Correspondence

Fast ThêoBR: A Method for Long Data Set Stability Analysis

Jennifer A. Taylor and
David A. Howe, *Senior Member, IEEE*

Abstract—**ThêoBR is a high-confidence statistic that evaluates frequency stability at long τ values. However, for real-world data sets that contain thousands of points or more, the calculation of ThêoBR can take hours, days, or even weeks on a typical PC. To make the calculation of ThêoBR faster for these data sets, a method of averaging points together within the data set is developed. The error introduced by this technique is analyzed and compared with the exact value, and a correction formula is developed to minimize this error for FM noise types. Finally, the technique is applied to real data sets and determines stability at the longest τ values in seconds as opposed to weeks.**

I. Introduction

Theoretical variance #1 (Thêo1) is a time domain statistic with the ability to evaluate frequency stability for a longer portion of a data run with greater confidence than the similar Allan variance (Avar) [1]–[3]. However, because of its novel sampling, Thêo1 is biased with respect to Avar, where the bias is the ratio of the expected values of Avar and Thêo1 for all mutual τ values [1]–[4]. This can be corrected by using the bias functions if the noise type is known and there is only one noise type present [2]–[4]. Typically, however, the noise types affecting the data are not known or there are mixed noise types. In this case, an estimate bias value is computed from the data run. Thêo1 with bias removed (ThêoBR) can be used effectively to estimate Avar at very long-term τ, even where Avar cannot be calculated [2]–[4]. For a data run of N_x closely spaced measurements with a sampling period between adjacent observations given by τ_0, ThêoBR is defined as

$$\text{ThêoBR}(m,\tau_0,N_x) = \left[\frac{1}{n+1}\sum_{i=0}^{n}\frac{\text{Avar}(m=9+3i,\tau_0,N_x)}{\text{Thêo1}(m=12+4i,\tau_0,N_x)}\right] \times \text{Thêo1}(m,\tau_0,N_x), \quad (1)$$

where m is the averaging factor and $n = \lfloor(0.1N_x/3) - 3\rfloor$ ($\lfloor\cdot\rfloor$denotes the floor function, which returns the integer part of a number) [2]–[4]. ThêoBR removes a computed bias between Thêo1 and Avar by multiplying Thêo1 by the value of the computed bias. The accuracy of the removed bias is proportional to the number of points N_x. Thêo hybrid (ThêoH) is a hybrid statistic that combines Avar and ThêoBR to obtain, respectively, short-term and long-term frequency stability information with high confidence.

Although this is an effective method of removing the bias between Thêo1 and Avar and creating a useful frequency stability plot, the computation time of ThêoBR increases dramatically with increasing N_x. Using individually coded algorithms or commercially available software, the time for a calculation of ThêoBR can range from a few seconds for 1000 points to nearly one hour for 16 000 points. Often, data sets with over 100 000 points need to be analyzed; however, using this method would tie up computer resources and might take weeks to complete. This effectively makes the exact ThêoBR algorithm in (1) too time-consuming for long-term, closely spaced measurements.

II. A Practical Averaging Technique

To expand the usefulness of ThêoBR and ThêoH to these large data runs, we investigated ways of speeding up the calculation. One way to calculate ThêoBR for a long data set in a reasonable time is to decrease the number of N_x phase points while still keeping the general properties of the noise process; we find that this can be achieved by averaging. For instance, if one has a set of 1000 points and takes the average of the first 100 consecutive points, one now has a single number representing those 100 points. The same can be done for the remaining sets of 100 numbers until a final data set of 10 numbers is generated. Then the ThêoBR algorithm can be applied and a result for the bias value achieved almost instantaneously. The number of points that can be averaged together, which we call the bias average or BA, ranges from BA = N, resulting in one single value, to BA = 2, resulting in a run of half the original length. BA = 1 returns the entire original data set and produces the actual value of the bias.

This averaging technique dramatically improves in the speed of the bias calculation, which is the main goal of this study. For example, calculating the actual bias for 16 384 points takes nearly an hour on commercially available software. Averaging five numbers at a time, however, reduces the data set length to 3277 points and the calculation time to 20 s with an estimate bias value, discussed in the next section, that differs by less than 1% from the original value.

III. Correction Formula for FM Noise Types

As one might expect, the value of this estimated bias approaches the actual value when fewer numbers of points

Manuscript received September 25, 2007; accepted April 30, 2010.

The authors are with the National Institute for Standards and Technology, Time and Frequency Division, Boulder, CO (e-mail: mcgeeja@boulder.nist.gov).
Digital Object Identifier 10.1109/TUFFC.2010.1656

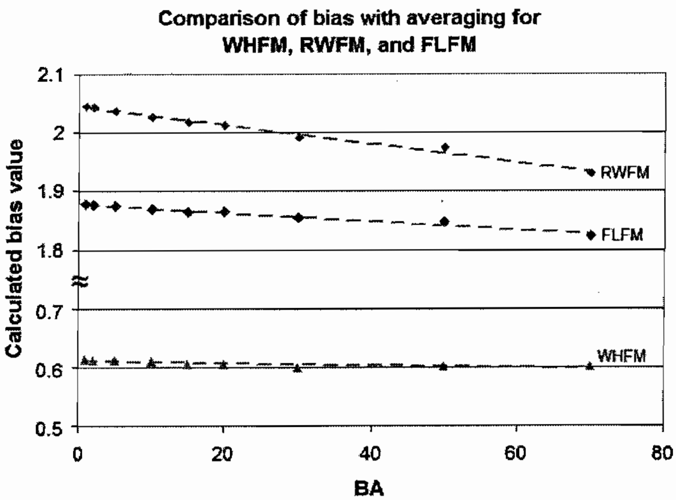

Fig. 1. A plot of bias calculated at various BA for simulated white FM, flicker FM, and random walk FM. A linear trend line (noted by the dashed line) can be fitted to each noise type, and an empirical formula for prediction of the bias can be determined by examining the slopes of the fits to these simulations of noise.

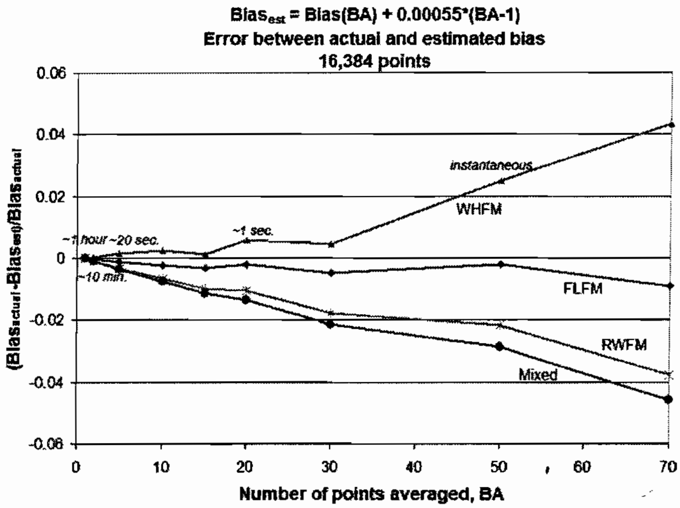

Fig. 2. The linear correction formula is used to calculate the bias from Bias(BA), and the departure from the actual bias value is plotted as a function of number of points averaged, BA. Approximate calculation speed is also indicated at select values of BA.

are averaged together. Fig. 1 shows the bias, calculated using a commercially available software program with a built-in ThêoBR and bias algorithm, plotted as a function of the number of averages, BA. Again, the actual bias value is given at BA = 1.

For the FM noise types of random walk, flicker, and white, the divergence of the bias from the actual value is approximately linear, as shown in Fig. 1, so a best-fit linear trend line with a slope of δbias can be drawn through the points. By finding δbias, one can calculate Bias(BA), the bias calculated with BA points averaged together, and apply a correction term to obtain Bias_{est}, the estimate of the actual unaveraged bias, or $\text{Bias}_{\text{actual}}$. Using the definition of slope, rise over run,

$$\delta\text{bias} = \frac{\text{rise}}{\text{run}} = \frac{\text{Bias(BA)} - \text{Bias}_{\text{est}}}{\text{BA} - 1}. \qquad (2)$$

This yields the correction formula for Bias(BA):

$$\text{Bias}_{\text{est}} = \text{Bias(BA)} - \delta\text{bias} \cdot (\text{BA} - 1). \qquad (3)$$

Because they have a different characteristic slope, PM noise types must be considered separately; however, unless white PM and flicker PM are dominant at long-term τ, we can ignore their effect because FM noises will dominate in most cases. An identical study may be done focused on PM noises.

The slope of this trend line, δbias, was recorded for many data sets of varying single and mixed white, flicker, and random walk FM noise types of different sizes. An average value of these slopes was determined and then adjusted to minimize the error between the calculated and exact bias. A δbias value of -0.00055 was found to be optimal for estimating the bias from the linear fit. Thus, our correction formula is

$$\text{Bias}_{\text{est}} = \text{Bias(BA)} + 0.00055 \cdot (\text{BA} - 1). \qquad (4)$$

IV. Comparison of Actual and Estimated Bias

Fig. 2 shows the error, that is, how much the predicted value of the bias from the formula differs from the actual value:

$$\text{error} = \frac{(\text{Bias}_{\text{actual}} - \text{Bias}_{\text{est}})}{\text{Bias}_{\text{actual}}}. \qquad (5)$$

The plot shows that as BA increases, the error increases in magnitude and the calculation time decreases. But even a small average of BA = 5 for a 16 384-sample data set turns an hour-long calculation into a 20-s one with an error of less than 1%.

Table I compares the actual bias to corrected bias on various segments of generated mixed-FM noise sets of different lengths. It also shows the error between the two. BA was chosen for each data set so that the calculation lasted less than one minute on a typical PC.

We can also compute the adjusted and actual ThêoBR values and compare the difference between them relative to the width of the one-sigma confidence interval associated with that point [5], [6]. One way to analyze this is to calculate what percentage of the total confidence width is the difference between these two ThêoBR values, or ΔThêoBR. This is shown in Table II for the same generated FM-noise sets in Table I. As shown in the table, any error associated with estimating the bias for ThêoBR is so small as to be negligible when compared with the allowed confidence interval.

V. Fast ThêoBR for Real Data Sets

The benefit of this method is that it reduces the time it takes to analyze real data. For example, Fig. 3(a) shows

TABLE I. ERROR BETWEEN ACTUAL AND ESTIMATED BIAS.

Number of points	BA for $Bias_{est}$	$Bias_{actual}$	$Bias_{est}$	Error
8000	3	3.0185	3.1073	0.0294
10000	4	2.1459	2.1396	0.0029
15000	4	0.7357	0.7354	0.0004
20000	7	2.7188	2.7193	0.0002

TABLE II. COMPARISON TO WIDTH OF CONFIDENCE INTERVAL.

Number of points	Confidence interval spanned by ΔThêoBR (%)
8000	0.053
10000	0.063
15000	0.818
20000	0.008

the plot of 15 octave-τ levels of the Allan deviation (Adev) for preliminary data from a high-quality space-qualified hydrogen maser (TEMEX, Sophia-Antipolis, France).[1] This data run has 223 130 points, representing one second time-error data taken for 2.57095 d. A calculation of Adev takes only a few seconds; however, because of the properties of Adev, the longest τ averaging time of the clock's stability is limited to less than half of the length of the data set for ample confidence [7]. Using ThêoH, we would be able to see the stability up to 3/4 of the length of the data set; however, a calculation of the actual bias on over 200 000 points would take weeks to compute on a present-day PC.

The fast ThêoBR method allows us to calculate ThêoH for this data set in minutes, not weeks, on the same PC. Fig. 3(b) shows a plot of ThêoH using this method for the bias calculation. The onset of a different noise type is visible at the expanded averaging time with the addition of two very large τ values.

Fig. 4 shows another example of the benefit of a usable ThêoH method for any data length with agreement between the exact and fast methods. The plot shows the frequency stability of a free-running 563-nm laser ultimately locked to a mechanically stable optical cavity [3]. The presence of steep τ^{+1} slope in short-term might be interpreted as frequency drift; however, systematic drift is not indicated in the raw frequency data. By using ThêoH, we can see a long-term slope change toward apparent random walk FM whose level can be estimated, even for this limited data run.

VI. CONCLUSION

The properties of ThêoBR and ThêoH are particularly useful for evaluating frequency stability at large τ values.

[1] No endorsement is implied. Products are available from other manufacturers.

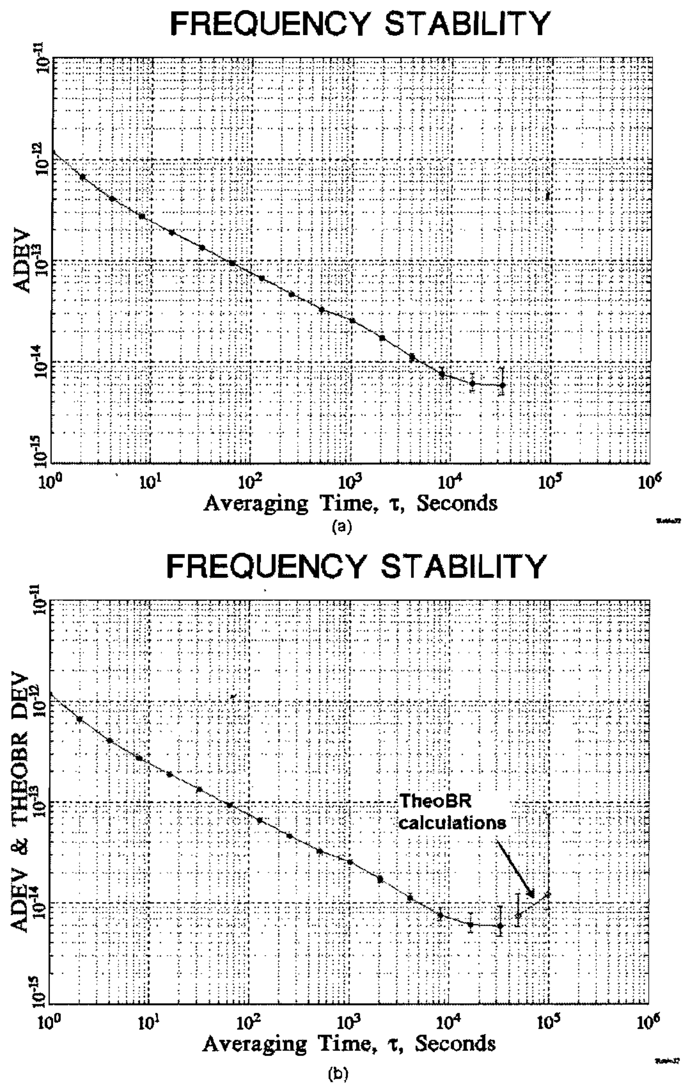

Fig. 3. (a) Plot of the Allan deviation for a compact, space-qualified H-maser. (b) Plot of ThêoH using the Fast-ThêoBR method showing its additional calculations of long-term stability.

This is advantageous for large data sets; however, calculating the bias for these data sets takes a very long time. We have shown a method of averaging data points that dramatically decreases this calculation time and makes ThêoBR, and hence ThêoH, practical for long data runs. A correction formula was also introduced that estimates the actual bias value by an error of less than 1%, well within the typical one-sigma confidence interval. Although not exact, the benefit of dramatically decreasing the calculation time using fast ThêoBR outweighs the small, essentially negligible introduced difference.

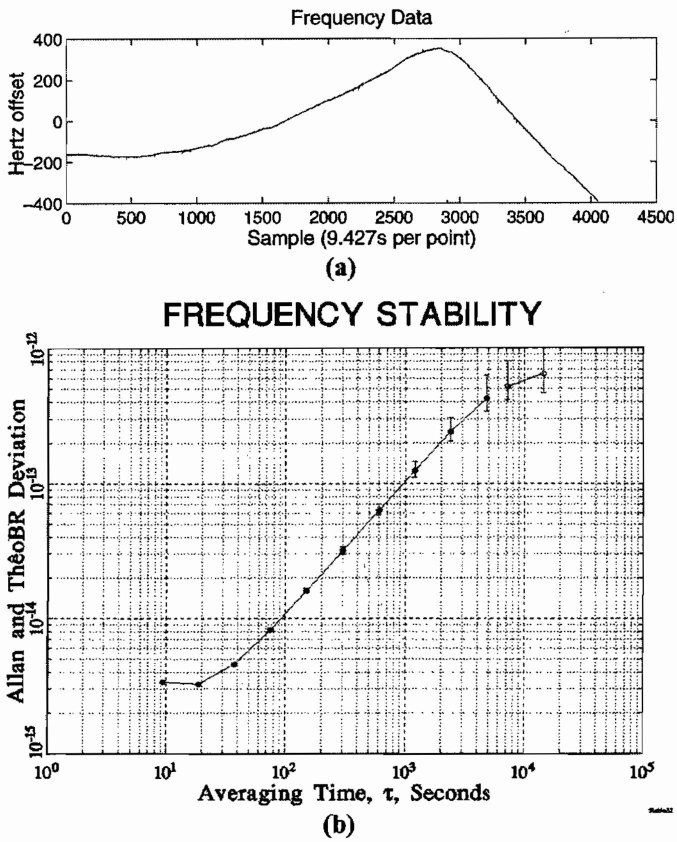

Fig. 4. ThêoH plot using the Fast TheoBR technique of a laser frequency that is locked to a stable optical cavity (below); the raw frequency data are shown above. Courtesy of J. Bergquist and S. Diddams, June 2006.

Acknowledgments

The authors thank W. J. Riley of Hamilton Technical Services for making commercial software updates for testing this new Fast ThêoBR method. We gratefully acknowledge raw clock data provided by TEMEX and the NIST Optical Frequency Group.

References

[1] D. A. Howe and T. K. Peppler, "Very long-term frequency stability: Estimation using a special-purpose statistic," in *Proc. 2003 Joint Meeting IEEE Int. Frequency Control Symp. and European Frequency and Time Forum Conf.*, May 2003, pp. 233–238.

[2] D. A. Howe and T. N. Tasset, "Thêo1: Characterization of very long-term frequency stability," in *Proc. 18th European Frequency and Time Forum*, Apr. 2004, pp. 581–587.

[3] D. A. Howe, "ThêoH: A hybrid, high-confidence statistic that improves on the Allan deviation," *Metrologia*, vol. 43, pp. 322–331, Aug. 4, 2006.

[4] D. A. Howe, D. B. Percival, J. A. McGee, and T. N. Tasset, "ThêoH bias-removal method," in *Proc. 2006 IEEE Frequency Contol Symp.*, pp. 788–792.

[5] T. N. Tasset, D. A. Howe, and D. B. Percival, "Thêo1 confidence intervals," in *Proc. 2004 IEEE Frequency Control Symp.*, Aug. 2004, pp. 725–728.

[6] J. A. McGee and D. A. Howe, "ThêoH and Allan deviation as power-law noise estimators," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 2, pp. 448–452, Feb. 2007.

[7] *Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities*, IEEE Std. 1139–1999, 1999.