Modeling atmospheric phase corruptions in high-frequency VLBI using Gaussian processes

Uri Rolls,[1] Dominic W. Pesce,[2,3] Paul Tiede,[2,3] Lindy Blackburn,[2,3] Iniyan Natarajan,[2,3] and Sheperd S. Doeleman[2,3]

[1] *Harvard College, Harvard University, Cambridge, MA 02138, USA*

[2] *Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*

[3] *Black Hole Initiative, Harvard University, 20 Garden Street, Cambridge, MA 02138, USA*

ABSTRACT

Using very long baseline interferometry (VLBI) observations at (sub)millimeter wavelengths, the Event Horizon Telescope (EHT) currently achieves the finest angular resolution of any astronomical facility, necessary for imaging the horizon-scale structure around supermassive black holes. A significant calibration challenge for high-frequency VLBI stems from rapid variations in the atmospheric water vapor content above each telescope in the array, which induce corresponding fluctuations in the phase of the correlated signal that limit the coherent integration time and thus the achievable sensitivity. In this paper, we introduce a model that describes station-based phase corruptions jointly with a parameterization for the source structure. We adopt a Gaussian Process (GP) prescription for the time evolution of these phase corruptions, which provides sufficient flexibility to capture even highly erratic phase behavior. The use of GPs permits the application of a Kalman filtering algorithm for numerical marginalization of these phase corruptions, which permits efficient exploration of the remaining parameter space. Our model also removes the need to specify an arbitrary "reference station" during calibration, instead establishing a global phase zeropoint by enforcing the GPs at all stations to have fixed mean and finite variance. We validate our method using a real EHT observation of the blazar 3C 279, demonstrating that our approach yields calibration solutions that are consistent with those determined by the EHT Collaboration. The model presented here can be straightforwardly extended to incorporate frequency-dependent phase behavior, such as is relevant for the frequency phase transfer calibration technique.

1. INTRODUCTION

A common issue faced in the calibration of radio interferometric data is the presence of substantial and time-variable corruptions in the primary measurement quantity, the complex visibility. For high-frequency very long baseline interferometry (VLBI) observations in particular, such as those conducted by the Event Horizon Telescope (EHT), these corruptions are especially significant and rapid in the phase of the complex visibility (Event Horizon Telescope Collaboration et al. 2019a). These phase variations – primarily introduced by fluctuations in the distribution of water vapor in the troposphere above each station participating in the array – limit the coherence time of observations and thus the length of time over which VLBI signals can be coherently integrated. Effective mitigation of these time-variable phase corruptions is essential for achieving imaging capabilities with VLBI arrays.

At observing frequencies above ∼100 GHz, traditional "phase referencing" calibration (e.g., Beasley & Conway 1995) becomes impractical, because the atmospheric variability timescales are too short and the typical separation between any given science target and a sufficiently bright calibrator is too large. Instead, "self-calibration" techniques are employed that determine phase solutions using the science target itself (e.g., Readhead & Wilkinson 1978; Pearson & Readhead 1984). Such techniques take advantage of the station-based origin of the dominant phase variations to decouple the corruptions from the underlying source visibility phase structure. Self-calibration as applied to EHT data has to date followed an algorithmic inverse-modeling approach, whereby visibility phase measurements on baselines to a sensitive "reference station" are treated as station phases and removed in a manner that stabilizes the time variability while preserving phase closure (e.g., Event Horizon Telescope Collaboration et al. 2019a; Blackburn et al. 2019).

In this paper, we present an alternative approach for VLBI phase calibration that forward-models the phase contributions from each station alongside a prescription for the source visibility phase. We model the time-variable phases as Gaussian Processes (GPs) with a parameterized covariance structure that is free to vary across stations. Our approach is conceptually aligned with a number of other recent advances in forward-modeling techniques for VLBI, particularly those that jointly model corrupting effects alongside the source structure within a Bayesian framework (e.g., Natarajan et al. 2017, 2020; Arras et al. 2019; Broderick et al. 2020; Pesce 2021; Tiede 2022), as well as modeling of tropospheric turbulence as a Gaussian process for the purpose

Corresponding author: Uri Rolls
urirolls@college.harvard.edu

of data simulation (Blecher et al. 2017; Natarajan et al. 2022). The primary difference we advance here is in calibrating data that are further "upstream" than most previous approaches; i.e., our calibration model does not require any pre-processing to stabilize nonlinear visibility phase variations in time.

This paper is structured as follows. In Section 2, we detail our phase model and describe its implementation using a Kalman filtering algorithm. In Section 3 we validate our method using synthetic VLBI datasets, and in Section 4 we apply our approach to EHT observations and compare our results with the original EHT calibration. In Section 5, we present a discussion of our model, and we summarize and conclude in Section 6.

Throughout this paper, we adopt the notation for Gaussian distributions used in Petersen & Pedersen (2006), where a normally-distributed vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is denoted as

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{1}$$

When necessary, we explicitly indicate the random variable in the subscript, e.g., $\mathcal{N}_{\boldsymbol{x}}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, to avoid ambiguity.

2. MODEL DESCRIPTION

An interferometric baseline between stations i and j is sensitive to the complex visibility V_{ij} of the on-sky source structure (Thompson et al. 2017), which is related to the complex correlation of the electric fields E_i and E_j incident on the two stations,

$$V_{ij} \propto \left\langle E_i E_j^* \right\rangle. \tag{2}$$

Here, the asterisk denotes complex conjugation, and the proportionality constant depends on the sensitivities of the participating stations (see, e.g., Blackburn et al. 2019). In this paper, we work exclusively with the phase of the complex visibility,

$$\phi_{ij} = \arg(V_{ij}). \tag{3}$$

2.1. *Phase model*

The relationship in Equation 2 is complicated in practice by the presence of corrupting effects in the data, the most severe of which are typically station-based and known as "complex gain" corruptions. For high-frequency VLBI, the gain phase introduced by variations in the amount of tropospheric water vapor can dominate the signal, causing the visibility phase to vary by more than a radian on timescales of a few to tens of seconds (e.g., Event Horizon Telescope Collaboration et al. 2019b).

Our basic phase model is given by

$$\hat{\phi}_{ij}(t) = \phi_{ij} + \theta_i(t) - \theta_j(t) + \varepsilon_{ij}(t), \tag{4}$$

where ϕ_{ij} represents the true source visibility phase that would be observed in the absence of gain corruptions, $\hat{\phi}_{ij}(t)$ is the measured visibility phase on baseline ij at time t, $\theta_i(t)$ and $\theta_j(t)$ are the gain phases at stations i and j, and $\varepsilon_{ij}(t)$ captures the statistical uncertainty (or "thermal noise") in the phase measurement $\hat{\phi}_{ij}(t)$. We assume that the thermal noise contribution is Gaussian-distributed with variance $\sigma_{ij}^2(t)$,

$$\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma_{ij}^2(t)), \tag{5}$$

and that they are independent across stations and timestamps. This assumption only strictly holds in the limit of high signal-to-noise ratio, but it remains approximately valid even down to signal-to-noise ratios as low as $\sim$2 (Blackburn et al. 2020).

We model the time-dependent gain phases $\theta_i(t)$ and $\theta_j(t)$ as GPs. Specifically, the time series of gain phases at each station is assumed to follow a multivariate normal distribution with a parameterized covariance structure,

$$\begin{pmatrix} \theta_i(t_1) \\ \theta_i(t_2) \\ \vdots \\ \theta_i(t_T) \end{pmatrix} \equiv \boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{6}$$

where $\boldsymbol{\mu}_i$ is a vector representing the mean phase at each timestamp and $\boldsymbol{\Sigma}_i$ is the covariance matrix describing the gain phase structure for station i. For the analyses presented in this paper, all elements of $\boldsymbol{\mu}_i$ are fixed to be zero. The covariance structure of each GP is described by a kernel function, which can in principle take many forms. For the analyses presented in this paper, we use a Matérn-1/2 kernel, such that the elements of the covariance matrix are given by

$$(\Sigma_i)_{mn} = \sigma_i^2 \exp\left(-\frac{|t_m - t_n|}{\tau_i}\right). \tag{7}$$

Here, τ_i sets the variation timescale and σ_i sets the magnitude of the phase fluctuations. In Appendix C, we assess the sensitivity of our measurements to the particular choice of kernel function.

For an array containing N stations and making visibility phase measurements on each of T timestamps, our model contains the following free parameters:

- There are $N(N-1)/2$ visibility phase parameters ϕ_{ij}, which we assume to be constant in time.
- There are NT gain phase parameters $\theta_i(t_k)$, one for each station i at each timestamp t_k.
- There are $2N$ GP kernel parameters, a timescale τ_i and variance σ_i^2 for each station, which we assume to be constant in time.

We note that the source visibility phases ϕ_{ij} could in principle be inherited from a more general model of the source structure – such as might be available during image reconstruction – rather than being assumed constant. However, for the analyses presented in this paper, we assume that the duration of the observation is sufficiently short that a constant-valued assumption for ϕ_{ij} is a reasonable approximation.

2.2. *Kalman filtering*

For many applications, the gain phases θ_i are nuisance parameters that we model only for the sake of accessing the visibility phases ϕ_{ij}, which are themselves the primary parameters of interest. Because our phase model is linear in the gain phases (Equation 4) and our measurement uncertainties are taken to be Gaussian (Equation 5), we can analytically marginalize over the θ_i parameters, yielding a posterior distribution for the remaining ϕ_{ij}, τ_i, and σ_i parameters. We derive the marginal posterior distribution for our model in Appendix A, but unfortunately this distribution is inefficient to work with in practice because it involves the construction and inversion of large (typical dimension $\gg 10^3$) matrices.

Fortunately, our model is linear and Gaussian, and because we have restricted our covariance structure to the half-integer Matérn family (see Equation 7) it also admits a so-called "state-space representation" (see Rozanov 1977, 1982 and Appendix B for details). This representation enables application of the efficient Kalman filtering algorithm to numerically solve and marginalize over the gain phases. Kalman filtering is commonly employed for modeling time-series data (e.g., Särkkä 2013), where it takes advantage of the sequential nature of such data to iteratively refine a running solution using each data point consecutively. I.e., given a solution derived from data points 1 through k, the Kalman filtering algorithm incorporates data point $k+1$ to update the existing solution rather than recomputing an entirely new one. In this way, the complexity of the Kalman filtering algorithm grows only linearly with the number of data points, providing an efficient method for estimating the gain phases in our model. We detail our implementation of the Kalman filtering algorithm in Appendix B.

2.3. *Parameter space exploration*

In our framework, the Kalman filter operates as an "inner loop" within a broader "outer loop" that explores the full parameter space of the model. This outer loop samples over the kernel parameters τ_i and σ_i for each station, as well as over the visibility phases ϕ_{ij} for each baseline. For each sample of τ_i, σ_i, and ϕ_{ij}, the Kalman filter inner loop provides the marginal likelihood – i.e., the likelihood of the data (given these model parameters) after marginalizing over the gain phase parameters – specified in Equation B.23. This marginal likelihood value is then combined (via Bayes' Theorem) with prior information for each parameter to update the sampler. The priors for our model parameters are as follows:

Table 1. Input values for synthetic data generation

Station/ Baseline	τ **(Timescale)**	σ^2 **(Variance)**	ϕ **(Phase)**
1	20	1.0	-
2	25	2.0	-
3	30	1.5	-
4	35	0.5	-
12	-	-	1.0
13	-	-	0.5
14	-	-	2.0
23	-	-	1.5
24	-	-	0.0
34	-	-	1.0

- The variation timescale τ_i for each station follows a truncated normal distribution with mean 0 s and standard deviation 30 s; the distribution is truncated at zero such that negative values are not permitted.
- The magnitude of the phase fluctuations σ_i for each station follows a truncated normal distribution with a mean of 0 radians and a standard deviation of 2 radians; the distribution is truncated at zero such that negative values are not permitted.
- The visibility phases ϕ_{ij} follow a uniform prior over the range $(-4\pi, 4\pi)$ radians.

For the analyses presented in this paper, we use `dynesty` to explore the model parameter space and to produce posterior distributions (Speagle 2020).[1] The `dynesty` code implements the nested sampling algorithm (Skilling 2004, 2006), which was originally developed to evaluate Bayesian evidence integrals but which has since been broadly applied as a posterior estimation tool.

3. DEMONSTRATION USING SYNTHETIC DATA

In this section, we provide a demonstration of our phase modeling approach using synthetic data. We generate synthetic visibility phase data appropriate for a 4-station array following the model described in Section 2.1.

For each station, we generate a time series of 300 gain phases via a random draw from a GP with τ_i and σ_i parameters as specified in Table 1. The parameters for the synthetic datasets aim to reflect typical values observed in high-frequency VLBI. The τ values of $\sim$tens of

[1] `Dynesty.jl`, a Julia wrapper for `dynesty`, is available at https://github.com/tkoolen/Dynesty.jl.

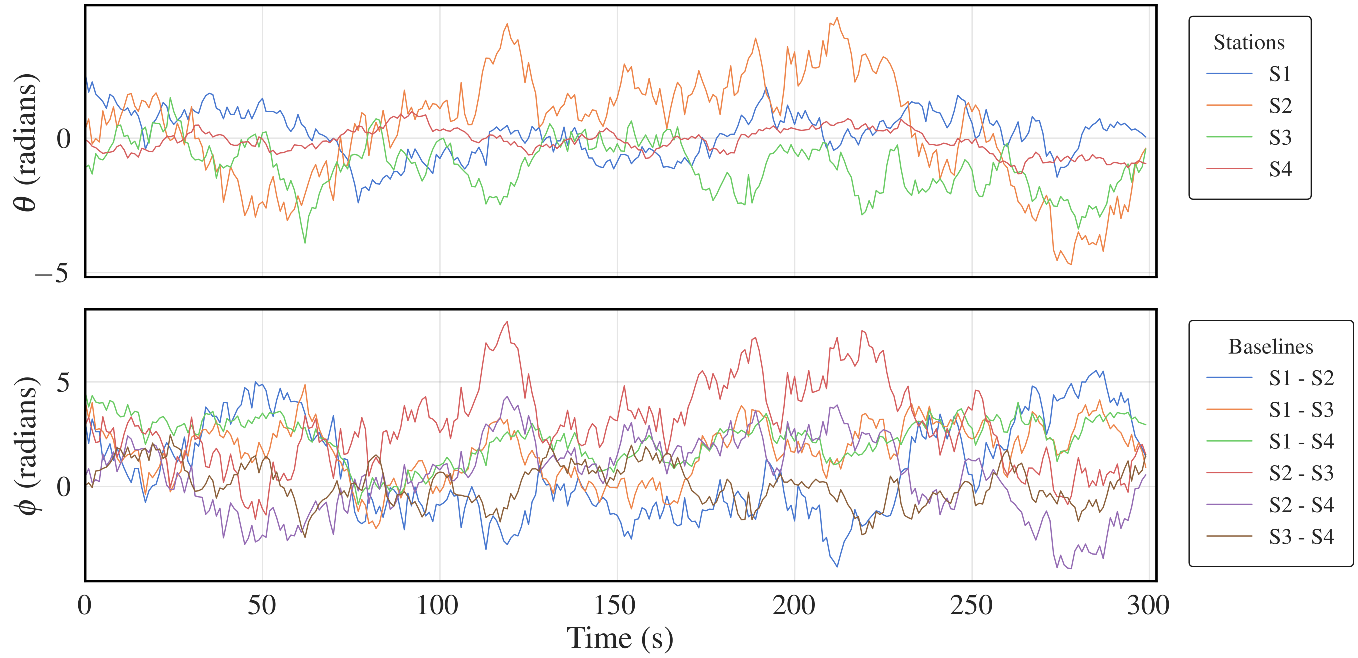

Figure 1. Synthetic data gain phases (top) and visibility phases (bottom) versus time for each station and baseline, respectively.

seconds represent common atmospheric coherence times at 230 GHz (Event Horizon Telescope Collaboration et al. 2019b), and the σ values ($\sim$ 1 rad) generate total phase fluctuations that are qualitatively similar to those in the 3C 279 data shown in Figure 2. We use the `KernelFunctions.jl`[2] Julia package to sample the GPs, and we assume a Matérn-1/2 kernel for all stations (see Equation 7). Each baseline in the array is assigned an intrinsic ϕ_{ij} visibility phase (specified in Table 1), whose value is modified by the gain phases per Equation 4. We then add to each timestamp on each baseline an independent random phase drawn from a Gaussian distribution with variance σ_{ij}^2, to simulate thermal noise. The value chosen for the thermal noise is 0.05. The generated gain phases for each station and the corresponding visibility phases on each baseline are shown in the top and bottom panels of Figure 1, respectively.

After generating the synthetic data, we sample the posterior parameter space using the scheme described in Section 2.3. Figure 2 shows the posterior distributions for the τ_i and σ_i parameters describing the GP for each station, and the posterior distributions for the ϕ_{ij} visibility phase values. We find accurate recovery of all model parameters.

4. APPLICATION TO REAL DATA

In this section, we apply our phase model to real high-frequency VLBI data collected by the EHT. We use a dataset corresponding to an EHT observation of the blazar 3C 279 carried out on 2017 April 5. This blazar is a bright calibrator source with many high signal-to-noise ratio detections in the EHT data, which are important for reliably unwrapping visibility phases. Details about the original processing and calibration of this dataset can be found in Event Horizon Telescope Collaboration et al. (2019a), and the scientific utility has been presented in Kim et al. (2020).

For our analysis, we work with a version of the original EHT dataset that has been partially calibrated using the `EHT-HOPS` pipeline (Blackburn et al. 2019). Specifically, we only carry out calibration through "stage 2" of the typical `EHT-HOPS` pipeline, such that phase bandpass solutions have been applied but no time-dependent phase calibration has yet taken place beyond a single delay-rate correction per scan. We coherently average the data in frequency across the entire 2 GHz band, but we retain the original post-correlation time segmentation of 0.4 s. For the proof-of-concept demonstration presented in this paper, we work with only the "low-band" data (centered on an observing frequency of 227 GHz) and only with the left circularly polarized parallel-hand correlation products (LL). We further restrict our analysis to just 5 of the 7 stations originally participating in the observation – retaining ALMA, APEX, LMT, SMT, and the IRAM 30 m but flagging SMA and JCMT – so that the signal-to-noise ratio of the remaining data are high enough to disambiguate phase wraps at the native time sampling

[2] https://juliagaussianprocesses.github.io/KernelFunctions.jl

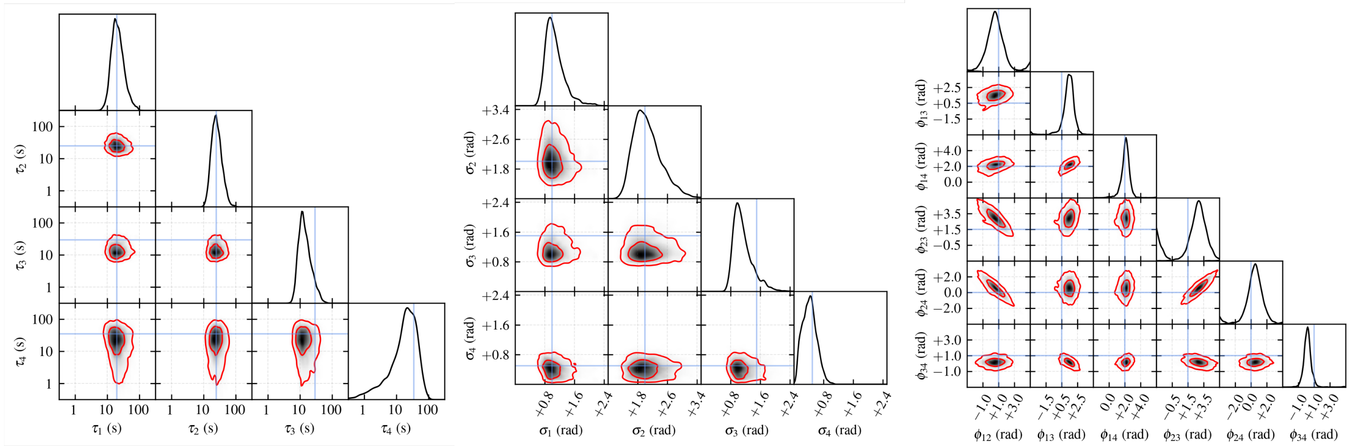

Figure 2. Posterior distributions for the model parameters determined from fitting synthetic data. Left: Pairwise distributions for the coherence timescale parameters τ_i. Center: Pairwise distributions for the phase fluctuation magnitude parameters σ_i. Right: Posterior distributions for the visibility phase parameters ϕ_{ij}. For all 2D distributions, the red contours enclose 50% and 90% of the posterior probability. Blue lines represent the input values used when generating the synthetic data.

of 0.4 s.[3] We unwrap the visibility phases prior to fitting using the `numpy` library's `unwrap` function. Figure 3 shows the visibility phases versus time for all scans in this dataset, where individual scans can be identified as groups of data points separated by visible gaps in time. Figure 4 shows the unwrapped visibility phases versus time for a single scan, and Figure 5 shows our gain phase solution for that same scan.

The 3C 279 dataset is divided into 19 individual scans, each of which has a duration of several minutes. We expect the constant-valued visibility phase assumption to hold for each scan individually, but not across many scans, so we independently fit our model to each individual scan. The resulting calibrated visibility phases are shown in Figure 6, where they are compared against the scan-averaged output of the full `EHT-HOPS` calibration from Event Horizon Telescope Collaboration et al. (2019a).[4] We find quantitative agreement between the results of our fitting routine and the fully-calibrated EHT data, both in the values of the visibility phases and in the magnitudes of their statistical uncertainties.

[3] Proper handling of the phase unwrapping for the baselines to the SMA and JCMT stations – which are located on Mauna Kea in Hawaii, and thus only participate in long baselines that have a generically lower SNR – would require using a different averaging timescale compared to other baselines, which complicates the analysis without adding any insight. So while the baselines contributed by these stations would be useful for, e.g., an imaging analysis, for the purposes of the calibration demonstration presented here, we have opted instead to stick with a more uniform subset of high-SNR baselines.

[4] We note that the data calibrated with `EHT-HOPS` are phase-referenced to the ALMA station, such that visibility phases on baselines to ALMA are zero-valued while retaining phase closure. For the comparison in Figure 6, we thus carry out a post-fitting phase referencing to the ALMA station, though we note that no such referencing is required during the fitting procedure itself.

5. DISCUSSION

Our phase model and associated fitting routine provides a new tool for calibrating visibility phase measurements. The choice of a GP prescription for the individual station gains affords the model substantial flexibility in describing phase variations, which is necessary for capturing the rapid phase fluctuations typically introduced by the atmosphere during high-frequency VLBI observations.

For both the synthetic data and real data demonstrations presented in this paper, we have adopted a Matérn-1/2 covariance kernel for the GPs describing each of the station gain phases. This specific choice of kernel is not intrinsic to our phase model or our fitting procedure, and in principle many different kernels could be selected and employed in an analogous manner. In Appendix C, we assess the sensitivity of our parameter measurements to the particular choice of kernel function, demonstrating that for the primary parameters of interest – i.e., the visibility phases – the Matérn-1/2 kernel is a sufficiently flexible choice. An additional benefit of the Matérn family of kernels is that they permit use of the efficient Kalman filtering algorithm, which is not a property that is shared by all possible kernels (see, e.g., Hartikainen & Särkkä 2010).

We have implemented our model using the Julia scientific programming language, which will permit its straightforward incorporation into the `Comrade.jl` Bayesian modeling framework for radio interferometry (Tiede 2022). Within the `Comrade.jl` framework, our phase model can serve as a modular component of a comprehensive instrument model, and the constant-valued visibility phase assumption (see Section 2.1) can be relaxed in favor of visibility phase values that are self-consistently informed by a parameterized model (e.g., image) for the source structure. Furthermore, forward-

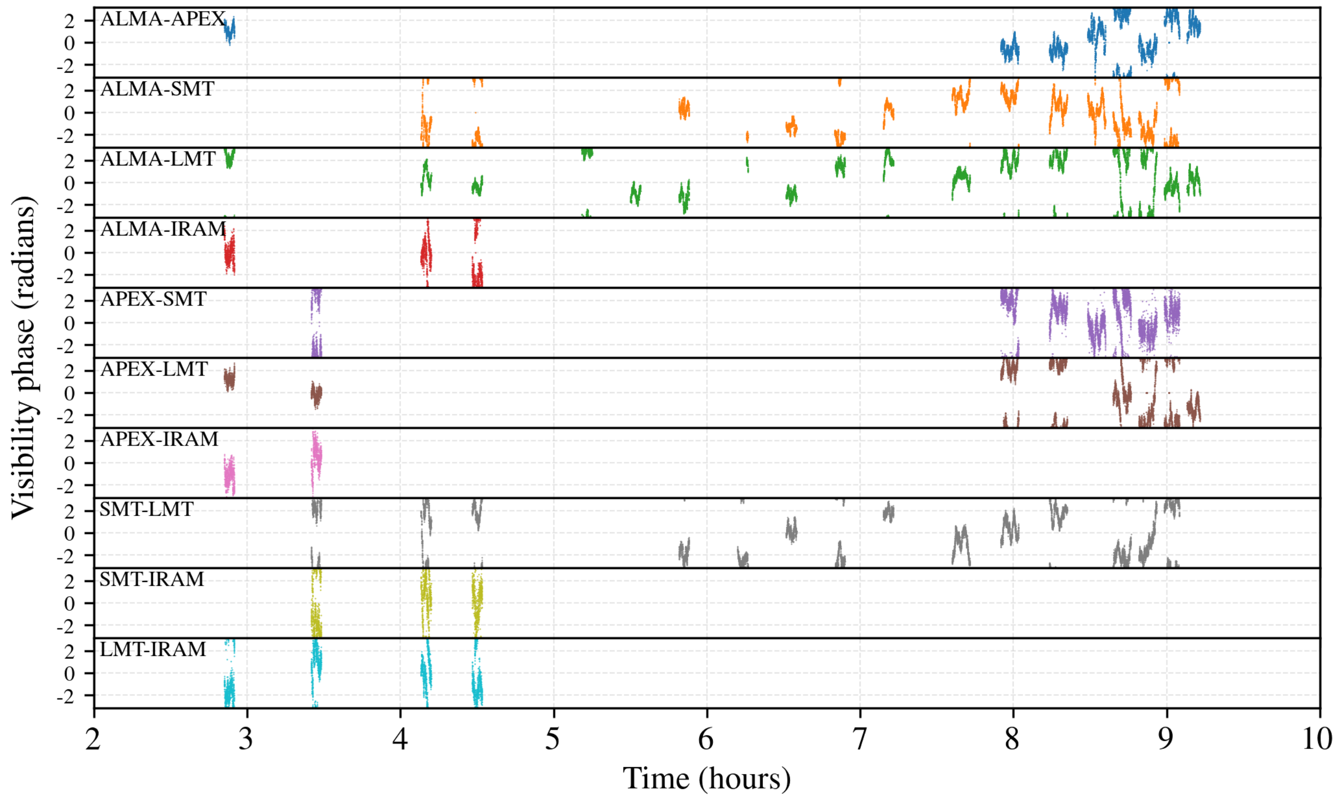

Figure 3. Uncalibrated visibility phase versus time on each of the baselines in the EHT 3C 279 dataset; in each panel, the baseline is specified in the upper left-hand corner. Time is measured in UT hours since the start of 2017 April 5.

modeling of the gain phases in this manner permits straightforward generalizations, such as inclusion of the deterministic frequency dependence expected for non-dispersive tropospheric phase contributions (e.g., Carilli & Holdaway 1999); we expect such model extensions to be the subject of future development.

5.1. *Reference station considerations*

One useful feature of our approach is that it does away with the notion of a reference station, the specification of which is a typical requirement in other phase calibration methods to remove trivial phase degeneracies.

For a fully-connected array containing N stations, there are $B = N(N-1)/2$ baseline-based visibility phase parameters and N station-based gain phase parameters at each time (see Equation 4). However, there are only B visibility phase measurements acting as constraints on the system, leaving an additional N degrees of freedom unconstrained. A standard approach for removing these additional degrees of freedom is to select one station – the reference station – and to fix (typically to zero) the values of all $(N-1)$ visibility phases containing that station as well as the (single) gain phase for that station (e.g., Blackburn et al. 2019).

There are a number of conceptual and practical drawbacks associated with the reference station approach. Conceptually, the identification of a single station to receive special treatment requires an arbitrary choice to be made. Furthermore, imposing fixed gain phase values for any single station effectively forces a calibration solution to redistribute any actual variations in that station's gain phases among the other stations in the array, potentially in conflict with the assumptions of a physical model for the phase gain behavior. Selection of a reference station also runs into practical issues with real-world observations, where individual stations can join late or leave early; discontinuous phase jumps can result when the reference station changes partway through an observation.

In our approach, the GP specification of the gain phases acts to regularize the system and break the phase degeneracies in a global manner, without needing to select out a single station to be treated differently from the others. For fixed GP parameters, unique solutions for the gain phases can be determined through, e.g., likelihood maximization (see Section A.3). Even when fitting for the GP parameters, fixing the mean of each GP (which we set to zero) and imposing a finite-width prior on its variance (e.g., σ_i^2 in Equation 7) is sufficient to break the phase degeneracies.

5.2. *Limitations*

We note that the current formulation of our model and fitting routine suffers from a number of limitations,

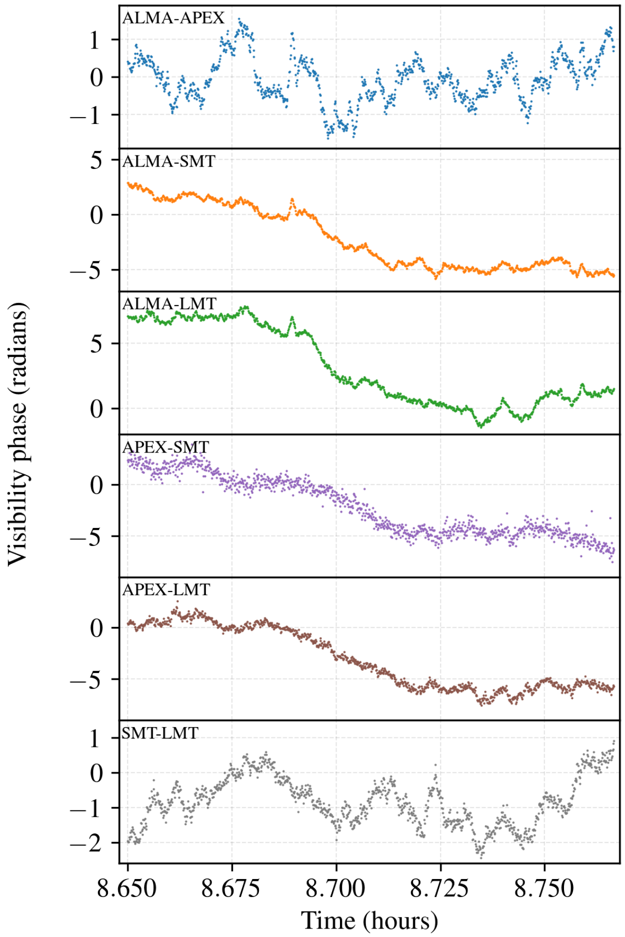

Figure 4. Same as Figure 3, but showing now the unwrapped visibility phases for a single scan. Note that no single scan contained all 10 baselines shown in Figure 3, so we show here a scan containing only 6 baselines.

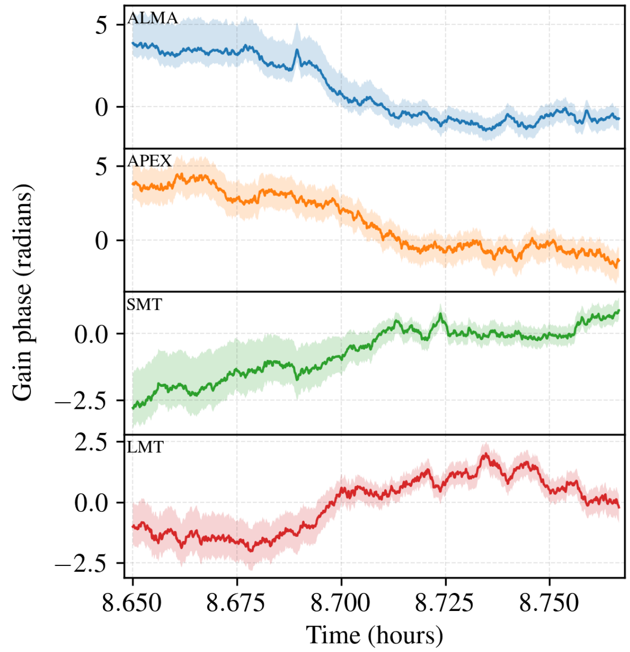

Figure 5. Recovered station gain phases versus time for each of the stations participating in the scan whose visibility phases are shown in Figure 4. The solid curve in each panel shows the median posterior value, and the shaded region indicates 90% confidence interval.

primarily associated with our choice of visibility phase as the measured quantity of interest. Interferometers directly measure complex visibilities, not phases, and the association of a unique phase value for any particular complex visibility measurement can only be made when the signal-to-noise ratio is sufficiently large.

For reliably "unwrapping" phases, we find that signal-to-noise ratios $\gtrsim$3 are often sufficient, but this threshold should be understood as a rule of thumb rather than a strict cutoff. The critical requirement is the ability to distinguish phases near $+180°$ from those near $-180°$. At SNR $\sim$1, the RMS phase variation between consecutive data points is approximately 1 radian, making it difficult to uniquely identify phase wraps. We expect the performance to degrade approximately continuously with decreasing SNR rather than exhibiting a sharp failure at some particular threshold.

Relatedly, our assumption in Equation 5 that the thermal noise in the visibility phase measurements is Gaussian-distributed will also only hold in the limit of large signal-to-noise ratio (Blackburn et al. 2020), meaning that for data with low or moderate signal-to-noise ratios the likelihood function we've adopted during parameter estimation is not statistically appropriate. A generalization of our approach that directly models the complex visibilities (which are themselves expected to be Gaussian-distributed in the complex plane; see Thompson et al. 2017) rather than just the visibility phases could plausibly overcome both of these limitations, though such an approach would also break the linearity of Equation 4 and thus require a more sophisticated parameter estimation routine than we have employed here.

6. SUMMARY AND CONCLUSIONS

We present a model and associated parameter estimation routine for calibrating radio interferometric visibility phase measurements, appropriate for tackling the strong atmospheric phase variability typically impacting high-frequency VLBI observations. We use a Gaussian Process (GP) prescription to model the individual station gain phases, which provides sufficient flexibility to capture even highly erratic phase behavior. The GP prescription further permits use of the Kalman filtering algorithm for efficient gain marginalization, which we employ during parameter estimation. By enforcing the GPs for all stations to have fixed mean and finite

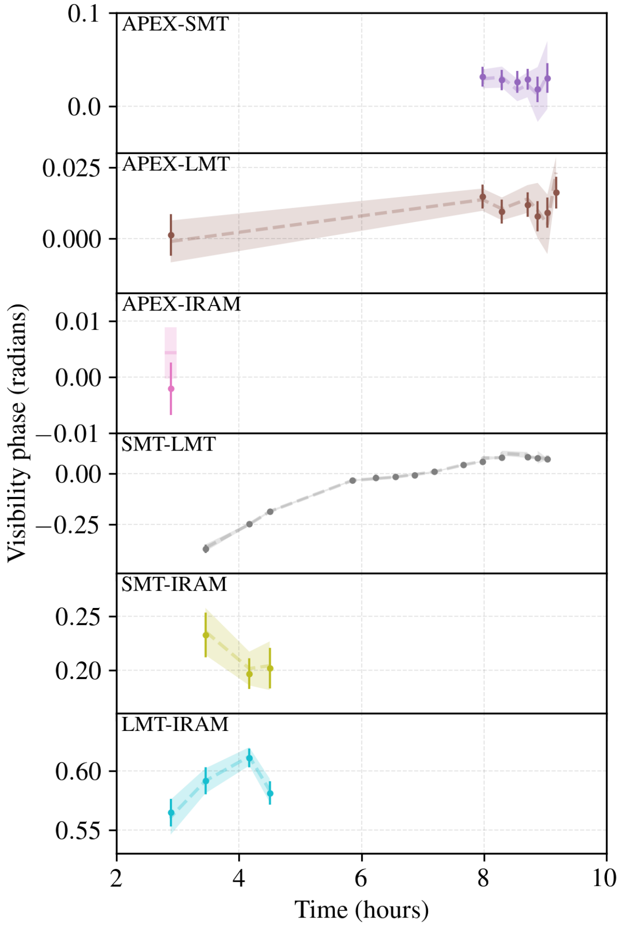

Figure 6. Calibrated and scan-averaged visibility phase versus time on each of the non-ALMA baselines in the EHT 3C 279 dataset; in each panel, the baseline is specified in the upper left-hand corner. The data calibrated by the EHT Collaboration are shown as dashed lines with shading indicating 1σ confidence regions, while the results of the calibration carried out in this paper are shown as circular markers with error bars (again indicating 1σ confidence regions).

variance, our model establishes a global phase zeropoint and removes the need to specify an arbitrary reference station during calibration.

We have applied our model to both synthetic data and to a real high-frequency (∼230 GHz) VLBI dataset from the EHT. For the synthetic data, we demonstrate that the model faithfully recovers the input parameter values describing both the individual station GPs as well as the visibility phases (Figure 2). For the real data, we find that our recovered visibility phase parameters are quantitatively consistent with the results of `EHT-HOPS` calibration carried out by the EHT Collaboration (Event Horizon Telescope Collaboration et al. 2019a), both in the values of the phases and in the magnitudes of their uncertainties (Figure 6). These applications demonstrate that our phase model provides a viable method for calibrating high-frequency VLBI data.

Looking ahead, one of the main development priorities for high-frequency VLBI projects – including the EHT (The Event Horizon Telescope Collaboration 2024), the next-generation EHT (Doeleman et al. 2023), and the Black Hole Explorer (Johnson et al. 2024) – is the inclusion of multi-frequency observing capabilities. A driving factor behind this push is a desire to leverage the "frequency phase transfer" (FPT) calibration technique (Rioja et al. 2023; Issaoun et al. 2025), whereby atmospheric phase solutions derived at a lower frequency are used to stabilize the phase and increase coherent integration times at a higher frequency. Because the expected phase relationship between the two observing frequencies is known, an FPT calibration scheme is straightforward to implement within a forward-modeling framework such as the one we present here.

ur current model suffers from some limitations, discussed in Section 5.2. Most notably, our method requires the ability to reliably unwrap visibility phases, which is typically possible only for SNR $\gtrsim$3. In practice, we expect that our phase model will serve most effectively as one calibration component within more comprehensive forward-modeling image reconstruction tools such as `ehtim` (Chael et al. 2016, 2018) or `Comrade.jl` (Tiede 2022). Future work will focus on generalizing our model from visibility phases to complex visibilities, and on expanding it to include frequency dependence of the phase signal. We expect the resulting improvements to the quality and efficiency of interferometric phase calibration to afford corresponding benefits to downstream image reconstructions and associated scientific analyses.

Software: `ChatGPT` (OpenAI 2022), `dynesty` (Speagle 2020), `EHT-HOPS` (Blackburn et al. 2019), `KernelFunctions.jl`, `matplotlib` (Hunter 2007), `numpy` (van der Walt et al. 2011; Harris et al. 2020)

ACKNOWLEDGMENTS

We are grateful to Laurent Loinard for constructive comments that improved the quality of this paper. We have used `ChatGPT` (OpenAI 2022) to assist with code and figure generation for this paper.

Support for this work was provided by the NSF through grants AST-1935980 and AST-2034306, and by the Gordon and Betty Moore Foundation through grants GBMF5278 and GBMF10423. This work has been supported in part by the Black Hole Initiative at Harvard University, which is funded by the John Templeton Foundation (grants 60477, 61479, and 62286) and the Gordon and Betty Moore Foundation (grant GBMF8273).

This paper makes use of the following ALMA data: 2016.1.01114.V, 2016.1.01176.V, 2016.1.01404.V, and

2016.1.01154.V. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

REFERENCES

Arras, P., Frank, P., Leike, R., Westermann, R., & Enßlin, T. A. 2019, A&A, 627, A134, doi: 10.1051/0004-6361/201935555

Beasley, A. J., & Conway, J. E. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 82, Very Long Baseline Interferometry and the VLBA, ed. J. A. Zensus, P. J. Diamond, & P. J. Napier, 327

Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, ApJ, 894, 31, doi: 10.3847/1538-4357/ab8469

Blackburn, L., Chan, C.-k., Crew, G. B., et al. 2019, ApJ, 882, 23, doi: 10.3847/1538-4357/ab328d

Blecher, T., Deane, R., Bernardi, G., & Smirnov, O. 2017, MNRAS, 464, 143, doi: 10.1093/mnras/stw2311

Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020, ApJ, 898, 9, doi: 10.3847/1538-4357/ab9c1f

Carilli, C. L., & Holdaway, M. A. 1999, Radio Science, 34, 817, doi: 10.1029/1999RS900048

Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, ApJ, 857, 23, doi: 10.3847/1538-4357/aab6a8

Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, ApJ, 829, 11, doi: 10.3847/0004-637X/829/1/11

Doeleman, S. S., Barrett, J., Blackburn, L., et al. 2023, Galaxies, 11, 107, doi: 10.3390/galaxies11050107

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJL, 875, L3, doi: 10.3847/2041-8213/ab0c57

—. 2019b, ApJL, 875, L2, doi: 10.3847/2041-8213/ab0c96

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hartikainen, J., & Särkkä, S. 2010, in 2010 IEEE International Workshop on Machine Learning for Signal Processing, 379–384, doi: 10.1109/MLSP.2010.5589113

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Issaoun, S., Pesce, D. W., Rioja, M. J., et al. 2025. https://arxiv.org/abs/2502.18776

Johnson, M. D., Akiyama, K., Baturin, R., et al. 2024, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 13092, Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave, ed. L. E. Coyle, S. Matsuura, & M. D. Perrin, 130922D, doi: 10.1117/12.3019835

Kim, J.-Y., Krichbaum, T. P., Broderick, A. E., et al. 2020, A&A, 640, A69, doi: 10.1051/0004-6361/202037493

Natarajan, I., Paragi, Z., Zwart, J., et al. 2017, MNRAS, 464, 4306, doi: 10.1093/mnras/stw2653

Natarajan, I., Deane, R., van Bemmel, I., et al. 2020, MNRAS, 496, 801, doi: 10.1093/mnras/staa1503

Natarajan, I., Deane, R., Martí-Vidal, I., et al. 2022, MNRAS, 512, 490, doi: 10.1093/mnras/stac531

OpenAI. 2022, Introducing ChatGPT. https://openai.com/blog/chatgpt/

Pearson, T. J., & Readhead, A. C. S. 1984, ARA&A, 22, 97, doi: 10.1146/annurev.aa.22.090184.000525

Pesce, D. W. 2021, AJ, 161, 178, doi: 10.3847/1538-3881/abe3f8

Petersen, K., & Pedersen, M. 2006, The Matrix Cookbook (Technical University of Denmark)

Readhead, A. C. S., & Wilkinson, P. N. 1978, ApJ, 223, 25, doi: 10.1086/156232

Rioja, M. J., Dodson, R., & Asaki, Y. 2023, Galaxies, 11, 16, doi: 10.3390/galaxies11010016

Rozanov, J. A. 1977, Mathematics of the USSR-Sbornik, 32, 515, doi: 10.1070/SM1977v032n04ABEH002404

Rozanov, Y. A. 1982, Markov random fields (Springer)

Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (AIP), 395–405, doi: 10.1063/1.1835238

Skilling, J. 2006, Bayesian Analysis, 1, 833, doi: 10.1214/06-BA127

Speagle, J. S. 2020, MNRAS, 493, 3132, doi: 10.1093/mnras/staa278

Särkkä, S. 2013, Bayesian Filtering and Smoothing, Institute of Mathematical Statistics Textbooks (Cambridge University Press)

The Event Horizon Telescope Collaboration. 2024, arXiv e-prints, arXiv:2410.02986, doi: 10.48550/arXiv.2410.02986

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition (Springer International Publishing), doi: 10.1007/978-3-319-44431-4

Tiede, P. 2022, The Journal of Open Source Software, 7, 4457, doi: 10.21105/joss.04457

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, doi: 10.1109/MCSE.2011.37

APPENDIX

A. ANALYTIC EXPRESSIONS

Following Blackburn et al. (2020), we can compactly write Equation 4 for all measurements simultaneously using a vectorized notation,

$$\hat{\boldsymbol{\phi}} = \boldsymbol{\phi} + \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{A.1}$$

Here, $\hat{\boldsymbol{\phi}}$ denotes a vector of measured visibility phases, $\boldsymbol{\phi}$ denotes a vector of modeled visibility phases, $\boldsymbol{\theta}$ denotes a vector of modeled gain phases, $\boldsymbol{\varepsilon}$ denotes a vector of the thermal noise values, and $\boldsymbol{\Phi}$ is a "design matrix" that maps the gain phases to the visibility phases. The design matrix $\boldsymbol{\Phi}$ contains as its elements only values of 1, -1, and 0, and the specific construction of $\boldsymbol{\Phi}$ depends on how the θ_i and ϕ_{ij} parameters are ordered within the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, respectively. For example, given a 3-station array with 2 timestamps of data, we could expand out Equation A.1 as

$$\begin{pmatrix} \hat{\phi}_{12}(t_1) \\ \hat{\phi}_{12}(t_2) \\ \hat{\phi}_{13}(t_1) \\ \hat{\phi}_{13}(t_2) \\ \hat{\phi}_{23}(t_1) \\ \hat{\phi}_{23}(t_2) \end{pmatrix} = \begin{pmatrix} \phi_{12} \\ \phi_{12} \\ \phi_{13} \\ \phi_{13} \\ \phi_{23} \\ \phi_{23} \end{pmatrix} + \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \theta_1(t_1) \\ \theta_1(t_2) \\ \theta_2(t_1) \\ \theta_2(t_2) \\ \theta_3(t_1) \\ \theta_3(t_2) \end{pmatrix} + \begin{pmatrix} \varepsilon_{12}(t_1) \\ \varepsilon_{12}(t_2) \\ \varepsilon_{13}(t_1) \\ \varepsilon_{13}(t_2) \\ \varepsilon_{23}(t_1) \\ \varepsilon_{23}(t_2) \end{pmatrix}, \tag{A.2}$$

where the design matrix appears as the 6×6 square matrix in the middle term on the right-hand side of the equation. Note that because we assume that the ϕ_{ij} parameters are constant-valued in time, there are repeated elements in the first term on the right-hand side of Equation A.2.

We can similarly cast Equation 5 as a multivariate Gaussian distribution,

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{S}), \tag{A.3}$$

where $\mathbf{S}$ is a diagonal matrix containing as its entries the $\sigma_{ij}^2(t)$ thermal noise variances. We can combine Equation A.1 and Equation A.3 and rearrange to construct a likelihood function for the model parameters,

$$\mathcal{L}(\hat{\boldsymbol{\phi}}|\boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\theta}) = \mathcal{N}_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi} + \boldsymbol{\Phi}\boldsymbol{\theta}, \mathbf{S}). \tag{A.4}$$

A.1. *Gain phase marginalization*

Following the procedure detailed in Appendix C.5 of Blackburn et al. (2020) and assuming uninformative (i.e., wide, flat) priors on each of the model parameters, we can apply Bayes' Theorem using this likelihood function and integrate over the $\boldsymbol{\theta}$ nuisance parameters to determine the marginalized posterior distribution over the remaining parameters,

$$p(\boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{\sigma}|\hat{\boldsymbol{\phi}}) = \mathcal{N}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}, \boldsymbol{\Sigma}). \tag{A.5}$$

Here, the posterior covariance matrix $\boldsymbol{\Sigma}$ can be constructed as a block matrix in terms of the contributions from individual baselines. For instance, for a three-station array we could write

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{S}_{12} + \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 & \boldsymbol{\Sigma}_1 & -\boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_1 & \mathbf{S}_{13} + \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_3 & \boldsymbol{\Sigma}_3 \\ -\boldsymbol{\Sigma}_2 & \boldsymbol{\Sigma}_3 & \mathbf{S}_{23} + \boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_3 \end{pmatrix}, \tag{A.6}$$

where the $\mathbf{S}_{ij}$ sub-matrices are diagonal and contain the as their entries the $\sigma_{ij}^2(t)$ thermal noise variances, and the $\boldsymbol{\Sigma}_i$ sub-matrices are the individual station GP covariances whose entries are given by Equation 7.

Although Equation A.5 is Gaussian, we note that the parameters describing the GP kernels for each station – i.e., the τ_i and σ_i parameters – are not themselves Gaussian-distributed. These parameters appear in the individual $\boldsymbol{\Sigma}_i$ sub-matrices, and are thereby subject to nonlinear operations (e.g., the determinant of $\boldsymbol{\Sigma}$) that result in a nontrivial posterior distribution. Furthermore, $\boldsymbol{\Sigma}$ – which is a $NT \times NT$ matrix – has typical dimensions that are well in excess of 10^3, making numerical sampling from Equation A.5 extremely inefficient in practice.

A.2. *Maximum a posteriori visibility phase solution*

Given fixed values of the GP kernel parameters τ_i and σ_i for each station i, the maximum a posteriori (MAP) solution for $\boldsymbol{\phi}$ can be determined by setting all ϕ-derivatives of Equation A.5 (e.g., $\frac{\partial p}{\partial \phi_{12}}$, $\frac{\partial p}{\partial \phi_{13}}$, etc.) to zero. The result can be expressed as

$$\boldsymbol{\phi}_{\mathrm{MAP}} = \begin{pmatrix} \mathbf{1}_{12}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{12} & \mathbf{1}_{12}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{13} & \dots & \mathbf{1}_{12}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{N-1,N} \\ \mathbf{1}_{13}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{12} & \mathbf{1}_{13}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{13} & \dots & \mathbf{1}_{13}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{N-1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{N-1,N}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{12} & \mathbf{1}_{N-1,N}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{13} & \dots & \mathbf{1}_{N-1,N}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}_{N-1,N} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}_{12}^{\top}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\phi}} \\ \mathbf{1}_{13}^{\top}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\phi}} \\ \vdots \\ \mathbf{1}_{N-1,N}^{\top}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\phi}} \end{pmatrix}, \tag{A.7}$$

where $\boldsymbol{\Sigma}$ is the posterior covariance (Equation A.5), and the notation $\mathbf{1}_{ij}$ indicates a vector whose entries are 1 wherever $\boldsymbol{\phi}$ is equal to ϕ_{ij} and zero everywhere else.

A.3. *Maximum likelihood gain phase solution*

Rather than marginalizing the gain phases, we can alternatively ask what the maximum-likelihood solution for $\boldsymbol{\theta}$ is, given some guess or solution for the other model parameters (i.e., the visibility phases $\boldsymbol{\phi}$ and the GP kernel parameters τ_i, σ_i for each station). To identify a unique solution in the face of phase degeneracies (e.g., associated with an arbitrary phase zeropoint; see discussion in Section 5.1), the likelihood function Equation A.4 needs to be augmented with a prior; in our case, the gain phase prior is determined by the GPs at each station. The effective likelihood function is then given by

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{N}_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi} + \boldsymbol{\Phi}\boldsymbol{\theta}, \mathbf{S})\mathcal{N}_{\boldsymbol{\theta}}\left(0, \boldsymbol{\Sigma}_{\theta}\right), \tag{A.8}$$

where $\boldsymbol{\Sigma}_{\theta}$ encodes the covariance structure for all stations. For instance, $\boldsymbol{\Sigma}_{\theta}$ for a three-station array could be constructed as a block-diagonal matrix,

$$\boldsymbol{\Sigma}_{\theta} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_3 \end{pmatrix},$$

where $\boldsymbol{\Sigma}_i$ is the covariance matrix describing the GP gain phase structure for station i (see Equation 6 and Equation 7) and $\mathbf{0}$ indicates a matrix whose entries are all zero.

As in Section A.2, the maximum-likelihood solution for $\boldsymbol{\theta}$ can be determined by setting all θ-derivatives of Equation A.8 to zero. The result is given by

$$\boldsymbol{\theta} = \left(\boldsymbol{\Phi}^{\top}\mathbf{S}^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\theta}^{-1}\right)^{-1}\boldsymbol{\Phi}^{\top}\mathbf{S}^{-1}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right), \tag{A.9}$$

which can also be understood as a standard generalized least squares solution to the system of linear equations described by Equation A.1 (plus terms associated with the prior).

B. KALMAN FILTER IMPLEMENTATION

In this appendix, we provide a description of the Kalman filter implementation we use to estimate atmospheric gain phases. We begin by establishing the model formulation, and we then describe the filter's prediction and update steps.

B.1. *Parameters and measurements*

For a VLBI array with N stations, we define the "state vector" $\boldsymbol{\theta}_k$ at time t_k to be the collection of gain phases at each station,

$$\boldsymbol{\theta}_k = \begin{pmatrix} \theta_1(t_k) \\ \theta_2(t_k) \\ \vdots \\ \theta_N(t_k) \end{pmatrix}. \tag{B.1}$$

The "measurement vector" $\tilde{\boldsymbol{\phi}}_k$ consists of the residual phases on each baseline after subtracting the model visibility phases,

$$\tilde{\boldsymbol{\phi}}_k = \begin{pmatrix} \hat{\phi}_{12}(t_k) - \phi_{12}(t_k) \\ \hat{\phi}_{13}(t_k) - \phi_{13}(t_k) \\ \vdots \\ \hat{\phi}_{N-1,N}(t_k) - \phi_{N-1,N}(t_k) \end{pmatrix}. \tag{B.2}$$

Our measurement model relates the measurement vector to the state vector,

$$p(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k) = \mathcal{N}_{\tilde{\boldsymbol{\phi}}_k}(\boldsymbol{\Phi}_k\boldsymbol{\theta}_k, \mathbf{S}_k), \tag{B.3}$$

where $\boldsymbol{\Phi}_k$ is the design matrix mapping station phases to baseline phases at timestamp t_k (Blackburn et al. 2020), and $\mathbf{S}_k$ is a diagonal matrix containing as its nonzero elements the thermal noise variances for each baseline.

B.2. *Necessary conditions for Kalman filtering*

The Kalman filtering algorithm is only applicable to linear Gaussian "state-space" models. In the context of our phase model, this state-space requirement is equivalent to the assumption that the gain phases are a Markov process; i.e., the set of gain phases at time t_k must depend only on the set of gain phases at time t_{k-1}. Because the measurements at each timestamp are independent (i.e., the measurement noise is assumed to be uncorrelated in time), we can write the general posterior conditioned on the observations at $\{t_1, t_2, \ldots, t_T\}$ as

$$p(\boldsymbol{\theta}_{1:T}|\tilde{\boldsymbol{\phi}}_{1:T}) \propto p(\boldsymbol{\theta}_{1:T}) \prod_{k=1}^{T} p_k(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k), \tag{B.4}$$

where we use the notation $p_k(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k) \equiv p(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k)$ to indicate that the likelihood is specific to the measurement at time t_k. The Markovian condition permits us to factorize the prior $p(\boldsymbol{\theta}_{1:T})$ as

$$\begin{aligned} p(\boldsymbol{\theta}_{1:T}) &= p(\boldsymbol{\theta}_T|\boldsymbol{\theta}_{1:T-1})\, p(\boldsymbol{\theta}_{1:T-1}) \\ &= p(\boldsymbol{\theta}_T|\boldsymbol{\theta}_{T-1})\, p(\boldsymbol{\theta}_{1:T-1}) \\ &= \prod_{k=1}^{T} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}), \qquad \textit{(by recursion)}, \end{aligned} \tag{B.5}$$

where we let $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_1)$ for notational simplicity. The posterior can then be written as

$$p(\boldsymbol{\theta}_{1:T}|\tilde{\boldsymbol{\phi}}_{1:T}) = \prod_{k=1}^{T} p(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}). \tag{B.6}$$

With this formulation, one can show (e.g., Särkkä 2013) that the marginal posterior at time t_k can be written as

$$p(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\phi}}_{1:k}) = \frac{p(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\phi}}_{1:k-1})}{\int d\boldsymbol{\theta}_k\, p(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\phi}}_{1:k-1})}. \tag{B.7}$$

This structure enables fast, sequential updates to the marginal posterior that "build up" the full posterior one timestamp at a time.

Not all possible GP kernels admit a state-space formulation, but it can be shown (Rozanov 1977, 1982) that all half-integer Matérn kernels (e.g., Matérn-1/2, Matérn-3/2, etc.) do adhere to this condition.

B.3. *Transition distribution*

A key distribution in Kalman filtering is the "transition distribution" $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})$, which in our case describes how the station gain phases evolve between two consecutive timestamps. Each gain phase θ_i evolves according to a Gaussian process specific to station i. Consequently, the covariance matrix $\boldsymbol{\Sigma}(t_n, t_m)$ for the joint distribution of station gains at times t_n and t_m can be constructed as a block matrix,

$$
\mathbf{\Sigma}(t_n, t_m) = \begin{bmatrix} \begin{pmatrix} \Sigma_1(t_n,t_n) & 0 & \dots & 0 \\ 0 & \Sigma_2(t_n,t_n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_N(t_n,t_n) \end{pmatrix} & \begin{pmatrix} \Sigma_1(t_n,t_m) & 0 & \dots & 0 \\ 0 & \Sigma_2(t_n,t_m) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_N(t_n,t_m) \end{pmatrix} \\ \begin{pmatrix} \Sigma_1(t_m,t_n) & 0 & \dots & 0 \\ 0 & \Sigma_2(t_m,t_n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_N(t_m,t_n) \end{pmatrix} & \begin{pmatrix} \Sigma_1(t_m,t_m) & 0 & \dots & 0 \\ 0 & \Sigma_2(t_m,t_m) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_N(t_m,t_m) \end{pmatrix} \end{bmatrix}
$$

$$
\equiv \begin{bmatrix} \mathbf{\Sigma}_{\boldsymbol{\theta}} & \mathbf{T}_{nm} \\ \mathbf{T}_{nm} & \mathbf{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix}, \tag{B.8}
$$

where in the second line we've introduced some shorthand notation and used the fact that $\Sigma_i(t_m, t_n) = \Sigma_i(t_n, t_m)$ and $\Sigma_i(t_n, t_n) = \Sigma_i(t_m, t_m)$. For the Matérn-1/2 kernel (see Equation 7), the two sub-matrices are diagonal with elements

$$
\{\mathbf{\Sigma}_{\boldsymbol{\theta}}\}_{ii} = \sigma_i^2 \tag{B.9a}
$$

$$
\{\mathbf{T}_{nm}\}_{ii} = \sigma_i^2 \exp\left(-\frac{|t_n - t_m|}{\tau_i}\right). \tag{B.9b}
$$

The joint distribution for the gain phases on two consecutive timestamps t_{k-1} and t_k is then given by

$$
\begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_{k-1} \end{bmatrix} \sim \mathcal{N}\Big(0, \mathbf{\Sigma}(t_k, t_{k-1})\Big), \tag{B.10}
$$

from which the transition distribution $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})$ can be determined as the conditional distribution (see, e.g., Petersen & Pedersen 2006)

$$
p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) = \mathcal{N}_{\boldsymbol{\theta}_k}\left(\mathbf{T}_{k,k-1}\mathbf{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta}_{k-1}, \mathbf{\Sigma}_{\boldsymbol{\theta}} - \mathbf{T}_{k,k-1}\mathbf{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{T}_{k,k-1}\right). \tag{B.11}
$$

This transition distribution can also be derived within the state-space model formulation, as in Hartikainen & Särkkä (2010).

B.4. *Algorithmic implementation*

For our Kalman filter implementation, we follow the algorithmic approach outlined in Särkkä (2013). Here we provide key mappings between the Särkkä (2013) notation (left) and that of our model (right):

$$
\boldsymbol{x}_k \longrightarrow \boldsymbol{\theta}_k \tag{B.12a}
$$

$$
\boldsymbol{y}_k \longrightarrow \tilde{\boldsymbol{\phi}}_k \tag{B.12b}
$$

$$
\mathbf{A}_k \longrightarrow \mathbf{T}_{k+1,k}\mathbf{\Sigma}_{\boldsymbol{\theta}}^{-1} \tag{B.12c}
$$

$$
\mathbf{Q}_k \longrightarrow \mathbf{\Sigma}_{\boldsymbol{\theta}} - \mathbf{T}_{k+1,k}\mathbf{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{T}_{k+1,k} \tag{B.12d}
$$

$$
\mathbf{H}_k \longrightarrow \mathbf{\Phi}_k \tag{B.12e}
$$

$$
\mathbf{R}_k \longrightarrow \mathbf{S}_k \tag{B.12f}
$$

$$
\mathbf{P}_0 \longrightarrow \mathbf{\Sigma}_{\boldsymbol{\theta}}. \tag{B.12g}
$$

Given some set of kernel parameters $\{\sigma_i, \tau_i\}$ describing the Gaussian processes at each of the stations, we initialize the gain phases $\boldsymbol{\theta}_0$ using a mean-zero Gaussian prior:

$$
p(\boldsymbol{\theta}_0) = \mathcal{N}_{\boldsymbol{\theta}_0}(0, \mathbf{\Sigma}_{\boldsymbol{\theta}}), \tag{B.13}
$$

with $\mathbf{\Sigma}_{\boldsymbol{\theta}}$ as defined in Equation B.9a. For the first timestamp ($k = 1$), we perform an initial prediction step:

$$
\boldsymbol{m}_1^- = 0 \tag{B.14a}
$$

$$
\mathbf{P}_1^- = \mathbf{\Sigma}_{\boldsymbol{\theta}}. \tag{B.14b}
$$

This prediction step is followed by the first update step, which computes:

1. the "innovation vector,"
$$\boldsymbol{v}_1 = \boldsymbol{y}_1 - \mathbf{H}_1\mathbf{m}_1^- = \tilde{\boldsymbol{\phi}}_1; \tag{B.15}$$

2. the "innovation covariance,"
$$\mathbf{J}_1 = \mathbf{H}_1\mathbf{P}_1^-\mathbf{H}_1^\top + \mathbf{R}_1 = \mathbf{\Phi}_1\mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top + \mathbf{S}_1; \tag{B.16}$$

3. the "Kalman gain,"
$$\mathbf{K}_1 = \mathbf{P}_1^-\mathbf{H}_1^\top\mathbf{J}_1^{-1} = \mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top(\mathbf{\Phi}_1\mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top + \mathbf{S}_1)^{-1}; \tag{B.17}$$

4. the updated state estimate,
$$\boldsymbol{m}_1 = \boldsymbol{m}_1^- + \mathbf{K}_1\boldsymbol{v}_1 = \mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top(\mathbf{\Phi}_1\mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top + \mathbf{S}_1)^{-1}\tilde{\boldsymbol{\phi}}_1; \tag{B.18}$$

5. and the updated state covariance,
$$\mathbf{P}_1 = \mathbf{P}_1^- - \mathbf{K}_1\mathbf{J}_1\mathbf{K}_1^\top = \mathbf{\Sigma}_{\boldsymbol{\theta}} - \mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top(\mathbf{\Phi}_1\mathbf{\Sigma}_{\boldsymbol{\theta}}\mathbf{\Phi}_1^\top + \mathbf{S}_1)^{-\top}\mathbf{\Phi}_1\mathbf{\Sigma}_{\boldsymbol{\theta}}^\top. \tag{B.19}$$

For subsequent timestamps ($k > 1$), we recursively apply the following steps:

$$\text{Predict:} \qquad \boldsymbol{m}_k^- = \mathbf{A}_{k-1}\boldsymbol{m}_{k-1} \tag{B.20a}$$
$$\mathbf{P}_k^- = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^\top + \mathbf{Q}_{k-1} \tag{B.20b}$$

$$\text{Update:} \qquad \boldsymbol{v}_k = \boldsymbol{y}_k - \mathbf{H}_k\mathbf{m}_k^- \tag{B.21a}$$
$$\mathbf{J}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^\top + \mathbf{R}_k \tag{B.21b}$$
$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^\top\mathbf{J}_k^{-1} \tag{B.21c}$$
$$\boldsymbol{m}_k = \boldsymbol{m}_k^- + \mathbf{K}_k\boldsymbol{v}_k \tag{B.21d}$$
$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k\mathbf{J}_k\mathbf{K}_k^\top \tag{B.21e}$$

This procedure yields the predictive distribution for the measurements at each time t_k given all measurements at $t < t_k$,
$$p(\tilde{\boldsymbol{\phi}}_k|\tilde{\boldsymbol{\phi}}_{1:k-1}) = \mathcal{N}_{\tilde{\boldsymbol{\phi}}_k}(\mathbf{H}_k\mathbf{m}_k^-, \mathbf{J}_k). \tag{B.22}$$
At the completion of filtering ($k = T$), we obtain the marginal likelihood,
$$p(\tilde{\boldsymbol{\phi}}_{1:T}|\boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\phi}) = \prod_{k=1}^{T} p(\tilde{\boldsymbol{\phi}}_k|\tilde{\boldsymbol{\phi}}_{1:k-1}), \tag{B.23}$$
which is then used during parameter space exploration (see Section 2.3).

C. RESILIENCE TO CHOICE OF KERNEL FUNCTION

In this section, we explore the resilience of our model to the choice of kernel function assumed by the Kalman filter.

As detailed in the main text and expanded upon in Appendix B, our implementation of the Kalman filter assumes a Matérn-1/2 kernel. However, we expect that real data will in general not adhere to this assumption. To investigate the effects of this model choice, we generated synthetic datasets using the same approach described in Section 3, but with the underlying kernel changed to Matérn-3/2 and Matérn-5/2. We continue to fit the data in all cases using a Kalman filter that assumes a Matérn-1/2 covariance structure.

For this experiment, we used the same input parameters as in Section 3 for τ, σ, and ϕ.

Because of the mismatch in model specification, we expect that the recovered parameter values will not in general match the input parameter values at the level nominally indicated by the posterior distribution. However, of most interest for downstream use of the calibrated visibility phases is the fidelity with which the underlying astrophysical information is recovered. We thus evaluate fit fidelity via comparison of the input and recovered closure phases ψ_{ijk} on a triangle of baselines connecting stations i, j and k, which are constructed from the visibility phases as

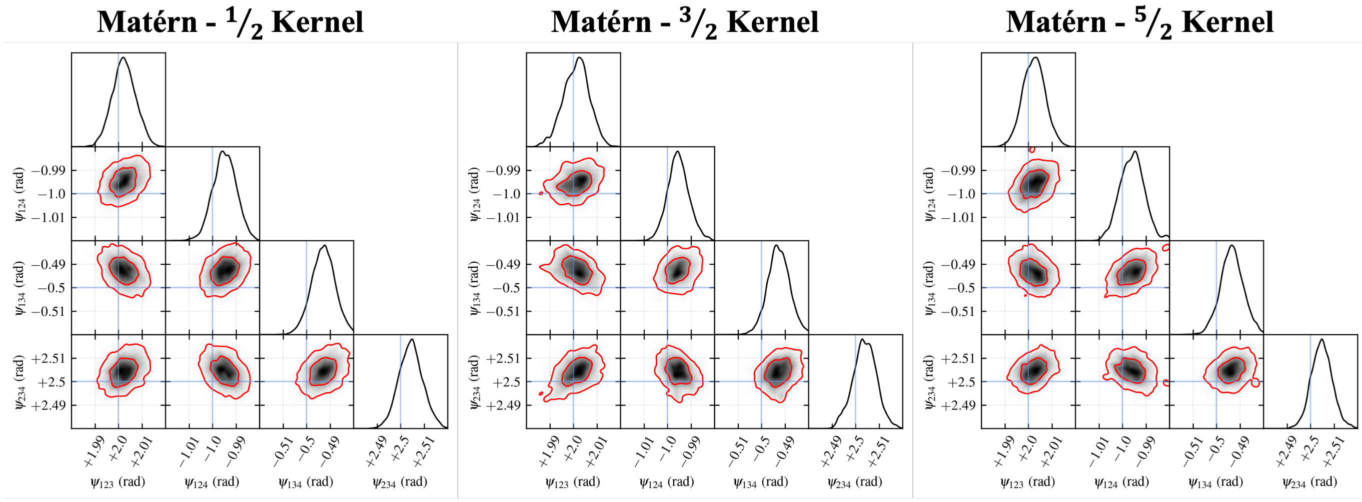

Figure 7. Posterior distributions for the modeled closure phases (ψ), when fitting synthetic data generated using different GP kernels with a model that assumes the same kernel. From left to right, the GP kernel used to generate the synthetic data is Matérn-1/2, Matérn-3/2, and Matérn-5/2. Blue lines indicate the input ψ values used to generate the data.

$$\psi_{ijk} = \phi_{ij} + \phi_{jk} + \phi_{ki}. \tag{C.1}$$

By construction, closure phases are independent of any station gain phases present in the data, and thus represent a "robust" observable quantity.

Figure 7 shows the recovered closure phase posterior distributions for each of our three test datasets. We find that the recovered closure phases show good agreement with the input closure phases, even when the GP kernels are mismatched.