Characterizing Frequency Stability Measurements Having Multiple Data Gaps

David A. Howe, *Fellow, IEEE*, and Noah Schlossberger

***Abstract*—Time series measurements with data gaps (dead times) prevent accurate computations of frequency stability variances such as the Allan variance (AVAR) and its square-root the Allan deviation (ADEV). To extract frequency distributions, time-series data must be sequentially ordered and equally spaced. Data gaps, particularly large ones, make ADEV estimates unreliable. Gap imputation by interpolation, zero-padding, or adjoining live segments, all fail in various ways. We have devised an algorithm that fills gaps by imputing an extension of preceding live data and explaining its advantages. To demonstrate the effectiveness of the algorithm, we have implemented it on 513-length original datasets and have removed 30% (150 values). The resulting data is consistent with the original in all three major criteria: the noise characteristic, the distribution, and the ADEV levels and slopes. Of special importance is that all ADEV measurements on the imputed dataset lie within 90% confidence of the statistic for the original dataset.**

***Index Terms*—Allan deviation (ADEV), Allan variance (AVAR), clock, deviation, frequency, imputation, modified, noise, oscillator, power-law, stability, standard, time, total imputer, variance.**

I. Introduction

TIME-SERIES measurements of clocks and oscillators must be equally spaced to characterize noise models using Allan deviation (ADEV). Random gaps ("dead times") cause ADEV and related statistics to mischaracterize noise or to fail outright for gaps greater than 10% of the dataset. Trying to characterize noise with gaps introduces biases and significant ADEV uncertainty. Section II reviews frequency-time statistics and shows how data extensions by reflection instead of by common periodic extensions are independent. Section III gives the imputation algorithm. Section IV shows simulation results of ADEV on data before a 30% segment is removed and after the imputation algorithm is applied to the gapped data.

II. Frequency Stability

Frequency stability is characterized by the Allan deviation, designated as $\sigma_y(\tau)$, ADEV or sq-rt AVAR and its related statistical estimators include the modified-Allan deviation (MDEV), and TotalDEV, Total-MDEV, and Theo with lower uncertainty. These estimators are in the class of "frequency-time" statistics that distinguish between different noise types [1]–[3]. Discrete-continuous functions of either time error $x(t)$ or fractional-frequency error $y(t)$ produce power spectral density (PSD) functions $S_x(f)$ and $S_y(f)$, respectively, by use of fast Fourier transform (FFT) and discrete Fourier transform (DFT) to extract in-phase (AM) and quadrature (PM) noise levels normalized to 1-Hz bandwidth [1]. The ADEV related statistics are different in that they extract distributions of frequency-noise in 1/2 -octave constant-Q filter bins centered at $f = 1/(2\tau)$ as shown in Fig. 1 [4]. Powers-of-two values of AVAR at $\tau = \tau_0, 2\tau_0, 4\tau_0, 8\tau_0, 16\tau_0$, etc., are sufficient to distinguish power-law noise models [5].

Let N be the number of y data, $N = 2^J$. Overlapping AVAR$(2^j\tau_0)$, $j = 0, 1, 2, 3, \ldots$, is (for brevity, τ_0 is omitted)

$$\mathrm{AVAR}\left(2^j\right) \equiv \frac{1}{2^{J-j-1}} \sum_{k=2}^{2^{J-j}} \left(\bar{y}_{2^j k}\left(2^j\right) - \bar{y}_{2^j(k-1)}\left(2^j\right)\right)^2 \quad (1)$$

where $\bar{y}_n(2^j)$ is the average of the last 2^j points at location $n\tau_0$ and

$$\bar{y}_n\left(2^j\right) \equiv \frac{1}{2^j} \sum_{l=0}^{2^j-1} y_{n-l}. \quad (2)$$

From (1) we see that if there are any data gaps, then the number of equally spaced 1st differences in the summand for each 2^j level becomes either zero or not representative of the live data. If the raw data are simply joined through gaps, this causes substantial errors and ambiguous results [6]. Thus, imputed data in spaces of dead time is routinely but wrongly by interpolation based on time-domain regression coefficients or frequency-domain models of data before and after a gap. They include Kalman, maximum-likelihood estimation, ARIMA, and min-max entropy analysis [7]. The problems with these methods are: 1) each can become unstable, hence, unreliable, for creating any data for large gaps beyond about 10% of the live-time data [8] and 2) interpolations do not capture noise properties of live data. The above methods become generally ineffectual at accurately producing ADEV with 10% or more dead time. In summary, the absence of stochastic noise in gaps: 1) causes instability in the above methods as dead time grows, leading to erratic, divergent

Manuscript received October 15, 2021; accepted December 17, 2021. Date of publication December 22, 2021; date of current version January 26, 2022. This work was supported in part by the National Aeronautics and Space Administration, in part by the University of Colorado, and in part by the National Institute of Standards and Technology. *(Corresponding author: David A. Howe.)*

The authors are with the Department of Physics, University of Colorado at Boulder, Boulder, CO 80303 USA (e-mail: david.a.howe@colorado.edu; noah.schlossberger@colorado.edu).

Digital Object Identifier 10.1109/TUFFC.2021.3137425

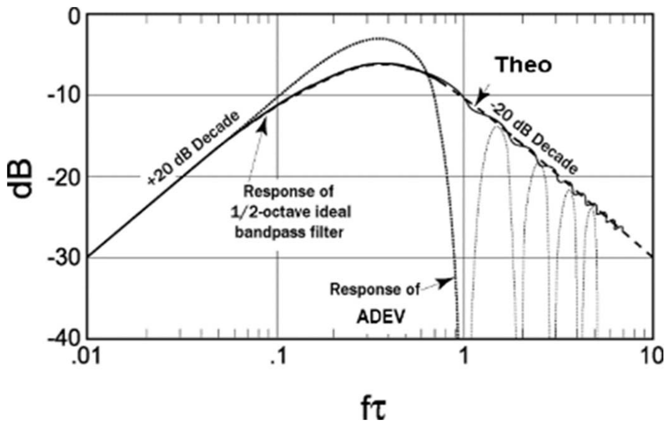

Fig. 1. Digital frequency filter of Allan variance (AVAR), hence, its square-root ADEV. Also shown is Theo1, a statistic created to eliminate the nulls of standard ADEV and match the ideal bandpass response of two cascaded filters, a single-pole high-pass followed by a low-pass with identical break points at $RC = \tau/2$ (dashed line).

results; 2) does not represent the live noise; and 3) causes ADEV bias and errors.

AVAR(τ) treats 1st-differences of $\bar{y}_n$ summand terms in (1) as ergodic [9]. The squared-difference terms of $\bar{y}_n$ in (1) are simply averaged, so are stationary ergodic for non-convergent red $1/f$ and $1/f^2$ spectral noises. The 1/2 -octave bandpass filter caused by the 1st-diffencing operation itself makes levels separated by octaves in (1) independent [5]. In conclusion, AVAR(τ) is, in essence, a digital power-spectrum analyzer whose response is from a filter with a 1/2 -octave distribution of signal components that occur around center-$f = 1/(2\tau)$ as shown in Fig. 1 [4]. This property together with ergodicity means that discrete-continuous signals are regarded as infinitely repeating. The FFT and DFT by way of common examples treat a T-length block of data as infinitely extended, or circular [3]. The imputation strategy in this writing is based on circular convolution.

AVAR(τ) is the mean-square of all sets of a second τ-average frequency subtracted from a prior τ-average frequency. Subtraction is the multiplication of a one-period ± 1 square wave with a 2τ-length interval within the time-series, called its sampling function, that's squared and averaged. We reason that by a conservation principle that the variance cannot diminish for intervals $2\tau >$ T, justifying circularization of the data [10].

Circularization, where the next point after the last point in the time series is the first point, is convolution in the frequency-domain. We remove endpoint discontinuities by reflecting the data run (creating a mirror image) at its last point since ADEV responds identically to data either backward or forward. The TotalDEV is ADEV with this same data extension [11]. Theo applies multiple sampling functions that are the equivalent of circularizing the data for each τ-value without actually circularizing the data for each τ-value. Theo is added to ADEV calculations at τ-values beyond ADEV's longest-possible τ-value at T/2 to create a combination plot called Theo-H [4]. Computations are as fast as standard ADEV [12]. ADEV fundamentally measures "self-similarity" or underlying long-term autocorrelations. White-noise FM is

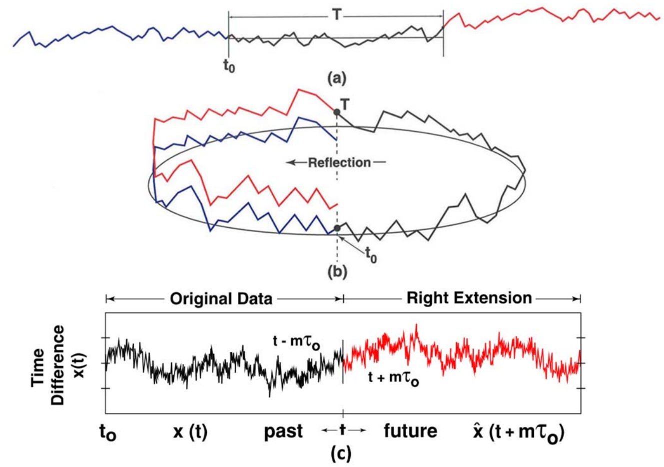

Fig. 2. (a) and (b) Illustration of circularization of data. (c) Right extension of original data by reflection and inversion. Gaps (dead-times) can be on either side of the original "live-time" data shown between as t_0 to $t_0 + T$. Gaps are filled in intervals $t_0 - T$ to t_0 to the left of live data and $t_0 + T$ to $t_0 + 2T$ to the right. New imputed terms in gaps are i.i.d., that is, independent and identically distributed, as the live data and inherently correlated (self-similar) to the degree of the live data.

uncorrelated, whereas random-walk is fully correlated since it is an integral of white noise. Flicker-noise data is correlated by virtue of being a fractional integral [13]. The gap-filling algorithm works with mixed noises [14]. This is important for today's atomic-clock comparisons by satellite that are costly and often have gaps making estimates of a single-clock's ADEV very difficult [15], [16]. The algorithm makes no distinction between random gaps and periodic or deterministic gaps. In keeping with tradition, the name "Total Imputer" is suggested.

III. Strategy for Filling in Data Gaps

A method of data imputation that avoids pitfalls and is effective for up to 100% dead time is to reflect live data around its end and introduce a phase-slope to match the resumption of live data. Fig. 2(a) and (b) illustrates circularized live data. With some care, a dataset with gaps filled by imputing extended data preserves the noise characteristic of the live dataset. The original inherent degree of short- and long-term behavior of a time series when measurements have gaps works best by using inverted-reflected extensions for a "right extension" as shown in Fig. 2(c). Simulations of noises are in the next section.

Some might object to the method here because the reflected data is seemingly not independent identically distributed (i.i.d.) because, while it is identically distributed, it is not clear that reversed sequences are independent. Independence is determined by how the correlation between any two values of the time series changes as their separation, or "lag-$n\tau_0$," increases [17]. ADEV(τ) reports a two-point correlation for lag-τ, and τ cannot exceed $T/2$ set by the definition of ADEV(τ). Processes correlate or decorrelate in proportion to $\tau = \tau_0$, $2\tau_0$, $4\tau_0$, $8\tau_0$, $16\tau_0$, etc., and we find this behavior to continue into gaps when circularized-reflected data is imputed in gaps [14]. We note that extended data is not made-up. It is actual measured data, albeit carefully and strategically manipulated.

An additional argument is by the equivalent degrees of freedom (EDF), i.e., the number of independent values in a statistical average. EDF is chiefly used to calculate the confidence of an average. We find that EDF always increases with extensions relative to un-extended ADEV(τ) calculations, and most dramatically increases for white noise by up to 6X, while less so for flicker and random-walk noise [4].

For a given gap in a dataset of size n missing points, our algorithm is as follows: We take the previous n measurements before the gap (or the following n measurements if there is not enough data before the gap). We then take these n points, reverse the temporal indices, and insert them into the gap. Finally, we add a linear slope to this dataset to match the last point of the inserted data matches up with the point of the data after the gap. A naïve single-point endpoint-matching can produce an artificial sawtooth pattern in the presence of white PM noise, so a more nuanced approach is used and detailed in the Appendix.

For datasets with multiple gaps of dead time, the algorithm that works well is as follows. Find the largest continuous run of data in the set, then impute the gap to its right. Continue to impute consecutive right gaps until the end of the dataset is reached. Then spatially reverse the dataset and continue this algorithm until you reach the beginning of the dataset. In this manner, one can fill all gaps in the dataset provided that the maximum gap size is sufficiently small compared to the initial run of data. In situations where this is not the case, we can impute from the left and right sides and match averages in the middle [14]. If there is a systematic periodic fluctuation embedded in the noise, such as a diurnal, it may be necessary to move the point of reflection to match the phase of the periodic.

IV. Simulations

In this article, we impute a "right extension" of live data that is inverted and reversed in a sequence that decorrelates it while preserving the noise distribution and power-law characteristics of the live data. To demonstrate the effectiveness of this approach, we have implemented the right extensions on 513-length datasets with 150 values missing for the range of clock noise types. The results are shown in Fig. 3. The resulting data is similar to the original in all three major criteria: the noise characteristic, the distribution, and the ADEV curve. Of special importance is that all of these ADEV measurements on the imputed set lie within error bars for 90% confidence of the original dataset.

Monte Carlo simulations of the aforementioned scenario of 150 values missing out of 513 were performed for each noise, shown in Fig. 4. The percentage of trials in which there was overlap in the 90% confidence intervals of the filled and original AVAR showed good consistency despite the large nature of the gap present.

V. Algorithm and Opportunity

The algorithm, called "fillgaps.py" written in Python, is available at: https://zenodo.org/record/5594587. A GitHub repository is at https://github.com/nkschlos/time-series-imputation. An executable version with a simple GUI that takes a .csv input file is at: https://zenodo.org/record/5595200. The .exe is large (55 MB) because it includes all of the Python dependencies. Rather than create analyses of different gap

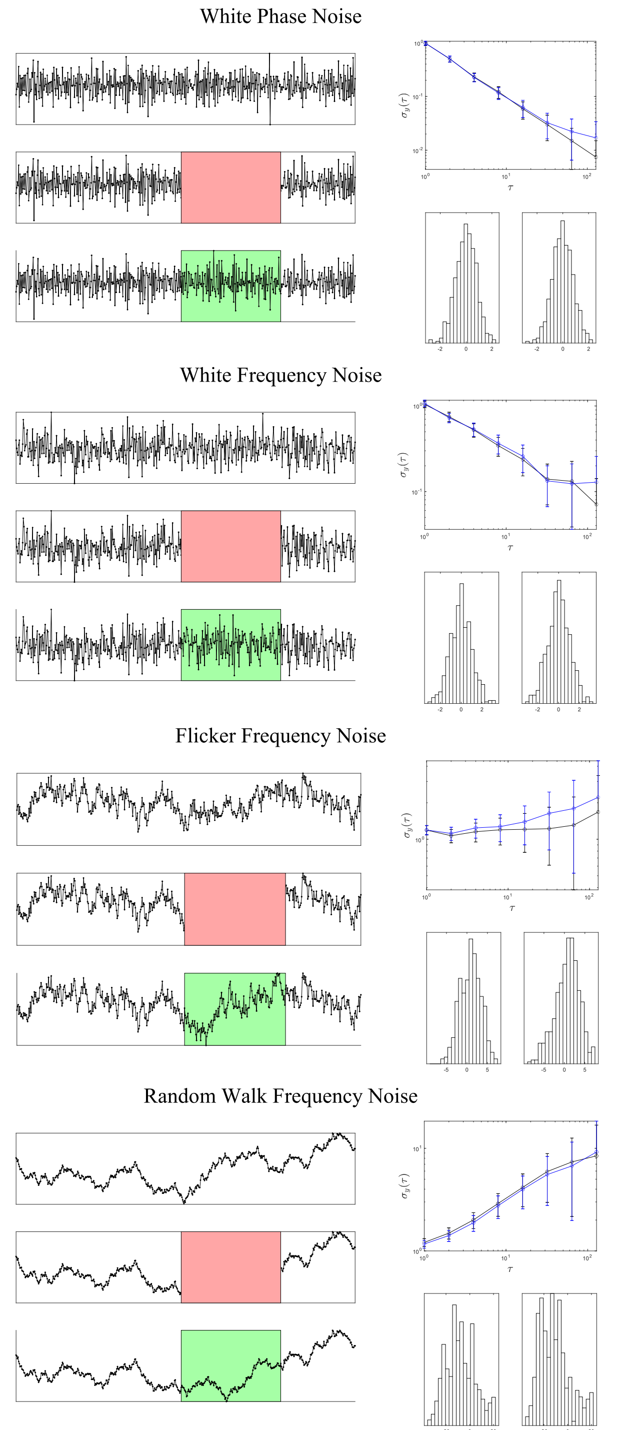

Fig. 3. Application of data imputation algorithm on simulated data of the four predominant noise types. Left: The data reconstruction algorithm. On the top left is the original data, in the middle left the dataset with values removed, and on the bottom left the dataset with imputed values. Top Right: The Allan deviations for the original (black) and recovered (blue) datasets. Bottom Right: The distribution for the original is in the left box and for the recovered data is in the right box.

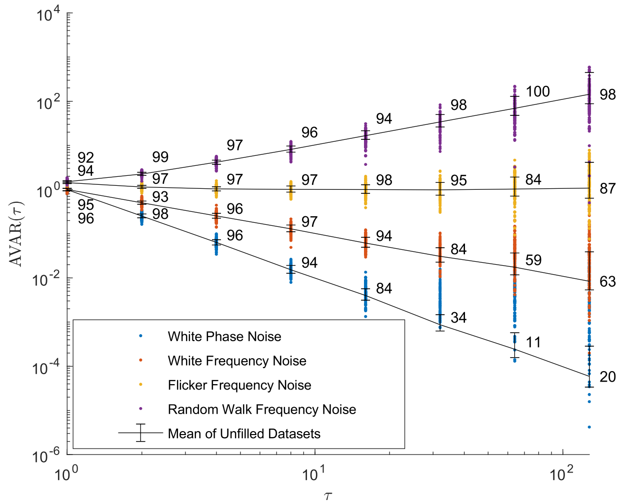

Fig. 4. Monte Carlo simulations of the data imputation algorithm on noise of length 513 points in which points 250–400 were removed and then imputed with our algorithm. 100 trials were done for each noise type. The AVARs of each individual filled dataset are shown scattered in color, while the mean of the original datasets is shown in black with their 90 percent confidence intervals. Next to each point we display the number of trials (out of 100) in which the 90% confidence intervals of the AVAR of the filled dataset overlapped with the 90% confidence intervals of the AVAR of the original dataset.

situations, readers are equipped to run their own simulation, adjust gaps, modify the code to suit, etc. This enables readers to simply try the methodology on their own gapped data and run an analysis to any depth as necessary to see if it meets specific criteria.

The imputation algorithm in the link above is not meant to be used as a substitute for a measurement deficiency. Analysts should take seriously their measurement methodologies to eliminate gaps if at all feasible. If that is not possible or practical, the presence of gaps and the reason for them should be clearly stated along with the analysis results.

VI. Conclusion

Using interpolation, adjoining, or zero-padding to fill gaps fail to provide acceptable ADEV estimates. The imputation algorithm presented here proved successful in simulation trials of ADEV using white frequency modulation (FM) and phase modulation (PM), flicker FM, and random walk FM in which a 30% segment was removed. In Monte Carlo simulations of AVAR, our algorithm, dubbed Total Imputer, working on 30% gaps agreed with the original data within 90% confidence limits in a majority of trials for FM noises.

Appendix

A. Endpoint-Matching Procedure

FM noises originate from a clock's frequency-determining element, or fundamental resonator. PM noise is wideband jitter from later stages of analog and digital components. To demonstrate why adding a linear slope that simply matches the endpoints of the imputed and original data does not work, consider the case of white noise. If the endpoints used to match differ from the local mean, the result is an artificial "sawtooth" feature in the data, as demonstrated in Fig. 5.

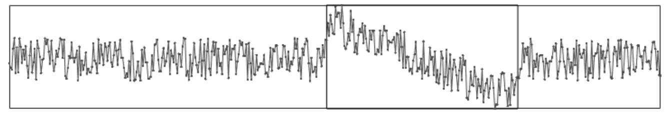

Fig. 5. With simple endpoint matching, the reflect + invert extension on a sample of white noise can result in a visibly incorrect slope associated with imputated data into a gap that's shown as the boxed segment.

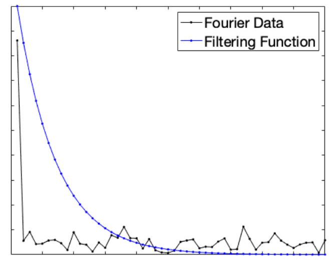

Fig. 6. Plot of the filtering function in the frequency domain, alongside an example Fourier spectrum of white noise.

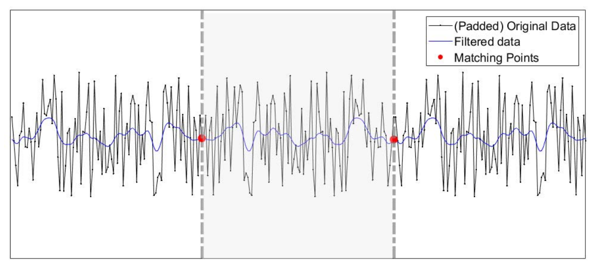

Fig. 7. Plot of the filtering function acting on the data. The middle section is the data segment, padded on either side of the gray dashed line with its own reflection. The thin blue line is then the filtered data with the two filtered are used for endpoint matching. Filtered points are highlighted by the dots in red.

Instead, an approach that is suitable for PM and FM noises is to match endpoints based on a filtered version of the data that ignores higher frequency fluctuations. The filter takes the form of a decaying exponential scaled to the length of the dataset x in the frequency domain as

$$x_{\text{filtered}} = \text{ifft}\left(\text{fft}(x) \cdot e^{-2^3 \frac{|f|}{T} \tau_0^2}\right) \tag{3}$$

where "fft" denotes fast Fourier transform and "ifft" denotes the inverse fast Fourier transform, T is the total length of the imputed data, τ_0 is the difference between points in the time series, and f is the Fourier-frequency variable in operator fft(x). A qualitative idea of what this filter does can be inferred from its plot in Fig. 6.

A remaining issue is that the fft in (3) treats the data as periodic, so when filtered, this will create periodic slopes with steps at the ends that join the first with the last point of the data. The way to avoid these steps is to pad the data with reflected, i.e., flipped, copies of itself on either end. The result is shown in Fig. 7. Most importantly, this technique correctly matches endpoints for not only white PM noise but also all FM noises in Fig. 3.

The algorithm is then as follows.

Use the filter on the block of data to the right and to the left of the gap. Then take the imputation of the filtered dataset and add a linear slope to the unfiltered imputed data such that its difference of endpoints is equal to the difference of the filtered endpoints.

References

[1] *P1139?/D1 (Draft) Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities, IEEE-SA Standards Board, The Institute of Electrical and Electronics Engineers*, Also, Int'l Telecom Union Recommendation ITU-R TF.538-5—Measures for Random Instabilities in Frequency and Time (Phase), New York, NY, USA, 2015.

[2] D. B Sullivan, D. W Allan, D. A Howe, and F. L Walls, "Characterization of clocks and oscillators," Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, Tech. Rep. 1337, Mar. 1990. [Online]. Available: http://tf.nist.gov/timefreq/general/pdf/868.pdf

[3] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of signal sources and measurement methods," in *Proc. IEEE Int. Symp. Freq. Control*, May 1981, pp. 1–47.

[4] D. A. Howe, "TheoH: A hybrid, high-confidence, statistic that improves on the Allan deviation," *Metrologia*, vol. 43, pp. S322–S331, Dec. 2006.

[5] N. Schlossberger and D. Howe, "Analysis of powers-of-two calculations of the Allan variance and their relation to the standard variance," in *Proc. Int. Freq. Control Symp. Eur. Freq. Time Forum*, Apr. 2019, pp. 1–5.

[6] P. Lesage, "Characterization of frequency stability: Bias due to the juxtaposition of time-interval measurements," *IEEE Trans. Instrum. Meas.*, vol. IM-32, no. 1, pp. 204–207, Mar. 1983.

[7] D. Mondal and D. Percival, "Wavelet variance analysis for gappy time series," *Ann. Inst. Stat. Math.*, vol. 62, no. 5, pp. 943–966, 2010.

[8] G. Box and G. Jenkins, *Time Series Analysis, forecasting and control*. San Francisco, CA, USA: Holden-Day, 1970.

[9] P. Halmos, *The Mathematical Society* (Letures on Ergodic Theory), Tokyo, Japan: Kenkyushu Printing, 1956.

[10] D. A. Howe, "Circular representation of infinitely extended sequences," in *Proc. EEE Int. Freq. Control Symp.*, Feb. 1999, pp. 337–346.

[11] C. A. Greenhall, D. A. Howe, and D. B. Percival, "Total variance, an estimator of long-term frequency stability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control.*, vol. 46, no. 5, pp. 1183–1191, Sep. 1999.

[12] B. Lewis, "A fast algorithm for calculation of Thêo1," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, vol. 10, pp. 2187–2190, Oct. 2020.

[13] B. Mandelbrot, "Statistical self-similarity and fractional dimension," *Science*, vol. 156, no. 3775, pp. 636–638, sMay 1967.

[14] D. A. Howe, C. Champagne, and N. Schlossberger, "A 'total' imputation algorithm that fills gaps in time series measurements for ADEV and phase noise characterizations of power-law noise models," in *Proc. Joint IEEE Int. Freq. Control Symp. Eur. Freq. Time Forum (IFCS-EFTF) Symp.*, Paris, France, Apr. 2022.

[15] C. Hackman and T. E. Parker, "Noise analysis of unevenly spaced time series data," *Metrologia 33D*, vol. 33, pp. 457–466, Oct. 1996.

[16] D. Yu, M. A. Weiss, and T. E. Parker, "Uncertainty of a frequency comparison with distributed dead time and measurement interval offset," *Metrologia*, vol. 44, pp. 91–96, Feb. 2017.

[17] J. Beran, "Statistics for data with long-range dependence," *Stat. Sci.*, vol. 7, pp. 404–427, Apr. 1996.

David A. Howe (Fellow, IEEE) worked as a Statistical Theorist with the Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO, USA, from 1994 to 1999. He is currently an Adjunct Professor with the University of Colorado at Boulder, Boulder.

Prof. Howe is also the Creator of the Total and TheoH frequency stability variances. From 1987 to 1994, he led and implemented several global high-accuracy satellite-based two-way time-synchronization experiments with other national laboratories for which he was awarded the Gold Medal from NIST.

Noah Schlossberger received the B.S. degree in physics and mathematics from Indiana University at Bloomington, Bloomington, IN, USA, in 2018, and the M.S. degree in physics from the University of Colorado (CU) at Boulder, Boulder, CO, USA, in 2021, where he is currently pursuing the Graduate Certificate in biomedical engineering and the Ph.D. degree in physics.

His research interests include atomic and molecular physics, imaging science, biomedical technologies, and signal processing.