The Evolution of Frequency Stability Analysis Software

W.J. Riley, IEEE Life Fellow
Hamilton Technical Services
Beaufort, SC 29907 USA
bill@wriley.com

Abstract

This paper is part of an effort to commemorate the 50th anniversary of the Allan variance for the analysis of frequency stability. It traces the evolution of the stability analysis software that implements the Allan variance and related statistics that are now universally used for that purpose. The parallel advances in atomic clocks, the modeling of their performance, and the supporting computational hardware and software have resulted in remarkable progress in the art and science of timekeeping that has influenced both scientific endeavors and the communications and navigation aspects of everyday life. The Allan variance family of statistics and their associated software are important part of that progress.

Introduction

This paper (really more of a historical note) describes the evolution of frequency stability analysis software from the perspective of one of its practitioners. I was fortunate to have started my professional engineering career in early 1960's at the beginning of the commercialization of atomic frequency standards and a time of rapid progress in the characterization of those devices and the computational means for doing so. Foremost among that evolution was the development of the Allan variance and related statistical measures. Much progress has been made in this field over the last half-century, and one aspect of that progress has been our ability to better measure, analyze, understand and thereby improve the stability of our frequency sources.

The characterization of frequency stability (really instability) involves describing noise processes. In the time domain, the time series analyses we perform, while similar to those in other fields, emphasize the underlying noise types and their physical origins. In the frequency domain, our analyses emphasize quantifying noise spectral densities rather than identifying discrete components. Fundamental to both domains are the power law noise models and statistical measures used to characterize the instabilities of clocks and oscillators. A general guide to frequency stability analysis will be found in Reference [1].

Terminology

The time and frequency field, like most, has its own conventions and standards, and these, essential for the uniform exchange of information, can pose an initial hurdle for newcomers. For example, the most frequently asked questions involve our use of phase data, x, in the form of time error in seconds and frequency data, y, in the form of dimensionless fractional frequency. For a set of equivalent phase and frequency data there are N phase points and one fewer, M=N-1 frequency points. The confusion is then compounded by the use of unfamiliar statistical measures and descriptions of noise types. This requires that we, and time and frequency practitioners, be prepared to explain our methodologies and that specialized frequency stability analysis software be supported by tutorial information.

We usually call the time between data samples the sampling or measurement time, τ_0, and the analysis tau, called the averaging time, is an integer multiple of that, $\tau=n\cdot\tau_0$. Averaging is indeed the process applied to frequency data at a longer tau, but for phase data it is downsampling that is applied, using every

n^{th} point, a process we often wrongly call decimation, and which has implications when there are discrete spectral components [2].

Other aspects of frequency stability analysis that can cause confusion are the difference between the Allan variance and Allan deviation, our use of the word "clock" to refer to a frequency source, the subtle distinction between "drift" and "aging", and the usual non-distinction between "drift" and "drift rate".

It is important to distinguish between a statistical measure such as the Allan variance and its estimate. The former is expressed as an expectation (< >) while the latter (a computation formula) is subject to statistical uncertainty. It is also important to distinguish between biased and unbiased estimates of the Allan variance, where the latter can be advantageous because of higher confidence but generally requires knowledge of the underlying noise process in order to apply a bias correction.

The non-specialist can be relieved of most of those considerations, along with such details as power law noise identification, χ^2 statistics, equivalent degrees of freedom, confidence intervals and the like by using appropriate analysis software, but it is still important to understand the underlying principles.

Progress

It is perhaps best to start with some examples showing the need for the Allan variance and the progress that has been made in estimating it.

Figure 1 compares the standard and Allan deviations for several sample sizes of flicker FM noise. The standard deviation value depends on the sample size while the Allan deviation does not. The problem can be thought of as being associated with the use of the mean in calculating the standard deviation; the mean is poorly-defined for divergent noise, and the Allan deviation uses 1st differences of the frequency instead. The non-convergence of the standard deviation shows the need for the Allan variance statistic when analyzing flicker and other divergent power law noise types commonly associated with frequency sources.

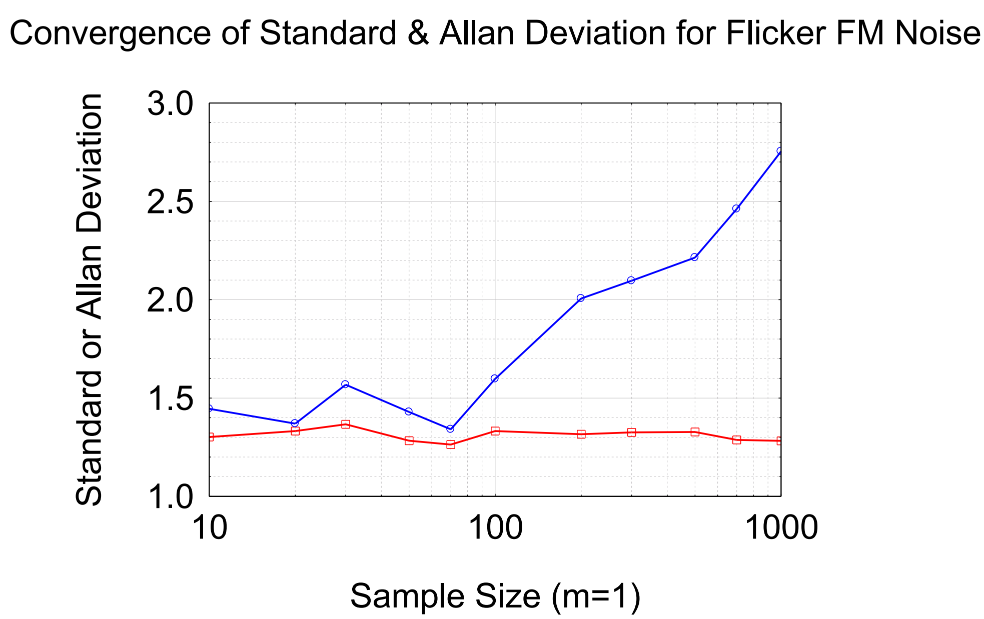
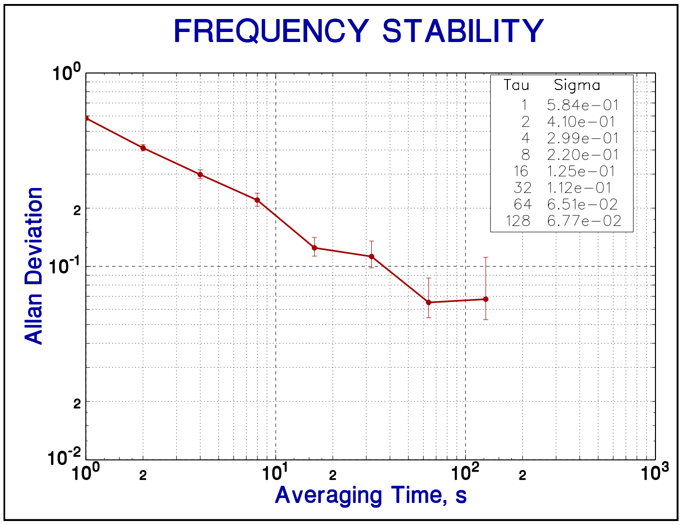
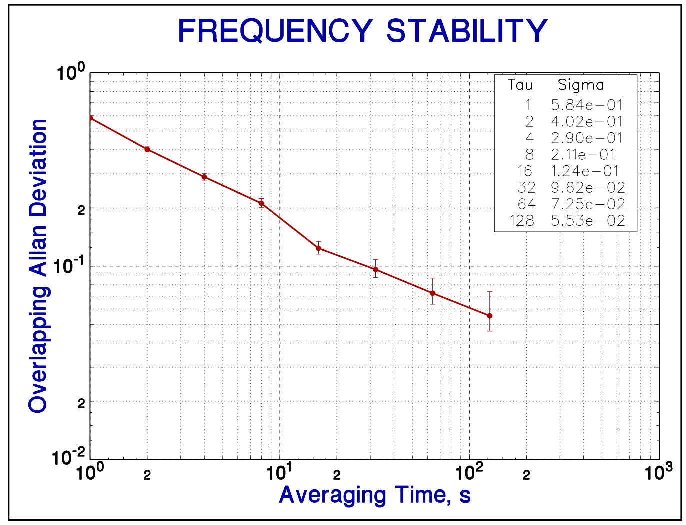


Figure 1. Convergence of Standard and Allan Deviation of Flicker FM Noise.
The standard deviation depends on the sample size for flicker FM noise, while the Allan deviation does not and is therefore a better measure of frequency stability.

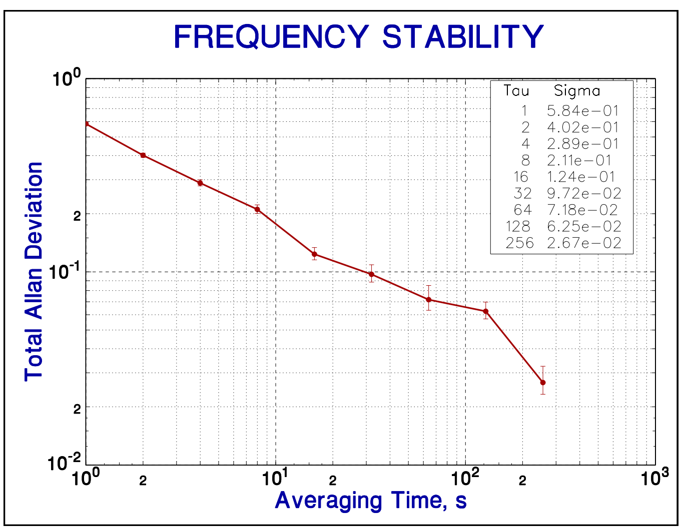
Figure 2 shows the progress made in devising an Allan variance estimator with better confidence at longer averaging factors. The progression from the original Allan variance to its fully overlapping counterpart, the Total variance, the finally the Thêo1 statistic shows the narrowing of the error bars and the extension to longer averaging times provided by the more advanced estimators. Improved confidence for a certain sample size is an obvious advantage, but the ability to obtain stability estimates at larger averaging factors is perhaps an even bigger advantage, especially when long measuring times are involved. While computational time is increased for the more complex statistics, that is negligible compared with the days or weeks that may be necessary to obtain another octave in tau with a longer run.



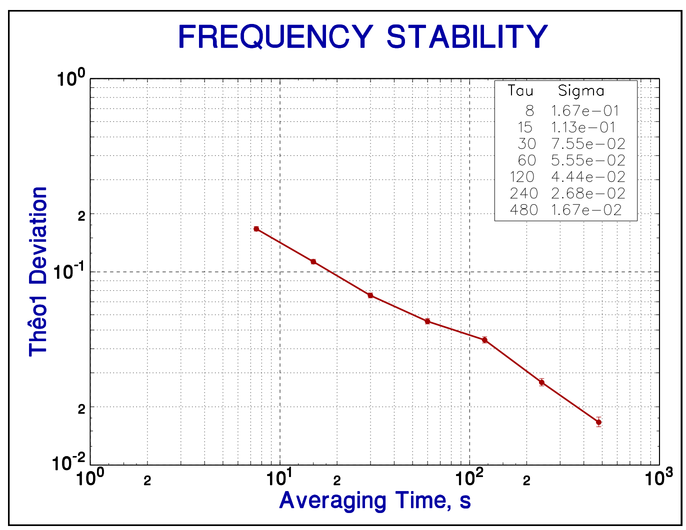
(a) Original Allan Deviation



(b) Overlapping Allan Deviation



(c) Total Deviation



(d) Thêo1

Figure 2. Progress in Frequency Stability Analysis
The confidence improves, the error bars shrink and tau range expands as the stability measure evolves from the original Allan to the overlapping Allan, Total and Thêo1 statistics.

Pre-Allan Variance

Prior to the introduction of the Allan variance circa 1966, frequency stability was characterized by general-purpose statistics such as the standard deviation. Because noise was being measured, there was recognition that the instability of a frequency source varied with the averaging time. Occasionally, particularly in the case of academic papers and formal product specifications, confidence limits were applied, but there was considerable confusion as to how to set them, especially since the nominal values often depended on the number of samples.

The general convention was, and still is, to specify a stability requirement as a nominal instability value, not the upper limit at some confidence factor. But reported measurements should, whenever possible, include (most commonly 1-sigma) error bars; a good general rule is that even poorly-defined error bars are better than none at all. One of the areas of considerable progress has been techniques for determining these confidence limits depending on the particular statistic and dominant noise process involved. The complexity of several of the more elaborate recent statistics has led from analytical to empirical methods for this, using Monte Carlo simulations.

The IEEE-NASA Symposium on Short-Term Frequency Stability [3] held in November 1964 can be considered as the driving force that led to the development of the Allan variance and its associated measurement techniques. This was followed by a special issue of the Proceedings of the IEEE devoted to this subject in February 1966 which included papers by J.A. Barnes [4] and D.W. Allan [5] that introduced the Allan variance concept. By 1970, the two-sample Allan variance had become the preferred stability measure [6].

Post-Allan Variance

The Allan variance was quickly adopted by the time and frequency community because its advantages were immediately appreciated (and, as an aside, one early reason for this might have been because in most cases it made frequency sources appear more stable). The original Allan deviation estimator was no more difficult to calculate than the standard deviation, and it was soon implemented on a variety of platforms including mainframe and time-sharing computers, computing counters and programmable calculators [7].

Other Members of the Allan Variance Family

Over the years since the Allan variance was introduced, several closely-related statistics have been added to the Allan variance family. Perhaps most important is the use of a fully-overlapping unbiased estimation formula that has the same expected value but provides better confidence by utilizing a stride of one data point instead of a stride equal to the averaging factor [8]. This provides more equivalent χ^2 degrees of freedom and significantly tighter error bars. The original Allan deviation estimator has therefore been largely replaced by the overlapping version, and the same stability specifications apply.

D.A Howe continued the search for even better Allan variance estimators and introduced the Total variance [9] and then the Thêo1 statistic [10], which has both better confidence and the ability to analyze data out to 75% of the record length. Thêo1 is a biased estimator, which means that it requires a correction factor to yield the same expected value as the classic Allan variance, and the bias correction depends on both the dominant power law noise type of the data being analyzed and the averaging factor. Thus we are faced with a more complex analysis situation, but one that software can make largely transparent to the user. A particularly attractive way to analyze a stability record to obtain results over the widest possible range of averaging times with the best possible confidence is the combination of the overlapping Allan–called ThêoH (for hybrid). It uses the overlapping Allan deviations at short and medium tau and Thêo1 at long tau, automatically determining the noise type and bias correction, in most cases providing a smooth overall stability characteristic. An example of the use of ThêoH is shown in Figure 3 showing the drift-removed stability of a GPS rubidium clock. The Thêo1 statistic is able to estimate the stability out to about 21 days for a 30-day test compared with only about 5 days for the Allan deviation, thereby saving much expensive test time.

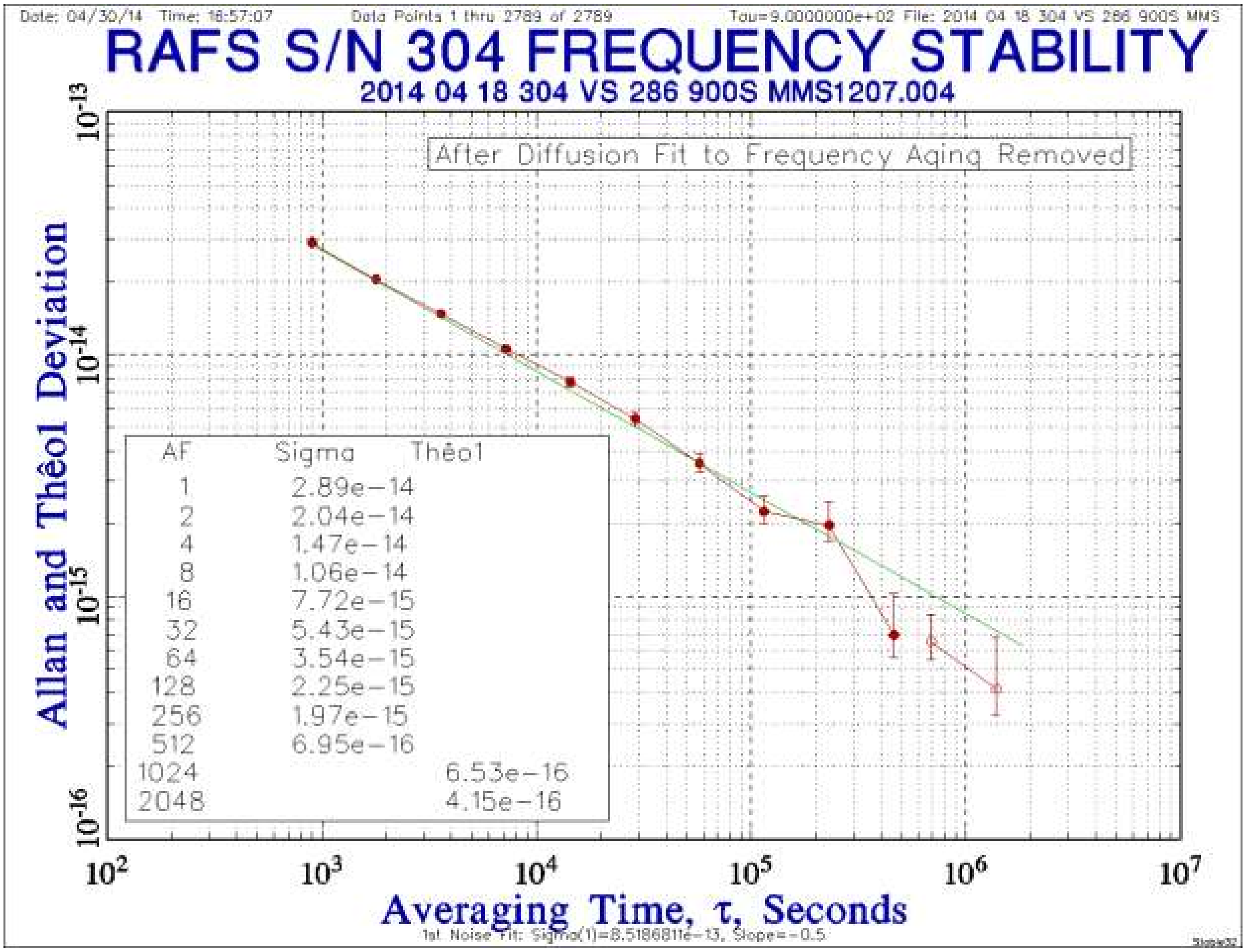


Figure 3. ThêoH Stability Plot for a GPS Rubidium Clock
The Thêo1 statistic provides a stability estimate out to nearly 75% of the 30-day record length.

Another method for frequency stability analysis is the Dynamic Allan deviation whereby the frequency or phase record is divided into sections that are analyzed separately so as to be able to detect stability changes over time.

Regarding the error bars for the various stability measures, one needs to take sufficient data to provide the required confidence at the longest tau of interest. One cannot expect to obtain useful results at large averaging factors where there are an insufficient number of samples, a quantity that depends on the particular statistic. Allan deviation and related stability results tend to "collapse" there and are not meaningful. Another source of confusion is the asymmetrical error bars associated with behavior of the χ^2 distribution.

Other Variances and Frequency Stability Measures

The main alternative to the Allan variance used for frequency stability analysis is the Hadamard variance. The Hadamard variance is a three-sample variance (as compared to the two frequency samples used by the Allan variance). It uses an unbiased estimator with the same expected value as the Allan variance for white FM noise, and has the advantage of convergence for flicker walk power noise and insensitivity to linear frequency drift, but at the expense of wider confidence limits. It is utilized in original, overlapping, total and dynamic forms, mainly so that frequency drift need not be removed before performing a noise analysis. The modified Allan variance [11] includes additional phase averaging that allows distinction between white and flicker PM noise; it is also the basis of the time variance [12]. MTIE and TIE rms are clock stability measures used principally by the telecommunication industry [13].

Phase averaging during the measurement process can reduce the white PM noise floor of the measurement system but that filtering also changes the noise spectral type in a fashion similar to the modified Allan deviation and is generally not recommended.

Frequency Stability Analysis Software

The relatively large number of samples required for a frequency stability analysis means that manual calculation of stability statistics is impractical except is very small test cases.

Table I shows some of the functions typically provided by a stability analysis software package or suite of separate programs. Preprocessing prepares the raw data for analysis by plotting it for visual inspection, scaling it to the proper units, detecting and removing outliers and jumps, and filling gaps or making the points regularly spaced using their associated timetags [14, 15, 16]. The data may be converted between phase and frequency, averaged or down-sampled to a longer sampling interval, be normalized by removing their mean value, filtered in the frequency domain and/or have their drift modeled and removed. The preprocessed data are then ready for stability, spectral, histogram or autocorrelation analysis. The stability analysis may use cross-correlation techniques to lower the instrumental noise or apply the 3-cornered hat method to determine the individual stabilities of several sources [17]. Finally, the results of the analysis are shown by stability or power spectral density plots in a report that describes the analysis methodology. In addition, simulated power law noise may be generated to clarify the requirements, which can be expressed in either the time or frequency domain. Domain conversions can be performed by either closed-form expressions for power law noises or by integration of the PSD characteristic. Stability specifications must be consistent between the two domains.

Of particular importance are decisions made by the analyst regarding outlier and drift removal and the choice of analysis tools and parameters so as to provide the best information about the device under test. One should always keep in mind R.W. Hamming's admonition that "the purpose of computing is insight, not numbers". The usual analysis procedure is to examine the data, remove any outliers, analyze and remove the deterministic drift, analyze the stochastic noise and then report the results.

Table I Functions for Frequency Stability Analysis		
Preprocessing	**Analysis**	**Reporting/Other**
Data Plotting	Stability Analysis	Stability Plotting
Scaling	Spectral Analysis	PSD Plotting
Outlier Detection and Removal	Autocorrelation	Report Preparation
Jump Detection and Removal	Cross-correlation	Simulation
Gap Filling and Regularization	Dead-Time Correction	Domain Conversions
Phase-Frequency Conversions	3-Cornered Hat Analysis	
Averaging	Histogram	
Drift Modeling and Removal		
Normalizing		
Filtering		

The Stable32 Program

The Stable32 program is a popular tool for frequency stability analysis [18]. It began as a set of programs for the HP-85 computer in 1980. When the IBM PC was introduced in 1981, a DOS version of Stable was written for it in 1982. Its initial application was to support the development of rubidium clocks for the GPS navigation satellites. Stable was ported to Microsoft Windows as Stable/Win in 1992, and then to its current Stable32 form in 1997 when Windows evolved to 32 bits.

The basic Stable32 idiom is a pair of equally-spaced convertible phase and frequency data arrays that can be subjected to a complete suite of analysis functions as shown in Table II.

Table II Statistics for Frequency Stability Analysis		
Allan Variance Family	**Hadamard Variance Family**	**Other**
Original	Original	Standard Deviation
Overlapping	Overlapping	TIE rms
Modified		MTIE
Total	Total	
Total Modified		
Time		
Dynamic	Dynamic	
Thêo1		
Overlapping- Thêo1 Hybrid		
Cross-correlation		

The determination of power law noise type, the number of equivalent χ^2 degrees of freedom, and the setting of confidence limits for the various variance types can be quite complicated, particularly as several techniques have been devised over the years [19, 20].

During the DOS/16-bit Windows era, RAM space was a significant limitation. Because of that, and the simple Allan variance types, computation speed was no a severe problem if a math coprocessor was available. Memory space ceased to be an issue with 32-bit operating systems and speed is generally not much of a problem with today's fast processors. Most of the stability calculations use phase data, with background conversion from frequency data as needed. Figure 4 shows the typical computation run times for a set of octave-spaced Stable32 stability calculations [21].

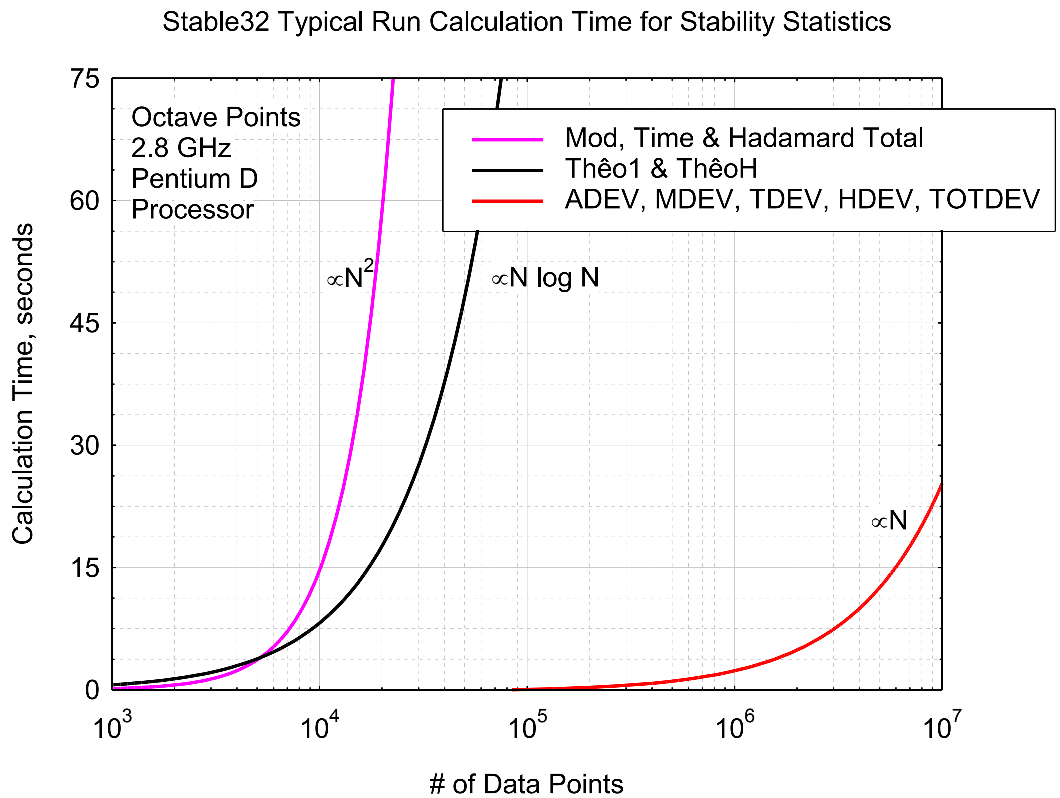


Figure 4. Stable32 Typical Run Calculation Times
The time for a set of octave-spaced stability calculations depends on the number of data points and the variance type.

The simpler statistics of the Allan and Hadamard families and the Total deviation have a calculation time proportional to the number of points, N, while the Thêo1 and ThêoH statistics have a N log N dependence, and the Modified, Time and Hadamard Total deviations have a stronger N^2 dependence.

Stable32 octave and decade stability runs use a maximum averaging factor determined by the number of data points and the variance type. Each variance type is assigned a stop ratio as shown in Table III, and the maximum averaging factor for a given stability run is determined by the number of data points divided by this stop ratio. The number of stability points produced by the run is approximately $\log_2$(N/Stop Ratio).

Table III Stable32 Stop Ratios	
Variance Type	**Stop Ratio**
Allan	5
Overlapping Allan	4
Modified Allan	4
Time	4
Overlap & Mod Allan	4
Hadamard	5
Overlapping Hadamard	4
Total	2
TIE rms	4
Hadamard Total	3
MTIE	2
Thêo1	1
ThêoH	1
Modified Total	2
Time Total	2

Other Stability Analysis Programs

Innumerable software functions and complete suites have been written over the years after the introduction of the Allan variance by many persons and organizations to perform frequency stability analysis. Most of these are customized for a particular application, and have evolved from mainframe computers, time-sharing computing services, minicomputer systems and programmable calculators to personal computers and (perhaps) cloud-based services. The user interfaces have evolved from batch processing of punch cards, teletype machines and CRT terminals to personal computers, first with text and then color graphical user interfaces. As with many such applications, advances in computer hardware and operating systems now allow interactive analysis of large data sets providing greatly-enhanced functionality.

We will mention several other general-purpose frequency stability analysis programs that can serve as examples of the way such software can be implemented:

1. CANVAS is a package of MATLAB routines for frequency stability analysis written by Ken Senior/USNO [22].
2. Plotter is a fairly complete PC freeware suite for frequency stability analysis and plotting written by Ulrich Bangert [23]. It is quite popular amongst time and frequency hobbyists.
3. ALAVAR is another free PC program for frequency stability analysis and plotting written by Alaa Makdissi [24].
4. TimeLab is an excellent clock data acquisition and analysis program that supports John Miles' TimePod clock measurement module and also as a stand-alone application [25].
5. R is a large, general-purpose open-source statistical analysis program that includes the Allan variance (allanvar) and could, presumably, be utilized for the other more complex statistics [26].
6. MATLAB can be used as the basis of an analysis suite (e.g., CANVAS) or as a way to experiment with and check algorithms [27].
7. Excel or other general-purpose spreadsheet programs are not particularly useful for calculating frequency stability statistics.

8. A dynamic link library (DLL) written in C or another language is a way to reuse functions for frequency stability analysis, and, indeed, the Stable32 program uses one called FrequenC.dll for its core analysis functionality [28].

It is quite straightforward to write a function to calculate the Allan variance, and not much harder to do the same for most of the other statistics, although implementing Thêo1 and the associated noise type identification and bias corrections gets a bit complicated. It is a major undertaking to create a complete suite of frequency stability analysis functions along with the associated user interface, especially when one considers the debugging and validation effort. The user interface should be simple enough for a new or untrained operator to use while being flexible enough to allow an experienced analyst the flexibility required for special situations. The software must be supported by effective documentation and tutorial information.

Database management is an important consideration when measuring a large number of clocks over an extended period of time, especially if monitor data is also acquired and stored. Automated stability analysis becomes increasingly important as the number of clocks increases and when consistent processing is required.

Embedded Firmware

The trend in recent years has been to empower instruments with either embedded firmware or as so-called "virtual" instruments whose functionality resides in external computer software. Generally, that functionality is limited to basic user measurement control and display with numeric data outputted and captured for more detailed external analysis.

Stability Analysis Software Validation and Testing

Considerable effort is required to ensure that the results of a frequency stability analysis are correct. Mature software and proven functions should be used whenever possible instead of developing custom software. The analyst must learn the proper operation of his/her analysis tools. The following methods are recommended to validate frequency stability analysis software.

1. <u>Manual Analysis:</u> The results obtained by manual analysis of small data sets such as [7] can be compared to the program output. It is always good to get a "feel" for the process, and small data sets are effective for verifying algorithms and detecting "off-by-one" errors.
2. <u>Published Results:</u> The results of a published analysis or test suite such as [29] or [30] can be compared with the new program output.
3. <u>Other Stability Analysis Programs:</u> The results from other stability analysis programs can be compared with the new program output.
4. <u>General Purpose Programs:</u> The results from standard, general purpose programs such as MathCAD, MATLAB and Excel can be compared with the new program output.
5. <u>Consistency Checks:</u> The new program should be verified for internal consistency, such as producing the same results from phase or frequency data. The standard and 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 about one-half the normal Allan variance for white FM noise, and the normal and overlapping Allan deviations should be approximately equal, with the latter having higher confidence. The various method of drift removal should produce similar results.

6. Simulated Data: Simulated clock data can serve as a useful cross check. Known values of power law noise can be generated, analyzed, plotted and modeled. Known values of frequency offset, drift and jumps can be inserted, analyzed and removed.

Time Scales

Local, national and international time scales and clock ensembles underlie today's timekeeping and navigation systems. Those time scales depend on the Allan variance and related statistics to characterize and model the instabilities of the associated clocks.

Conclusions

The Allan variance concept put frequency stability analysis on a firm footing 50 years ago, and led to the development of other specialized statistics to characterize high-precision clocks and oscillators. Atomic frequency standards have become commonplace devices, especially as time references at telecommunications sites. Higher performance atomic clocks support global navigation systems and a few even higher performance devices at major timing laboratories maintain international time scales. In all these applications and more, the Allan variance serves as the primary means for assessing their stability.

Specialized technical software and its documentation are time consuming and expensive to write, fast and cheap to produce, inexpensive to buy and use, but can be difficult to learn and understand. One of the biggest benefits of well-written software is that it can encapsulate complex procedures like frequency stability analysis, allowing their successful use by non-specialists. One can expect that this trend will continue into the future.

References

1. W.J. Riley, *Handbook of Frequency Stability Analysis*, NIST Special Publication 1065, July 2008. *
2. W.J. Riley, "The Averaging of Phase and Frequency Data". Hamilton Technical Services, February 2011. *
3. *Proceedings of the IEEE-NASA Symposium on the Definition and Measurement of Short-Term Frequency Stability*, Goddard Space Flight Center, Greenbelt, Maryland, November 1964, NASA SP-80, 1965.
4. J.A. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators", *Proceedings of the IEEE*, Volume 34, Number 2, February 1966, pp.207-220.
5. D.W. Allan, "Statistics of Atomic Frequency Standards", *Proceedings of the IEEE*, Volume 34, Number 2, February 1966, pp.221-230.
6. J.A. Barnes, et al, *The Characterization of Frequency Stability*, NBS Technical Note 394, October 1970.
7. B.E. Blair, Editor, *Time and Frequency: Theory and Fundamentals*, NBS Monograph 140, May 1974, Annexes 8.E and 8.F.
8. D.A. Howe, D.W. Allan and J.A. Barnes, "Properties of Signal Sources and Measurement Methods", *Proceedings of the 35th Annual Symposium on Frequency Control*, May 1981, Appendix 1, pp. 669-716.
9. D.A. Howe, "An Extension of the Allan Variance with Increased Confidence at Long Term", *Proceedings of the 1995 IEEE International Frequency Control Symposium*, June 1995, pp.321-329.
10. D.A. Howe and T.K. Peppler, "Very Long-Term Frequency Stability: Estimation Using a Special-Purpose Statistic", *Proceedings of the 2003 IEEE International Frequency Control Symposium*, May 2003, pp. 233-238.

11. D.W. Allan and J.A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability", *Proceedings of the 35th Annual Symposium on Frequency Control*, May 1981, pp. 470-474.
12. D.W. Allan, et. al, "New Inexpensive Frequency Calibration Service from NIST", *Proceedings of the 44th Annual Symposium on Frequency Control*, June 1990, pp. 107-116.
13. S. Bregni, "Clock Stability Characterization and Measurement in Telecommunications", *IEEE Transactions on Instrumentation and Measurement*, Volume 46, Issue 6, December 1997, pp. 1284-1294.
14. W.J. Riley, "Outliers in Time and Frequency Measurements", Hamilton Technical Services, July 2013. *
15. W.J. Riley, "Algorithms for Frequency Jump Detection", *Metrologia*, Volume 45, 2008, pp. S154-S161. *
16. W.J. Riley, "Gaps, Outliers, Dead Time and Uneven Spacing in Frequency Stability Data", Hamilton Technical Services, November 2006. *
17. W.J. Riley, "Application of the 3-Cornered Hat Method to the Analysis of frequency Stability", Hamilton Technical Services, January 2003. *
18. W.J. Riley, *Stable32 User Manual*, Hamilton Technical Services, September 2008. *
19. W.J. Riley and C.A. Greenhall, "Power Law Noise Identification Using the Lag 1 Autocorrelation", *Proceedings of the 18th European Frequency and Time Forum*, April 2004.
20. W.J. Riley. "Confidence Intervals and Bias Corrections for the Stable32 Variance Functions", Hamilton Technical Services, August 2003. *
21. W.J. Riley, "Calculation Times in Stable32, Hamilton Technical Services, September 2008. *
22. K. Senior, R. Beard and J. White, "CANVAS: Clock Analysis, Visualization and Archiving System. A New Software Package for the Efficient Management of Clock/Oscillator Data", *Proceedings of the 37th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, December 2005, pp. 623-627.
23. http://www.ulrich-bangert.de/html/downloads.html.
24. http://www.alamath.com.
25. http://www.ke5fx.com/timelab/readme.htm.
26. http://www.r-project.org.
27. http://www.mathworks.com/products/matlab.
28. W.J. Riley, "The FrequenC Library", Hamilton Technical Services, January 2007. *
29. W.J. Riley, "A Test Suite for the Calculation of Time Domain Frequency Stability", *Proceedings of the 1995 IEEE International Frequency Control Symposium*, June 1995, pp. 360-366, and "Addendum to a Test Suite for the Calculation of Time Domain Frequency Stability", *Proceedings of the 1996 IEEE International Frequency Control Symposium*, June 1996, pp. 880-882.
30. "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology – Random Instabilities", *IEEE Std 1139-1999*, July 1999.

Note: The references marked * are available on the Hamilton Technical Services web site at www.wriley.com.

File: The Evolution of Frequency Stability Analysis Software.doc
W.J. Riley
Hamilton Technical Services
July 26, 2014