithm by considering six different scenarios for the availability of the optical clock, spanning from the ideal case, where the optical clock operates continuously, to the worst one, where frequency measurements are non-uniformly distributed with long unavailability periods. We also repeated the simulations with different noise levels for the AHM clock model. We evaluated the performances by considering the time offset of the steered time scales with respect to an ideal time reference, and we compared and discussed the results obtained within the different scenarios. We also discussed the performances degradation due to a non-nominal maser behaviour, such as an unexpectedly high level of frequency noise or other non-stationarities.

To the best of our knowledge, such a detailed study of robustness has not been carried out yet for an optical time scale. In contrast with other published papers on the same topic [6, 8, 9], focussing only on a specific scenario for the availability of the optical clock, the main focus of our work is on the comparison among the different possible scenarios, including the more realistic ones, as well as on the robustness of the steering algorithm within such scenarios. A similar work has been presented in [10, 11], but on a reduced set of scenarios and with other relevant differences, including the chosen steering algorithm, the nature of the flywheel oscillator, and the absence of a time accuracy correction for the generated time scale.

The paper is organized as follows: the selected steering algorithm is described in section 2; the simulation of all the required data sets, including details on the clock model, is described in section 3.1, whereas the six simulated scenarios are presented in section 3.2; the test results are reported and discussed in detail in section 4 and, finally, the conclusions are presented in section 5.

2. The steering algorithm

A robust, simple and effective steering algorithm is the one described in [7], which is currently used for generating the official Italian time scale UTC(IT) [12]. It can be easily adapted to any possible type of steering reference, including an optical clock. With such an algorithm, the frequency steering correction Δf applied to the master clock is the sum of three terms, Δf_0, Δf_1, and Δf_2. The first one is the main component, correcting the frequency offset of the master clock with respect to the steering reference. The other two terms correct the residual frequency and time offset of the generated time scale with respect to a reference time scale, e.g. the coordinated universal time (UTC). In particular, Δf_2 provides time accuracy to the resulting time scale. In the original version of the algorithm presented in [7], the three components are computed as follows: Δf_0 is obtained by extrapolating to the current epoch a linear fit performed on N_{fit} days of past frequency offset data; Δf_1 is obtained as an average of the residual frequency offset of the time scale over a recent period; finally, Δf_2 is obtained by dividing the most recently measured time offset of the time scale by the number of days after which that offset has to be reduced to zero, indicated as N_{acc}. According to [7] and to the results achieved during the first year of autonomous generation of UTC(IT), with a time offset versus UTC smaller than 5 ns during 2020, the algorithm turns out to be highly performing even in its simplified form, where the Δf_1 component of the steering correction is set to 0. Therefore, the simulations discussed in the present work have been performed in the case $\Delta f_1 = 0$.

When the steering reference is an optical clock, the frequency offset of the master clock is measured with respect to such a reference, and the measured frequency data are used to compute Δf_0. In principle, the batch of past data used for the computation is regularly updated to include the most recent measurement results. However, when the optical clock becomes unavailable or, more in general, when the last N_{fit} days contain less data than a configurable threshold, the Δf_0 component of the steering correction is extrapolated from the last valid data batch, as further discussed in section 4.3.

Note that, theoretically, a simple linear fit is not always the best strategy to compute the steering correction. Indeed, under well-defined conditions, the optimal method depends on the master clock model and on the availability of the steering reference. Nonetheless, the selected algorithm is so simple and robust that, in case of critical applications in a real environment, it is an effective choice for those scenarios in which the optical clock uptime is very small, and in such scenarios it can be applied without substantial modifications.

When the uptime is instead high, the selected algorithm needs to be refined to fully exploit the high data availability, to take full advantage from the exceptional stability and accuracy of the optical clock. In particular, in the refined version of algorithm the Δf_0 component is obtained as $y_0 + dt$, where y_0 is the result of the last frequency offset measurement, d is an estimate of the linear frequency drift of the master clock, obtained by a linear fit on a window of N_{fit} days as previously defined, and t is the time elapsed from the measurement epoch (mid of the y_0 measurement period) to the correction epoch (mid of the Δf_0 correction period).

3. Simulated data and scenarios

3.1. Data sets simulation

This section discusses the simulation of all the necessary data sets, which are the input of the steering algorithm and of the tools for validating the performance of the obtained time scale. First, the clock model used for simulating the master clock is presented and validated in section 3.1.1. Then, the simulation of the frequency offset between the master

and the optical clock, and of the time offset between the master clock and an ideal time reference, is presented in section 3.1.2.

3.1.1. Clock model. The master clock to be steered is assumed to be an AHM. As a reference for the performances of a commercial AHM, the model CH1-75A manufactured by KVARZ has been considered. The main reason for this choice is that, among the various AHMs hosted at the time laboratory of the Italian National Institute of Metrological Research (INRiM), such model is the one which has been actually measured with respect to the ytterbium optical clock developed by INRiM, IT-Yb1 [13]. Hereafter, the selected maser is referred to as HM3, its code used at INRiM.

The nominal stability of HM3 has been experimentally evaluated on a period of almost 5 months, carefully selected to avoid any major clock non-stationarity or measurement issue. Figure 1 shows the Allan deviation (ADEV) of HM3, estimated with respect to UTC and its rapid version, UTCr, both on the original measurement data and on the detrended data, where a constant frequency drift has been estimated and removed. The measured frequency offset between HM3 and IT-Yb1 has been used to infer the noise coefficients of the clock model, which is necessary for simulating the data. The noise coefficients have been determined by the group working on IT-Yb1. The model includes four stochastic components, namely a white phase modulation (WPM), a white frequency modulation (WFM), a Flicker frequency modulation (FFM), and also a random walk frequency modulation (RWFM). The coefficients of these noise components are reported in table 1. The FFM is likely the most important stochastic component limiting the medium-to-long-term stability of the clock, as it can be deduced from the ADEV plotted in figure 1, where the two drift-removed curves flatten for averaging times larger than about 4×10^5 s. However, an RWFM has been also included in the model, with a conservative estimate of its coefficient, so that the robustness of the optical time scale can be tested under realistic conditions, in which the long-term stability of the AHM can be degraded by several factors, such as variable environmental conditions or other frequency non-stationarities. Figure 2 shows the ADEV of the stochastic part of the clock model (solid curve), along with the ADEV of all the individual noise components (dashed lines). Note that the complete clock model also includes a deterministic component, i.e. a second-degree polynomial affecting its time offset. This deterministic component corresponds to a linear frequency drift, whose value is also reported in table 1. The long-term ADEV of a simulated data series, after removal of the frequency drift, is also plotted in figure 2 (solid curve with dots), and it is in perfect agreement with the theoretical ADEV of the clock model.

Finally, in order to validate the clock model, the ADEV of the model and of the simulated data, taken from figure 2, are compared with the measured drift-removed ADEV, reported in figure 1, and with the clock specifications declared by the manufacturer [14]. As it can be observed from figure 3, the model is in agreement both with the specifications, which only give an upper limit for the ADEV (dashed curve),

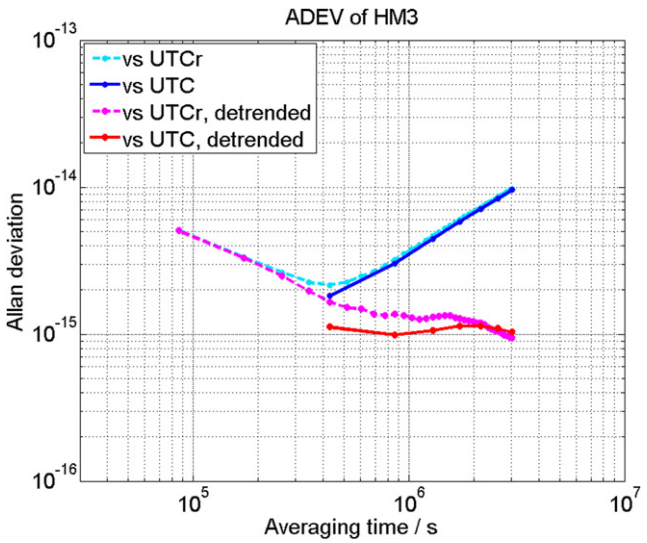

Figure 1. Stability of HM3 estimated versus UTC and UTCr, both on original and on detrended data.

Table 1. Noise coefficients and frequency drift of the HM3 clock model used for the simulations.

Stochastic Component	Coefficient (ADEV at 1 s)
WPM	1.5×10^{-13}
WFM	4×10^{-14}
FFM	5.5×10^{-16}
RWFM	1×10^{-18}
Deterministic component	**Value**
Linear frequency drift	5×10^{-16}/day

and with the measurements (dash-dotted curve). It can also be observed that the model is conservative with respect to the long-term stability, which is actually slightly worse than the measured one. With such observations, the clock model and the clock simulation tools can be considered as successfully validated.

The one described above, is the reference clock model for the simulations presented in this work. However, in order to cover a wider spectrum of possible realistic scenarios, we also performed simulations with different values for the coefficients of the stochastic components. First of all, we repeated all the simulations with a smaller coefficient for the RWFM, i.e. 2×10^{-19}, which is less conservative but probably more representative of a well-behaving AHM in a controlled environment, according to the observations carried out on the four masers hosted at INRiM's time laboratory (1 × KVARZ CH1-75A, 1 × T4S iMASER 3000, 2 × Microsemi MHM 2010). Moreover, we performed additional simulations and dedicated studies for some non-nominal AHM behaviour, whose results are presented in section 4.8. Specifically, we considered the effect of a larger coefficient for the WFM, and also the effect of a frequency jump, a common clock non-stationarity.

3.1.2. Frequency and time offset simulation. We used the clock model discussed in the previous section to simulate the

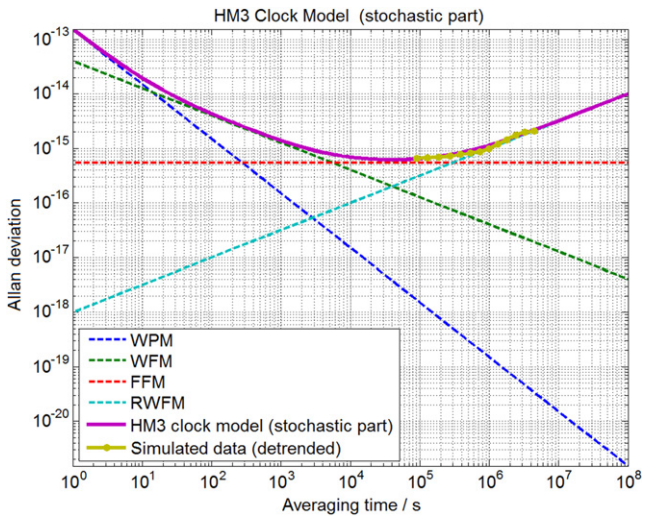

Figure 2. ADEV of the individual noise components of the clock model (dashed curves), of the whole stochastic part of the model (solid curve), and of a drift-removed simulated data series (solid curve with dots).

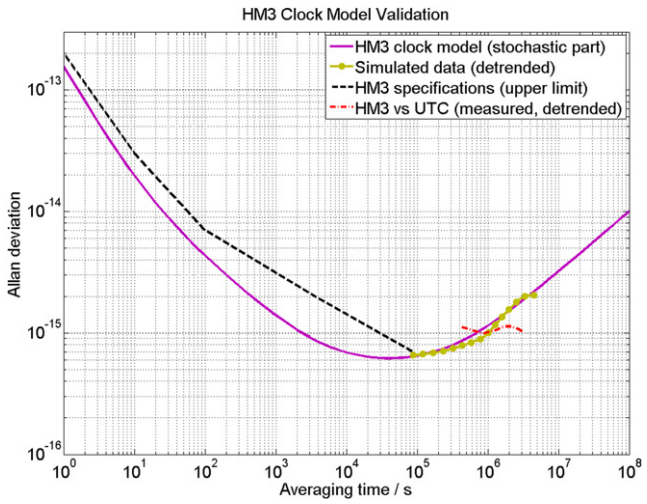

Figure 3. ADEV of the stochastic part of the clock model (solid curve) and of a drift-removed simulated data series (solid curve with dots), compared with the clock specifications (dashed curve) and with the measured stability (dash-dotted curve).

fractional frequency offset between HM3 and IT-Yb1. In particular, we simulated four five-months-long data series for each of the six scenarios described in section 3.2, including a frequency drift of 5×10^{-16} per day, corresponding to the upper limit declared by the manufacturer [14]. We used a sampling time of one day, so that each simulated data sample represents the average fractional frequency offset over one day, unless differently specified. The simulated frequency data are the fundamental input of the steering algorithm, used for the computation of the Δf_0 component of the steering correction. Note that the algorithm also requires, as an additional input, the uncertainty of the frequency measurements, in order to perform a weighted linear fit for computing Δf_0. However, for the simulations reported in this work, the same default uncertainty has been given to all the measurements, as all of them are average values obtained under the same conditions, over a period of time with the same duration. Therefore, the relative weight is just the same for all the measurements and has no influence on the final result of the linear fit.

Then, in order to compute the Δf_2 component and also to have a reference for the evaluation of the steered time scale, the time offset with respect to a reference time scale is also necessary, and we simulated it as follows. First of all, we assumed that IT-Yb1 can be considered as an ideal frequency reference, whereas UTC can be considered as an ideal time reference. The first assumption is justified by the far better stability and accuracy of IT-Yb1 with respect to HM3. The second one is justified if we think about a future UTC which will be steered via optical clocks instead of caesium fountains, and we compare it with a single local time scale. Moreover, we stress the fact that the validity of our assumptions is not impacting the general conclusions of this work, as our main objective is to test the robustness and compare the performances of optical time scales obtained within different scenarios, rather than a precise assessment of the time scale performances in a given scenario. Then, we simulated the free-clock time offset with respect to UTC by integrating the simulated fractional frequency offset between HM3 and IT-Yb1. Before integrating, we added a WFM measurement noise with ADEV of 1×10^{-15} at one day to the frequency data series. Finally, we simulated the free-clock time offset with respect to UTCr by interpolating the time offset versus UTC, reducing the sampling time from five days to one day, and adding a WPM noise with a standard deviation of 0.5 ns, which takes into account the fact that UTCr is less stable than UTC. The simulated time offset versus UTC is used to compute the time offset between the optical time scale and UTC, whereas the time offset versus UTCr is used to compute the Δf_2 component of the steering correction, as explained in [7]. Actually, Δf_2 can be computed also from UTC, but the use of UTCr gives comparable or better results, with the advantage of a faster response to any possible deviation of the optical time scale from UTC, thanks to the reduced latency of UTCr (one week instead of one month), as discussed in [7]. Therefore, we selected UTCr as the reference for the computation of Δf_2. Based on previously performed trade-off analyses and on the experience gained with the automated generation of UTC(IT) during 2020, we set the value of the corresponding configuration parameter N_{acc}, to 20 days.

3.2. Simulated scenarios

We considered six different scenarios for the availability of the optical clock, with four five-months-long simulated data series for each scenario, corresponding to 24 simulations for each value of the RWFM coefficient, i.e. the reference one reported in table 1 and the reduced one equal to 2×10^{-19}. Henceforth, the word 'scenario' will be used to distinguish among the main six scenarios, whereas the expression 'data set #x', with $x = 1, 2, 3, 4$, will be used to distinguish among the different simulations within a given scenario.

Table 2. Simulated scenarios.

Scenario	Description
Ideal	Daily data, no gaps, average over 24 h
Short measurement	Daily data, average over 2 h
Long gaps	Daily data, two long gaps (30 and 15 days)
Weekly data	Weekly data, on a fixed day of the week
Weekly data with jitter	Weekly data, with a jitter on the day of the week
Weekly data with long gaps	Weekly data, with two long gaps (30 and 15 days)

The basic scenario, according to the description given in section 3.1.2, has a sampling time of one day and no data gaps. With a 100% data availability, we named it 'ideal' scenario. The additional five scenarios have been derived from the ideal one, by adding data gaps and/or changing the sampling time to the four simulated data sets of the ideal scenario. Specifically: in the 'short measurement' scenario, a noise has been added to the data series of the ideal scenario, so that the resulting data represent a measurement performed over 2 h only, instead of 24; in the 'long gaps' scenario, two long data gaps have been added to the ideal scenario, namely a one-month gap corresponding to the 3rd month and a two-weeks gap at the beginning of the 5th month of the simulation period; in the 'weekly data' scenario, the data from the ideal scenario have been down-sampled to only one fixed day per week, corresponding to a constant sampling time of seven days; then, the 'weekly data with jitter' scenario has been obtained from the weekly data scenario by randomly displacing the day of the week in which the measurement is available, with an approximate Gaussian distribution, so that the day remains the same with 68% probability, it changes of ± 1 day with 27% probability, and of ± 2 days with 5% probability; finally, the 'weekly data with long gaps' scenario is a straightforward combination of the long gaps scenario with the weekly data scenario. For the ease of reading, the six simulated scenarios are summarized in table 2.

4. Test results

In this section we present the tests results in terms of time offset of the realized optical time scale versus UTC, for each of the six simulated scenarios listed in section 3.2, and we compare the results obtained within the different scenarios. In sections 4.1 to 4.6, as an example, we plot the results for only one of the four simulated data sets, namely the data set #4, whereas in section 4.7 we compare such results and, after defining a metric for assessing the performances of the time scale, we report the results for all the data sets and also an average value characterizing each scenario. Finally, in section 4.8 we show the results for two non-nominal AHM behaviours, namely, a large WFM and a frequency jump, and we compare them to the corresponding results obtained in the previous sections.

Note that the main configuration parameters, N_{fit} and N_{acc}, have the same values for all the simulations, namely 29 and 20 days, respectively.

4.1. Ideal scenario

Within this high-availability scenario we apply the refined version of the steering algorithm, as described in section 2. Note that, in this particular case, the value of the variable t is always one, so that Δf_0 is just equal to $y_0 + d$.

For the data set #4 and the reference RWFM value, figure 4 shows the simulated input frequency data (blue curve with dots) along with the daily computed Δf_0 (green curve with triangles), so that the agreement between the two data series can be directly observed. The effectiveness of the steering corrections is also confirmed by the good performances of the resulting time scale, which is plotted in figure 5. Specifically, figure 5 shows the time offset of the time scale obtained with the total Δf (red curve with dots), compared with the one obtained with Δf_0 only (grey curve with circles). As expected, there is a visible improvement in using the total steering correction rather than the Δf_0 component only. Indeed, adding the Δf_2 component prevents the time scale from slowly diverging from UTC by providing time accuracy to it, whereas Δf_0 provides frequency stability and frequency accuracy only. The difference would have been much more evident if a longer simulation period had been used.

Finally, figure 6 compares the time scale obtained with the reference RWFM value (red curve with dots, the same of figure 5) to the one obtained with the reduced RWFM value (green curve with squares). As expected, the latter is slightly better than the former, due to the higher predictability of the AHM with reduced RWFM. Such an improvement is more rigorously quantified in section 4.7, table 3, where we summarized the results from all of the four data sets for each of the six scenarios.

4.2. Short measurement scenario

As in all the other scenarios, also in the present scenario a new steering correction is computed and applied once per day. However, in this case, due to the shorter duration of the daily measurements (2 h instead of 24), the prediction error is slightly larger than in the ideal scenario, resulting in a slightly worse time scale, which is plotted in figure 7 (red curve with dots) and compared with the one obtained in the ideal scenario (grey curve with circles). The time scale obtained with the reduced RWFM in the short measurement scenario is also plotted for comparison (green curve with squares).

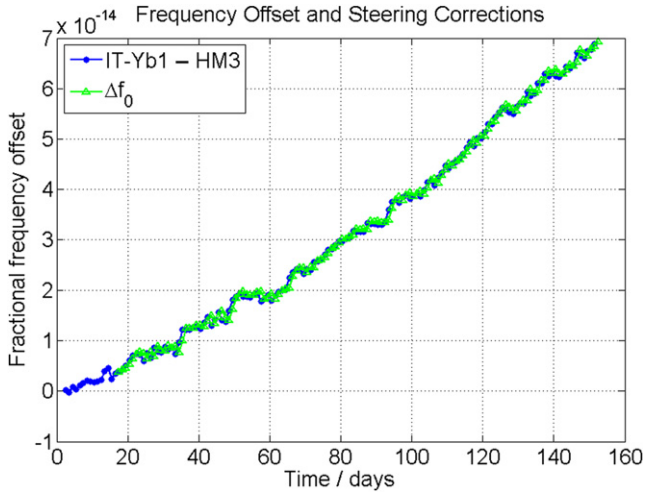

Figure 4. Ideal scenario: simulated input frequency offset data (blue dots) and daily computed Δf_0 (green triangles).

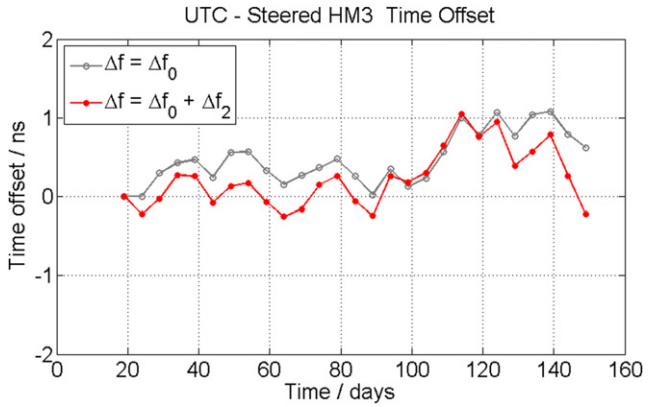

Figure 5. Ideal scenario: optical time scale obtained with $\Delta f = \Delta f_0 + \Delta f_2$ (red dots) compared with the one obtained with Δf_0 only (grey circles).

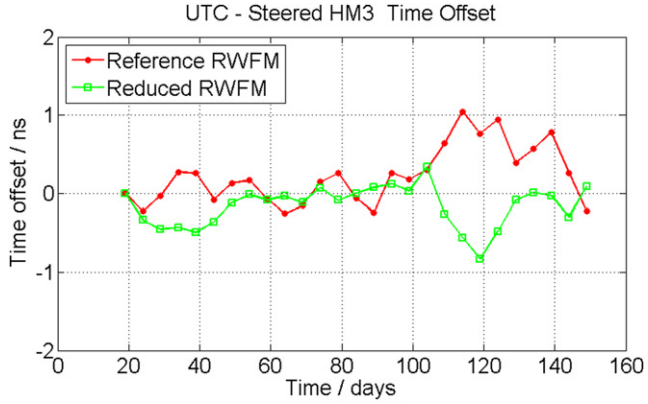

Figure 6. Ideal scenario: optical time scale obtained with the reference RWFM (red dots) compared with the one obtained with the reduced RWFM (green squares).

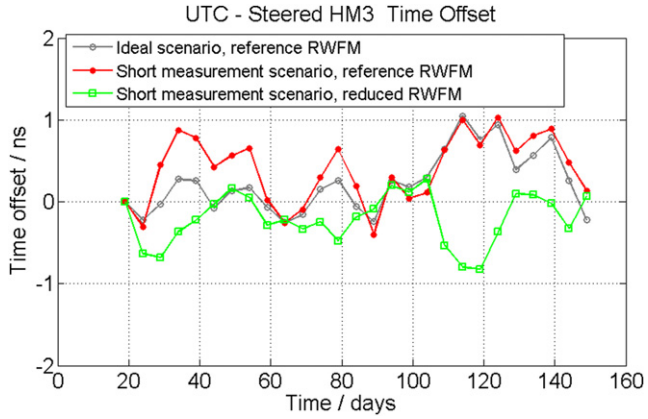

Figure 7. Short measurement scenario: optical time scale obtained with the reference RWFM (red dots) compared with the corresponding one from the ideal scenario (grey circles), and with the one obtained with the reduced RWFM (green squares).

4.3. Long gaps scenario

When long data gaps, or multiple short gaps, are present within the input data sets, another configuration parameter is important for the computation of Δf_0, i.e. the minimum number of measurements, $N_{\min}$, which must be included within the fitting window of length N_{fit}, to allow for a meaningful computation of the new steering correction. Indeed, until the data are regularly updated (no gaps), the fitting window slides forward each day, to include the latest measurement. However, when a data gap occurs, the fitting window stops sliding and, for the whole duration of the gap, the steering correction is linearly extrapolated from the data within that fixed window. Then, when the gap ends and new measurements become available, the fitting window is updated if the following condition is satisfied: the window ending with the most recent measurement and expanding backwards for N_{fit} days must include at least $N_{\min}$ data. Until such minimum number is reached, the Δf_0 component of the steering correction is obtained in a way that depends on the particular version of the steering algorithm: for the original version (presented in [7]), it is still extrapolated from the old window, as during the gap; instead, for the refined version presented in section 2, the old window is used only for the estimation of the frequency drift d, whereas the variable y_0 is immediately updated as soon as a new measurement becomes available after the gap.

For the simulations within the present scenario, in which, besides the long gaps, the optical clock uptime is still 100% as in the ideal scenario, we use the refined version of the algorithm. While the value of N_{fit} remains of 29 days, the value of $N_{\min}$ is set to 15. The simulation results for the data set #4 are shown in figure 8, where the two shaded regions indicate the two long gaps injected in the simulation period, with length of one month and two weeks, respectively. Figure 8 is the analogous of figure 7, comparing the time scale obtained in the present scenario (red curve with dots) with the one obtained in the ideal scenario (grey curve with circles), and also with the one obtained with the reduced RWFM in the long

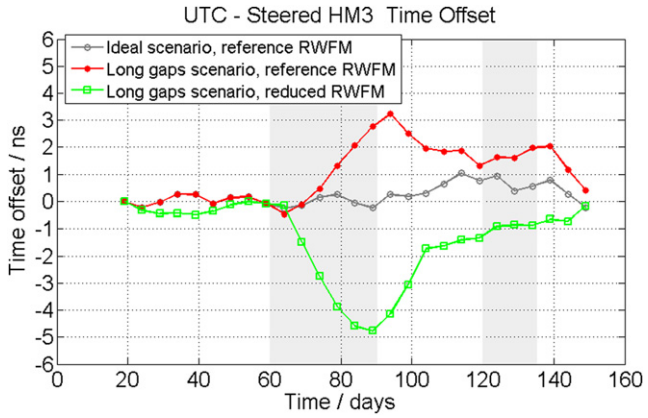

Figure 8. Long gaps scenario: optical time scale obtained with the reference RWFM (red dots) compared with the corresponding one from the ideal scenario (grey circles), and with the one obtained with the reduced RWFM (green squares). The shaded regions indicate the two data gaps.

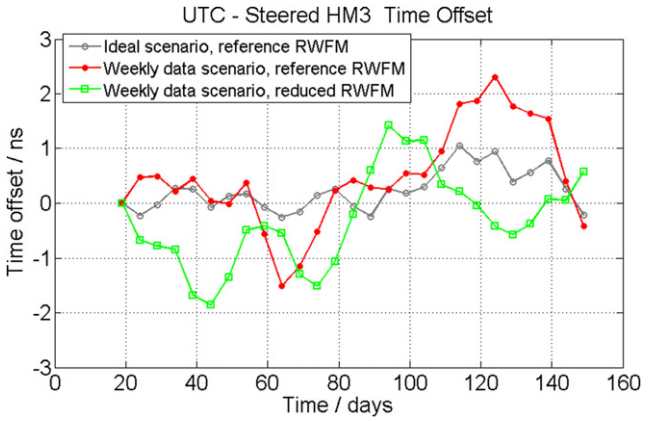

Figure 9. Weekly data scenario: optical time scale obtained with the reference RWFM (red dots) compared with the corresponding one from the ideal scenario (grey circles), and with the one obtained with the reduced RWFM (green squares).

gaps scenario (green curve with squares). Note that the prediction error increases with time during the gaps, due to the extrapolation procedure described above. Therefore, in figure 8, we can observe the time scale slowly diverging from UTC during the longer gap, until the updated steering correction stops such trend, thus allowing the Δf_2 component to bring the time offset back towards zero.

In order to reduce as much as possible the time offset accumulated during the gaps, it is necessary to perform a fine tuning of the configuration parameter N_{fit}. For example, an increase of N_{fit} might improve the prediction error, at least until the effect of the RWFM becomes relevant over the fitting window, and only if the frequency of the master clock does not suffer from other non-stationarities. Moreover, a larger N_{fit} means that the steering algorithm is less reactive to any possible change in the frequency drift of the master clock, eventually degrading the overall performances of the generated time scale. The value of 29 days chosen for the simulations turned out to be effective and a good trade-off between the different needs. Regarding the configuration parameter N_{min}, reducing its value would fasten the update of the steering correction (or of the frequency drift coefficient only, depending on the algorithm version) after a long gap, but the chosen value cannot be too small if a good estimate of d is required. In the present scenario, where the refined algorithm is used and y_0 is immediately updated at the end of the gap, there is no real need to set N_{min} smaller than 15, which is a value enabling a sufficiently good estimate of the frequency drift.

Finally, note that such a minimum value is the same used in all the simulations to initialize the optical time scale. Indeed, if the only relevant parameter was N_{fit}, 29 days would be necessary for the first computation and application of the steering correction, whereas, by implementing N_{min}, we can reduce the initialization period to 15 days only. This is evident e.g. from figure 4, where the first steering correction (green triangle) is actually plotted on day 16, or from all the following figures, where the time offset of the optical time scale is plotted starting from day 19, i.e. on the first day on which a five-days-sampled UTC measurement is available after day 16.

4.4. Weekly data scenario

In the weekly data scenario, the master clock is measured versus the optical clock only once a week, always on the same day of the week. Within this low-availability scenario we apply the original version of the steering algorithm, as described in section 2. With only one measurement per week, a 29-days-long sliding window contains only five measurements, assuming no data gaps. As a consequence, in the weekly data scenario, the value of N_{min} is reduced to 3. Figure 9 compares the time scale obtained in the present scenario (red curve with dots) with the one obtained in the ideal scenario (grey curve with circles), and also with the one obtained with the reduced RWFM in the weekly data scenario (green curve with squares).

4.5. Weekly data with jitter scenario

With respect to the scenario discussed in section 4.4, in the present one a jitter is added to the day of the week in which the measurement is available. As an effect of the probability distribution chosen for simulating the jitter (the time tag remains the same with 68% probability, it changes of ± 1 day with 27% probability, and of ± 2 days with 5% probability), the number of measurements in a 29-days fitting window can be even smaller than in the previous scenario, reaching a minimum of four. Note that there is no need to further lower the minimum number of measurements N_{min} which must be included within the fitting window, which was already set to three within the previous scenario. Figure 10 shows the resulting time scale (red curve with dots) and compares it with the one obtained in the weekly data scenario (grey curve with circles), in order to highlight the effect of the jitter on the final performances. The time scale obtained with the reduced RWFM is also plotted for comparison.

4.6. Weekly data with long gaps scenario

The present scenario is the straightforward combination of the weekly data and the long gaps scenarios. Figure 11 shows the resulting time scale (red curve with dots) and compares it with the one obtained in the weekly data scenario (grey curve with

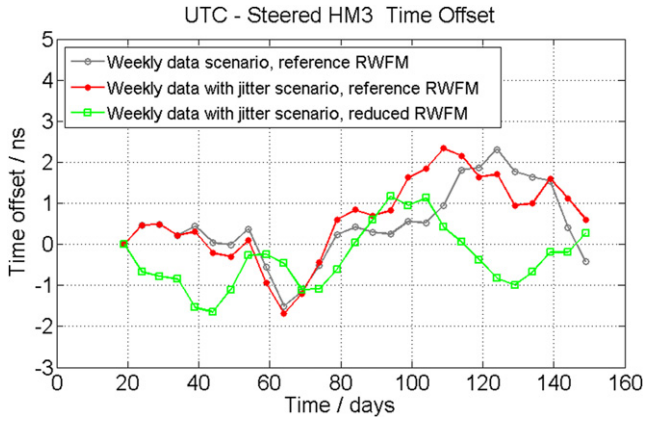

Figure 10. Weekly data with jitter scenario: optical time scale obtained with the reference RWFM (red dots) compared with the corresponding one from the weekly data scenario (grey circles), and with the one obtained with the reduced RWFM (green squares).

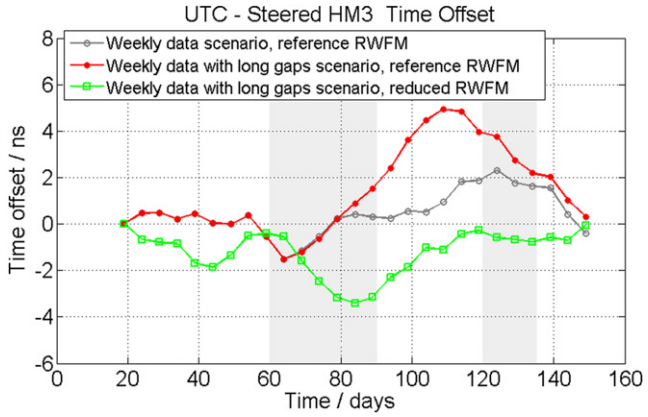

Figure 11. Weekly data with long gaps scenario: optical time scale obtained with the reference RWFM (red dots) compared with the corresponding one from the weekly data scenario (grey circles), and with the one obtained with the reduced RWFM (green squares). The shaded regions indicate the two data gaps.

circles), in order to highlight the performances degradation due to the addition of the long gaps. The time scale obtained with the reduced RWFM is also plotted for comparison.

Note that the red curve continues diverging from UTC for about 15 days after the end of the longer data gap: this is due to the fact that $N_{\min}$ measurements need to be accumulated after the gap before properly updating the steering correction, and this takes about 15 days. Theoretically, the same applies to the green curve, even if the effect is not visible in this particular simulation, probably masked by the Δf_2 component of the steering correction which, at the end of the gap, is larger for the green curve than for the red one.

4.7. Performances summary and comparison

First of all, a summary of all the graphical results shown in sections 4.1 to 4.6 is given in figures 12 and 13, respectively for the reference RWFM and for the reduced RWFM coefficient of the master clock, to allow a direct comparison of the time scales obtained within all of the different scenarios, with

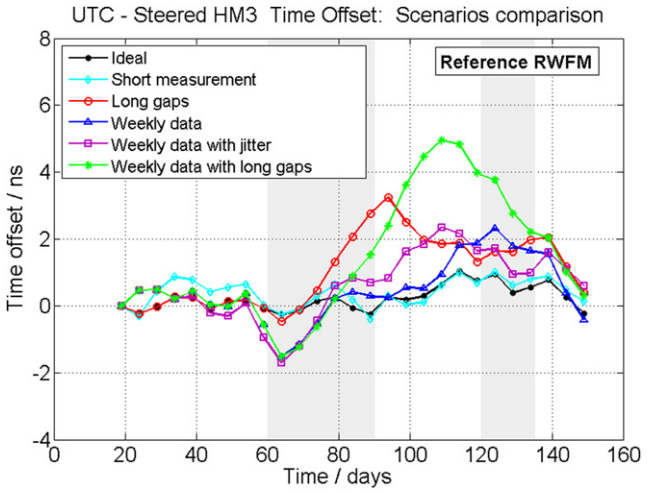

Figure 12. All-scenarios comparison with reference RWFM for the master clock.

the same data set #4. These figures are rich of useful information: the ideal scenario (black curves with dots) is the one leading to the best time-scale performances (i.e. the smaller time deviation from UTC), as expected; the two scenarios including long gaps (red curves with circles and green curves with stars) are those leading to the worst performances, because of the frequency prediction error increasing during the data gaps; the short measurement scenario (cyan curves with diamonds) gives results that are only slightly worse than those of the ideal scenario, suggesting that a short daily run of the optical clock is almost as effective as a continuous operation when the steering correction is updated and applied only once per day (see section 5 for further comments on this point); the weekly data and weekly data with jitter scenarios (blue curves with triangles and purple curves with squares, respectively) give similar results, suggesting that a realistic jitter like the simulated one does not degrade too much the resulting time scale; moreover, the blue curves (weekly data) remains within a couple of nanoseconds from the black curves (ideal scenario), suggesting that having only one measurement per week does not degrade too much the performances obtained with daily measurements, provided that the master clock does not suffer from anomalies during the dead time between two consecutive measurements (see section 4.8 for further comments on this point).

Then, the performances of the optical time scales obtained within the different scenarios, including all of the four simulated data sets, are evaluated in terms of the 95th percentile of the time offset with respect to UTC, over the whole five-months simulation period. The results are reported in tables 3 and 4, respectively for the reference RWFM and for the reduced RWFM coefficient of the master clock: for each scenario, the 95th percentile is given for each of the four data sets and, in the last column, the average of these four results is reported. It can be observed from the average values that, for the two scenarios with daily measurements, regardless of the simulated RWFM, the 95th percentile of the time offset remains well below 1 ns, which is a remarkable result. The low-uptime scenarios have of course worse performances but,

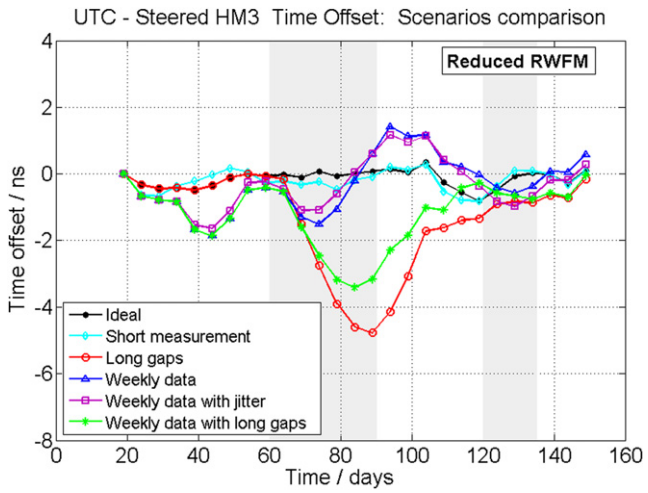

Figure 13. All-scenarios comparison with reduced RWFM for the master clock.

Table 3. 95th percentile of the time offset/ns of the optical time scale with respect to UTC, over the five-months simulation period, for all the scenarios and for the reference RWFM of the master clock.

Scenario	Set 1	Set 2	Set 3	Set 4	Average
Ideal	0.87	0.85	0.55	0.96	0.81
Short meas.	1.03	0.72	0.71	1.00	0.87
Long gaps	2.9	1.1	4.4	2.8	2.8
Weekly data	1.9	2.3	3.1	1.9	2.3
Weekly data with jitter	1.8	1.5	3.1	2.2	2.2
Weekly data with long gaps	6.9	7.5	6.1	4.8	6.3

nonetheless, with the only exception of the weekly data with long gaps scenario, the 95th percentile remains well below 3 ns, which is the same level of performances reached today by the best local realizations of UTC, UTC(k). The weekly data with long gaps scenario is the one giving the worst performances, as expected, as it is the one with the lowest optical clock availability. Nonetheless, even in this case, the 95th percentile remains below 7 ns.

As a final comment, let us consider the average values reported in the last column of the two tables. As expected, better results are obtained with the reduced RWFM, which makes the frequency of the AHM more predictable and easier to correct. However, this is not so evident for the long gaps scenario, where the average values are 2.8 ns with the reference RWFM and 2.9 ns with the reduced RWFM. The reason is twofold: first, it must be noted that those values are averages obtained from only four realizations of a stochastic process, so that the corresponding uncertainty is quite large; second, in the long gaps scenario, the 95th percentile of the time offset is determined by the large time offset caused by the one-month-long data gap, which basically depends on the quality of the prediction made at the last computation of Δf_0 before the gap. By chance, within a given simulation, such a prediction error can be larger than that obtained with a smaller RWFM, even if, on the average over a large number of simulations, it should be smaller.

Table 4. 95th percentile of the time offset/ns of the optical time scale with respect to UTC, over the five-months simulation period, for all the scenarios and for the reduced RWFM of the master clock.

Scenario	Set 1	Set 2	Set 3	Set 4	Average
Ideal	0.81	0.62	0.57	0.60	0.65
Short meas.	0.95	0.72	0.72	0.80	0.80
Long gaps	3.7	1.3	2.1	4.6	2.9
Weekly data	2.7	0.9	2.0	1.7	1.8
Weekly data with jitter	1.9	1.0	1.9	1.6	1.6
Weekly data with long gaps	3.5	2.5	3.4	3.2	3.2

Table 5. 95th percentile of the time offset/ns of the optical time scale with respect to UTC, over the five-months simulation period, for nominal and increased WFM of the master clock.

Scenario	Set 1	Set 2	Set 3	Set 4	Average
Ideal, nominal WFM	0.81	0.62	0.57	0.60	0.65
Ideal, increas. WFM	0.63	0.61	0.70	1.03	0.74

4.8. Non-nominal AHM: high WFM & frequency jumps

In this section we show what might happen to the optical time scale when the AHM suffers from a non-nominal behaviour. In particular, we considered two cases. In the first one, we evaluated the impact of a noise increase, such that the WFM coefficient becomes larger than the one declared by the manufacturer. After a general theoretical discussion of the problem, we verified our hypothesis with a WFM coefficient of 4×10^{-13}, which is an order of magnitude larger than the reference value reported in table 1. In the second case, we evaluated the impact of a frequency jump with the shape of a Heaviside step function affecting the AHM. Also in this case, we discussed the problem both from the theoretical point of view and by simulations, with particular attention to the relation between the uptime of the optical clock and the time offset accumulated due to the frequency jump. Specifically, we simulated the effect of a frequency jump with an amplitude of 1×10^{-14}, comparable with those occasionally observed in our laboratory.

In the first case, the increased WFM of course degrades the short-term stability of the optical time scale, at least up to an averaging time of the order of τ_0 days, where τ_0 is the time interval between two consecutive measurements of the AHM versus the optical clock, e.g. one day in the ideal scenario and seven days in the weekly data scenario. However, in the ideal scenario or, more in general, when the uptime of the optical clock is sufficiently high that the refined version of the steering algorithm can be applied, the increased WFM should not affect too much the time offset accumulated by the optical time scale over time periods larger than τ_0, thanks to the fact that the y_0 component of Δf_0 corrects the WFM averaged over τ_0. This has been verified by simulations in the ideal scenario, with a WFM coefficient of 4×10^{-13}, and the results are reported in table 5, in terms of the 95th percentile of the time offset with respect to UTC: as expected, the average over four simulations is fully consistent with the one obtained with the nominal WFM coefficient.

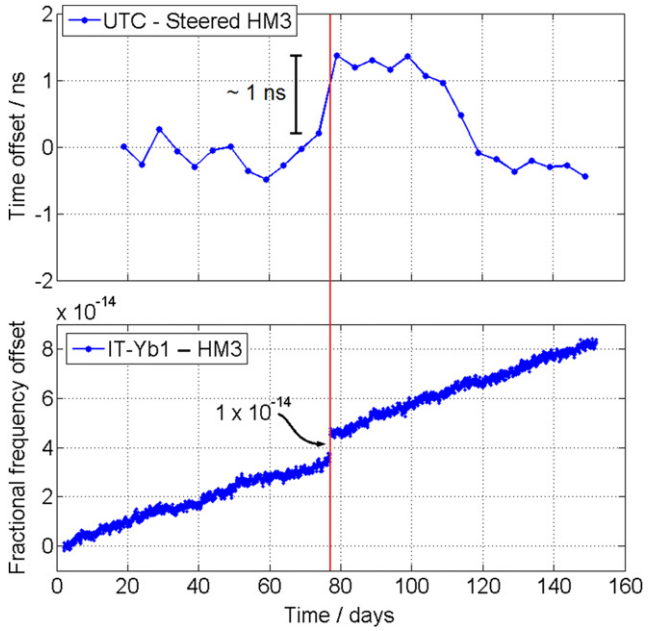

Figure 14. Ideal scenario: effect of a frequency jump on the optical time scale. Bottom: AHM frequency with jump. Up: optical time scale with accumulated phase offset due to the frequency jump.

We now consider the effect of the frequency jump. In the worst-case scenario, the jump occurs immediately after the latest application of a steering correction, so that the time offset of the time scale linearly diverges from the one in the jump-free scenario for the whole duration of the interval between two consecutive corrections. The accumulated phase offset due to the frequency jump is then $A\tau_0$, where A is the amplitude of the jump. Therefore, it is evident how the effect of the jump is inversely proportional to the rate with witch the frequency measurements are performed and the steering corrections are applied. The final effect also depends on the particular version of the steering algorithm, and on the possibility to detect the frequency jump, e.g. by using the clock anomalies detector described in [15]. As an example, we considered the ideal scenario and the use of the refined version of the steering algorithm, coupled with a frequency jump detector like the one described in [15], so that the estimation of the frequency drift, which is necessary for the computation of Δf_0, is not performed on the data batches including the jump (in that case, the algorithm uses the most recent value of d estimated before the jump). In this scenario, the phase offset due to the jump builds up for one day before the next steering correction is applied, with its y_0 component correcting the frequency jump. In particular, we simulated a jump with $A = 1 \times 10^{-14}$, so that we expect an accumulated phase offset of about 1 ns. The simulation results are shown in figure 14: in the bottom part of the figure, the AHM frequency is plotted along with the frequency jump indicated by the arrow, whereas in the upper part the optical time scale is plotted. As expected, as a consequence of the frequency jump, the time scale accumulates an additional time offset of about 1 ns, which is also highlighted in the figure. Moreover, note that the accumulated phase offset is recovered after the jump, thanks to the Δf_2 component of the steering correction.

5. Conclusions

We have presented the results of a series of robustness tests for a time scale algorithm based on an optical clock, by simulating several possible scenarios for its availability.

The results prove that the selected steering algorithm is robust and effective despite its very simple implementation. As expected, the scenarios giving the best and worst performances are the ideal one (continuous operation) and those with long unavailability periods, respectively. In particular, the optical time scale obtained in the ideal scenario shows a time accuracy of a few hundreds of picoseconds, even if the steering correction is computed and applied only once per day. Moreover, it is interesting to note that the realistic scenario with only one measurement per week, gives results which are nonetheless within a couple of nanoseconds from the ideal scenario. This is remarkable as it means that, even with a much smaller effort dedicated to the optical clock, the final performances of the time scale are already close to those of the best time scales currently realized worldwide with the classical methods. Even more interesting are the results obtained in the scenario in which the optical clock is operated only for a few hours per day, where the optical time scale shows a time accuracy of a few hundreds of picoseconds, as in the ideal scenario. Once again this is remarkable, as it means that, even with a smaller effort dedicated to the optical clock, the final performances of the optical time scale can be pushed to an unprecedented level of accuracy. As a final remark, we note that in the ideal scenario with 100% uptime, the hypothesis of computing and applying the steering correction only once per day, even if easier to implement in the laboratory, is not the most performing one. With the refined algorithm, in fact, it is possible to further improve the performances by using a more frequent computation and application of the steering correction, e.g. on an hourly basis.

Currently, at the INRiM time and frequency laboratory, we are working to integrate the IT-Yb1 optical clock as a steering reference within the time scale algorithm described in [7, 12], used for the generation of UTC(IT), by applying the same steering algorithm described in the present work. First, we will perform an off-line test of the algorithm by using the IT-Yb1 experimental data collected from October 2018 to February 2019 [13]. Second, we will test the algorithm in real time to generate an optical test time scale in parallel with UTC(IT), without affecting the official Italian time scale. Then, if the results will be at least as good as those presented in this article, we will actually use IT-Yb1 for the generation of UTC(IT) along with all the other available steering references, as described in [7, 12]. The results presented in this work will be useful to understand the effort that should be dedicated to the optical clock operation in order to improve the current performances of UTC(IT). More in general, we believe the results presented in this work will be useful to drive the development

of optical time scales in the near future, in any kind of time laboratory equipped with an optical clock.

Acknowledgments

This work has been carried out within the 18SIB05 ROCIT project, which 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. We are deeply grateful to the anonymous reviewers, whose comments played a fundamental role in improving the quality and accuracy of the presented results. We also thank Dr Marco Pizzocaro for the useful discussions on INRiM's optical clock IT-Yb1.

References

[1] Bloom B J, Nicholson T L, Williams J R, Campbell S L, Bishof M, Zhang X, Zhang W, Bromley S L and Ye J 2014 An optical lattice clock with accuracy and stability at the 1×10^{-18} level *Nature* **506** 71–5

[2] Lodewyck J 2019 On a definition of the SI second with a set of optical clock transitions *Metrologia* **56** 055009

[3] Lodewyck J *et al* 2016 Optical to microwave clock frequency ratios with a nearly continuous strontium optical lattice clock *Metrologia* **53** 1123

[4] Kobayashi T, Akamatsu D, Hosaka K, Hisai Y, Wada M, Inaba H, Suzuyama T, Hong F-L and Yasuda M 2020 Demonstration of the nearly continuous operation of an ^{171}Yb optical lattice clock for half a year *Metrologia* **57** 065021

[5] Baynham C F A *et al* 2017 Absolute frequency measurement of the $^2S_{1/2}$–$^2F_{7/2}$ optical clock transition in $^{171}Yb^+$ with an uncertainty of 4×10^{-16} using a frequency link to international atomic time *J. Mod. Opt.* **65** 585–91

[6] Hachisu H, Nakagawa F, Hanado Y and Ido T 2018 Months-long real-time generation of a time scale based on an optical clock *Sci. Rep.* **8** 4243

[7] Galleani L, Signorile G, Formichella V and Sesia I 2020 Generating a real-time time scale making full use of the available frequency standards *Metrologia* **57** 065015

[8] Grebing C *et al* 2016 Realization of a timescale with an accurate optical lattice clock *Optica* **3** 563–9

[9] Milner W R *et al* 2019 Demonstration of a timescale based on a stable optical carrier *Phys. Rev. Lett.* **123** 173201

[10] Yao J, Parker T E, Ashby N and Levine J 2018 Incorporating an optical clock into a time scale *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.* **65** 127–34

[11] Yao J *et al* 2019 Optical-clock-based time scale *Phys. Rev. Appl.* **12** 044069

[12] Formichella V *et al* 2020 Reliable and robust real-time time scale generation: developments and experimental results at INRiM *Proc. 2020 Precise Time and Time Interval Systems and Applications Meeting*

[13] Pizzocaro M, Bregolin F, Barbieri P, Rauf B, Levi F and Calonico D 2020 Absolute frequency measurement of the $^1S_0$–$^3P_0$ transition of ^{171}Yb with a link to international atomic time *Metrologia* **57** 035007

[14] CH1-75A active hydrogen maser specifications http://kvarz.com/general/1-75E.html (accessed 17 September 2020)

[15] Galleani L and Tavella P 2017 Robust detection of fast and slow frequency jumps of atomic clocks *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.* **64** 475–85