

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM
A TEST SUITE FOR THE CALCULATION OF TIME DOMAIN
FREQUENCY STABILITY

W.J. Riley, EG&G, Inc., Salem, MA 01970

ABSTRACT

This paper describes a test suite to check the accuracy of frequency stability analysis software. It contains the values of several common statistics and frequency stability measures for two data sets, a small one suitable for manual entry, and a larger one produced by a portable pseudo-random number generator. The paper also discusses related issues such as data gaps and outliers, phase-frequency conversions, drift removal, numerical precision, plotting, and simulation.

INTRODUCTION

Specialized calculations are necessary to express the results of time domain frequency stability measurements [1,2,3]. A common example is the Allan variance for a set of fractional frequency data. Such calculations are generally performed by a computer, for which a custom program may need to be written and debugged. Each generation of computer hardware and operating system usually requires an update of the software, which must then be validated before use. A suite of test data, for which correct values of common frequency stability measures are known, can be a valuable tool for checking the accuracy of frequency stability analysis software.

DATA TYPES AND TERMINOLOGY

The time domain stability of a frequency source can be measured by either phase or frequency data. The former is normally expressed as $x(t) = \phi(t)/2\pi\nu_0$, where ν_0 is the nominal frequency. This quantity has units of time, but is generally called "phase" to avoid confusion with the independent time variable, t . Frequency is normally expressed as fractional frequency, $y(t) = [\nu(t) - \nu_0]/\nu_0 = x'(t)$, which is dimensionless. When making stability measurements, it is preferable to take phase data, since it is the more fundamental quantity. The terms "frequency standard" and "clock" are often used interchangeably. A frequency source may be called a clock even though it does not contain any actual clock hardware, especially if it is used for timing purposes. The term "oscillator" is best used for an active source like a crystal oscillator rather than a passive device such as a cesium beam tube or rubidium cell frequency reference.

DATA PREPROCESSING

Preprocessing of the measurement data is often necessary before performing the actual analysis, which may require data averaging, or removal of outliers, frequency offset, and drift [4].

Phase data can be modified for a longer averaging time by simply removing the intermediate points. This is accomplished for frequency data by calculating the average of each group of points.

Frequency offset may be removed from phase data by subtracting a line determined by the average of the first differences, or by a least-squares linear fit. An offset may be removed from frequency data by normalizing it to have an average value of zero.

Frequency drift may be removed from phase data by a least-squares or 3-point quadratic fit [5], or by subtracting the average of the second differences. Frequency drift may be removed from frequency data by subtracting a least-squares linear fit, by subtracting a line determined by the mean of the first-differences of the frequency data, or by calculating the drift from the difference between the average of the two halves of the data. The latter, called here the bisection drift, is equivalent to the 3-point fit to the phase data.

Other, more specialized, methods of drift removal may also be used. For example, the frequency data may be fit to a particular model such as the log function of MIL-O-55310B [6]. The latter is particularly useful to describe the stabilization of a frequency source. In general, the objective is to remove as much of the deterministic behavior as possible, obtaining random residuals for subsequent noise analysis.

Phase data may be converted to frequency data by calculating the first differences of the phase data and dividing by the measurement interval or averaging time. Frequency data may be converted to phase data by piecewise integration, using the averaging time as the integration interval.

FREQUENCY STABILITY MEASURES

The most common time-domain measures of frequency stability are as follows:

1. The normal Allan variance, $\sigma_y^2(\tau)$ [7].
2. The normal Allan variance calculated from fully overlapping samples, $\sigma_y^2(\tau)$ [8].
3. The modified Allan variance, Mod $\sigma_y^2(\tau)$ [9].
4. The time variance, $\sigma_x^2(\tau)$ [10,11].

where the variances are functions of the averaging time, τ . They are often denoted as AVAR, MVAR and TVAR, respectively. These quantities are not affected (except possibly for reasons of numerical precision) by the nominal frequency offset. They are usually calculated after removal of frequency drift, and are expressed as their square roots, ADEV, MDEV and TDEV. Each can be calculated from either phase or frequency data, which give the same result.

These calculations are reasonably fast on a modern computer with a math coprocessor, except, perhaps, for the modified Allan variance (and the related time variance). For a data set of several thousand points, calculation of the normal Allan variance is practically instantaneous, and the overlapping Allan variance calculation takes only a few seconds. Calculation of the modified Allan variance from phase data also takes only seconds, while its calculation from frequency data requires a triply-nested loop that can take several minutes. Algorithms faster than the obvious ones are available [12,13], particularly if the data have no gaps. A complete stability run is commonly done over a range of averaging times by doubling the τ at each successive analysis point for which there is sufficient data.

TEST DATA

A "classic" suite of frequency stability test data is the set of nine 3-digit numbers from Annex 8.E of NBS Monograph 140 [14] shown in Table I. Those numbers were used as an early example of an Allan variance calculation. This frequency data is also normalized to zero mean by subtracting the average value, and then integrated to obtain phase values. A listing of the properties of this data set is shown in Table II. While nine data points are insufficient to calculate large frequency averages, they are, nevertheless, a very useful starting point to verify frequency stability calculations, since a small data set can easily be entered and analyzed manually. A small data set is also an advantage for detecting "off-by-one" errors.

The larger frequency data test suite used here consists of 1000 pseudo-random frequency data points. It is produced by the following prime modulus linear congruential random number generator [15]:

$$n_{i+1} = (16807 \cdot n_i) \text{ Mod } 2147483647$$

This expression produces a series of pseudo-random integers ranging in value from 1 to 2147483646 (the prime modulus, $2^{31} - 1$, avoids a collapse to zero).

When started with the seed $n_0 = 1234567890$, it produces the sequence $n_1 = 395529916$, $n_2 = 1209410747$, $n_3 = 633705974$, etc. These numbers may be divided by 2147483647 to obtain a set of normalized floating-point test data ranging from 0 to 1. Thus the normalized value of n_0 is 0.5748904732. A spreadsheet program is a convenient and reasonably universal way to generate this data. The frequency data set may be converted to phase data by assuming an averaging time of 1, yielding a set of 1001 phase data points. Similarly, frequency offset and/or drift terms may be added to the data. These conversions can also be done by a spreadsheet program.

The values of this data set will be uniformly distributed between 0 and 1. While a data set with a normal (Gaussian) distribution would be more realistic, and could be produced by summing a number of independent uniformly distributed data sets, or by the Box-Muller method [16], this simple data set is adequate for software validation.

RESULTS

Statistics for the 1000-point test suite are shown in Table III. These values, reported to 7 significant figures, were obtained using IEEE 754 64-bit double-precision (15-digit) floating point arithmetic. While this reported precision is unwarranted for actual stability measures, it is useful for software validation. The theoretical expected value for the mean of a random variable uniformly distributed over the interval (0,1) is 0.5, and is independent of the averaging factor. The linear slope (per data interval) and the intercept are calculated as a least-squares linear regression fit. The standard deviation is that for the sample (not the population). The theoretical expected value for the standard deviation is $1/\sqrt{12} = 0.2886751$. The normal Allan deviation, $\sigma_y(\tau)$, is calculated for the full data set without averaging ($\tau = \tau_0$), using adjacent differences. For white frequency noise, it is equal to the standard deviation. The modified Allan deviation, Mod $\sigma_y(\tau)$, is, by definition, equal to the normal Allan deviation for an averaging time equal to that of the basic data interval, τ_0 . The overlapping Allan deviation, calculated using fully overlapping samples (every available difference at a certain averaging time) is also equal to the normal Allan deviation for $\tau = \tau_0$. The time deviation, $\sigma_x(\tau)$, is simply the modified Allan deviation multiplied by $\tau/\sqrt{3}$.

GAPS, JUMPS, and OUTLIERS

It is not uncommon to have gaps and outliers in a set of raw frequency stability data. Missing or erroneous data may occur due to power outages, equipment malfunctions, and interference. For long-term tests, it may be impossible or impractical to repeat the run, or to otherwise avoid such bad data points. Usually the rea-

son for the gap or outlier is known. It is particularly important to explain all phase discontinuities. Plotting the data will often show the bad points, which may have to be removed before further analysis to obtain meaningful results.

A simple, yet effective, technique for finding outliers is to compare each frequency data point, y_i , with the median value, m , of the data set plus or minus some multiple of the median absolute deviation (MAD):

$$MAD = \text{Median} \{|y_i - m| / 0.6745\}$$

where $m = \text{Median} \{y_i\}$, and the factor 0.6745 makes the MAD equal to the standard deviation for normally distributed data [17]. These median statistics are more robust because they are insensitive to the size of the outliers. Outlier detection is normally applied only to frequency data.

More elaborate techniques exist for the recognition of outliers in marginal cases [17,18]. A particularly effective means is statistical comparison of each data point against an optimum predictor based on an appropriate noise model.

Often a bad data point is replaced with a gap. The gaps should be kept in the record because they serve as “place-keepers” in time, and because “truth-in-packaging” may require them to be identified. A value of zero is often used as an obvious and unique way to indicate a gap, especially in fractional frequency data, where zero almost never appears. It is also an easy value to test for. However, a value of zero does occur at the beginning and end of a set of normalized phase data, so, while a zero is suitable to indicate an embedded gap in phase data, it is not unique as the first or last point.

Stability analysis algorithms can be modified to handle gaps. Two sample variances can be formed for the available pairs, taking into account the actual number of analysis points. Averages can be formed where there is at least one point within the group. Otherwise a gap is inserted into the averaged data. Phase-frequency conversions can also be written to handle gaps.

Optionally, a gap may be replaced by an interpolated value. While this may be desirable in some cases, it masks the existence of the missing data, and creates a fictitious value. Filling in gaps has the advantage, however, that the plotting and stability analysis methods do not need provisions for handling gaps.

Frequency jumps can also be a problem for stability analysis. Intuitively, a frequency jump is an indication that the statistics of the device being measured are not stationary. It may be necessary to divide the record into two portions before and after the jump and analyze them separately. A jump may be defined

and recognized by moving a sliding window through the frequency data, looking for a change in the average value between the two halves of the window. The magnitude of the jump can be judged in relation to the scatter of the data.

DATA ARRAYS

A one-column vector is all that is required for a phase or frequency data array. Because the data points are equally spaced, no time tags are necessary. While time tagging may be needed for archival storage of clock measurements, a vector of extracted gap-filled data is sufficient for analysis.

NUMERICAL PRECISION

There are relatively few numerical precision issues relating to the analysis of frequency stability data. One obvious case, however, is phase data for a highly stable frequency source having a relatively large frequency offset. The raw phase data will be essentially a straight line (representing the frequency offset), and the instability information is contained in the small deviations from the line. A large number of digits is necessary unless the frequency offset is removed by subtracting a linear term from the raw phase data. Similar considerations may also apply to the quadratic phase term (linear frequency drift).

Many frequency stability measures involve averages of first or second differences. Thus, while their numerical precision depends upon the variable digits of the data set, there is little error propagation in forming the summary statistics.

DATA PLOTTING

Data plotting is often the most important step in the analysis of frequency stability. Visual inspection can provide vital insight into the results, and is an important “preprocessor” before numerical analysis. A plot also shows much about the validity of a curve fit.

Phase data is generally plotted with line segments connecting the data points. This presentation properly conveys the integral nature of the phase data. Frequency data is often plotted the same way, simply because that is the way plotting is usually done. But a better presentation is a flat, horizontal line between the frequency data points. This shows the averaging time associated with the frequency measurement, and mimics the analog record from a frequency counter. As the density of the data points increases, there is essentially no difference between the two plotting methods. In a plot, missing data should be shown as a gap without a line connecting the adjacent points.

Stability plots generally take the form of graphs of $\log \sigma$ versus $\log \tau$, often with error bars to show the

precision of the results. The slope of the $\sigma_y(\tau)$ characteristic depends on the type of noise. It is customary to show points at octave increments of τ . These are equally spaced on the log scale, and are the result of successive averaging by a factor of two. Such a run usually ends when there are too few analysis points (say <7) for reasonable confidence. A run for all possible τ values, while slow, can provide valuable information since it is, in effect, a form of spectral analysis that can show power-law spectra and periodic instabilities such as environmental effects.

ERROR BARS

Several approaches exist for the setting of error bars on a stability plot to indicate the precision of the results. A rough approximation to a standard 1-sigma confidence interval can be obtained by simply dividing the point value by the square root of the number of analysis points. This approximation degenerates for a small number of degrees of freedom. More exact single- and double-sided confidence intervals can be determined by the noise and variance type [2,8,19,20], as shown in Table IV.

NOISE RECOGNITION AND MODELING

It is often desirable to have a means for identifying the type of noise that is being analyzed, and to be able to fit it to a power-law noise model. Noise recognition can generally be accomplished by using the B1 bias function (the ratio of the standard variance to the Allan variance) and the R(n) ratio function (the ratio of the modified Allan variance to the normal Allan variance). These functions can be calculated by the methods described in Reference 1. Power-law noise can be modeled over a range of averaging times by fitting a line to the results of an Allan variance stability analysis on a log σ versus log τ plot.

SIMULATED CLOCK NOISE

It is valuable to have a means of generating simulated power-law clock noise having the desired noise type (white phase, flicker phase, white frequency, flicker frequency, and random walk frequency), Allan deviation, frequency offset and frequency drift. This can serve as an additional means to validate stability analysis software, particularly for checking numerical precision and noise recognition and modeling. An excellent method for power-law noise generation is described in Reference 21. This reference also provides very useful analysis insights.

VALIDATION METHODS

Several methods are available to validate frequency stability analysis software:

1. Manual Analysis: The results obtained by manual analysis of small data sets (such as in NBS Monograph 140 Annex 8.E) can be compared with the new program output. This is always good to do to get a "feel" for the process.
2. Published Results: The results of a published analysis can be compared with the new program output. The purpose of this paper is to serve as such a reference.
3. Other Programs: The results obtained from other specialized stability analysis programs (such as that from a previous generation computer or operating system) can be compared with the new program output.
4. General Purpose Programs: The results obtained from industry standard, general purpose mathematical and spreadsheet programs (such as MathCAD and Lotus 1-2-3) can be compared with the new program output.
5. Consistency Checks: The new program should be verified for internal consistency, such as producing the same stability results from phase and frequency data. The standard and normal Allan variances should be approximately equal for white FM noise. The normal and modified Allan variances should be identical for an averaging factor of 1. For other averaging factors, the modified Allan variance should be approximately one-half the normal Allan variance for white FM noise and $\tau \gg \tau_0$. The normal and overlapping Allan variances should be approximately equal. The overlapping method provides better confidence of the stability estimates. The various methods of drift removal should yield similar results.
6. Simulated Data: Simulated clock data can also serve as a useful cross check. Known values of frequency offset and drift can be inserted, analyzed, and removed. Known values of power-law noise can be generated, analyzed, plotted, and modeled.

CONCLUSIONS

There is a continuing need to validate the custom software used to analyze time domain frequency stability. The suite of test data presented here, along with the other suggestions made, can help ensure that correct results are obtained.

ACKNOWLEDGMENTS

The author wishes to thank Messrs. David W. Allan, Leonard S. Cutler, Samuel R. Stein, and Fred L. Walls for their help in checking the results presented here. Any remaining errors are my responsibility alone.

REFERENCES

1. D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls (Editors), *Characterization of Clocks and Oscillators, NIST Technical Note 1937*, U.S. Department of Commerce, National Institute of Standards and Technology, March 1990.
2. S.R. Stein, "Frequency and Time - Their Measurement and Characterization," Chapter 12, pp.191-416, *Precision Frequency Control*, Vol. 2, Edited by E.A. Gerber and A. Ballato, Academic Press, New York, 1985.
3. "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology," *IEEE Std. 1139-1988*.
4. J.A. Barnes, "The Measurement of Linear Frequency Drift in Oscillators," *Proc. 15th Annu. PTTI Meeting*, pp. 551-582, Dec. 1983.
5. M.A. Weiss and C. Hackman, "Confidence on the Three-Point Estimator of Frequency Drift," *Proc. 24th Annu. PTTI Meeting*, pp. 451-460, Dec. 1992.
6. MIL-O-55310B, *Oscillators, Crystal, General Specification For*.
7. D.W. Allan, "The Statistics of Atomic Frequency Standards," *Proc. IEEE*, Vol. 54, No. 2, pp. 221-230, Feb. 1966.
8. D.A. Howe, D.W. Allan, and J.A. Barnes, "Properties of Signal Sources and Measurement Methods," *Proc. 35th Annu. Symp. on Freq. Contrl.*, pp. 1-47, May 1981.
9. D.W. Allan and J.A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability," *Proc. 35th Annu. Symp. on Freq. Contrl.*, pp. 470-474, May 1981.
10. D.W. Allan, D.D. Davis, J. Levine, M.A. Weiss, N. Hironaka, and D. Okayama, "New Inexpensive Frequency Calibration Service From NIST," *Proc. 44th Annu. Symp. on Freq. Contrl.*, pp. 107-116, June 1990.
11. D.W. Allan, M.A. Weiss, and J.L. Jespersen, "A Frequency-Domain View of Time-Domain Characterization of Clocks and Time and Frequency Distribution Systems," *Proc. 45th Annu. Symp. on Freq. Contrl.*, pp. 667-678, May 1991.
12. D.W. Allan, "Time and Frequency Metrology: Current Status and Future Considerations," *Proc. 5th European Freq. and Time Forum*, pp. 1-9, March 1991.
13. C.A. Greenhall, "A Shortcut for Computing the Modified Allan Variance," *Proc. 1992 IEEE Freq. Contrl. Symp.*, pp. 262-264, May 1992.
14. B.E. Blair (Editor), *Time and Frequency: Theory and Fundamentals, NBS Monograph 140, Annex 8.E*, p. 181, May 1974.
15. S.K. Park and K.W. Miller, "Random Number Generators: Good Ones are Hard to Find," *Comm. ACM*, Vol. 31, No. 10, pp. 1192-1201, Oct. 1988.
16. W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*, Cambridge Univ. Press, Cambridge, U.K., 1988, pp. 216-217.
17. D.B. Percival, "Use of Robust Statistical Techniques in Time Scale Formation," *Preliminary Report*, U.S. Naval Observatory Contract No. N70092-82-M-0579, 1982.
18. Gernot M.R. Winkler, "Introduction to Robust Statistics and Data Filtering," *Tutorial at 1993 IEEE Freq. Contrl. Symp.*, Sessions 3D and 4D, June 1, 1993.
19. C.A. Greenhall, "Recipes for Degrees of Freedom of Frequency Stability Estimators," *IEEE Trans. Instrum. Meas.*, Vol. 40, No. 6, pp. 994-999, Dec. 1991.
20. C.A. Greenhall, "Estimating the Modified Allan Variance," *Proc. IEEE 1995 Freq. Contrl. Symp.*, (to be published).
21. N.J. Kasdin and T. Walter, "Discrete Simulation of Power Law Noise," *Proc. 1992 IEEE Freq. Contrl. Symp.*, pp. 274-283, May 1992.

Table I. NBS Monograph 140, Annex 8.E Test Data			
Data Point #	Frequency Data	Normalized Frequency	Phase Data ($\tau=1$)
1	892	103.11111	0.00000
2	809	20.11111	103.11111
3	823	34.11111	123.22222
4	798	9.11111	157.33333
5	671	-117.88889	166.44444
6	644	-144.88889	48.55555
7	883	94.11111	-96.33333
8	903	114.11111	-2.22222
9	677	-111.88889	111.88889
10			0.00000

Table II. NBS Monograph 140, Annex 8.E Test Data Statistics		
Averaging Factor	1	2
# Data Points	9	4
Maximum	903	893.0
Minimum	644	657.5
Average	788.8889	802.875
Median	809	830.5
Linear Slope	- 10.20000	-2.55
Intercept	839.8889	809.25
Standard Deviation [1]	100.9770	102.6039
Normal Allan Deviation	91.22945	115.8082
Overlapping Allan Deviation	91.22945	85.95287
Modified Allan Deviation	91.22945	74.78849
Time Deviation	52.67135	86.35831

Table II Notes:

[1] Sample (not population) standard deviation.

Table III. 1000-Point Frequency Data Set			
Averaging Factor	1	10	100
# Data Points	1000	100	10
Maximum	9.957453e-01	7.003371e-01	5.489368e-01
Minimum	1.371760e-03	2.545924e-01	4.533354e-01
Average [1]	4.897745e-01	4.897745e-01	4.897745e-01
Median	4.798849e-01	5.047888e-01	4.807261e-01
Linear Slope [2,3]	6.490910e-06	5.979804e-05	1.056376e-03
Intercept [3]	4.865258e-01	4.867547e-01	4.839644e-01
Bisection Slope [2]	-6.104214e-06	-6.104214e-05	-6.104214e-04
1st Difference Slope [2]	1.517561e-04	9.648320e-04	1.011791e-03
Log Fit [2,4]			
$y(t) = a \ln(bt + 1) + c$	a	5.577220e-03	5.248477e-03
	b	9.737500e-01	4.594973e+00
	c	4.570469e-01	4.442759e-01
$y'(t) = ab/(bt + 1)$	Slope at end	5.571498e-06	5.237080e-05
Standard Deviation [5]	2.884664e-01	9.296352e-02	3.206656e-02
Normal Allan Deviation [6]	2.922319e-01	9.965736e-02	3.897804e-02
Overlapping Allan Deviation [8]	2.922319e-01	9.159953e-02	3.241343e-02
Modified Allan Deviation [7,8]	2.922319e-01	6.172376e-02	3.897804e-02
Time Deviation [8]	1.687202e-01	3.563623e-01	1.253382e-02

Table III Notes:

- [1] Expected value = 0.5.
- [2] All slopes are per interval.
- [3] Least-squares linear fit.
- [4] Exact results will depend on iterative algorithm used. Data not suited to log fit.
- [5] Sample (not population) standard deviation. Expected value = $1/\sqrt{12} = 0.2886751$.
- [6] Expected value equal to standard deviation for white FM noise.
- [7] Equal to normal Allan deviation for averaging factor = 1.
- [8] Calculated with listed averaging factors from the basic $\tau = 1$ data set.

Table IV. Error Bars for n=1000 Point =1 Data Set with Averaging Factor=10							
Allan Deviation			Noise		Confidence Interval		
Type	#	Value	Ratio	Type	Value	χ^2	Remarks
Normal	99	9.965736e-02	B1=0.870	W FM [1]	CI= 8.713870e-03 [2,3]	-	$\pm 1\sigma$ CI. [6]
Overlapping	981	9.159953e-02	-	W FM	Max $\sigma_y(\tau) = 1.014923e-01$	119.07	95% CI. [7]
				# χ^2 df =	Max $\sigma_y(\tau) = 1.035201e-01$	114.45	95% CI. [8]
				146.177	Min $\sigma_y(\tau) = 8.223942e-02$	181.34	
Modified [4]	972	6.172376e-02	R(n)=0.384	W FM [5]	Max $\sigma_y(\tau) = 7.044412e-02$	72.64	95% CI. [7]
				# χ^2 df =	Max $\sigma_y(\tau) = 7.224944e-02$	69.06	95% CI. [8]
				94.620	Min $\sigma_y(\tau) = 5.419961e-02$	122.71	

Table IV Notes:

- [1] Theoretical B1=1.000 for W FM noise and 0.667 for F and W PM noise.
- [2] Simple, noise-independent CI estimate = $\sigma_y(\tau)/\sqrt{N} = 1.001594e-02$.
- [3] This CI includes κ_α factor that depends on noise type:

Noise	α	κ_α
W PM	2	0.99
F PM	1	0.99
W FM	0	0.87
F FM	-1	0.77
RW FM	-2	0.75

- [4] BW factor = $2\pi f_h \tau_0 = 10$. Applies only to F PM noise.
- [5] Theoretical R(n) for W FM noise = 0.500 and 0.262 for F PM noise.
- [6] Double-sided 68.3% confidence interval: p = 0.158 and 0.842.
- [7] Single-sided 95% confidence interval: p = 0.950.
- [8] Double-sided 95% confidence interval: p = 0.025 and 0.975.