2022 Joint Conference of the European Frequency and Time Forum and IEEE International Frequency Control Symposium (EFTF/IFCS) | 978-1-6654-9718-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/EFTF/IFCS54560.2022.9850921

# A 'Total' Imputation Algorithm that Fills Gaps in Time Series Measurements for ADEV and Phase Noise Characterizations of Power-law Noise Models

D. A. Howe, *Fellow, IEEE*
National Institute of Standards and Technology, University of Colorado, Boulder, CO 80305 USA

C. Champagne
Space Science Division, U.S. Naval Research Lab, Wash. DC 20375 USA

N. Schlossberger
University of Colorado, Boulder, CO 80303 USA

***Abstract*—In this paper, we introduce an algorithm that uses the 'Total' approach as a means of filling gaps between live measurements. The Total approach extends a T-length data run by a backward, inverted replica before and after it, which in essence triples the data length to 3T as in Fig. 1. This operation is a way of obtaining long-term ADEV($\tau$) beyond its $\tau$ = T/2 limit with better overall confidence [1]. Ostensibly, the extensions convert the linear convolution of ADEV($\tau$) to circular but with an additional nuance: Total extensions are *iid*, i.e., independent, identically distributed while furthermore preserving intrinsic clock noise models [2]. We aptly call the gap-filling algorithm a 'Total imputer.' It has been tested on sets of 513-point simulated data runs in which 150 points (30%) were removed halfway into the runs. In Monte Carlo trials of ADEV($\tau$) plots for each of five $f^{\alpha}$, integer power-law noise models, i.e., where $-2 \le \alpha \le +2$, the imputation algorithm proved successful in yielding consistent ADEV($\tau$) responses in recovering the 150-point gaps with 90% confidence between original and recovered data [3]. We provide well-developed Python code and example data that enables analysts to test the algorithm on data which they may have or wish to study with simulation on specific patterns of missing data.***

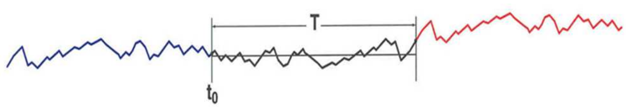


Fig. 1. Total extensions before and after T-length data in the middle.

**Keywords—ADEV, dead-time, gaps, imputation, missing data, power-law noise models, Python, sparce, time, time-series, Total.**

## I. Introduction

The Allan deviation (ADEV), phase noise ($\mathcal{L}(f)$) and general PSD characterizations of clock and oscillator noise, rely on equally-spaced time series data without gaps (dead time) [4]. A large gap or several gaps extending over many data points presents a severe obstacle because they can render ADEV($\tau$) and $\mathcal{L}(f)$ estimation with substantial errors especially at large $\tau$ or low $f$, respectively.

Every effort should be made to obtain clock and oscillator phase measurements with no gaps. If measurements have gaps, the analyst should seriously consider retaking the data or revising the measurement methodology. But when this is not possible, the Total-derived algorithm provided in this paper is the first practical imputer for obtaining sample ADEVs and PSDs that can readily fill large and multiple gaps without making any assumptions as to the data's missingness patterns nor mixtures of different noise models.

## II. Limits of Traditional Remedies for Gaps

If gaps exist in measurements, the reason for them should be made clear. A few small gaps (such as those caused by outlier removal) can be handled by simply ignoring them. But traditional solutions have diminished value for large gaps as studied in [5,6]. Ref. [5] requires a specific ratio of dead-to-live time. All methods need some *a priori* knowledge or a suitable determination as to which power-law noise model is dominant during gaps.

## III. Procedure

Total extensions of the live data are by an inverted mirror reflection at each of its ends as shown in Fig. 1. The extensions are used to fill gaps. A linear slope is added to an extension such that its last point matches the local mean of the start of live data. If not matched, the result is an artificial "sawtooth" pattern where a gap is filled [3]. Reflected + inverted Total extensions applied in this manner are *iid* because: (1) for phase noise and ADEV, blocks of T-length windows are processed while Total extensions produce a period = 2T which is outside window T, (2) extended data are not made up but are actual measured data, albeit carefully and strategically manipulated, and (3) *iid* persists in extensions to the same degree as the measured noise is inherently *iid*, quantified as self-similarity [7,8].

## IV. Large Gaps

Total imputer breaks traditional limits of Sec. II and can be applied to missingness patterns which have total gaps of over 100% of a data run. A case of filling gaps of over 100% of a data run that can use the Total approach would be the following. Suppose we have 10 measurements in the beginning of a time series, followed by a gap of 100 points, followed by 100 points of measurements, followed by another gap of 100 points, and ending with 10 measurements. This scenario would have 120 points of live measurements and 200 points of gaps, or a data run with 166% gaps. Total imputer can be applied to these large data gaps. Scenarios can vary, the main point being that imputation can be applied to patterns of small and large gaps.

For example, referring to Fig. 2, the top plot is original data that are the time differences between a NIST H-maser and the NIST time scale, UTC(NIST). Here, 3.5 x $10^3$ measurements are taken with 240s between each measurement. Four large segments of the original data have been removed in the middle plot. The imputation algorithm applied to the middle plot produced the bottom plot.

## V. Compararison of Results of $\mathcal{L}$(f) and ADEV

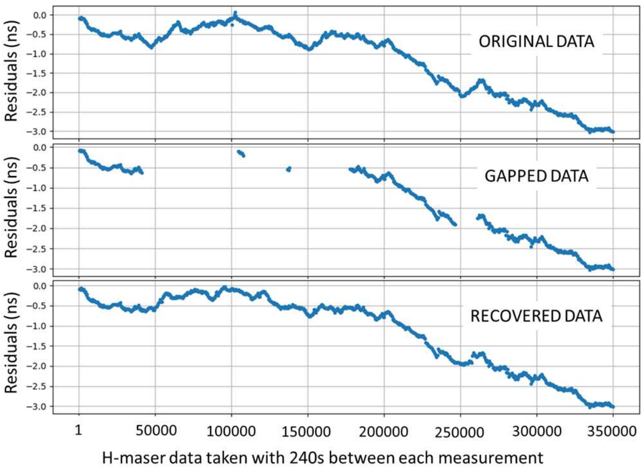


Fig. 2. TOP: Time-series measurements of NIST H-maser vs. UTC(NIST); MIDDLE: Four large gaps >100% of live data are intentionally created; BOTTOM: Test showing the gaps are filled in by the imputation algorithm.

We compare $\mathcal{L}(f)$ and ADEV of the original and recovered data as shown in Fig. 2. The results are shown and described in Fig. 3. These results of comparisons of $\mathcal{L}(f)$ to the lowest original $f$ of 8 μHz and full ADEV($\tau\sim 1d$) of samples of original and recovered data are indeed exemplary. To a significant degree, testing by you, the reader, is encouraged rather than compiling exhaustive tests of the myriad of gap patterns that would be broad, consuming and would not cover every case. Python code that is well-developed is provided to enable readers to try it on their data or on included examples and to run simulation.

## VI. Access to the Algorithm

The algorithm, called "fillgaps.py" written in Python, is available at: https://zenodo.org/record/5594587 and well-developed for Ref. [9]. An executable version with a first-draft GUI that runs on Windows and takes a .csv input file is at: https://zenodo.org/record/5595200. The .exe is large (55Mb) because it includes all of the Python dependencies. A GitHub repository https://github.com/nkschlos/time-series-imputation.

## VII. Conclusions

We introduce the Total imputer as perhaps the most effective method yet devised in filling data gaps for computations of ADEV($\tau$) and phase noise levels over the fullest possible range of $\tau$-values and Fourier-frequencies, respectively. We include the algorithm in a Python repository so that analysts can use it directly on their data.

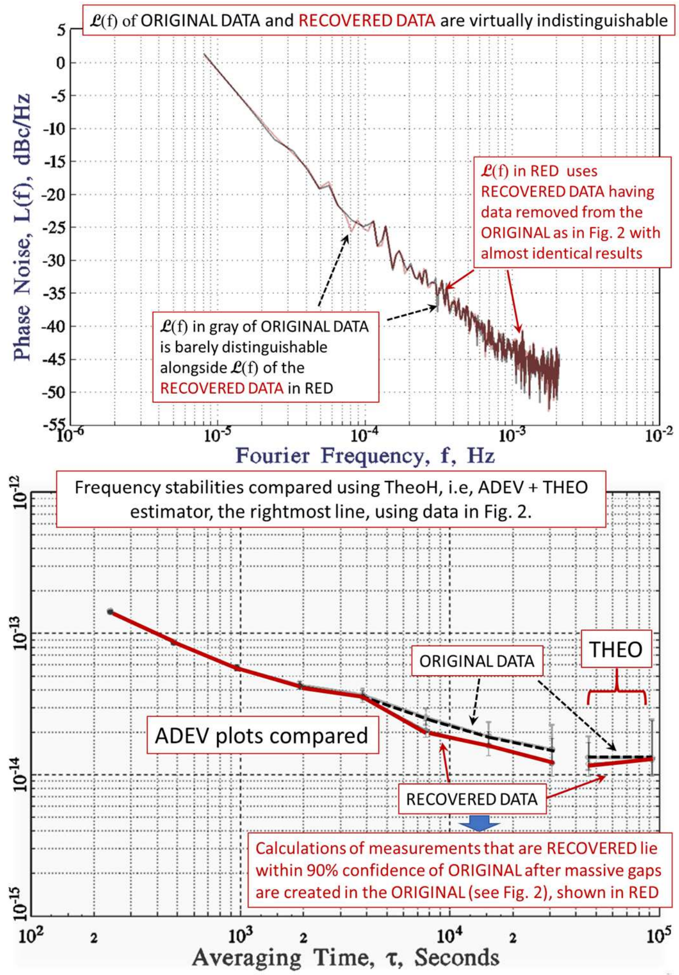


Fig. 3. TOP: $\mathcal{L}(f)$ shows that original vs. recovered phase noise are virtually identical and that the recovered data extends to the same low Fourier frequency as the original. BOTTOM: Frequency stability using high-confidence TheoH of original vs. recovered data.

## References


[1] D. A. Howe, "Total Variance Explained," Proc. 1999 Joint Mtg. IEEE Intl. Freq. Cont. Symp. and EFTF Conf., pp. 1093-1099.

[2] C.A. Greenhall, D.A. Howe and D.B. Percival, "Total Variance, an Estimator of Long-Term Frequency Stability", *IEEE Trans. UFFC*, Vol. 46, No. 5, Sept. 1999, pp. 1183-1191.

[3] D. A. Howe and N. Schlossberger, "Characterizing frequency stability measurements having multiple data gaps," in process, dhowe@nist.gov.

[4] P1139™/D1 (Draft) Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology – Random Instabilities, IEEE-SA Standards Board, New York 10016-5997, USA (in process).

[5] J. A. Barnes, "Tables of Bias Functions, B1 and B2, for Variances Based On Finite Samples of Processes with Power Law Spectral Densities," NBS Tech. Note 375, 42 p.

[6] C. Hackman and T. E. Parker, "Noise analysis of unevenly spaced time series data," Metrologia **33D**, pp. 457-466.

[7] G. Box and G. Jenkins, *Time Series Analysis, forecasting and control*, Holden-Day, San Francisco, 1976.

[8] D. B. Percival, "The statistics of long memory processes," Ph.D. dissertation, Department of Statistics, University of Washington, Seattle, WA, 1983.

[9] J. S. Hazboun, et al., "A Detection of Red Noise in PSR J1824−2452A and Projections for PSR B1937+21 using NICER X-ray Timing Data," in process, http://arxiv.org/abs/2112.02160.