Offprint from "Metrologia"
Journal International de Métrologie scientifique — International Journal of Scientific Metrology
Internationale Zeitschrift für wissenschaftliche Metrologie. Vol. 7, No. 2, April 1971, Pages 79—82

Comments on the October 1970 Metrologia Paper "The U.S. Naval Observatory Clock Time Reference and the Performance of a Sample of Atomic Clocks"*

D. W. Allan and J. E. Gray

Atomic Frequency and Time Standards Section,
National Bureau of Standards, Boulder, Colorado 80302 USA

Received March 9, 1971

Abstract

The paper cited in the title by Winkler, Hall, and Percival (WHP) documents some important aspects of the widely referenced United States Naval Observatory (USNO) Time Scale system. We augment the text of WHP regarding the mutual synchronization to within 5 μs of the Time Scale, UTC(USNO), with another widely referenced time scale, the National Bureau of Standard UTC(NBS) Time Scale. We show that some of the information available to WHP can be utilized to give even more precision to the USNO Time Scale system and that some of that information has been improperly interpreted causing some errors by WHP in their conclusions regarding time scale precision, clock weighting factors, and drift in rate of the time scale.

1. Introduction

In the October 1970 issue of Metrologia, a description of the atomic time scale generation procedures at the United States Naval Observatory (USNO) was given by Winkler, Hall, and Percival (WHP) [1]. Because of the importance of the paper by WHP and its relevance to a large number of precise time and frequency users as well as to the Time and Frequency Division of the National Bureau of Standards (NBS), we offer the following comments.

2. USNO and NBS Time Coordination Effort

Our comments in this section as well as throughout our text are written in the spirit that while it is of significant practical importance to maintain, as far as possible, synchronization of all the distributed time scales, the optimum procedures for doing this are still evolving as new technical knowledge and experience become available. We believe, therefore, that open dialogue and criticism at the highest level of sophistication possible must be encouraged and carried on concurrently with a program to provide unambiguous properly-coordinated high-accuracy time scale services. We commend the work of the USNO and acknowledge the cooperation they have given in maintaining a "coordinated time scale". We offer the following comments as augmentation to the comments on page 126 of WHP regarding the same subject.

Two widely referenced time scales in the USA are UTC(USNO) and UTC(NBS). The USNO, along with their ponderous task of supplying astronomical information, generates the UTC(USNO) scale. One of the uses of the UTC(USNO) scale is as the common reference for precise time and time interval requirements within DOD[1]. The UTC(NBS) scale — because the NBS has the responsibility of the "custody, maintenance, and development of the national standards of measurement,"[2] — is generated for general public utilization and is directly related to the NBS standard of frequency and time. Because of the large number of users who, for their own convenience, reference both of these time scales and because of the desirability, if not necessity, of simply having one unified UTC time scale for distribution in the USA, the USNO and the NBS agreed to coordinate these two time scales starting 1 October 1968 (a date when UTC(USNO) and UTC(NBS) were nearly coincident) by keeping them synchronized to within $\pm 5\,\mu s$ [2]. Prior to that date there was a fractional frequency difference between the two scales of about 8×10^{-13} in such a direction that if USNO decreased the rate of its UTC scale by 4×10^{-13} and NBS increased the rate of its UTC scale by 4×10^{-13} the rates would be nearly identical; both organizations agreed to make this change so that synchronization could be maintained between the two scales. Since that date small incremental rate changes have been made on occasion by both organizations in equal amounts and in opposite directions with respect to the local atomic time scale of each so that the agreed upon synchronization could be perpetuated. Two such incremental rate changes are indicated by the second and third equations of WHP.

The question of optimum procedures for maintaining synchronization between clocks and time scales in the future must be left flexible and open to discussion. It is clear that the need for uniformity is of paramount practical importance; it is also clear, however, that it is of the utmost scientific importance to maintain accurate documentation of rates and dispersion of scales with reference to the defined unit of time interval.

3. USNO Time and Frequency Stability Considerations

A. Stability Measures

WHP have used the IEEE-recommended time domain measure of frequency stability (IEEE Subcommittee on Frequency Stability, [3]) throughout the text; however, they have used an undefined fre-

* Contribution of the National Bureau of Standards, not subject to copyright.

[1] DOD (Department of Defense) Directive 5160.51 of 1 February 1965.

[2] Title 15 United States Code Section 272.

quency domain measure (see Fig. 3 of WHP) different from that recommended [3].

The equation by WHP on p. 127 defining a "good measure" for time dispersion is approximately a factor of 1.3 too optimistic for flicker noise frequency modulation[3] — a common noise found in frequency standards [4, 5, 6] (also see Fig. 2 of WHP). For a clock which is perfectly calibrated the defined measure is valid only for white noise frequency modulation [7] — another common noise process encountered in cesium beam frequency standards.

Concerning the WHP claim that "there is no simple relationship" for $\sigma_{\frac{\Delta f}{f}}(N,\tau)$ and time dispersion[4], there is a simple relationship for white noise frequency modulation [6], namely:

$$\sigma_{\frac{\Delta f}{f}}(N,\tau) = \sigma_{\frac{\Delta f}{f}}(2,\tau) \tag{1}$$

and hence

$$\sigma_{\Delta t}(2,\tau) = \tau\sigma_{\frac{\Delta f}{f}}(N,\tau)\,. \tag{2}$$

Equation (1) is experimentally verified in the first row of Table 6 of WHP.

B. Reliability, Optimum Stability, and Weighting Factors

The following equation may be easily derived for a set of n independent standards

$$\sigma^2 = \sum_{i=1}^{n} w_i^2\sigma_i^2\,, \tag{3}$$

where σ is a stability measure of the weighted collection, σ_i is the corresponding stability measure of the i^{th} unit, and w_i is a normalized weighting factor for the i^{th} unit. The w_i may be picked by any criteria. We would like to point out that WHP do use unequal weighting factors. They simply round off to relative weights of 0 or 1 given by the following:

$$w_i = \begin{cases} 0, \text{ if clock is rejected,} \\ 1/m, \text{ if clock is not rejected,} \end{cases} \tag{4}$$

where m is the number of clocks remaining after the rejection routine has been performed. Combining equations (3) and (4) gives

$$\sigma_{(1/m,0)\text{weight}} = \frac{1}{\sqrt{m}}\sigma_{\text{rms}} \tag{5}$$

which is similar to the WHP equation on p. 128; however, the clocks need not have "nearly similar performance". If one chooses optimum weighting[5] of the clocks one obtains

$$\sigma_{\text{optimum weight}} = \left[\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right]^{-1/2}. \tag{6}$$

There are four points we wish to make regarding equations (5) and (6) as they pertain to the WHP text.

First, from equation (5) the stability of the "mean" improves as the square root of the number of unrejected clocks and *not* as WHP have calculated, i.e., as the square root of the total number of clocks potentially composing the mean (see WHP, p. 128). Because of the rejection criteria of WHP, the effective number of clocks m is only about nine or ten (see WHP, p. 133), and not the potential of 16 actually available. This results in a figure of merit, as defined on page 128 of WHP, of about 4 instead of "5.2." A figure of merit of better than 5.2 could be achieved by using the optimum weight approach and the same 16 clocks. This latter approach would be equivalent to adding at least eight additional similar clocks (cesium beam frequency standards and dividers) to the USNO system.

Second, the achievable stability given by equation (5) is very sensitive to clocks with poor stability as they enter into σ_{rms}, hence the criteria of WHP to reject the worst one-third of the clocks has some validity when using the $(1/m,0)$ weight approach. On the other hand, with this weighting approach, an excellent clock does not contribute to the improvement of the stability of the "mean" nearly as much as it does when the optimum weighting approach is used. Indeed, the marginal value of an excellent clock over an average quality clock is only about $1/(2m)$ when using the $(1/m,0)$ weight approach, regardless of how excellent the clock is. In fact, an excellent clock could have better stability than that of the "mean". Now, if the optimum weighting approach were used, the stability of the "mean" would be, in general, better than the best clock; and even a poor clock would make a positive contribution to the stability of the "mean". It is true that the "mean" would be more dependent on the best clocks, and in the absence of a proper rejection procedure the reliability could become a problem. However, a much more efficient rejection routine such as one which is also related to the quality (stability) of each clock may be used rather than as has been done by WHP, i.e., applying the same rejection limit to all clocks, excellent or poor. A proper rejection routine would have the effect of improving the reliability (i.e., safety) because then the more weight a clock receives the better it is required to perform to remain unrejected.

When WHP used the "unequal weighting" approach, the stability measure used (see Table 1 of WHP) was $\sigma_{\Delta t}$ ($N = 100$, $\tau = 1$ day). This stability measure has some peculiar properties that need to be considered when applying it to the noise processes encountered in cesium beam clocks. As may be seen from Fig. 2 of WHP, the stability may be limited typically by flicker noise frequency modulation for long sampling times. For such a noise process this type of stability measure is proportional to $N\tau$ [8], and furthermore with $N\tau = 100$ days the weighting factor derived by WHP using this stability measure would be primarily dependent upon fluctuations having an extremely long period of the order of 3 months and longer. Unfortunately, those clocks which have the best stability on a 3-months to 3-months basis are not necessarily the ones which are best on a day-to-day basis.

Another significant problem with the "unequal weighting" approach of WHP was that the reference used for determining the stability of each clock was not independent, since each clock participated in the stability of the reference by an amount proportional to its weighting factor. This would cause the stability of each clock to appear to be better than it actually was by a like amount. Hence, the clocks

---

[3] Flicker noise frequency modulation has the time domain stability characteristic that $\sigma_{\frac{\Delta f}{f}}(2,\tau) = a\tau^0$, where "$a$" depends on the noise level. (See footnote 4.)

[4] For consistency, we use the same symbols as used in WHP which are somewhat different from those used in reference 3.

[5] Optimum weighting is here defined as that set of weighting factors which gives a minimum for the left side of equation (3).

with the best stability would have "an overriding influence over the system" as indeed was found (see p. 127 of WHP). Even for the $(1/m,0)$ weighting approach, for the individual clocks the stabilities reported (see Tables 1, 5, 6, and 7 and Figs. 2 and 3 of WHP) would generally be biased to appear better than they actually were.

Third, in general a comparison of the different stabilities in Table 5 of WHP should not be made because of the variety of τ values listed. However, if one assumes specifically that, for each clock, $\sigma_{\frac{\Delta f}{f}}(2,\tau)$ is independent of τ for all τ values listed (i.e., flicker noise frequency modulation [6]), then the different stabilities may be directly compared. Even though this assumption is not exactly true (see Fig. 2), it appears to be a reasonable approximation for some of the clocks. Assuming for consistency that WHP used the above assumption and following their rejection criteria, one can pick the ten best stabilities from Table 5 to compute $\sigma_{(1/m,0)\,\mathrm{weight}}$ using equation (5). For comparison one can use the stabilities listed for all the clocks to compute $\sigma_{\mathrm{optimum\,weight}}$ using equation (6), and one obtains the interesting result that it again would take about eight more clocks ($\sim$ \$ 120,000 worth of commercial cesium beam clocks) using the $(1/m,0)$ weight approach to achieve the same stability obtained by the optimum weight approach.

The value of $\sigma_{(1/m,0)\mathrm{weight},\,\frac{\Delta f}{f}}(2,\tau)$ calculated above is still about 2×10^{-14}, but it should be noted no further improvement is obtained by increasing the sampling time, τ, — contrary to the statement by WHP on page 128. They state that stability should continue to improve because there are standards "always available which continue the $\tau^{-1/2}$ behavior to very long integration times (see Table 5)". In order to achieve this they must give the majority of the weight to *these few* clocks — contrary to their desire to have equal weight for each clock — because simultaneously there will be clocks whose stability is behaving as τ^0 or $\tau^{-1/2}$ (see Fig. 2 of WHP).

Fourth, the second reason given in their text for not using "unequal weighting" is that it depends on the "past performance" of the clocks. We must point out that their $(1/m,0)$ weighting is also based on the past 5 days performance. It has been shown [7] that good stability measures and hence optimum weighting factors are very constant with time, and hence past performance is an argument *for* optimum weighting.

C. Possible Drift in Rate of the USNO Time Scale

It is incorrect to conclude that the "total system drift" is less than 1×10^{-13} per 2 years from the cited data (see Section 3.2, p. 131 of WHP). If it is indeed true that the two samples taken 2 years apart are "different" and are from a supposed normal distribution (as evidently was assumed by WHP, p. 131 and Table 2), then the *best* that can be said from this data is that no system drift was discernable with a confidence (one-sigma) of $\pm \sqrt{2} \times 4.4 \times 10^{-13}$ per 2 years, i.e., $\pm 6.2 \times 10^{-13}$ per 2 years.

We question if the two samples were in fact "different" since at least six cesium beams appear to have the same serial numbers in 1968 and 1970 (see Table 1 and Table 2 of WHP). Even if the cesium beam tubes were replaced during the time that elapsed between the two measurements, this would not preclude the important consideration of frequency drift due to the electronics.

4. Comparisons with BIH

The discontinuity shown in Fig. 1 of WHP is very misleading, since it is simply a reassignment on paper of the origin of UTC(NBS) by the Bureau International de l'Heure (BIH). No such discontinuity exists in the UTC(NBS) scale. The BIH obtains the data for UTC (NBS) via Loran C, and the reassignment was basically due to uncertainties associated with the Loran C comparisons.

Another misleading aspect of Fig. 1 is the vertical scale. If one looks at the data in the BIH Circular D, one sees that UTC(NBS) is not the furthest off in time. In fact, in Fig. 1 of WHP the zero point for the ordinate has been arbitrarily chosen for each time scale plotted.

Also in another part of the text referring to Fig. 1 (see p. 128 of WHP) they conclude that "no drift in the USNO results is apparent" — meaning frequency drift. The basis for this conclusion is that in Fig. 1 "USNO's deviations from a straight line are very small and probably reflect irregularities of the BIH scale to a large part". Such a conclusion is not valid since a measurement of drift requires an independent observer. The BIH is not an independent observer as indeed WHP stated in the caption of Fig. 1. The BIH scale has been from 50% to 30% dependent on the USNO Time Scale since 1 January 1969. Significant difficulties in the USNO scale would have had to occur before the "line" could be other than "straight".

5. Least Squares Data Fitting and Efficiency Considerations

For white noise frequency modulation and for flicker noise frequency modulation — common noise processes in cesium beam clocks — a "least squares straight line fit" to the time data (WHP, p. 128, Section 2.3) is not nearly optimum for the estimation of average rates. The use of the least squares procedure has the disadvantage of reducing the effective data length from 5 days to about 4 days. The optimum procedure for the white noise FM case is simply to take the difference between the clock readings at the beginning and at the end ("2 point procedure") of the 5-day interval to compute the best estimate of average rate over that interval, and this procedure is nearly optimum for the flicker noise FM case. However, the choice by WHP of a sample time τ of 3 h is also non-optimum, and it causes their data to be measurement noise limited as may be seen by calculating the system stability from the data in Table 4 of WHP. This data yields an effective figure of merit of about 3 which is much worse than their hoped for value of 5.2. If WHP had increased τ to at least 12 h under the arrangement indicated in their text they would not have been limited by the measurement noise (see Fig. 6 of WHP). The use of $\tau \geq 12$ h and the end points measurement procedure would be not only more efficient statistically but also simpler than the use of $\tau = 3$ h and the least-squares-fit procedure.

6. Additional Comments

Our paper is not an exhaustive critique. The hope is to augment, clarify, and correct some of the many interesting and important points brought forth in the subject paper, which we feel is a valuable contribution to the technology of time keeping.

Acknowledgments. We express sincere appreciation to Drs. Ernest Ambler, James A. Barnes, Donald Halford, Helmut Hellwig, and Peter Kartaschoff for the many helpful comments and criticisms of our text.

References

1. Winkler, G. M. R., Hall, R. G., Percival, D. B.: Metrologia **6**, 126 (1970).
2. NBS Technical News Bulletin **53**, 34 (1969).
3. Barnes, J. A., Chi, A. R., Cutler, L. S., Healey, D. J., Leeson, D. B., McGunigal, T. E., Mullen, J. A., Smith, W. L., Sydnor, R., Vessot, R. F. C., Winkler, G. M. R.: NBS Technical Note 394, October 1970; also IEEE Trans. on Instrumentation and Measurement **IM-20** (to be published May 1971).
4. Cutler, L. S., Searle, C. L.: Proc. IEEE **54**, 136 (1966).
5. Vessot, R., Peters, H., Vanier, J., Beehler, R., Halford, D., Harrach, R., Allan, D., Glaze, D., Snider, C., Barnes, J., Cutler, L., Bodily, L.: IEEE Trans. on Instrumentation and Measurement **IM-15**, 165 (1966).
6. Allan, D. W.: Proc. IEEE **54**, 221 (1966).
7. — Gray, J. E., Machlan, H. E.: (Summary) Proc. 24th Annual Symposium on Frequency Control, Ft. Monmouth, 1970, p. 361.
8. — Barnes, J. A.: AGARD Conf. Proc. No. 33, Phase and Frequency Instabilities in Electromagnetic Wave Propagation (Proc. AGARD/EPC 13th Symp. Ankara, 1967), K. Davies, Ed., Chapter 15 (Technivision Services, Slough, England).