
**Event Horizon Telescope**
**Memo Series**

EHT Memo 2019-CE-02

*Calibration & Error Analysis WG*

# EHT data set validation and characterization of errors

M. Wielgus[1,2,*], L. Blackburn[1], S. Issaoun[3], M. Janssen[3], M. D. Johnson[1], J.-Y. Koay[4]

Apr 3, 2019 – Version 1.0

[1]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA*
[2]*Black Hole Initiative at Harvard University, 20 Garden St., Cambridge, MA 02138, USA*
[3]*Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands*
[4]*Institute of Astronomy & Astrophysics, Academia Sinica, AS/NTU No. 1, Sec. 4, Roosevelt Rd, Taipei 10617, Taiwan*

**Abstract**

This memo compiles different subjects related to the validation of the EHT 2017 data set, explaining in some more detail the performed tests, their results and interpretation. At present this memo describes consistency tests of the thermal noise level in the VLBI data and tests of the closure phase bias towards zero, originating from averaging of data fringe-fitted with a phase calibration source model assumption.

*maciek.wielgus@gmail.com

# 1 Thermal noise consistency

## 1.1 Introduction

Thermal errors play a fundamental role in the VLBI data interpretation, typically dominating the uncertainty budget. Visibility thermal errors follow circularly symmetric complex normal distribution with zero mean and standard deviation [1]. For a correlation coefficient (visibility prior to the flux calibration), the error is

$$\sigma_{\rm th} = \frac{1}{\eta_{\rm Q}\sqrt{2\,\Delta t\,\Delta\nu}}, \tag{1}$$

where $\eta_{\rm Q}$ is a quantization efficiency factor ($\sim$0.88 for a 2-bit quantization), $\Delta t$ is a time averaging window, $\Delta\nu$ is bandwidth. Tracking the value of $\sigma_{\rm th}$ throughout the VLBI data reduction pipeline, with multiple scalings, flagging and averaging proves difficult in practice, and it is common to settle for the weights $w$, where $\sigma_{\rm th} \propto 1/\sqrt{w}$, representing the thermal noise up to a multiplicative constant [2]. However, if any meaningful statistical inference is to be made using the data, the uncertainties need to be represented properly. Hence, there is a necessity for developing tools that would allow to test the quality of the noise component reported by the VLBI data processing pipeline.

## 1.2 Matching amplitude moments

Visibility amplitudes $A = |V|$ follow a Rice distribution,

$$f(A|A_0,\sigma_{\rm th}) = \frac{A}{\sigma_{\rm th}^2}\exp\left(\frac{-A^2-A_0^2}{2\sigma_{\rm th}^2}\right) I_0\left(\frac{AA_0}{\sigma_{\rm th}^2}\right), \tag{2}$$

for a true visibility amplitude $A_0$ and $I_0(x)$ denoting a modified Bessel function of the first kind with order zero. The ensemble of visibility measurements obtained during a single observing scan on a particular baseline can be used to estimate the Rice distribution parameters $(A_0, \sigma_{\rm th})$. Here we propose an algorithm to do that using a moment matching method, that is, using estimators derived from analytic formulae for second and fourth raw moment of the Rice distribution. We have

$$\mu_2 = 2\sigma_{\rm th}^2 + A_0^2 \approx m, \tag{3}$$

$$\mu_4 = 8\sigma_{\rm th}^4 + 8\sigma_{\rm th}^2 A_0^2 + A_0^4 \approx m^2 + d^2, \tag{4}$$

where $m$ and $d$ can be estimated from the ensemble of amplitudes $A_i$,

$$m = \frac{\sum_{i=1}^N A_i^2}{N}, \tag{5}$$

$$d^2 = \frac{\sum_{i=1}^N \left(A_i^2 - m\right)^2}{N}. \tag{6}$$

Solving Equations 3-4 for $(A_0, \sigma_{\rm th})$ the distribution parameters can be represented as functions of $m$ and $d$, resulting in the following empirical estimators

$$\hat{A}_0 = \left(m^2 - d^2\right)^{1/4}, \tag{7}$$

$$\hat{\sigma}_{\rm th} = \left[\frac{m - \left(m^2-d^2\right)^{1/2}}{2}\right]^{1/2}. \tag{8}$$

The moment matching estimator is implemented in the EHT Analysis Toolkit (`eat`) library framework [3] as the *unbiased_std* function.

## 1.3 Scaling closure phases variation

A different approach to the problem of thermal noise estimation is based on the analysis of the variation in closure phases. A closure phase $\psi_{\rm C}$ is an argument of a triple product of visibilities $V$ on baselines forming a closed triangle $ijk$, and by construction is free of the station-based complex gains uncertainty [1],

$$\psi_{{\rm C},ijk} = {\rm Arg}\left(V_{ij}V_{jk}V_{ki}\right). \tag{9}$$

The corresponding closure phase uncertainty is obtained by the propagation of reported uncertainties of the visibilities:

$$\sigma_{\psi_{\rm C},ijk} \approx \sqrt{S_{ij}^{-2} + S_{jk}^{-2} + S_{ki}^{-2}}, \tag{10}$$

where $S_{ij}$ is the estimated signal-to-noise ratio (S/N), associated with the visibility $V_{ij}$, that is

$$S_{ij} = \frac{|V_{ij}|}{\sigma_{ij}}\,, \tag{11}$$

where $\sigma_{ij}$ is an uncertainty on the $ij$-th visibility measurement. The approximation from Equation 10 is accurate in a high S/N regime, hence it is important to build up the S/N of closure products by coherent averaging and adequate filtration. Assuming that the reported visibility uncertainties $\sigma_{\rm rep}$ are correct up to a multiplicative constant $\kappa$, common for the entire data set, we notice that

$$\sigma_{\rm th} = \kappa\sigma_{\rm rep}\,, \tag{12}$$

$$\sigma_{\psi_{\rm C},{\rm true}} \approx \kappa\sigma_{\psi_{\rm C},{\rm rep}}\,. \tag{13}$$

Therefore, the constant scaling factor between the reported uncertainty and the true one can be determined by measuring the uncertainty in closure phases $\sigma_{\psi_{\rm C},{\rm true}}$ and comparing it with the reported one $\sigma_{\psi_{\rm C},{\rm rep}}$. In practice we measure variation in studentized differences between subsequent closure phase measurements, that is, for each triangle we inspect a variable $X$,

$$X_{\rm n} = \frac{{\rm Arg}\exp\left[{\rm i}(\psi_{{\rm C},{\rm n}+1} - \psi_{{\rm C},{\rm n}})\right]}{\sqrt{\sigma^2_{\psi_{\rm C},{\rm rep,\ n}} + \sigma^2_{\psi_{\rm C},{\rm rep,\ n+1}}}}. \tag{14}$$

For each triangle we compute a standard deviation of corresponding $X$, and finally we associate the scaling factor $\kappa$ with a median standard deviation measurement across all considered triangles,

$$\kappa = {\rm median}_{\rm triangles}\left({\rm std(X)}\right). \tag{15}$$

This method of determining the thermal error budget was used for the Astronomical Image Processing System (AIPS) data, for which a scaling factor $\kappa$ was determined observation by observation, that is separately for each night and source. It is implemented in the `eht-imaging` library [4] as the *estimate_noise_rescale_factor* function. The closure phase difference technique used here is motivated by a similar method to rescale AIPS errors presented in Ortiz et al.[5] For that 86 GHz VLBA data set on Sgr A$^*$, there were a sufficient number of short baselines for which the model closure phase was zero. For the EHT data set, the differencing technique allows us to make use of many long high-S/N baselines to ALMA under the assumption of a compact but non-trivial source. For the production of EHT data sets from 2017 observation campaign, this method is used routinely for scaling AIPS reported uncertainties.

## 1.4 Results

Here we show results of the thermal noise estimation on the Event Horizon Telescope Science Release 1 (SR1) 2017 data set [6], consisting of M87 and 3C 279 data. The tested visibilities are phase-calibrated using three different *fringe-fitting* pipelines, the Haystack Observatory Postprocessing System (`HOPS`) [7], the Common Astronomy Software Applications package (`CASA`) [8], and `AIPS` [6]. We are using data sets after the fringe fitting stage, prior to flux calibration.

### 1.4.1 Reported uncertainties

In Figure 1 we show the histograms of errors reported in scan-averaged SR1 data for the 3 EHT pipelines. Treating separately each scan, baseline, band and polarization (LL or RR), and selecting only a subset of detections available for all three pipelines, we obtain a data set of 5399 ensembles of visibility. The exact amount of data and S/N varies between the ensembles, but the median number of available samples in a selected ensemble is N = 900. We have scaled the errors reported for the scan-long averaging, $\sigma_{\rm rep}$, by the square root of the number of samples in the ensemble N. Ideally this quantity should represent the uncertainty on the correlation product in 0.4 seconds / 1.875 GHz averaging, and be close to a constant given by Equation 1, which for the EHT 2017 setup and 0.4 s / 1.875 GHz averaging is equal to $2.934\times10^{-5}$. Because of the channel-flagging in the frequency domain, Equation 1 serves as a lower limit on the errors. The large spread in AIPS data is a result of the error scaling being estimated with the closure phase scaling method. However, Figure 1 indicates that the closure phase scaling method results in errors consistent with other pipelines.

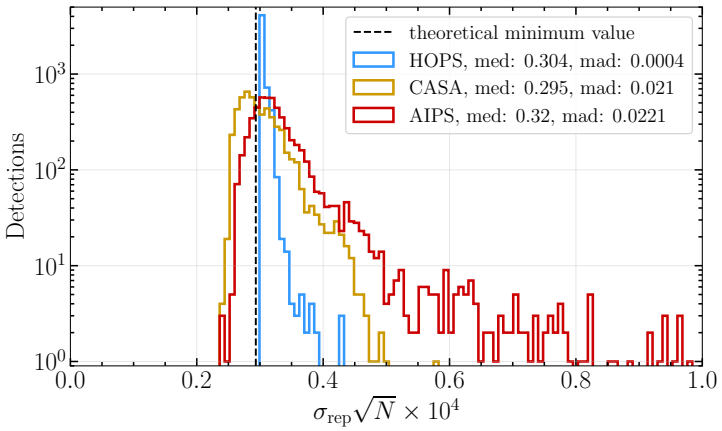


Figure 1: Reported thermal noise level in 0.4 s correlation coefficients in SR1 data set and the theoretical lower limit of $2.934\times10^{-5}$. Information about the median (med) and median absolute deviation (mad) is provided in the legend.

### 1.4.2 Consistency of reported and measured uncertainty

We compare the reported errors $\sigma_{\rm rep}$ with the ones measured by the moments matching algorithm, $\sigma_{\rm emp}$. Note that $\sigma_{\rm rep}$ denotes here a scan-averaged data uncertainty, while the moment matching estimator results in a sample standard deviation, hence for comparisons $\sigma_{\rm rep}$ is inflated by a factor $\sqrt{N}$, see Figure 2. For reference purposes, we give a theoretical limit on a moment matching estimator performance (black dashed line). This result is constructed using a Monte Carlo simulation under the following assumptions: (1) high S/N, (2) no gains variation, (3) all ensembles have $N = 900$ elements, (4) the reported noise $\sigma_{\rm rep}$ is ideal, so all uncertainties are coming from the spread in the $\sigma_{\rm emp}$ measurement.

Wide tails are seen in Figure 2 (left), indicating the presence of a data subset for which the measured variability is much larger than the reported thermal one. The subset was identified as high S/N data, for which thermal uncertainties are dwarfed by the fluctuations in complex station-based gains [6], varying on a timescale of single seconds for 1.3 mm VLBI, captured by the moment matching estimator. Indeed, when Figure 2 (left) is recreated with S/N < 50 threshold, the tail is no longer present (Figure 2, right).

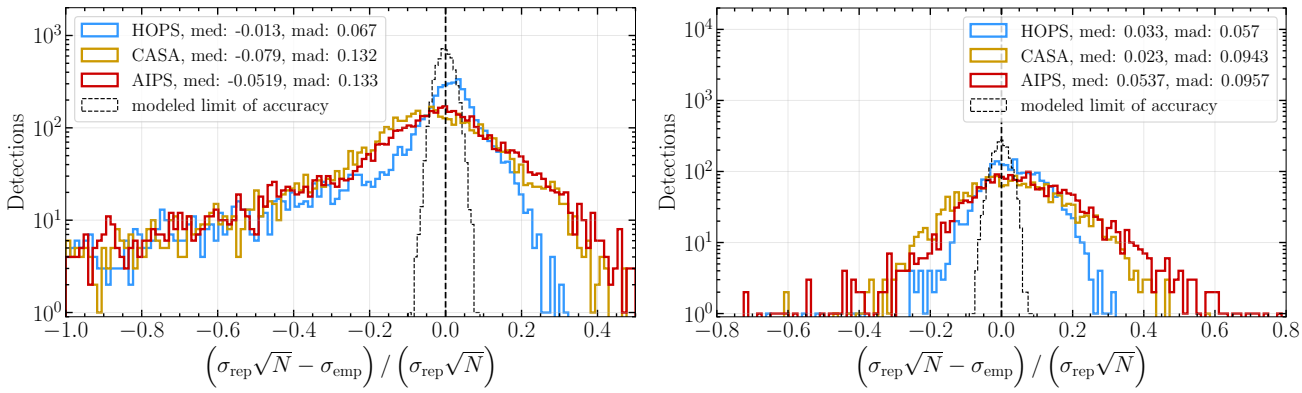


Figure 2: Left: histograms of relative difference between the reported uncertainty and uncertainty measured via the moment matching estimator for 5399 data ensembles shared in all 3 EHT pipelines. Information about the median (med) and median absolute deviation (mad) is provided in the legend. Right: same, but for 2038 ensembles with S/N < 50.

For the SR1 data set, consisting of 8 observations (unique source, unique night), the closure scaling factor $\kappa$, indicating the consistency of the reported noise with the one measured from closure phases variation, was found to be $1.03 \pm 0.02$ for the HOPS data set and $0.97 \pm 0.10$ for the CASA data set. We briefly summarize the properties of the two presented methods in Table 1.

Table 1: Comparison of properties of the two thermal noise estimation methods.

| **property** | **moment matching** | **closure phases scaling** |
|---|---|---|
| best performance | in (moderately) low S/N | in high S/N |
| robust against gain errors | no | yes |
| requires correct weights | no | yes |
| ensemble for estimation | individual ensemble | full observation |

# 2 Bias in closure phases

## 2.1 Introduction

At the fringe fitting stage of VLBI data calibration in CASA and AIPS, a source model is assumed during visibility phase calibration using the Schwab-Cotton global fringe fitting method [9]. The phase calibration for each solution interval is station-based, and does not affect the closure phase computed for a triangle formed from three connected baselines. However, the individual baseline phases themselves can exhibit a bias toward the source model used. A closure phase formed from the averaged baseline visibility phases taken over many intervals can then develop a bias toward the source structure model, which is zero for the commonly adopted point-source calibration model. The bias is present exclusively in case of visibility averaging over multiple fringe-fitting intervals prior to forming closure phases. The effect is more severe if shorter intervals are used for the fringe fitting solution interval [6], as this increases the number of free calibration parameters versus the total constraining S/N in the data. Hence, it is expected that AIPS should display a largest bias (short 2s intervals), the bias should be less severe for CASA data (varying solution interval depending on S/N, but typically several seconds or more) and HOPS data should not display a bias, since the point source model is not utilized for the HOPS baseline-based fringe search. While the SR1 data set consists of 10 s averaged visibilities, we consider scan-averaged data in this test. Since that implies reduced thermal uncertainties, the presence of this calibration/averaging related bias should be seen more clearly. The aim of these tests is to determine the importance of this bias and its dependence on closure phase magnitude.

## 2.2 Measuring closure phases bias

Here we show results of the closure phase bias estimation on the SR1 2017 data set [6], consisting of M87 and 3C 279 data. The tested visibilities are phase-calibrated using three different *fringe-fitting* pipelines, `HOPS` [7], `CASA` [8], and `AIPS` [6]. We are using scan-averaged, 2 GHz band-averaged Stokes I data.

For this test we select SR1 closure phases that are (1) at least $3\sigma$ away from 0°, (2) at least 20° away from zero (3) present in all 3 pipelines. There are in total 1645 closure phases compared in this test, drawn from maximum set of closure phases, for an even representations of baselines [6]. We then construct differences of closure phases HOPS–CASA and HOPS–AIPS and ratios HOPS/CASA and HOPS/AIPS. Results, shown in Figure 3, indicate the presence of a bias towards zero in both AIPS and CASA data. The median value of this bias is about 1.0° for both AIPS and CASA. This could be an important component of the total error budget if the scan-averaged quantities are used, particularly for bright sources like 3C 279, where we find a comparable magnitude of thermal noise and polarimetric leakage systematics [6]. The distribution of HOPS–AIPS differences indicates a heavier tail, with about 20% of AIPS closure phases shifted towards zero by more than 5° with respect to HOPS closure phases.

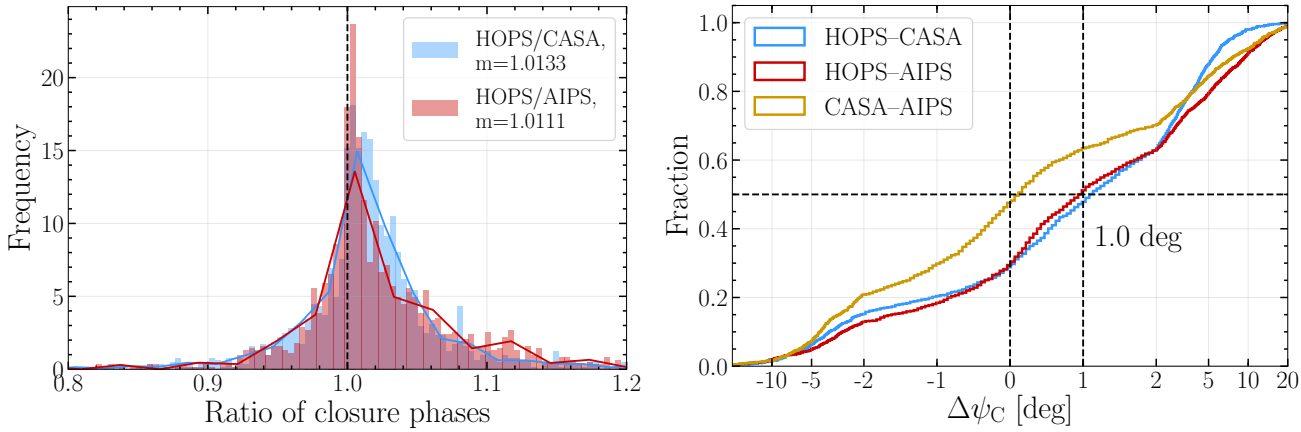


Figure 3: Left: Closure phase ratios between HOPS and CASA/AIPS pipeline results. The median values *m* of distributions are given, indicating a bias of about 1% towards zero in CASA/AIPS. Right: Closure phase differences between the fringe fitting pipelines. Similar distributions for HOPS-CASA and HOPS-AIPS with a 1.0° median bias.

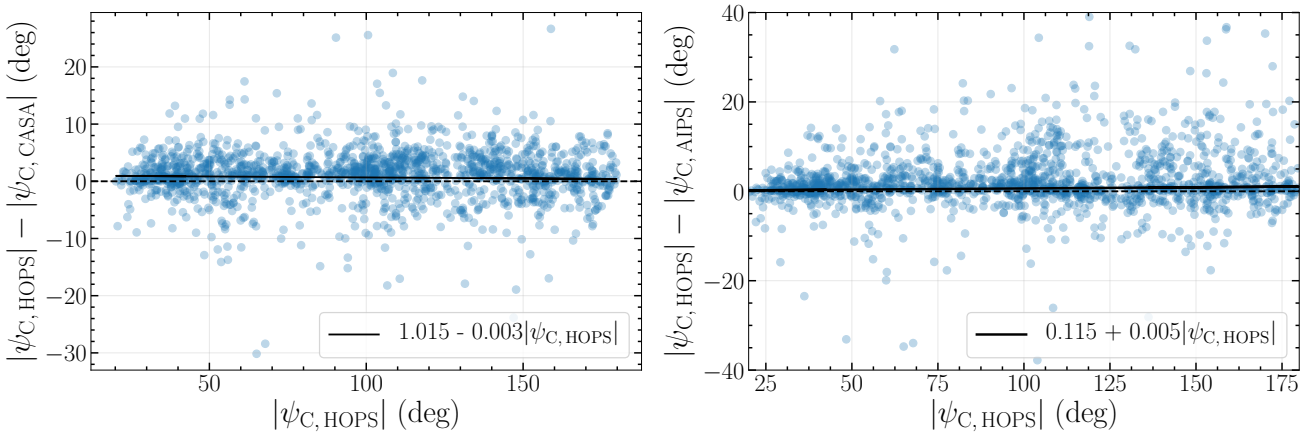


Figure 4: Left: Closure phase difference between HOPS and CASA and linear least squares fit. Right: Same for the HOPS–AIPS pair.

We further investigate the relation between the bias magnitude and closure phase distance from zero, attempting to model it with a linear regression. The dependence is not detected robustly and a constant offset model seems to be sufficient.

In conclusion, a small closure phase bias is present and clearly detected. It corresponds, on average, to about 1° shift of CASA and AIPS closure phases towards zero and should be negligible in most cases, but may be necessary to take it into account when modeling high S/N, strongly averaged data.

# References


[1] A. Richard Thompson, James M. Moran, and George W. Swenson Jr. *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. Springer International Publishing, 2017.

[2] AIPS. Memo 103. `http://www.aips.nrao.edu/aipsmemo.html`. [Online; accessed 04-March-2019].

[3] L. Blackburn, M. Wielgus, et al. eat library. `https://github.com/sao-eht/eat`. [Online; accessed 06-March-2019].

[4] A. Chael et al. eht-imaging library. `https://github.com/achael/eht-imaging`. [Online; accessed 06-March-2019].

[5] G. N. Ortiz-León, M. D. Johnson, S. S. Doeleman, L. Blackburn, V. L. Fish, L. Loinard, M. J. Reid, E. Castillo, A. A. Chael, A. Hernández-Gómez, D. H. Hughes, J. León-Tavares, R.-S. Lu, A. Montaña, G. Narayanan, K. Rosenfeld, D. Sánchez, F. P. Schloerb, Z.-q. Shen, H. Shiokawa, J. SooHoo, and L. Vertatschitsch. The Intrinsic Shape of Sagittarius A* at 3.5 mm Wavelength. *Astrophysical Journal*, 824:40, June 2016.

[6] EHT Collaboration et al. First M87 Event Horizon Telescope Results. III. Data Processing and Calibration. *Astrophysical Journal, Letters*, 875:L3 (Paper III), April 2019.

[7] L. Blackburn, C. Chan, G. B. Crew, Vincent L. Fish, S. Issaoun, M. D. Johnson, M. Wielgus, K. Akiyama, J. Barrett, K. L. Bouman, R. Cappallo, A. A. Chael, M. Janssen, C. J. Lonsdale, and S. S. Doeleman. EHT-HOPS pipeline for millimeter VLBI data reduction. *arXiv e-prints arXiv:1903.08832*, Mar 2019.

[8] M. Janssen, C. Goddi, I. M. van Bemmel, M. Kettenis, D. Small, E. Liuzzo, K. Rygl, I. Martí-Vidal, L. Blackburn, M. Wielgus, and H. Falcke. rPICARD: A CASA-based Calibration Pipeline for VLBI Data. Calibration and imaging of 7 mm VLBA observations of the AGN jet in M87. *arXiv e-prints*, February 2019.

[9] F. R. Schwab and W. D. Cotton. Global fringe search techniques for VLBI. *Astronomical Journal*, 88:688–694, May 1983.