FREQUENCY AND PHASE BREAK DETECTION

Scott Czopek
E-mail: ***scott.czopek@colorado.edu***

Abstract

A phase break detection method is proposed and successfully tested for the CANVAS analysis software for frequency standards. Phase breaks are detected above the noise floor by classifying a very large phase step as a phase break. This detection method succeeds in numerically quantifying a large visual jump in the phase plot without using input parameters other than the phase data. These phase breaks are automatically identified by the software, and then are removed from the data. To implement this solution, it is assumed that the frequency standard is well behaved. The most extreme phase steps (1% of the total data) are assumed to contain all phase breaks and other misbehaving data points, and this small subset of 1% is neglected during the detection method's preliminary analysis. If these assumptions are violated, then this phase-break detection method does not apply to the set. The phase-break detection algorithm still needs to be interfaced with the CANVAS user interface. Also, this method and the frequency-break detection method are intended for post-process use. A frequency-break detection method is also proposed, and the assumptions that invalidate the method are explained.

BACKGROUND

This work is the author's first paper. It was done under the supervision of Dr. Ken Senior at the Naval Research Laboratory, during the summer of 2009 and continued during the fall semester at the University of Colorado at Boulder. The author was an undergraduate when this paper was prepared, but graduated December 2009 with a Bachelor's degree in physics and a minor in mathematics. Clock data for this paper came from the GPS Block IIF Rubidium atomic clock #25, under life test at the Naval Research Lab, and a hydrogen maser. The GPS Rubidium Clock has been in operation since 22 August 2008 **[1]**.

PHASE BREAKS

DETECTION AND REMOVAL METHODS

A phase break's phase step is much more positive or much more negative than an average phase step. Therefore, to identify outliers is to identify phase breaks. These outliers are detected by determining the data's average phase step size and then determining which phase steps are much more positive or much more negative than the average phase step. The sample data's average phase step size is determined by eliminating the outliers and then taking the average. Assuming that a frequency standard is well behaved most of the time, so 1% of the data, the most extreme data points, are eliminated. And also, assuming that all the outliers are removed, the average is then taken. If the computed average is near the data's mode value, then the assumption that less than 1% of the data points are bad is valid. If not, the method does not apply to the data set. (The word *near* is quantified by using the standard deviation.)

How far each phase step lies beyond the average is quantified by the sample data's standard deviation. Assuming that the data set contains few outliers, the standard deviation is calculated by eliminating the same 1% of the data, then computing the standard deviation. If a phase step lies more than 2.5 standard deviations beyond the average phase step, then that phase step is classified as a phase break. (2.5 standard deviations worked very well on the Block IIF rubidium frequency standard's life test data provided by the Naval Research Laboratory.)

AUTOMATIC BREAK CORRECTION METHOD

Correcting phase breaks is reduced to loading the data, detecting the breaks, and correcting the data. For this paper, the trend line will be removed to show that the phase breaks are removed and are not masked by the frequency offset between standards.

LOAD PHASE BREAK DATA

The phase data were manually loaded into the CANVAS software from the Time and Frequency, Naval Research Lab database, as seen in Figure 1. As shown, the data for the Block IIF Rubidium 25 and the data for the P9 Maser both contain several phase breaks.

The Block IIF Rubidium 25 is a GPS frequency standard under life test at the Naval Research Laboratory. This atomic clock was built by Perkin-Elmer and has been in continuous operation since 22 August 2008 **[1]**. Whereas, the P9 Maser is one of the Naval Research Lab's Time and Frequency Division's reference clocks. It was built by the Smithsonian Astrophysical Observatory in the late 1970s and has been used by the Naval Research Lab since then.

DETECT PHASE BREAKS

Phase breaks, which are statistically outlying phase steps, are detected by determining the data's average phase step size and then determining which phase steps are much more positive or much more negative than the average phase step. The sample data's average phase step size is determined by eliminating the outliers and then taking the average. Assuming that a frequency standard is well behaved most of the time, so 1% of the data contains all of the outlying data points. This extreme 1% of the total data is eliminated, then the average is taken. If the computed average is near the data's mode value, then the assumption that less than 1% of the data points are bad is a valid assumption. If not, this method does not apply to the data set. (The word *near* is quantified by using the set's standard deviation.)

How far each phase step lies beyond the average is quantified by the sample data's standard deviation. Assuming that the data set contains few outliers, the standard deviation is calculated by eliminating the same 1% of the data, then computing the standard deviation. If a phase step lies more than 2.5 standard deviations beyond the average phase step, then that phase step is classified as a phase break. (2.5 standard deviations worked very well on the Block II F rubidium frequency standard's life test data provided by the Naval Research Laboratory.) The detected phase breaks can then be highlighted in light blue for easy visualization as seen in Figure 2.

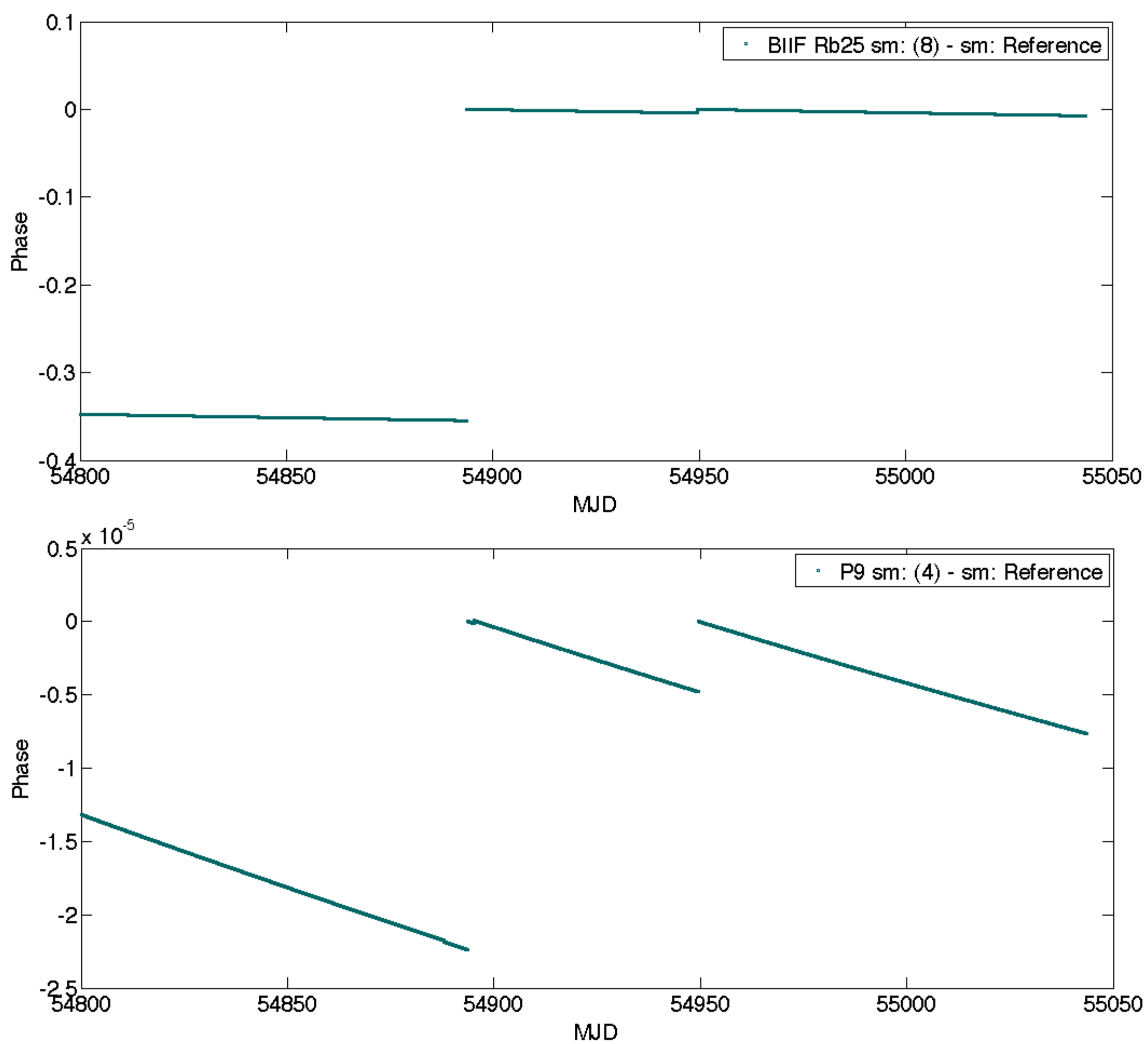


Figure 1. Loading the Data. The Block IIF Rubidium Clock 25 plot (top) and Maser P9 plot (bottom) shows 200 days worth of phase data – several phase breaks are evident in each plot.

CORRECT PHASE BREAKS

After phase break detection, the jumps in the trend line are removed. A quadratic trend line is fitted to 1 day's worth of data before the break. A quadratic curve is used to represent the contributions from the clock's phase offset, frequency, and drift. This trend line is projected forward in time to the breakpoint estimating the clock's phase if the break did not occur. A break did occur and the measured phase differs from the estimation. The difference between the phase measurement and the phase estimate is subtracted from the data points occurring after the phase break, removing the phase break from the subsequent data. Since the frequency offset between frequency standards is much, much larger than the drift offset between frequency standards, the phase data with the phase breaks removed is a straight trend line. This is shown in Figure 3.

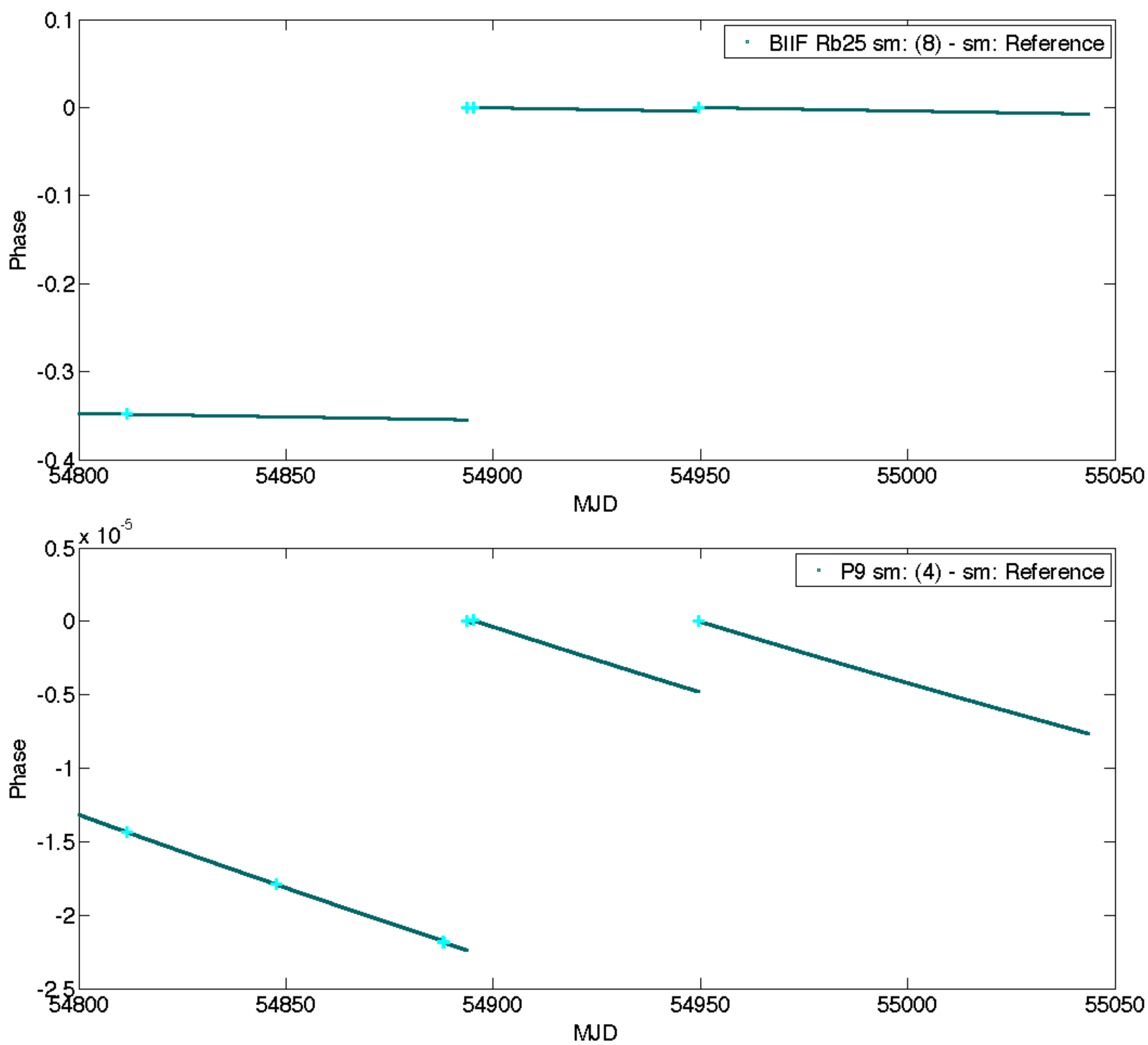
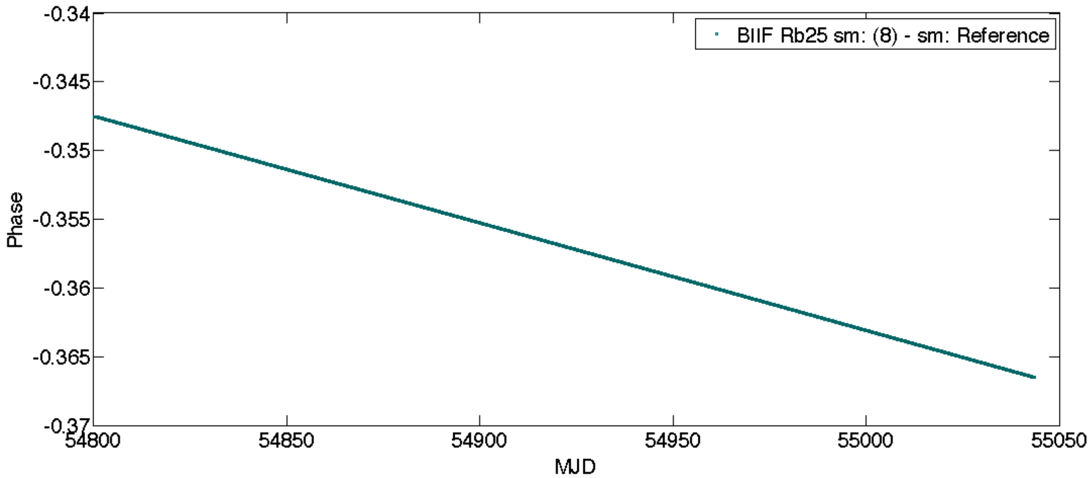


Figure 2. Detecting Phase Breaks. The phase breaks for the Block IIF Rubidium 25 (top) and the Maser P9 (bottom) data are highlighted in light blue.



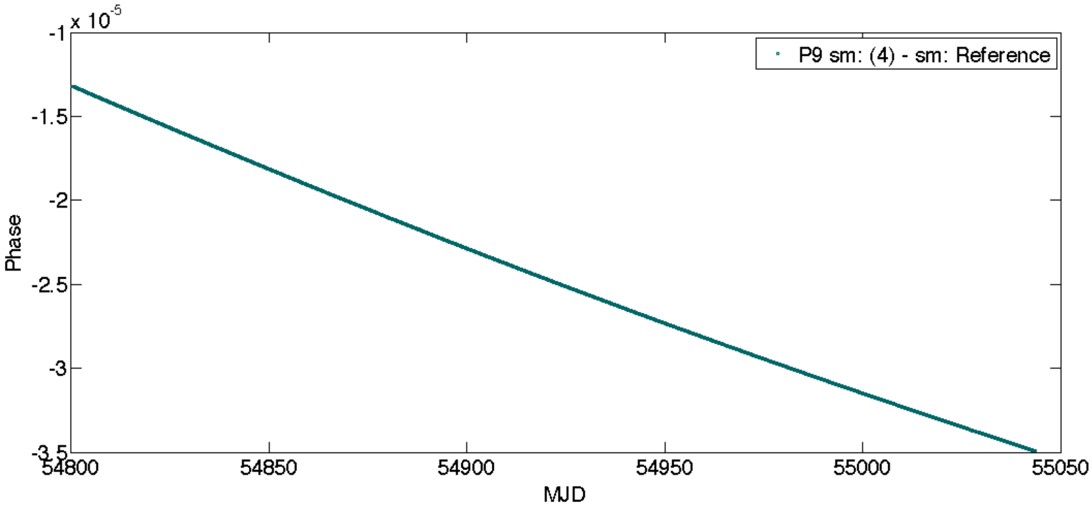
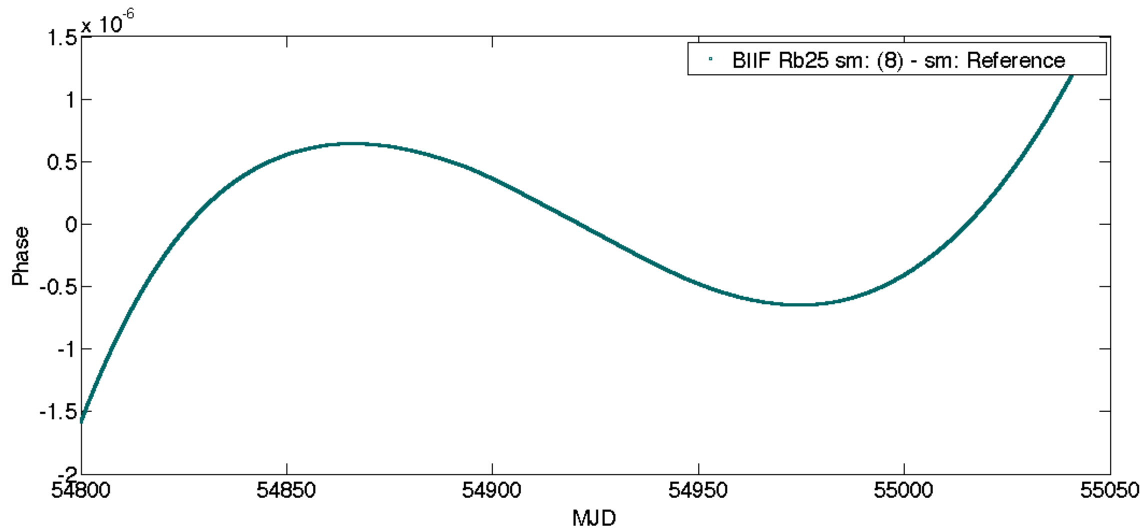


Figure 3. Correct phase breaks. The phase breaks are corrected and removed from both phase plot graphs.

CONFIRM METHOD WORKS

The trend line is removed to more closely inspect the data (Figure 4). It is assumed that the phase breaks are removed if they are not visually evident while viewing all the data at once and if a negligible jump is observable when examining a few data points surrounding each phase break (Figure 5).



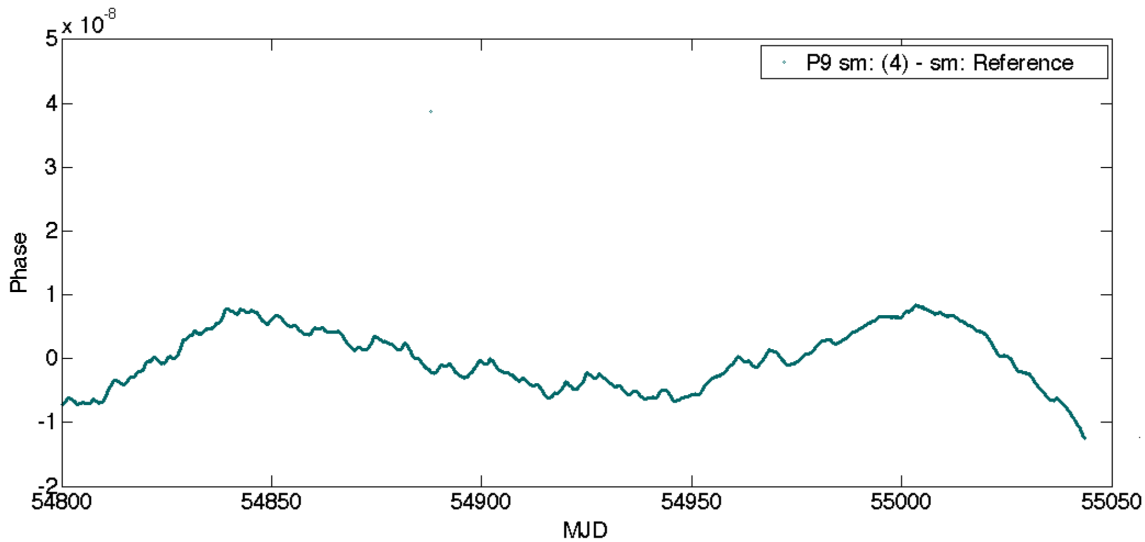


Figure 4. Remove Trend Line. The phase data trend line for both plots is removed. This allows the data to be more closely inspected to determine if the phase breaks have been removed.

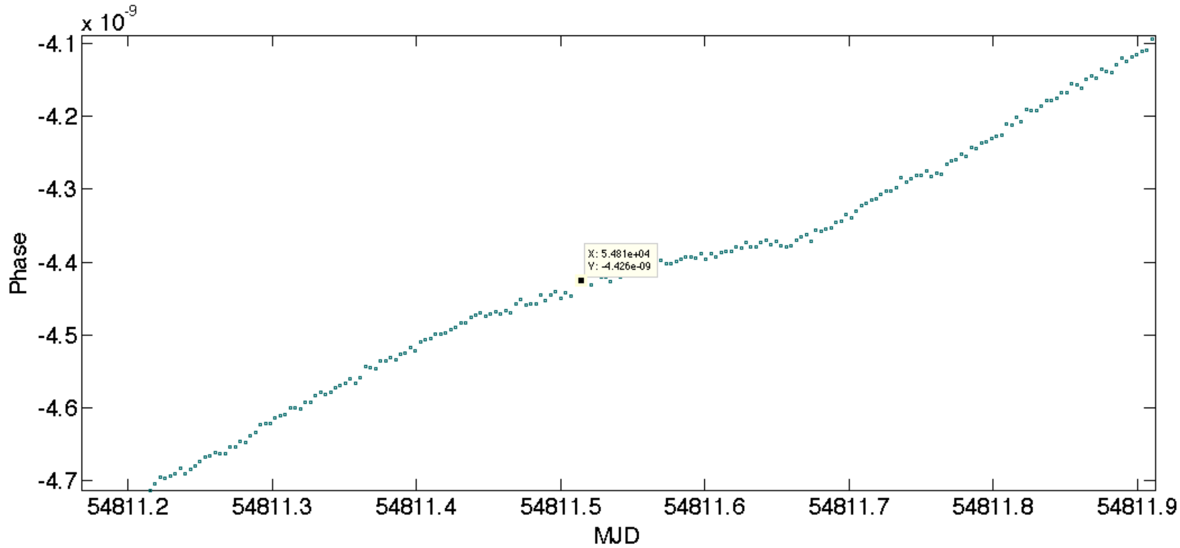


Figure 5. Visually inspect each phase break. Each phase break is individually inspected to confirm that all phase breaks are removed. This particular phase break for the P9 maser is flagged by a small yellow box.

FREQUENCY BREAKS

INVALID ASSUMPTIONS

The method that will be proposed to detect frequency breaks is invalid, because of poor simplifying assumptions, assuming that the frequency data is normally distributed. If the solution that detects frequency breaks for normally distributed data is applied to data with a power law distribution, an error occurs. The frequency break point needs to be known before applying the solution to data following the power law model. Since frequency data from atomic clocks and other frequency standards follows the power law model, requiring that the frequency-break timestamps be known in advance is unacceptable.

FREQUENCY BREAK DETECTION – UNCORRELATED TIME SERIES

If an uncorrelated time series distribution could be assumed for frequency data, then frequency breaks can be detected with a chi-squared analysis of the frequency values. The statistical distributions for data with no frequency breaks and for data with a single frequency break and for data with are different. If no break occurred, then the data are best modeled with one uncorrelated time series frequency distribution. If, instead, a single frequency break has occurred, then the data are best modeled by two distributions. One of the two uncorrelated time series distributions represents the frequency data before the break, while the other uncorrelated time series distribution represents the frequency after the break. Multiple breaks can occur and post-processing a large amount of data to detect an unknown number of frequency breaks is poor algorithm design. It simpler to look at a small sample of the data that is only likely to contain a single frequency break. This does not work if the smaller data sets contain multiple frequency breaks. If multiple breaks, exist the method is likely to indicate that a break is present, but provide a wrong frequency break timestamp. How well the single break and no break models fit the data is determined by Matlab's *chi2gof* function (chi-squared goodness-of-fit function). A judgment can then be made to determine which model is a better fit.

Selecting the model with the better quality of fit does not indicate if a frequency break has occurred. The fit with two distributions has at least one more degree of freedom than the fit with one distribution. Often, the more degrees of freedom give a better fit, so if both models fit the data well, a determination needs to be made if the second degree of freedom gives a substantially better fit. A threshold value quantifies the fit quality difference between models and determines if the frequency model is a substantially better fit.

The threshold value is determined with Matlab by examining many frequency data sets with a single frequency break and also with no frequency breaks. The single frequency break data sets were simulated by combining arrays of normally distributed data with separate mean values. (The normal distribution was chosen for simplicity.) The combined data array was compared with other data sets representing data with no frequency breaks. These data sets were simulated by arrays whose entries are filled with realizations of a single normal distribution. The threshold value is almost the mean difference in fit qualities between the single frequency break and no frequency break. To account for fluctuations above the mean, the threshold value is increased to two standard deviations above the mean.

FUTURE RESEARCH

PHASE BREAKS

The phase break detection algorithm is limited. When a quadratic trend line is removed from 200 days of Block II rubidium clock data, the residuals follow a deterministic pattern and forms a cubic polynomial. This causes the variances to be larger, because the variance calculation includes a pattern whose mean value evolves with time. And the sample's variance is central to detecting phase breaks, allowing only very large phase breaks to be found. If the data set could be subdivided to suppress this long-term trend or if the clock was better understood, allowing for better model, then perhaps the observed behavior would mostly be random and the variance would better describe the clocks internal noise. If the clock's internal noise were better described with this method, then smaller phase breaks would be detected.

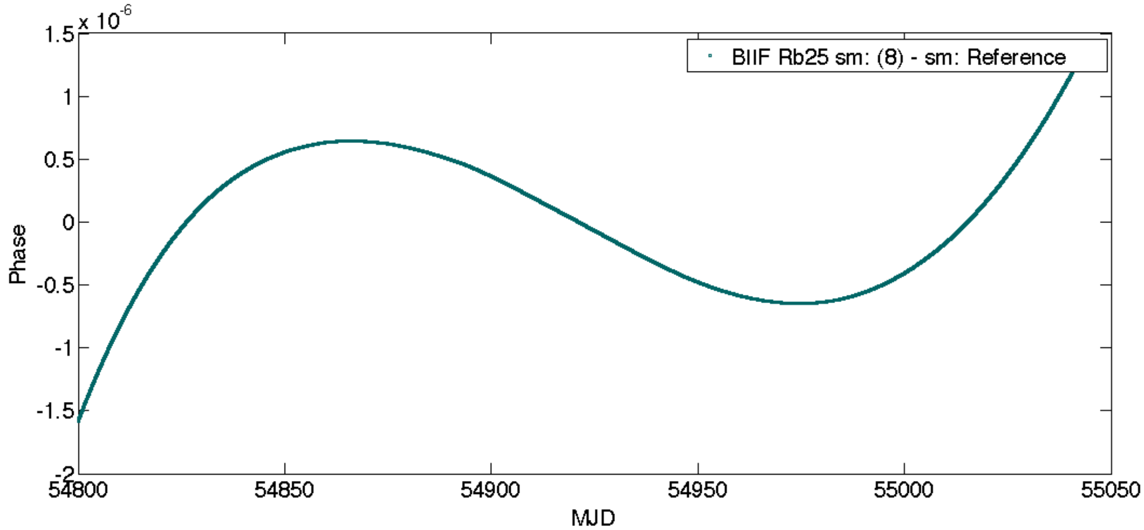


Figure 6. Deterministic behavior despite trend line removal.

FREQUENCY BREAKS

The frequency breaks detection is based upon invalid assumptions. To develop a successful frequency-break detection algorithm, these assumptions would be revised to include a power law model for clock processes. This improved model for clock noise would be tested under a Monte Carlo simulation using frequency data sets, where some sets contain a single frequency break and other sets contain no frequency breaks. Using this approach is difficult, because the data sets are strongly correlated time series. For a strongly correlated time series, it is difficult to select a point at which to divide the data set into pre- and post-frequency break data, without previously knowing the frequency break point. Tackling this problem or providing an alternative solution will be the focus of additional research.

CONCLUSIONS

The process for detecting very large phase breaks can successfully be automated and the process for automatically removing these breaks could be automated with more effort. These conclusions are supported with data from GPS flight-qualified rubidium and cesium clocks, and several hydrogen masers.

Frequency breaks were not successfully detected, because weakly correlated time series were assumed. This assumption is wrong, because the frequency data are a strongly time-correlated time series. Better understanding power law distributions and Kalman filters would allow this problem to be revisited.

ACKNOWLEDGMENTS

I sincerely thank Dr. Ken Senior for affording me this opportunity. It greatly increased my ability to code in Matlab. I also thank Dr. Leo Mallette for help with editing this paper. His turnaround time for each revision of this paper was amazing.

REFERENCES

[1] F. Vannicola, R. Beard, J. White, K. Senior, K. Kubik, D. Wilson, and A. Buisson, 2010, *"GPS Block IIF Rubidium Frequency Standards Life Test,"* in Proceedings of the 41st Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 16-19 November 2009, Santa Ana Pueblo, New Mexico, USA (U.S. Naval Observatory), pp. 449-456.