NASA/TM–20240006860

Automated detection of spurious signals in VLBI phase calibration data

Joe Skeens
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas

Leonid Petrov
NASA Goddard Space Flight Center, Greenbelt, Maryland

June 2024

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NTRS Registered and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI Program, see the following:

- Access the NASA STI program home page at `http://www.sti.nasa.gov`

- Help desk contact information: `https://www.sti.nasa.gov/sti-contact-form/` and select the "General" help request type.

NASA/TM–20240006860

Automated detection of spurious signals in VLBI phase calibration data

Joe Skeens
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas

Leonid Petrov
NASA Goddard Space Flight Center, Greenbelt, Maryland

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771-0001

June 2024

Available from:

NASA STI Program / Mail Stop 150
NASA Langley Research Center
Hampton, VA 23681-2199

Abstract

In this memorandum, a set of processing strategies for automatic masking of phase calibration tones is outlined as implemented in the software package PIMA in the task Generate Phase Calibration Mask, or GEPM. The task relies on a robust procedure of cleaning phase calibration data before employing several mathematical strategies designed to selectively identify spurious signals from phase calibration tones. These strategies were derived as more rigorous implementations of heuristics traditionally used by analysts in manually identifying problematic phase calibration data.

The task is intended to automate the process of generating a phase calibration mask and in so doing increase the speed and regularity of VLBI analysis. At the outset of the project, a series of goals were identified to evaluate the success of this mask generation. This included the development of an algorithm to identify and mask short-term (defined as less than 10 seconds in length) spurious signals affecting phase calibration data, the development of an algorithm to identify and mask phase calibration tones affected by constant radio-frequency interference, the implementation of a detection scheme for identifying large jumps in phase calibration phase caused by clock breaks, and finally a method of reporting phase calibration health metrics to the user.

Processing of over 10 experiments has demonstrated that the final form of the task GEPM as detailed in this technical memorandum satisfies each of these conditions and satisfactorily performs the task of automatic phase calibration data masking, although the wide variety in quality and characteristics of phase calibration data makes a single solution to the problem quite difficult. A series of user inputs have therefore been defined to assist analysts in tailoring automatic masking to specific stations and phase calibration generators.

The vast majority of applicable code was written in FORTRAN to increase execution speed and ease of interfacing with the existing code base in PIMA, but a wrapper function in Python was also written to allow for a simpler method of interacting with GEPM and inputting relevant parameters. In addition to this report, thorough documentation was added to the already existing repositories associated with PIMA as a whole.

Contents

List of Figures

1 Nomenclature

1.1 Symbols

θ	phase of calibration tone
ϕ	phase of calibration tone relative to reference tone
A	amplitude of calibration tone
C	complex plane representation of calibration tone
t	time
ω	angular frequency
τ_g	group delay

1.2 Acronyms

VLBI	very long baseline interferometry
IF	intermediate frequency
RMS	root-mean-square
FFT	fast Fourier transform

2 Introduction

A phase calibration system is a hardware signal generator that injects rails of narrowband signals, or phase calibration tones, into the receiving system of a VLBI dish in an effort to provide a calibration for instrumental delays due to various hardware components including cables, amplifiers, analog downconverters, and signal distributors. These signals are injected either into the feed horn of an observing antenna or directly into the feed just after.

The basic setup of the phase calibration hardware consists of a ground unit and an antenna unit. The ground unit is fed by the 1, 5, or 10 MHz signal from the frequency standard, which is amplified and distributed to several components in a ground circuit, including for example a phase shifter and comparator reference. [1]

The signal from the ground unit passes through a cable to the antenna unit. The antenna unit then creates a pulse train with a tunnel diode. This pulse sequence results in short impulses about 30-50 ps in length. In some cases, the tunnel diode has been replaced by a digital phase calibrator developed by MIT Haystack, as the older tunnel diode-based generator has become unavailable. [2] The pulses are typically generated at a rate of 1 or 5 MHz, which results in a phase calibration tone appearing at every integer frequency in the bandpass of recorded data. These tones appear as narrow spikes in the bandpass of less than 1 Hz width.

Phase calibration tones are then extracted either at the digital baseband converter of the receiving system or in a software correlator. In both methods, the phase and amplitude of each phase calibration tone in the recording band are computed. This phase and amplitude capture a number of instrumental signals, including the changing electrical length of cables due to stretching or thermal variation, phase delays through up and downconversion devices and distributors, changes in the local oscillator of the receiver, and gain and loss through the radio frequency chain of the antenna. These amplitude and phase variations caused by hardware can occur both in frequency and in time. The phase and amplitude of the phase calibration tones then serve as a calibration signal for amplitude changes as well as group and phase delays applied by instrumentation, which can then be removed during the fringe fitting process. In effect, the tones allow us to measure the phase response of the system and to sum the signal across the band coherently.

Because much of the electronics in every VLBI antenna and receiving system are all stabilized by a common frequency standard, and the receiving system includes a number of up and downconversions, harmonics that coincide with the narrowband phase calibration signals are commonly generated. These internally generated harmonics are of negligible significance for the broadband VLBI data, but they are a major source of distortion for the narrowband phase calibration tones, as they distort both the amplitude and the phase of the tone. Applying a phase calibration including a distorted phase calibration tone can introduce spurious delays and lead to decoherence in an accumulation, introducing noise in any subsequent imaging or geodetic application. For this reason, phase calibration tones must be monitored and masked out when they are affected by a spurious signal. This technical report details an attempt to implement an automatic procedure for phase calibration tone masking using several filters and statistical strategies. This analysis has been implemented in the software PIMA, which is a VLBI visibility data analysis package that is described in Petrov (2011). [3] The hypothesis that underlies the phase calibration masking done by analysts and in the automatic processing strategies described in this document is that the phase and amplitude of phase calibration tones should be a relatively smooth function in both time and frequency. Deviation from smoothness in time or frequency can therefore be used to identify when tones are distorted by a spurious signal.

Phase calibration tones are organized into intermediate frequency (IF) bands that are typically 4-128 MHz each. Depending on the phase calibration pulse frequency and the number of tones extracted in software correlation, there may be 1-128 tones per IF. In addition, VLBI observations are typically of 1-24 hr in length with data divided into 1-10 min observations of individual sources called scans. As the antennas slew between sources at scan boundaries, phase calibration data is not collected, making sampling of the phase calibration tones irregular. I represent the phase calibration data with three indices, i, j, k, with i running over time epochs as in Equation 1 and j and k running over frequency in tone and IF respectively (Equation 2). A conceptual view of the grouping of frequency bands is shown in Figure 1, where a recording frequency band is divided into 6 IFs with 8 spectral channels, each containing a single narrowband phase calibration tone.

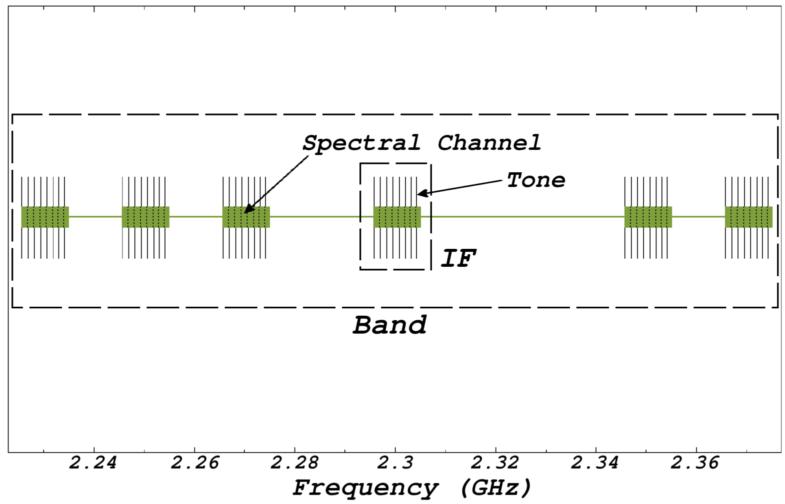

Figure 1: An S band frequency setup in the experiment r41056 showing a band with 6 IFs of 8 spectral channels and 8 phase calibration tones each.

$$t = t_i \in [0, T], \;\; i \in [1, N_{\mathrm{epochs}}] \tag{1}$$

$$\omega = \omega_{jk}, \;\; j \in [1, N_{\mathrm{tones}}], \;\; k \in [1, N_{\mathrm{freqs}}] \tag{2}$$

This leads to a phase, θ, and ampltiude, A, of the phase calibration tones expressed in these three indices:

$$\theta_{ijk} = \theta(t = t_i, \omega = \omega_{jk}) \tag{3}$$

$$A_{ijk} = A(t = t_i, \omega = \omega_{jk}) \tag{4}$$

An example of phase calibration tones at a single epoch for an experiment with 6 IFs of 8 tones each in S band data is shown in Figure 2, which was plotted with the PIMA command *mppl*. The phase calibration tones affected by spurious signals are circled in red. It is important to note that these spurious signals are characterized by a sudden jump in phase with respect to frequency, breaking the smooth curve that the tones would otherwise trace.

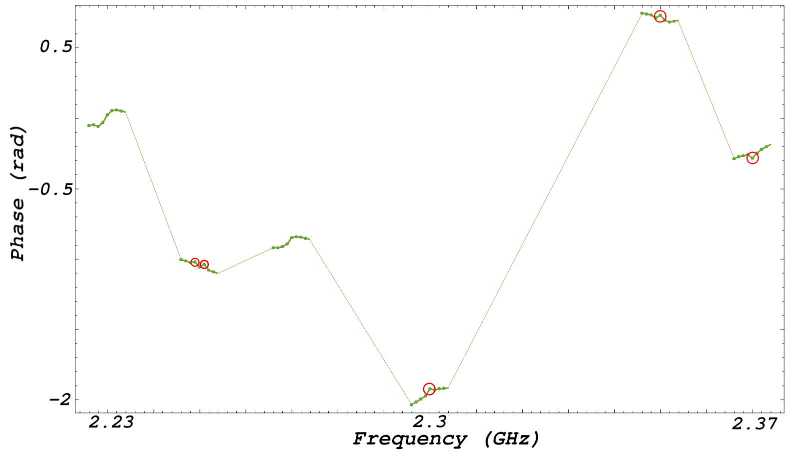

Figure 2: Phase calibration tones for a single epoch with phase unwrapped by frequency for 6 IFs. Spurious tones are circled in red.

Within a single IF, the wrapping phase of healthy phase calibration tones are plotted in Figure 3 by the PIMA command *pcpl*.

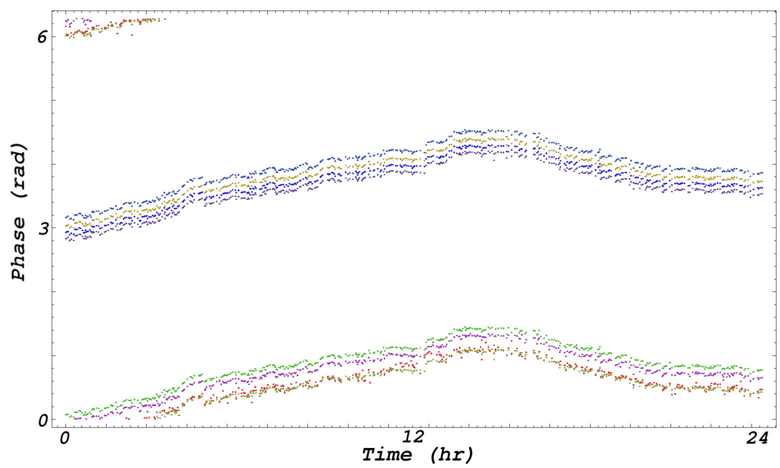

Figure 3: Phase calibration tones in a single IF plotted against time.

Phase calibration tones affected by spurious signals often appear in the time domain as having notably more phase jitter than other tones in the IF, making them unsuitable for calibration. Figure 4 shows this phenomenon, where the phase calibration tone in red displays notably more time domain jitter than its peers.

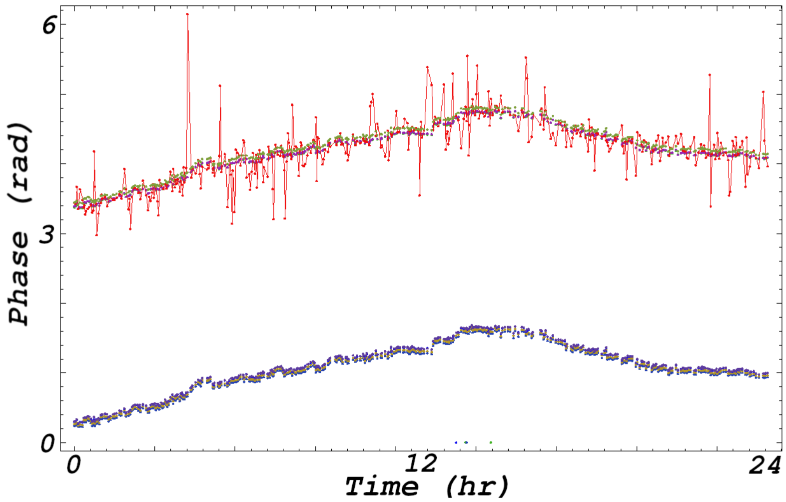

Figure 4: Phase calibration tones plotted against time with a spurious signal evident by rapid temporal variation.

Distortion by spurious signals is not, however, the only problem affecting phase calibration tones. Another common problem that must be identified and removed is discontinuous jumps in the phase calibration signal. This most commonly occurs when there is a jump in the frequency signal from the local oscillator in the ground unit of the phase calibration generator, which is called a clock break. Discontinuities in phase and group delay may also be caused by the distributor. These discontinuities may affect both phase and group delay, or either phase or group delay individually. Detection of jumps in phase calibration tones may also indicate a group delay bias in the complex visibilities. An example clock break is shown in Figure 5.

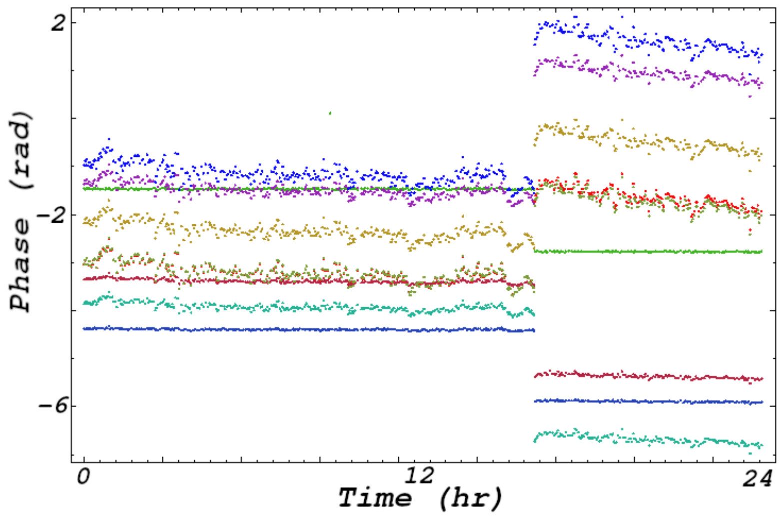

Figure 5: Unwrapped differenced phase showing a simultaneous phase calibration jump indicative of a clock break.

3 Design of Phase Calibration Masking Procedure

3.1 Group Delay Determination and Removal

In all IFs of a station's phase calibration tones, there is a common residual group delay that evolves over time, $\tau_g(t = t_i) = \tau_{g_i}$, where the phase of a phase calibration tone is approximately given by,

$$\theta_{ijk} \approx \theta_{i11} + \tau_{g_i}(\omega_{ijk} - \omega_{i11}) \tag{5}$$

To evaluate the smoothness of the phase calibration tones in frequency, this group delay must first be removed. This is done with a Fast Fourier Transform (FFT). The first step to this is casting the phase calibration tone phase and amplitude to the complex plane:

$$C_{ijk} = A_{ijk} \left(\cos(\theta_{ijk}) + i \sin(\theta_{ijk})\right) = A_{ijk} e^{i\theta_{ijk}} \tag{6}$$

Demonstrating the technique with a continuous function and the continuous Fourier transform,

$$F(\xi, t) = \int_{-\infty}^{\infty} C(\omega, t) e^{-i\xi\omega} \mathrm{d}\omega = C_1(t) \int_{-\infty}^{\infty} e^{-i\omega(\xi - \tau_g(t))} \mathrm{d}\omega = C_2(t) \cdot \delta(\xi - \tau_g(t)) \tag{7}$$

Where $C_1(t)$ and $C_2(t)$ are constants with respect to frequency composed of several terms, and $\delta(\xi)$ is the Dirac delta function. Thus, in the continuous, noise-free case, there is a single impulse that corresponds to the residual group delay at the epoch of evaluation. To find the group delay, it is only necessary to find the location of the peak in the Fourier transformed function,

$$\tau_g(t) = \underset{\xi}{\operatorname{argmax}}\ F(\xi, t) \tag{8}$$

In the discrete case, the group delay is selected as the argument of the maximum from a set number of trial values. The group delay is then removed from the phase calibration tones as,

$$\theta_{ijk} = \theta_{ijk} - \tau_{g_i}(\omega_{ijk} - \omega_{i11}) \tag{9}$$

Figure 6 shows a series of phase calibration tones wrapping in frequency, which is resolved by Equation 9 in Figure 7.

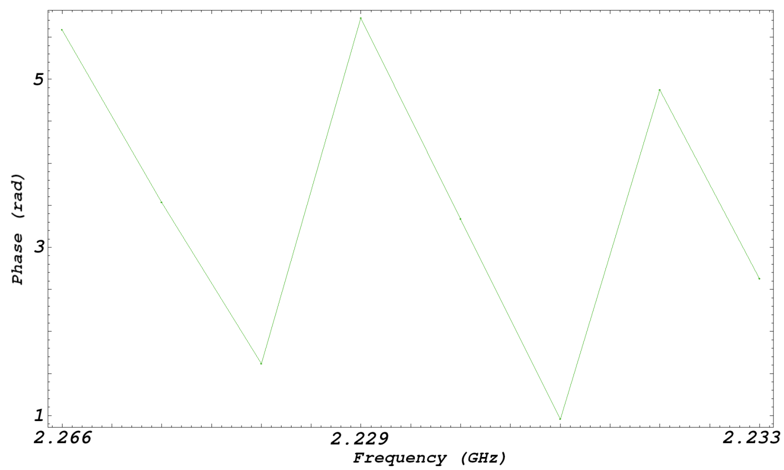

Figure 6: A phase calibration tone displaying a rapid wrapping pattern indicative of an unremoved group delay.

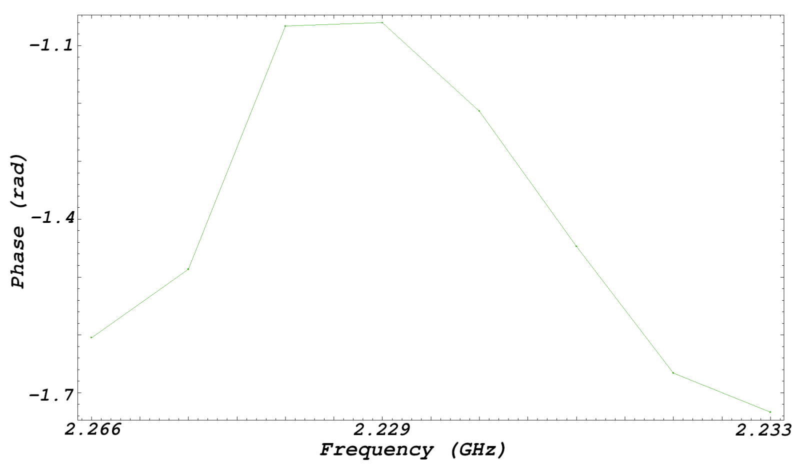

Figure 7: The above phase calibration tone after removing the group delay, now displaying a round-off indicative of bandpass structure.

Figure 7 now shows another problem inherent to analyzing phase calibration data–the tones in the IF have a distinct, bell-shaped bandpass structure in phase. Figure 8 shows that this same structure appears in every IF. The bell-shaped structure also appears in amplitude. This structure is due to the phase response of the filter applied to each IF.

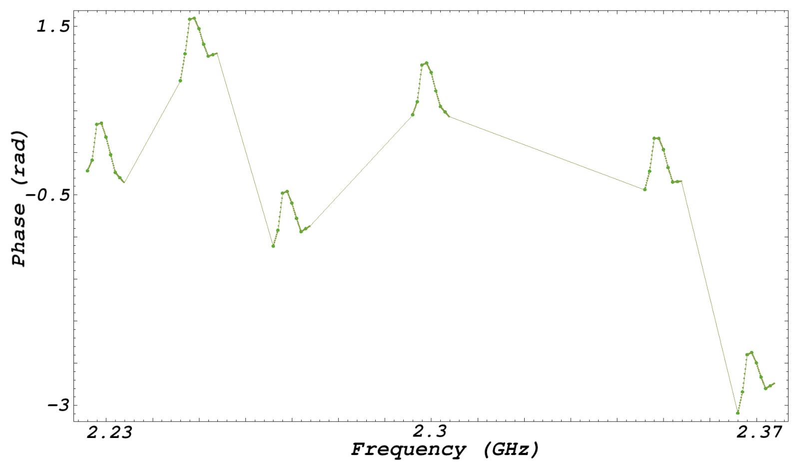

Figure 8: A group of 6 IFs all showing the same bandpass structure after group delay removal.

Before spurious signal analysis can begin, this bandpass structure must be removed so it is not mistaken as a sign of a spurious signal. This is done by analyzing the average phase and amplitude across all epochs (Equations 10 and 11).

$$\Phi^0_{jk} = \frac{1}{N_{\text{epochs}}}\left(\sum_{i=1}^{N_{\text{epochs}}} \theta_{ijk} - \frac{1}{N_{\text{tones}}}\sum_{i=1}^{N_{\text{epochs}}}\sum_{j=1}^{N_{\text{tones}}} \theta_{ijk}\right) \tag{10}$$

$$A^0_{jk} = \frac{1}{N_{\text{epochs}}}\left(\sum_{i=1}^{N_{\text{epochs}}} A_{ijk} - \frac{1}{N_{\text{tones}}}\sum_{i=1}^{N_{\text{epochs}}}\sum_{j=1}^{N_{\text{tones}}} A_{ijk}\right) \tag{11}$$

The bandpass structure curves, Φ^0_{jk} and A^0_{jk}, contain the average structure of phase and amplitude in the frequency domain across all epochs and have average values of 0 across all channels of each IF. These quantities are removed from the phase and amplitude of the phase calibration tones by simple subtraction:

$$\theta^*_{ijk} = \theta_{ijk} - \Phi^0_{jk} \tag{12}$$

$$A^*_{ijk} = A_{ijk} - A^0_{jk} \tag{13}$$

Figures 9 and 10 show the phase and amplitude of the tones in the example IF at one epoch before and after the bandpass structure is removed. Note that the removal of the bandpass structure is only to simplify spurious signal analysis. It would introduce phase and amplitude biases to remove these structures from the phase calibration tones when they are used to calibrate the interferometric data, so the phase calibration tones are left unmodified during fringe fitting.

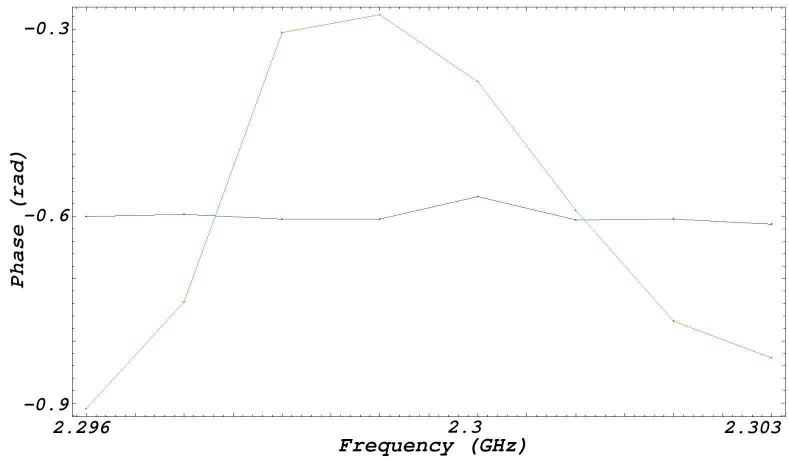

Figure 9: Phase calibration phases before (green) and after (blue) bandpass structure removal.

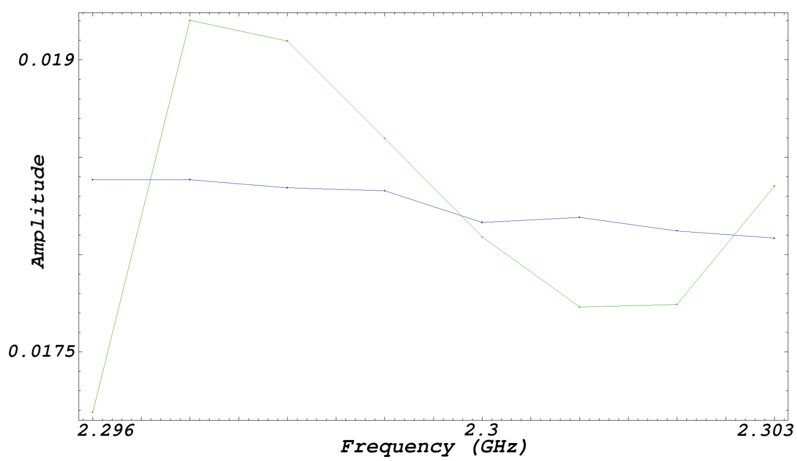

Figure 10: Phase calibration amplitudes before (green) and after (blue) bandpass structure removal.

3.2 Phase Calibration Tone Processing

Before filters can be applied and problematic tones detected and masked, the phase calibration data are processed such that phases and amplitudes are generally flat and regular in time and frequency. Figure 11 shows an initial, unprocessed phase calibration tone. This tone displays noticeable phase jitter and phase wrapping, defined by the phase jumping back to 0 when it exceeds 2π radians or the reverse.

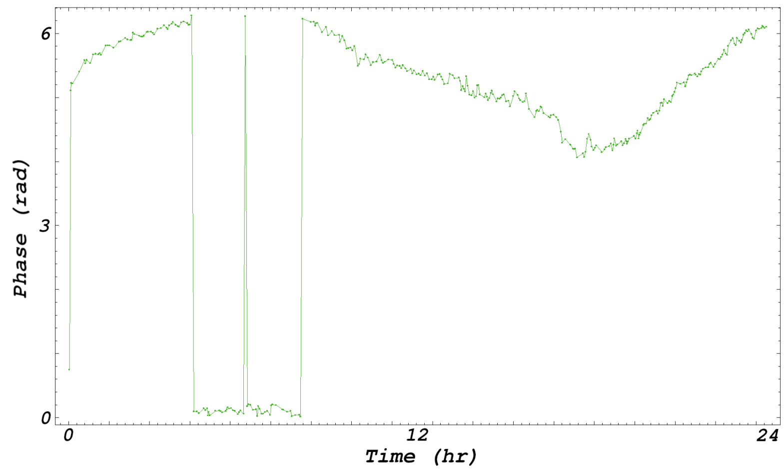

Figure 11: An unprocessed phase calibration tone showing phase jitter and wrapping throughout a 24-hour experiment.

To reduce natural phase jitter and provide phase calibration data that are easier to analyze, the phase calibration tones are time averaged within scans. For many phase calibration time series, the sampling of the phase calibration data is determined by scan boundaries and is thus irregular. Phase calibration data is also not recorded when the VLBI dish is slewing. The first step to averaging is thus to find the average sampling interval, Δt_{avg}:

$$\Delta t_{\text{avg}} = \frac{1}{N_{\text{epochs}} - 1} \sum_{i=1}^{N_{\text{epochs}}-1} t_{i+1} - t_i \tag{14}$$

A user inputs the desired average sampling interval $\Delta t'$, which is then used to find the number of data points that must be averaged to most closely approximate that desired sampling interval, where n_{avg} is the number of channels that must be averaged together. The CEIL and INT functions round their arguments to the next integer and closest integer respectively.

$$n_{\text{avg}} = \text{CEIL}\left[\frac{N_{\text{epochs}}}{\text{INT}\left[\frac{\Delta t'}{\Delta t_{\text{avg}}}\right]}\right] \tag{15}$$

Equations 16 and 17 show this averaging process. To average the phase calibration tones, the phase ambiguities must first be resolved. This is done through a phase ambiguity resolution process already implemented in PIMA. In the equations, the time epoch index i is redefined to i', reflecting a new set of epochs after averaging.

$$\theta_{i'jk} = \frac{1}{n_{\text{avg}}} \sum_{l=i}^{i+n_{\text{avg}}} \text{AMB}\left[\theta_{mjk},\ m \in [i,\ i+n_{\text{avg}}]\right]_l \tag{16}$$

$$A_{i'jk} = \frac{1}{n_{\text{avg}}} \sum_{l=i}^{i+n_{\text{avg}}} A_{ljk} \tag{17}$$

After the application of Equation 16, the phase wrapping is resolved, and there are fewer phase calibration epochs, as shown in Figure 12.

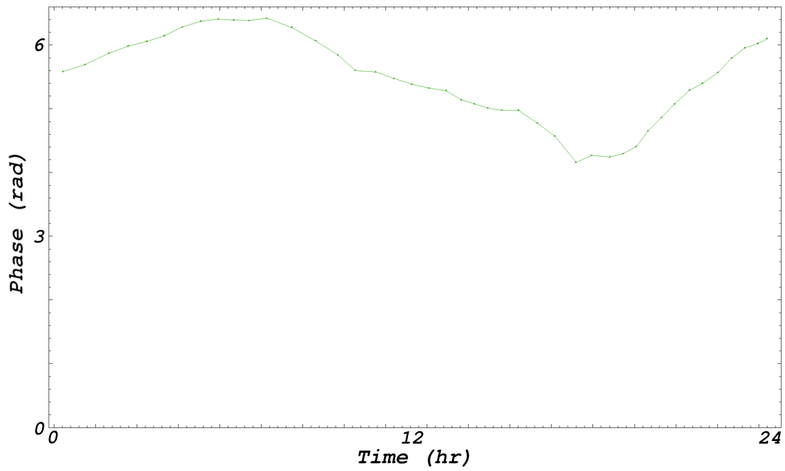

Figure 12: The phase calibration tone after resolving phase ambiguities and unwrapping.

One other variable that is quite useful for detecting spurious signals in phase calibration tones is the differenced phase calibration tone, ϕ_{ijk}. This is created by selecting a reference phase calibration tone, usually in the first IF, and differencing the corresponding tones in the other IFs by this reference tone. If a tone in the first IF is found to be spurious, the reference tone is instead set to one in the closest IF in which the tone is free from spurious signals. Equation 18 shows this differencing:

$$\phi_{ijk} = \theta_{ijk} - \theta_{ij}^{\text{REF}}, \; k \in [2, N_{\text{freqs}}] \tag{18}$$

The differenced tone is useful because much of the variation of the phase calibration data is common to all tones, as many of the noise processes present in phase calibration data, such as phase and amplitude variations in the wires, distributor, and downconversion hardware are reflected in all tones simultaneously. Figure 13 shows two somewhat noisy phase calibration tones before differencing with the reference tone in green. After differencing (Figure 14), much of the visible variation has been excised. This makes the differenced tone sensitive to distortions by a spurious signal.

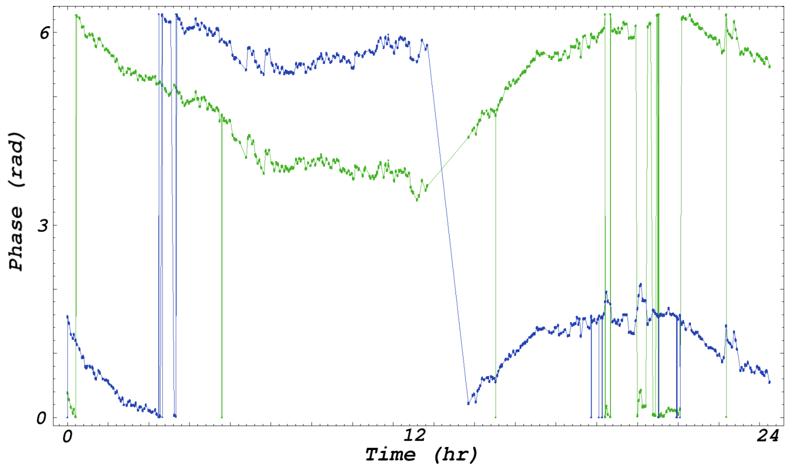

Figure 13: Two wrapping phase calibration tones in the same station from different IFs.

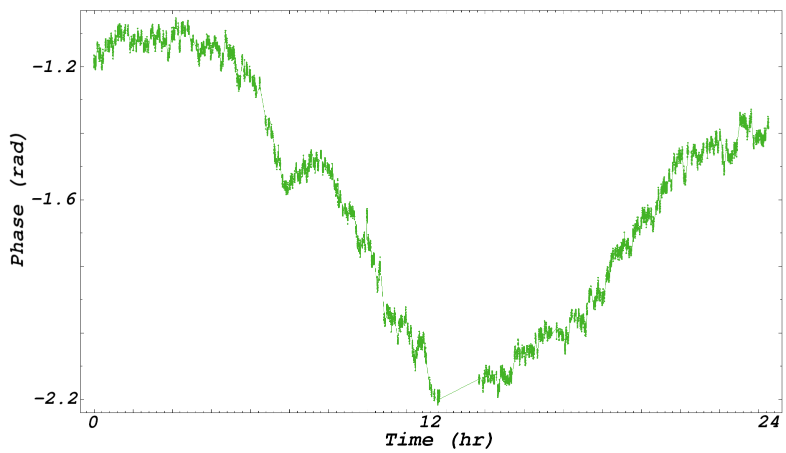

Figure 14: The differenced phase calibration tone that results from subtracting one tone from the other.

The differenced phase calibration tone ϕ_{ijk} is also averaged according to the method described in Equations 14-17, and phase ambiguities must be resolved:

$$\phi_{i'jk} = \frac{1}{n_{\text{avg}}} \sum_{l=i}^{i+n_{\text{avg}}} \text{AMB}\left[\phi_{mjk},\ m \in [i,\ i+n_{\text{avg}}]\right]_l \tag{19}$$

3.3 Root-Mean-Square (RMS) Characterization

The first method of health characterization and spurious signal detection for phase calibration tones is a relatively simple method utilizing a boxcar filter to find an estimate of the RMS phase jitter of each phase calibration tone. This method is intended to identify persistent spurious signals — those that are continuously affecting phase calibration tones throughout an experiment. Intermittent spurious signals may not be identified through this procedure. The boxcar filter is implemented as shown in Equation 20, in which n is by default 10 epochs.

$$\bar{\theta}_{ijk} = \frac{1}{2n+1} \sum_{l=i-n}^{i+n} \theta_{ljk} \tag{20}$$

An example phase calibration tone with its companion boxcar-averaged data is shown in Figure 15. The boxcar-averaged tone fits the coarse time evolution of the phase calibration tone, but omits much of the phase jitter superimposed.

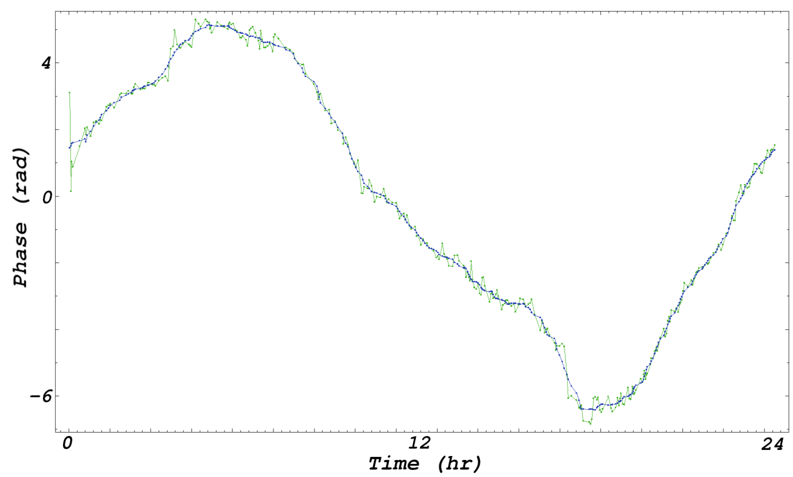

Figure 15: A phase calibration tone (green) and the same tone boxcar-filtered to be used in RMS phase jitter characterization in the time direction.

The boxcar-filtered tone is then used to compute the phase RMS in the time direction:

$$\theta_{\text{RMS-time},jk} = \sqrt{\frac{1}{N_{\text{epochs}}} \sum_{i=1}^{N_{\text{epochs}}} \left(\theta_{ijk} - \bar{\theta}_{ijk}\right)^2} \tag{21}$$

In addition, the sample variance of the time-direction phase RMS is computed across all phase calibration tones for each station as,

$$\sigma^2_{\text{RMS-time}} = \frac{1}{N_{\text{epochs}} N_{\text{tones}} N_{\text{freqs}} - 1} \cdot \sum_{i=1}^{N_{\text{epochs}}} \sum_{j=1}^{N_{\text{tones}}} \sum_{k=1}^{N_{\text{freqs}}} \left(\theta_{\text{RMS-time},jk} - \frac{1}{N_{\text{tones}} N_{\text{freqs}}} \sum_{j=1}^{N_{\text{tones}}} \sum_{k=1}^{N_{\text{freqs}}} \theta_{\text{RMS-time},jk} \right)^2 \tag{22}$$

This time-direction sample variance is then used to detect phase calibration tones distorted by spurious signals. Phase calibration tones are masked out if they display a phase RMS greater than two and a half standard deviations above the mean:

$$\theta_{\text{RMS-time},jk} > \frac{1}{N_{\text{tones}} N_{\text{freqs}}} \sum_{j=1}^{N_{\text{tones}}} \sum_{k=1}^{N_{\text{freqs}}} \theta_{\text{RMS-time},jk} + 2.5\sigma_{\text{RMS-time}} \tag{23}$$

This filter catches spurious signals typically of lower magnitude that are present throughout most or all of the experiment, a type of spurious signal commonly seen in VLBA data.

The phase RMS is also evaluated in the frequency-direction for each IF (Equation 24). This is primarily an indication of the quality of the station's phase calibration tones and is not used as a filter for spurious signals. The time-direction RMS is reported to the user by tone and the frequency-direction RMS by IF for analyst evaluation of phase calibration quality.

$$\theta_{\text{RMS-freq},ik} = \sqrt{\frac{1}{N_{\text{tones}}} \sum_{j=1}^{N_{\text{tones}}} \left(\theta_{ijk} - \frac{1}{N_{\text{tones}}} \sum_{j=1}^{N_{\text{tones}}} \theta_{ijk} \right)^2} \tag{24}$$

Typical frequency-direction phase calibration tone jitter across an IF is shown in Figure 16.

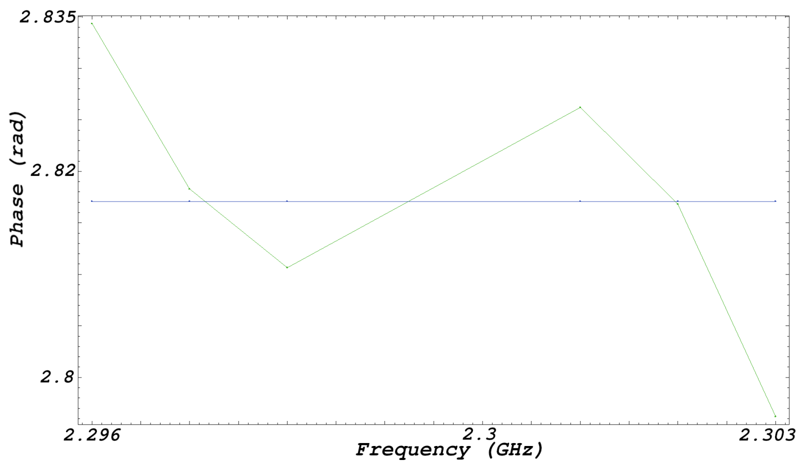

Figure 16: A phase calibration tone (green) and the average phase of the same tone to be used in RMS phase jitter characterization in the frequency direction.

3.4 Spurious Signal Detection

The main method of spurious signal detection employed is based on a series of B-splines of third degree. These B-splines, first discovered by Isaac Schoenberg [5], were implemented as described in Petrov (2007). [4] The knots of the B-spline occur at frequencies within the IF currently being analyzed. Outside the defined interval, the B-spline is simply 0 (Equation 25).

$$B_j^0(\omega) = \begin{cases} 1, & \text{if } \omega \in [\omega_j, \omega_{j+1}] \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

B-splines are defined recursively, where a spline of degree m is defined in terms of two B-splines of degree $m-1$:

$$B_j^m(\omega) = \frac{\omega - \omega_j}{\omega_{j+m} - \omega_j} B_j^{m-1}(\omega) + \frac{\omega - \omega_{j+m+1}}{\omega_{j+1} - \omega_{j+m+1}} B_{j+1}^{m-1}(\omega) \tag{26}$$

A B-spline is fit to the phase and amplitude of phase calibration tones in each IF and for each epoch in a least-squares optimization. The first and second derivatives are constrained to zero in the least-squares solution with weights selected to maintain smoothness. These constraints on the first and second derivative are intended to prevent the B-spline from following a sudden jump in phase and amplitude as would be typical of a tone distorted by a spurious signal. The derivative constraints must also not be too powerful as to cause the B-spline to be flat and therefore unresponsive to the phase calibration tone data. This strategy exploits the hypothesis that the phase calibration signal should be smooth in frequency and that a spurious signal distorts both the amplitude and phase of a phase calibration tone. I define the splined phase at time t_i and frequency ω_{jk} as θ_{ijk}^{sp} and the splined amplitude as A_{ijk}^{sp}.

An example of phase calibration tones in a single IF with a jump due to the presence of a spurious signal in one of the examined tones is shown in Figures 17 and 18, where the B-spline in green does not follow the sudden peak of the spurious signal because of its least squares fit to the broader curve and its limitation in first and second derivatives. The larger difference between the amplitude and phase of the spurious signal in the phase calibration data and its reference B-spline compared to its healthy neighbors is exploited to detect the spurious signal.

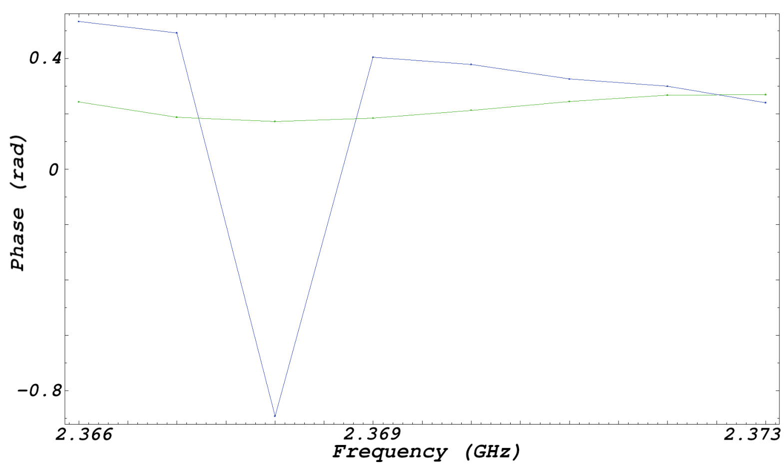

Figure 17: A phase calibration tone (blue) and the splined phase curve used in evaluating the presence of a spurious signal (green).

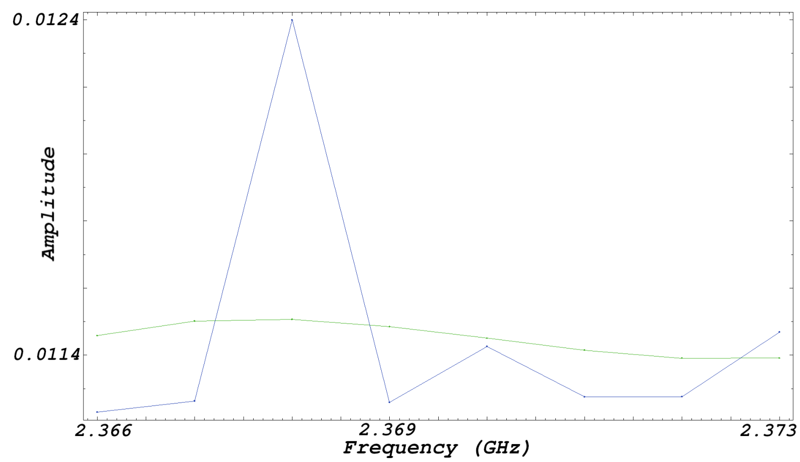

Figure 18: A phase calibration tone (blue) and the splined amplitude curve used in evaluating the presence of a spurious signal (green).

First, the phase calibration tone phase and amplitude are represented in the complex plane, and the same is done for the B-spline phase and amplitude (Equations 27 and 28).

$$C_{ijk} = A_{ijk} \left(\cos(\theta_{ijk}) + i \sin(\theta_{ijk})\right) \tag{27}$$

$$C^{\mathrm{sp}}_{ijk} = A^{\mathrm{sp}}_{ijk}(\cos\left(\theta^{\mathrm{sp}}_{ijk}\right) + i \sin\left(\theta^{\mathrm{sp}}_{ijk}\right)) \tag{28}$$

As shown in Figure 19, because the spurious signal causes a deviation in both phase and amplitude in the phase calibration tones, the jump in the complex plane is even more pronounced.

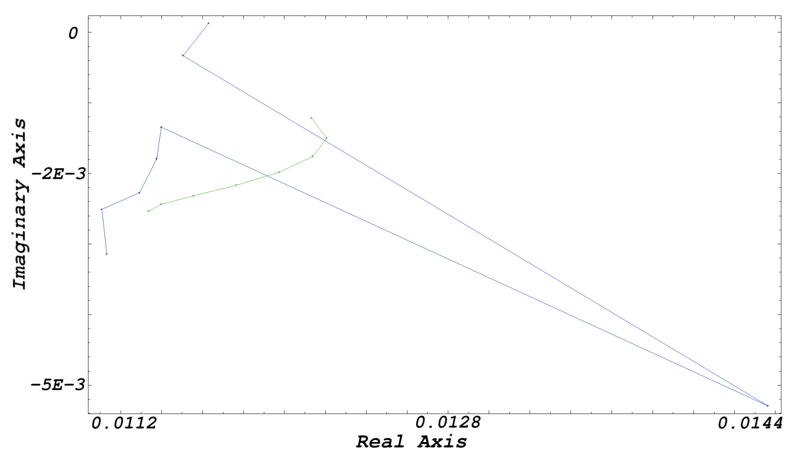

Figure 19: The phase calibration tone and B-spline evaluated in the complex plane, showing a large deviation due to the presence of a spurious signal.

To detect spurious signals on a tone-by-tone basis, the difference between the phase calibration tone complex representation and its corresponding splined point is evaluated using a two-norm:

$$d_{ijk} = \sqrt{\left(\mathrm{Re}\left(C_{ijk}\right) - \mathrm{Re}\left(C_{ijk}^{\mathrm{sp}}\right)\right)^2 + \left(\mathrm{Im}\left(C_{ijk}\right) - \mathrm{Im}\left(C_{ijk}^{\mathrm{sp}}\right)\right)^2} \tag{29}$$

Figure 20 shows d_{ijk} as evaluated from the curve in Figure 19. Note that the spike corresponding to a distorted tone shows clearly, with a magnitude more than twice that of any other tone.

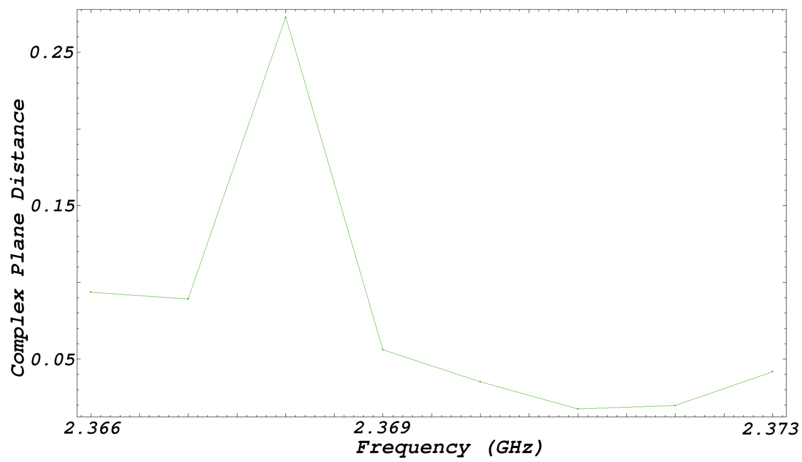

Figure 20: The distance between each phase calibration tone point and its corresponding splined point on the complex plane by frequency.

3.4.1 Multiple Spur Detection

A more challenging detection paradigm is when multiple tones within a single IF are affected by spurious signals. This is a relatively common occurrence. Figure 21 shows phase calibration tones unwrapped across an IF for a single epoch with two tones affected by spurious signals. The complex distance is shown to be degraded in this case, with multiple points significantly deviating from the B-spline.

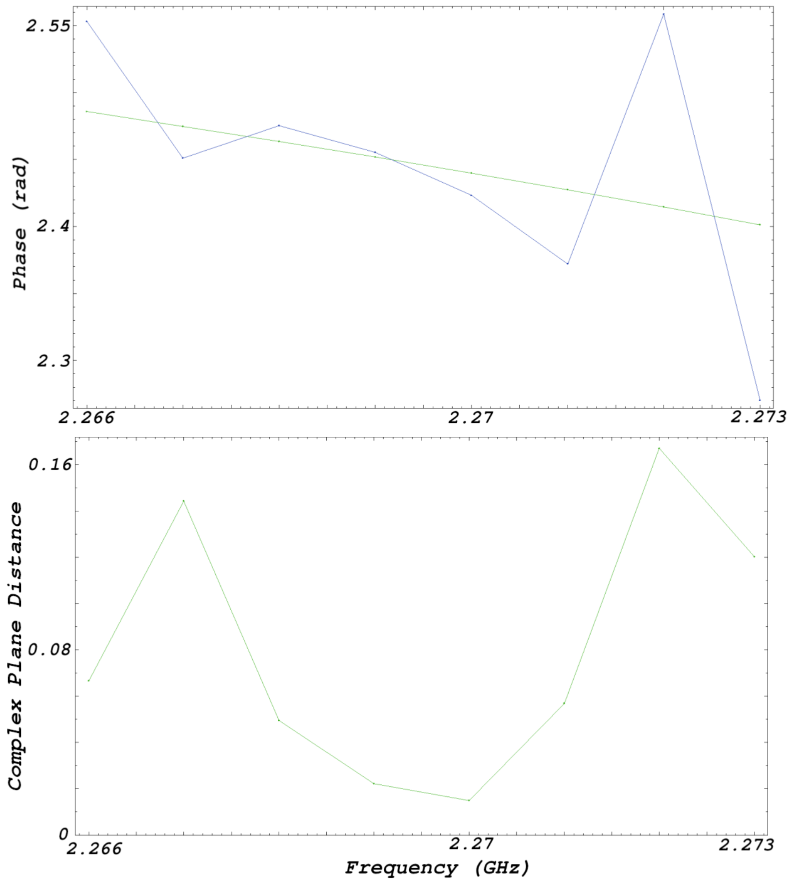

Figure 21: A phase calibration tone and B-spline with corresponding distance in the complex plane showing multiple peaks.

The approach to this problem is to remove one spurious signal at a time. The first and most prevalent distorted tone is flagged as spurious at the epoch, and a new B-spline is calculated. Equation 29 is then reevaluated with the identified spurious tone omitted. Figure 22 shows that the second spurious signal becomes clearly visible after the omission of the first flagged tone.

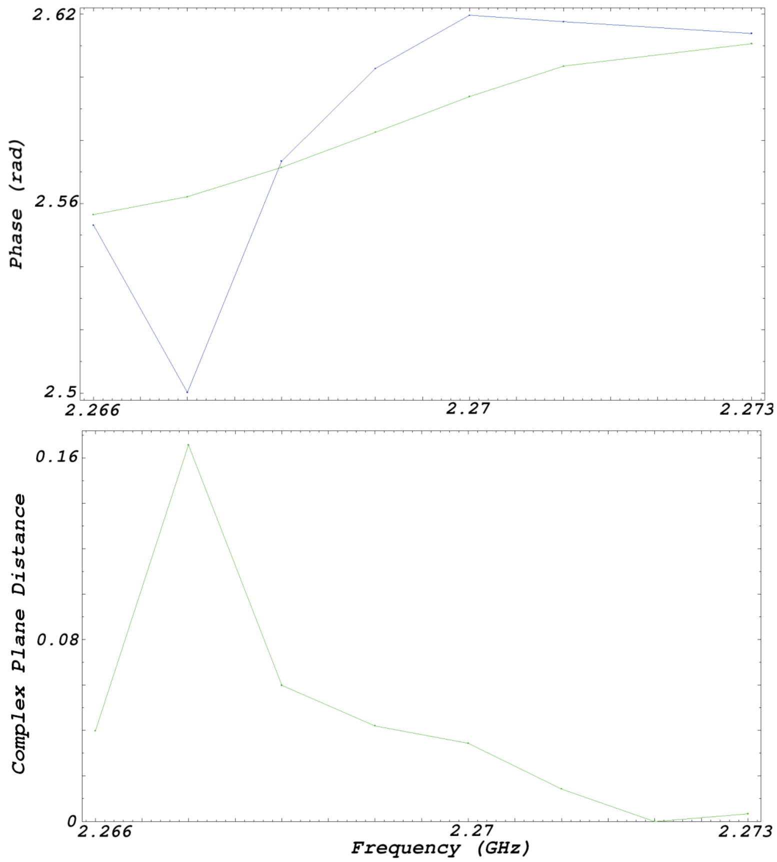

Figure 22: The same phase calibration tone with a new B-spline after removal of a spurious signal.

Finally, the second spurious signal is masked out, and a third spline is calculated to ensure that no further spurious signals remain in the IF. Figure 23 shows the close agreement between spline and data indicative of healthy phase calibration tones.

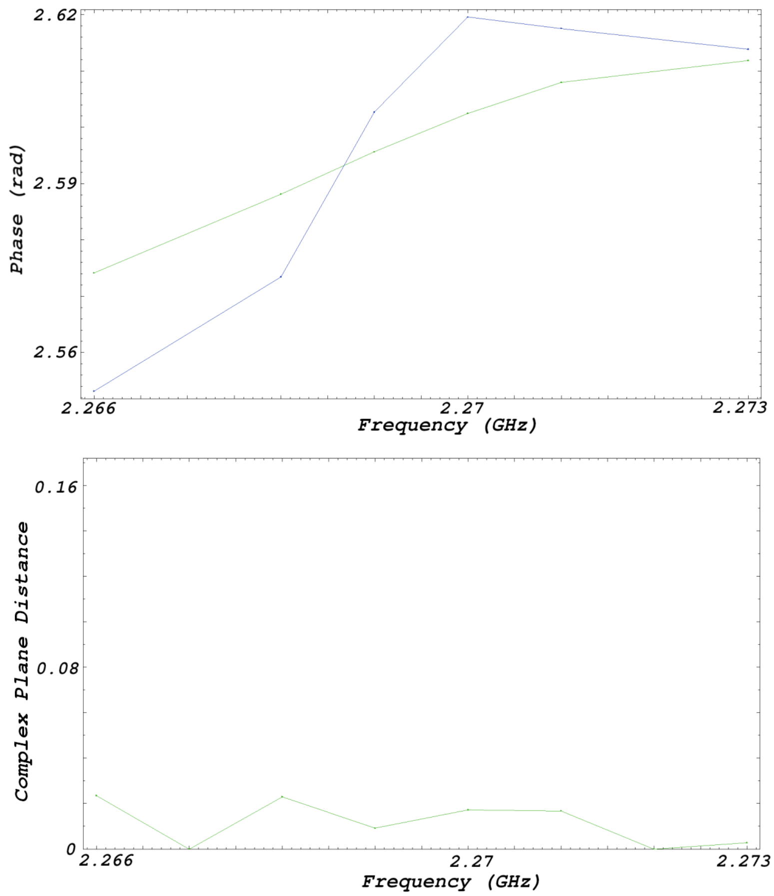

Figure 23: A final plot of phase calibration tone and corresponding spline showing all spurious signals have been removed with no peaks appearing in the complex plane distance.

By default, a tone at a specific epoch is judged healthy if the distance between it and its splined counterpart, d_{ijk}, is less than d_0 on the complex plane. A tone is masked out and added to a mask file automatically if it is flagged as affected by a spurious signal in more than $P\%$ of all epochs by default. The limits d_0 and P are user-defined and can be changed to suit specifics phase calibration generators or experiment types.

3.5 Detection Theory of Clock Breaks

To detect clock breaks and other phase jumps, a totally different detection strategy is used. This method of analysis depends on a statistical accounting of phase variability that arises due to an important property of the phase calibration data: the dominant source of phase noise affecting a differenced phase calibration tone ϕ_{ijk} is random. From epoch to epoch, the first order difference of the differenced phase tone, $\phi_{(i+1)jk} - \phi_{ijk}$, should be a random sample of a Gaussian distribution, and the differenced phase, ϕ_{ijk}, a Gaussian random walk. Figure 24 shows the distribution of first order differences in differenced phase for a single station compared to a standard Gaussian distribution of identical variance.

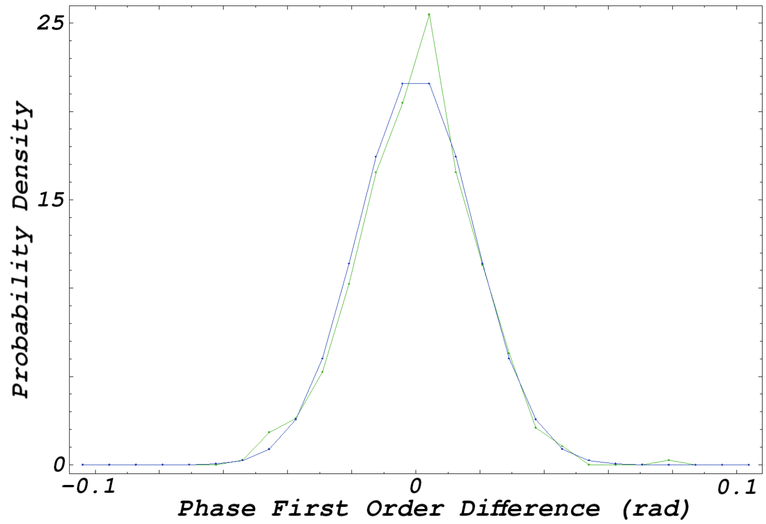

Figure 24: A histogram of first order differences in differenced phase (green) with a sampled Gaussian distribution with the same sample variance (blue).

Because a clock break occurs across all phase calibration tones in all IFs for a station simultaneously, it is statistically more powerful to use a summation of all of the phase differences across these tones. The assumption that the individual first order differences of the phase calibration tones are Gaussian-distributed means that the summation of N squared first order differences in a given epoch is a chi-squared distribution of order N:

$$x_{i^*} = \frac{1}{\sigma_\theta^2} \sum_{k=2}^{N_{\text{freqs}}} \sum_{j=1}^{N_{\text{tones}}} (\phi_{(i+1)jk} - \phi_{ijk})^2 \sim \chi^2(N_{\text{freqs}} + N_{\text{tones}} - 1) \tag{30}$$

In Equation 30, the summation across IFs (the k index) runs from 2 to N_{freqs} because the first IF is typically the source of the reference tones used in the computation of the differenced phase, thus ϕ_{1jk} does not exist. The random variable x_{i^*} is normalized by its sample variance, σ_θ^2:

$$\begin{aligned} \sigma_\theta^2 &= \max_{j,k} \frac{1}{N_{\text{epochs}} - 2} \sum_{i=1}^{N_{\text{epochs}}-1} ((\theta_{(i+1)jk} - \theta_{(i+1)j}^{\text{ref}}) - (\theta_{ijk} - \theta_{ij}^{\text{ref}}))^2 \\ &= \max_{j,k} \frac{1}{N_{\text{epochs}} - 2} \sum_{i=1}^{N_{\text{epochs}}-1} (\phi_{(i+1)jk} - \phi_{ijk})^2 \end{aligned} \tag{31}$$

To test for a clock break, I define a chi-squared variance test (as in, for example, Snedecor and Cochran, 1989, section 5.9 [6]) , in which the null hypothesis is that the sum of squared differences in each epoch is a chi-squared distribution with a known variance, and the alternative hypothesis is that the epoch includes a jump of unknown distribution and unknown variance not equal to that of the null hypothesis. To evaluate the test, we simply find the statistical likelihood of a jump being due to a sample of the chi-squared distribution and evaluate a cutoff probability for which the alternative hypothesis is chosen.

The false alarm threshold P_A is subtracted from unity and set equal to the cumulative distribution function (CDF) of the chi-squared distribution as in Equation 32.

$$F(x) = 1 - P_A = \frac{1}{2^{k/2}\Gamma(k/2)}\gamma(\frac{k}{2}, \frac{x}{2}) \tag{32}$$

Identifying a statistically significant jump is not enough on its own to declare a clock break, however, as short-term spurious signals also cause jumps of high statistical significance. The difference between the two is that after a spurious signal subsides, the phase returns to its original time evolution, whereas after a clock break the phase of the phase calibration tones is permanently altered. For this reason, a second random variable is defined using first order differences of differenced phase in non-consecutive epochs.

$$y_{i^*} = \frac{1}{\bar{\sigma}_\theta^2} \sum_{k=2}^{N_{\text{freqs}}} \sum_{j=1}^{N_{\text{tones}}} \left(\phi_{(i+n_{\text{shift}})jk} - \phi_{(i-n_{\text{shift}})jk}\right)^2 \tag{33}$$

In Equation 33, the two epochs of differenced phase are separated by $2n_{\text{shift}} + 1$ epochs, where n_{shift} is an integer number of epochs that represents as close as possible to 300 seconds, or 5 minutes of real time:

$$n_{\text{shift}} = \text{INT}\left(\frac{300 \text{ sec}}{\Delta t_{\text{avg}}}\right) \tag{34}$$

The random variable y_{i^*} represents a second chi-squared distribution of the same order, but the normalizing sample variance $\bar{\sigma}_\theta^2$ is now different. To find this sample variance, we examine the first order phase differences used in its creation. As discussed, the time series of differenced phase calibration tones is a Gaussian random walk, thus the state $\phi_{(i+n_{\text{shift}})jk}$ is dependent on the earlier state $\phi_{(i-n_{\text{shift}})jk}$ with the addition of $2n_{\text{shift}}$ Gaussian-distributed random variables:

$$\phi_{(i+n_{\text{shift}})jk} - \phi_{(i-n_{\text{shift}})jk} = \left(\phi_{(i-n_{\text{shift}})jk} + \sum_{l=1}^{2n_{\text{shift}}} X_l\right) - \phi_{(i-n_{\text{shift}})jk} = \sum_{l=1}^{2n_{\text{shift}}} X_l \tag{35}$$

A sum of $2n_{\text{shift}}$ Gaussian random variables X_l of variance σ_θ^2 is equivalent to a single Gaussian random variable Y with a variance of $2n_{\text{shift}}\sigma_\theta^2$:

$$\sum_{l=1}^{2n_{\text{shift}}} X_l = Y, \; X_l \sim N(0, \sigma_\theta^2), \; Y \sim N(0, 2n_{\text{shift}}\sigma_\theta^2) \tag{36}$$

This is equivalent to the new normalizing sample variance, $\bar{\sigma}_\theta^2$:

$$\bar{\sigma}_\theta^2 := 2n_{\text{shift}}\sigma_\theta^2 \tag{37}$$

When the threshold for x_{i^*} and y_{i^*} are both violated, a clock break is identified. If the threshold value for x_{i^*} is violated and not the threshold for y_{i^*}, the jump is determined to be a short-term spurious signal. Both conditions are reported to the PIMA user.

For implementation purposes, solving Equation 32 for each epoch and station carries a heavy computational burden due to its dependence on inverting a quotient of gamma functions. For this reason, a polynomial approximation was used to quickly calculate the x_{i^*} and y_{i^*} cutoffs for a given false alarm threshold, in this case 1 in 1 million for x_{i^*} and 1 in 100 thousand for y_{i^*}. The polynomial approximations for each are plotted on top of the numerical evaluations of Equation 32 in Figure 25.

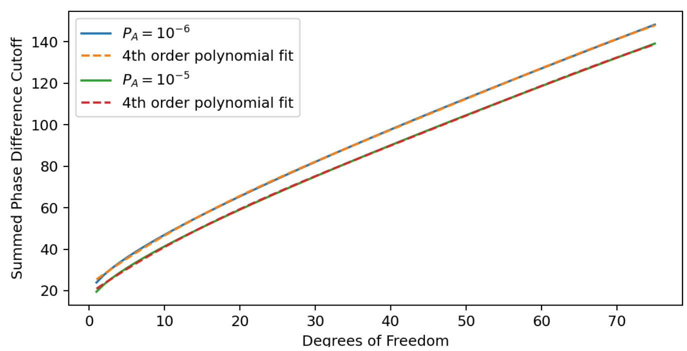

Figure 25: A 4th-order polynomial fit to the chi-squared cumulative distribution function for increasing degrees of freedom.

Equations 38 and 39 are the polynomial approximations in Figure 25 as used in PIMA.

$$x_c \approx -2.587075 \cdot 10^{-6} k^4 + 0.000495558 k^3 - 0.036278 k^2 + 2.692 k + 22.78 \quad (38)$$

$$y_c \approx -2.413742 \cdot 10^{-6} k^4 + 0.000461473 k^3 - 0.033647 k^2 + 2.550 k + 18.42 \quad (39)$$

Finally, the detection statistic x_{i^*} is used on a tone-by-tone basis (i.e. without summation across frequencies and IFs) as another measure of the quality of the phase calibration tones. By default, if a tone has 50 phase jumps violating the threshold for 1 degree of freedom in Equation 38, the tone is judged to be affected by an intermittent spurious signal and is turned off.

3.5.1 Testing of Detection Model

To test the detection models, I developed a simulation in Python using 50 independent Gaussian random walks consisting of cumulative summations of samples drawn from a normal distribution with standard deviation 0.05 radians. In the one-day, 2000-point datasets, one simulated phase jump and one simulated short-term spurious signal were added. Both the phase jump and the short-term spurious signal consisted of a single epoch of phase change drawn from a normal distribution with 0.4 radians standard deviation rather than 0.05 radians, but at the short-term spurious signal, the tones return to their previous phase values at the following epoch, while at the clock break, the abnormal variance jump is included in the cumulative summation of the random walk and thus the phase is permanently altered. Figure 26 shows 5 of the 50 Gaussian random walks, where the location of the spurious signal and clock break are given by black dashed lines.

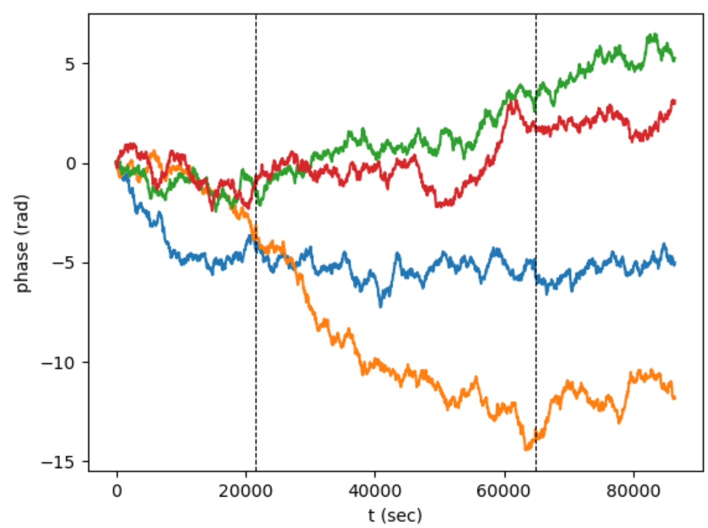

Figure 26: Multiple Gaussian random walks of standard deviation 0.05 radians with a spurious signal and clock break of standard deviation 0.4 included at 21,600 seconds and 64,800 seconds respectively.

Although the jumps are difficult to see in the random walk phase, Figure 27 displays that they are clearly visible in the variable x_{i^*} computed for the summation of the 50 tones as in Equation 30, rising far above the threshold value computed by Equation 38.

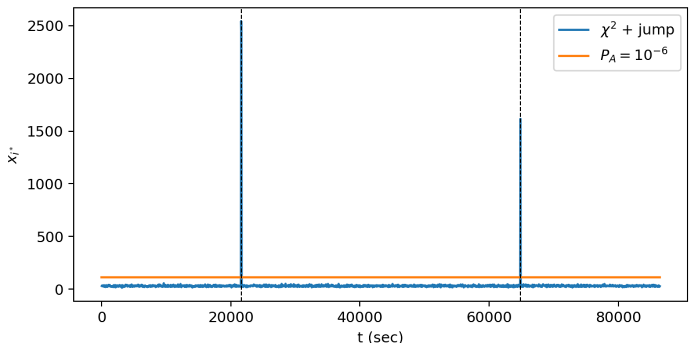

Figure 27: The value of the chi-squared random variable x_{i^*} for 50 summed-squared Gaussians of standard deviation 0.05 radians with two epochs of 0.4 radians standard deviation marked in black dashed lines.

Both the spurious signal and the clock break are flagged by the first test, so they must be differentiated by y_{i^*}. Figure 28 shows y_{i^*} in the neighborhood of the spurious signal, and Figure 29 shows the neighborhood of the clock break.

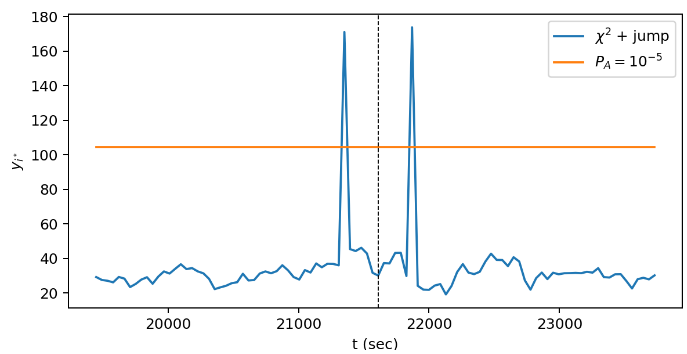

Figure 28: y_{i^*} in the neighborhood of a spurious signal.

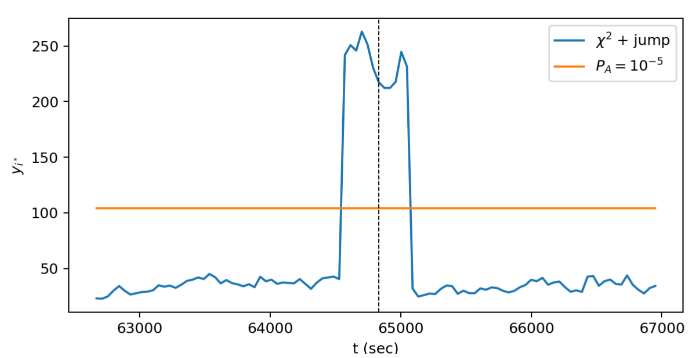

Figure 29: y_{i^*} in the neighborhood of a clock break.

y_{i^*} shows two separate peaks in the neighborhood of the spurious signal. These occur when one of the two differenced phases is exactly n_{shift} epochs from the phase jump. At the epoch of the clock break, however, the variable is at a typical level far below the threshold value. In contrast, in the neighborhood of the clock break, the epoch of interest is flagged far above the statistical threshold, indicating that y_{i^*} is successfully differentiating between the clock break and spurious signal.

4 Results and Discussion

4.1 Detection of Spurious Signals

Success in detection of spurious signals is difficult to evaluate, as there is no absolute measure of whether and to what degree a phase calibration tone is affected by spurious signals. GEPM currently identifies most tones that are also identified by analyst intuition, but it is by no means perfect. In addition to a phase calibration mask, the program generates a report file that can be easily added to by a VLBI analyst. This report file lists the reason for the tones' deactivation–it reports "phase jumps" if x_{i^*} reaches the threshold value more than 50 times, it reports "spurious signal" if it consistently deviates from the B-spline in its IF or has a time domain RMS more than 2.5σ above the average for the station, and it reports "low amplitude" if the amplitude of the tone is below the minimum threshold of 10^{-8}, indicating that there is no phase calibration data at all. The report file can then in turn be used to generate a new phase calibration mask after adding or removing tones from the list. An example report file is shown in Figure 30.

```
# PIMA PCAL_RPT   Format of 2022.07.07
#
# PCAL_RPT file for experiment r41056
#
# created by # PIMA PCAL_RPT_GEN  v  1.00 2022.07.07 on 2023.01.09-17:59:37
# using control file /vlbi/r41056/r41056_s_pima.cnt
#
PCAL  STA:  BADARY   IND_FRQ:     3-3   IND_TONE:     8-8    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  BADARY   IND_FRQ:     6-6   IND_TONE:     3-3    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  BADARY   IND_FRQ:     6-6   IND_TONE:     5-5    OFF !    Problem:  SPURIOUS SIGNAL
#
PCAL  STA:  HART15M  IND_FRQ:     4-4   IND_TONE:     6-6    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  HART15M  IND_FRQ:     5-5   IND_TONE:     7-7    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  HART15M  IND_FRQ:     6-6   IND_TONE:     7-7    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  HART15M  IND_FRQ:     6-6   IND_TONE:     8-8    OFF !    Problem:  SPURIOUS SIGNAL
#
PCAL  STA:  KOKEE    IND_FRQ:     4-4   IND_TONE:     2-2    OFF !    Problem:  PHASE JUMPS
PCAL  STA:  KOKEE    IND_FRQ:     4-4   IND_TONE:     5-5    OFF !    Problem:  PHASE JUMPS
PCAL  STA:  KOKEE    IND_FRQ:     5-5   IND_TONE:     6-6    OFF !    Problem:  PHASE JUMPS
#
PCAL  STA:  MEDICINA IND_FRQ:     1-1   IND_TONE:     1-1    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  MEDICINA IND_FRQ:     1-1   IND_TONE:     2-2    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  MEDICINA IND_FRQ:     1-1   IND_TONE:     3-3    OFF !    Problem:  SPURIOUS SIGNAL
PCAL  STA:  MEDICINA IND_FRQ:     1-1   IND_TONE:     4-4    OFF !    Problem:  SPURIOUS SIGNAL
```

Figure 30: A report file listing masked out tones and their reason for deactivation.

A second file produced by the program is mentioned first in Section 3.3. This is a characterization of the quality of the phase calibration tones by RMS of the phase jitter in the time and frequency directions. This report is generated for a station by IF in the frequency direction and tone-by-tone in the time direction. An example is given in Figure 31.

```
# PIMA PCAL_RMS   Format of 2022.06.30
#
# PCAL_RMS file for experiment r41056
#
# created by # PIMA PCAL_RMS_GEN  v  1.00 2022.06.30 on 2023.01.09-17:59:37
# using control file /vlbi/r41056/r41056_s_pima.cnt
#
AVERAGE TIME SPACING:      5.619 SEC
TIME DIRECTION RMS PHASE CALIBRATION JITTER (RAD) BY CHANNEL:
BADARY   IND_FRQ:      1 IND_TONE:      1 IND_ABS_CHN:       9 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      2 IND_ABS_CHN:      10 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      3 IND_ABS_CHN:      11 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      4 IND_ABS_CHN:      12 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      5 IND_ABS_CHN:      13 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      6 IND_ABS_CHN:      14 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      7 IND_ABS_CHN:      15 RMS:      0.009
BADARY   IND_FRQ:      1 IND_TONE:      8 IND_ABS_CHN:      16 RMS:      0.009
FREQ DIRECTION RMS PHASE CALIBRATION JITTER (RAD) BY IF:
BADARY   IND_FRQ:      1           0.017
BADARY   IND_FRQ:      2           0.008
BADARY   IND_FRQ:      3           0.005
BADARY   IND_FRQ:      4           0.012
BADARY   IND_FRQ:      5           0.013
BADARY   IND_FRQ:      6           0.034
```

Figure 31: A sample RMS file showing values in the time and frequency directions.

4.2 Detection of Clock Breaks

In real data, the phase jump detector x_{i^*} is typically not activated. The summed-squared first order differences of the differenced phase as evaluated by Equation 30 remain far below the statistical threshold of Equation 38 as in Figure 32.

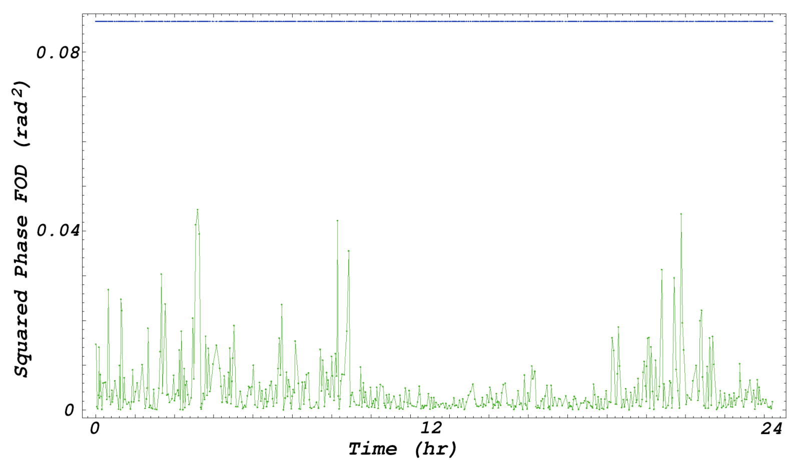

Figure 32: Summed-squared first order differences for a station with no clock breaks ($P_A = 10^{-6}$ limit in blue).

To show the success of the phase jump detection scheme, Figure 33 plots x_{i^*} for a dataset with one clock break and one spurious signal as in the simulations in Section 3.5.1. Both of these phase jumps are flagged as above the alternative hypothesis threshold in blue.

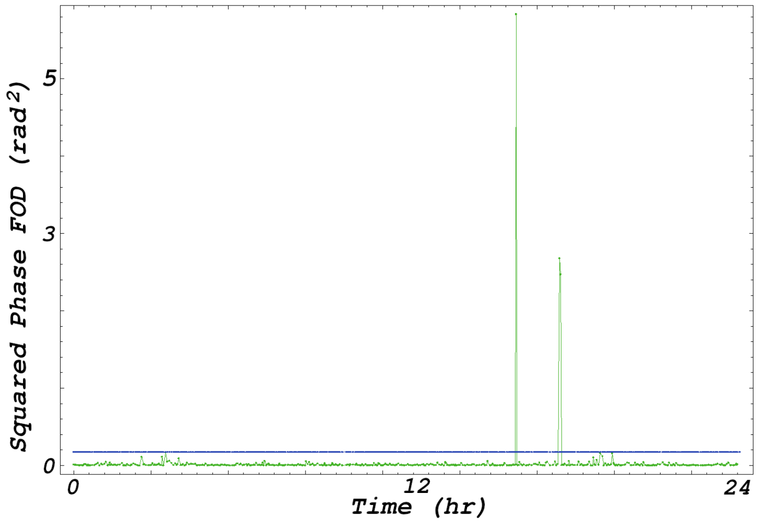

Figure 33: Summed-squared first order differences for a station with one clock break and one short-term spurious signal.

Figure 34 shows y_{i^*} evaluated in the neighborhood of the two phase jumps. The patterns in Figures 28 and 29 are replicated with notable accuracy, indicating a clock break in the neighborhood of 16 hours and a short-term spurious signal at about 17.5 hours in the dataset.

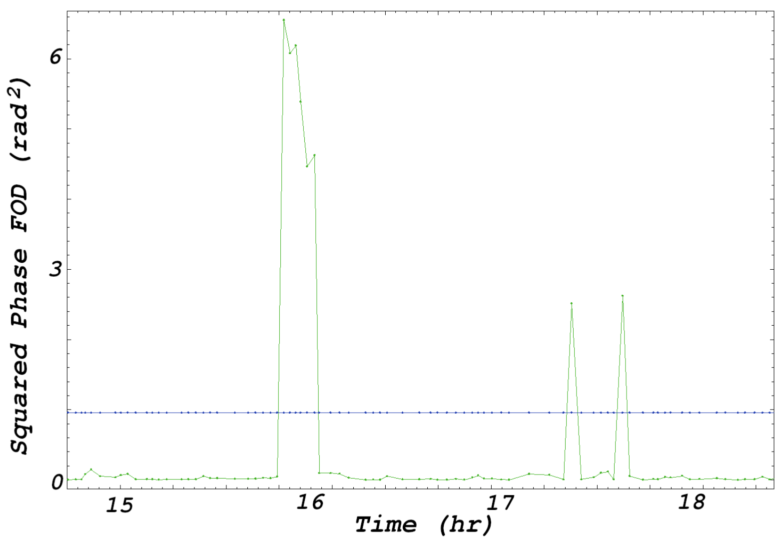

Figure 34: The same dataset zoomed to the epochs of interest and evaluated with the second random variable y_{i^*} .

Clock breaks and spurious signals are reported on the command line to the user with the epoch at which they occurred, but no further action is taken unless the tone is determined to be affected by a spurious signal in one of the methods discussed above.

5 Conclusion and Recommendations

A scheme for automatically detecting and masking phase calibration tones affected by spurious signals has been developed and implemented in the software package PIMA as the task GEPM. The task first cleans and prepares data for analysis by time averaging and ambiguity resolution in phase. Group delays are then removed in an FFT-based method along with any remaining bandpass structure before phase calibration health is determined. The health analysis process begins with a root-mean-square analysis of the phase calibration tones in the time and frequency directions, which is reported as a phase calibration health metric to the user.

The detection and masking of persistent spurious signals relies on complex plane analysis of a B-spline least-squares fit to the phase calibration tone phase and amplitude in each IF. The B-splines' first and second derivatives are constrained to zero with weights allowing them to remain both smooth and insensitive to the spikes in phase and amplitude caused by tones affected by spurious signals. A set of adjustable parameters defines the strictness of the automatic masking procedure, providing options to the user to tune the automatic masking for different phase calibration generators and experiments. These parameters include the sampling interval of the phase calibration data, the distance in the complex plane from the B-spline to the corresponding phase calibration tone at which a spurious signal is identified, and the percentage of flagged epochs at which a phase calibration tone is deactivated.

Finally, a detection method for phase jumps caused by short-term spurious signals and clock breaks is outlined along with a strategy to differentiate between them using a modified chi-squared variance test. A simulation of the effectiveness of these methods is reported, and finally an example is provided with real phase calibration data.

The most obvious next steps for GEPM and processing of phase calibration data in my mind are related to the processing of phase jumps. In the case of short-term spurious signals, a time-dependent phase calibration mask may be an effective way of retaining as much useful information provided by the phase calibration tone as possible while masking the tone in the specific epochs in which an intermittent noise source destroys its ability to effectively calibrate the phase of the complex visibilities of the bandpass data. This time-dependent mask could even be written automatically using the detection mechanisms already employed.

For phase calibration jumps caused by clock breaks, it is more difficult to implement a single solution, but multiple strategies appear viable. The first is that in a recognized clock break, a scan boundary may be created automatically, or the phase calibration tones masked out in the remainder of the given coherent accumulation interval. This would avoid having to entirely mask out the phase calibration tones or manually create a scan boundary by inspecting data. Another potentially more fraught strategy is to restore the phase calibration tones after the clock break, resetting the phase of each tone to the value it took in the time series just before the clock break occurred. If this scheme works, then clock breaks can be almost entirely corrected for, leading to no loss in phase calibration information except for the single epoch of the clock break.

References

1. A. E. E. Rogers. *A Receiver Phase and Group Delay Calibration System for Use in Very Long Baseline Interferometry.* Haystack Observatory Technical Note 1975-6. MIT Haystack Observatory.

2. G. Rajgapolan. *Phase Cal Basics.* 11th IVS Technical Operations Workshop, MIT Haystack Observatory.

3. L. Petrov, Y. Y. Kovalev, E. B. Fomalont, D. Gordon, 2011. *The Very Long Baseline Array galactic plane survey—VGaPS.* The Astronomical Journal; 142, 2, 35.

4. L. Petrov, 2007. *The empirical Earth rotation model from VLBI observations.* Astronomy and Astrophysics; 467, 359–369.

5. I. Schoenberg, 1946. *Contributions to the problem of approximation of equidistant data by analytic functions.* Quarterly of Applied Mathematics; 4, 1, 45-99.

6. Snedecor, G. W., Cochran, W. G. (1992). Statistical methods. Iowa State Univ. Press.