Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Royal
Astronomical
Society

MNRAS **535,** 3181–3197 (2024)
Advance Access publication 2024 November 7

https://doi.org/10.1093/mnras/stae2509

Testing Bayesian inference of GRMHD model parameters from VLBI data

A. I. Yfantis,[1]⋆ S. Zhao,[2] R. Gold,[3,4,5]⋆ M. Mościbrodzka[1] and A. E. Broderick[6,7,8]

[1] *Department of Astrophysics, IMAPP, Radboud University, NL-6500 GL Nijmegen, the Netherlands*
[2] *Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, People's Republic of China*
[3] *Institute for Mathematics and Interdisciplinary Center for Scientific Computing, Heidelberg University, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*
[4] *Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*
[5] *CP3-Origins, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark*
[6] *Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada*
[7] *Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada*
[8] *Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada*

Accepted 2024 November 1. Received 2024 November 1; in original form 2024 May 31

ABSTRACT

Recent observations by the Event Horizon Telescope (EHT) of supermassive black holes M87∗ and Sgr A∗ offer valuable insights into their space–time properties and astrophysical conditions. Utilizing a library of model images (~ 2 million for Sgr A∗) generated from general-relativistic magnetohydrodynamic (GRMHD) simulations, limited and coarse insights on key parameters such as black hole spin, magnetic flux, inclination angle, and electron temperature were gained. The image orientation and black hole mass estimates were obtained via a scoring and an approximate rescaling procedure. Lifting such approximations, probing the space of parameters continuously, and extending the parameter space of theoretical models is both desirable and computationally prohibitive with existing methods. To address this, we introduce a new Bayesian scheme that adaptively explores the parameter space of ray-traced, GRMHD models. The general relativistic radiative transfer code `IPOLE` is integrated with the EHT parameter estimation tool `THEMIS`. The pipeline produces a ray-traced model image from GRMHD data, computes predictions for very long baseline interferometric (VLBI) observables from the image for a specific VLBI array configuration and compares to data, thereby sampling the likelihood surface via a Markov chain Monte Carlo scheme. At this stage we focus on four parameters: accretion rate, electron thermodynamics, inclination, and source position angle. Our scheme faithfully recovers parameters from simulated VLBI data and accommodates time-variability via an inflated error budget. We highlight the impact of intrinsic variability on model fitting approaches. This work facilitates more informed inferences from GRMHD simulations and enables expansion of the model parameter space in a statistically robust and computationally efficient manner.

Key words: accretion, accretion discs – black hole physics – methods: data analysis – methods: statistical – techniques: high angular resolution – quasars: supermassive black holes.

1 INTRODUCTION

The Event Horizon Telescope (EHT) is a millimetre very long baseline interferometric (mm-VLBI) array capable of resolving compact (sizes of $\sim 20\,\mu$as) event horizon scale structures around supermassive black holes in M87 (called M87∗) (Doeleman et al. 2012; Event Horizon Telescope Collaboration 2019a, b, c, d, e, f) and in the Galactic Centre (called Sagittarius A∗, abbreviated Sgr A∗) (Doeleman et al. 2008; Fish et al. 2011; Johnson et al. 2015; Lu et al. 2018; Event Horizon Telescope Collaboration 2022a, b, c, d, e, f).

As recently demonstrated in Event Horizon Telescope Collaboration (2019e, f, 2022e, f), EHT also enables inferences of the astrophysical conditions present in the relativistic environment in the immediate vicinity of a black hole horizon by comparing the VLBI data to theoretical models that predict the on-sky emission map. For this purpose substantial libraries of diverse source models, model comparison techniques, and parameter estimation tools are being built and constantly improved.

In particular, general relativistic magnetohydrodynamics (GRMHD) numerical models of inefficiently radiating accretion flows onto a black hole combined with general relativistic radiative transfer (GRRT) models predict the appearance of the two EHT main targets (Dexter, Agol & Fragile 2009; Mościbrodzka et al. 2009, 2014; Mościbrodzka, Falcke & Shiokawa 2016; Gold et al. 2017; Jiménez-Rosales & Dexter 2018; Chael, Narayan & Johnson 2019). These numerical models of magnetized accretion flows depend on a few key physical parameters such as: (i) the spin of the black hole, (ii) magnetic flux threading the horizon, (iii) the mass accretion rate onto the black hole, (iv) the electron thermodynamics (here simply modelled via $R_{\rm high}$, and (v) the orientation of the system with respect to the observer. Constraining these free parameters of GRMHD simulations via EHT observations can give us quantitative estimates

⋆ E-mail: a.yfantis@astro.ru.nl (AIY); ; roman.gold@iwr.uni-heidelberg.de (RG)

of black hole mass and spin, insights into how gravitational energy is converted into radiation in strong gravity and what mechanism launches the astrophysical jets such as the one observed in M87$*$ (e.g. Hada et al. 2013; Kim et al. 2018).

To constrain physical parameters of M87$*$ and SgrA$*$, Event Horizon Telescope Collaboration (2019e, 2022e) created a *static* library of approximately 60 000 GRMHD model images for M87$*$ and about 1800 000 for Sgr A$*$ and then compared the libraries to the EHT data via various scoring procedures (Event Horizon Telescope Collaboration 2019e, f, 2022e, f). The two main ones are: (a) average image scoring (AIS, used for total intensity data), where snapshots from simulations are turned into synthetic VLBI observations and compared to real data; and (b) snapshot characterization (used mostly for polarimetric images), where a suite of image properties such as, e.g. resolved degree of linear and circular polarizations and their directions are compared to polarimetric characteristics of the reconstructed source images.

In the existing scoring procedure of the EHT to total intensity data (AIS scoring), (a) only the total flux of the image (for a fixed accretion rate), (b) the mass of the black hole, and (c) the position angle of the model are estimated in adaptive fashion over the entire possible range. In EHT terminology this is called snapshot scoring, and creates distributions using all snapshots available (see previous par.). After that, for each model (combination of GRMHD + GRRT parameters), consisting of ~ 500 snapshots an average image is created. This averaged image generates synthetic data that are compared with the real data, given a standard deviation from the spread of the snapshots. Then, given a passing criterion (e.g. cumulative distribution $\in$ [2.5 per cent, 97.5 per cent]) the models pass or fail the AIS test (see Event Horizon Telescope Collaboration 2019e). In this procedure parameters such as inclination angle, electron heating parameter R_{high}, black hole spin, and magnetic flux on the horizon are sampled sparsely in the limited range. Moreover, the two latter parameters are fixed for a given GRMHD simulation and changing them is computationally expensive as it requires running an entire GRMHD simulation.

A key aspect of accretion flows and GRMHD simulations, variability, is particularly challenging for inference pipelines. Variability refers to the inhomogeneity of an accretion flow in a spatial and temporal sense. These two aspects are often intertwined, since a spatial variability (a disc without azimuthal symmetry for example) is magnified by temporal variability, where the directions and particularities of this asymmetrical flow are changing direction and even structure over time.

This means that when comparing EHT data with simulations it is necessary to provide many snapshots of a simulation to test if any of them resembles the source, and even then it will be an approximation. Hence the inference pipelines need to be capable of matching two images that are a priori different. In AIS, this is done by the usage of the aforementioned large model libraries.

In this paper, we propose a new Bayesian parameter estimation procedure by integrating the GRRT code `ipole` (Mościbrodzka & Gammie 2018) with EHT/VLBI data analysis framework THEMIS [Broderick et al. (2020) and Section 2.3 in this paper] and enable the *adaptive* GRRT parameter estimation given an arbitrary GRMHD snapshot. In the improved parameter estimation scheme, the parameters defined in GRRT, e.g. the inclination angle, the accretion rate, and the plasma thermodynamics parameter will be *adaptively* sampled across the entire parameter space to compute the posterior distributions via Bayesian inference. Notice that in this approach the large amount of memory for statically storing the image library is not necessary. We also show that our pipeline could provide a robust framework to account for the variability challenges, see Section 3.4. Additionally, the pipeline is designed to be highly parallelized and extensible, which is the important first step towards the large-scale computation of *adaptive* parameter estimation from GRMHD simulations in the future.

To assess the accuracy and efficiency of our new parameter extraction scheme, we first pick an arbitrary set of GRRT model parameters and generate an image from a GRMHD simulation. Next, we simulate the EHT 2017 observation by assuming the above image has the same celestial position, mass, and distance as Sgr A$*$ (but the procedure can be also adopted for M87$*$) and generate synthetic mm-VLBI data, including visibility amplitudes and closure phases. Since, for our fitting routine we decide to use only the closure phases, the interstellar media scattering effect is not considered at this primary step (which is important and complicated for fitting archive Sgr A$*$ data, see e.g. Johnson et al. 2018; Issaoun et al. 2019, 2021). Additionally, we apply standard thermal noise (Chael et al. 2016, 2018) and systematic errors (1 per cent, 10 per cent, or 30 per cent depending on the case). Then we use a Markov chain Monte Carlo (MCMC) algorithm to sample the posteriors of the parameters from the underlying unknown distribution of all the physically possible models by comparing the synthetic data with the GRMHD + GRRT model. We perform well-controlled tests with a known 'truth' value first by fitting two parameters and then extend it to fit four parameters simultaneously. Such an incremental approach provides clarity when interpreting the results.

In the future, the pipeline can be improved, for instance, with more realistic observational corruptions of model images (e.g. Blecher et al. 2017; Janssen et al.2019) or polarimetric models (Event Horizon Telescope Collaboration 2021b).`THEMIS` as well as `IPOLE` can handle different observing frequencies. Therefore, the pipeline presented here can naturally handle model fitting to upcoming EHT, ngEHT, and non-EHT data sets, e.g. to longer wavelengths VLBI observations of the EHT targets or AGN sources (Kim et al. 2018; Issaoun et al. 2019; EHT MWL Science Working Group 2021).

The paper is organized as follows. In Section 2, we describe the pipeline which produces the GRMHD + GRRT models of Sgr A$*$ (or M87$*$) at millimetre waves (Section 2.1), the process of generating synthetic mm-VLBI observation data sets from the image (Section 2.2), the sampling methods (Section 2.3). In Section 3, we use the synthetic data (generated in Section 2.2) to test the adaptive parameter estimation pipeline by two parameter fitting (Section 3.1) as well as multiparameter fitting (Section 3.2). In Section 3.3 we introduce time variability in the fitting algorithm, and in 3.4 we propose a few ways to account for it. We summarize the results and conclude in Section 4.

2 PARAMETER ESTIMATION PIPELINE: DESCRIPTION

2.1 Physical model and model parameters

Extracting physical parameters from EHT observations requires a model for the accretion flow onto a compact object and a model for the arising emission.

Our model describes an accreting black hole within ideal-GRMHD simulation and is therefore intrinsically dynamical. The simulation starts with a torus of plasma in Keplerian, equatorial orbit around a Kerr black hole (Fishbone & Moncrief 1976) that would be in hydrostatic equilibrium in absence of magnetic fields. The torus is then seeded with weak magnetic fields and the evolution of such configuration is computed by solving the equations of ideal general relativistic magnetohydrodynamics (Gammie, McKinney & Tóth 2003). Simulations show that magnetic turbulence is developed, which acts as an effective source of viscosity (Hawley & Balbus 1991; Balbus & Hawley 1998), thereby causing matter to accrete onto the black hole. In this process some material escapes the system in the form of winds and jets that may or may not be visible in the model image depending on radiation properties and electron thermodynamics.

The free parameters of the GRMHD-informed model images can be divided into two categories: numerical parameters and physical parameters. The numerical parameters are e.g. numerical grid size and resolution of the simulation. Physical parameters include: black hole spin parameter $a_* \in (0, 1)$[1] or the normalized magnetic flux threading the horizon in the relaxed, steady state, which is the fundamental parameter describing the state of the magnetic field along the standard and normal evolution (SANE) and magnetically arrested disc (MAD) regimes (Tchekhovskoy, Narayan & McKinney 2011; Porth et al. 2019). In this paper, we utilize time-slices of an existing 3D GRMHD SANE simulation of Shiokawa (2013) (applied to model Sgr A∗ in Mościbrodzka et al. 2014), with the black hole spin parameter $a_* = 0.9375$ and the adiabatic index of 13/9. We pick snapshots where the turbulence is fully developed and the simulation exhibits roughly a steady-state behaviour in which at least the interior regions have not retained their initial conditions.

Next, we generate synthetic images (intensity maps) from this simulation snapshot using the ray-tracing and radiative transfer code IPOLE[2] (Mościbrodzka & Gammie 2018), which was tested against other radiative transfer codes used in the original IPOLE paper as well as in Gold et al. (2020) and Prather et al. (2023). The fast-light approach is used throughout this work for both synthetic data and Bayesian runs. The main radiative processes considered in computing images from the GRMHD simulation is synchrotron emission and synchrotron self-absorption. Generating images of a particular astrophysical source requires rescaling the dimension-less GRMHD simulations from geometrized unit system ($G = M = c = 1$) to c.g.s. units. The scaling requires providing the mass of the central black hole which will also set the length scale of the system according to $GM_{\rm BH}/c^2$ [cm] and time scale units $GM_{\rm BH}/c^3$ [s]. The scaling also requires providing the mass unit parameter $M_{\rm unit}$ that scales the density of the plasma around the black hole, i.e. the density of the matter in the accretion flow is $\rho_{\rm c.g.s.} = \rho_{\rm code} M_{\rm unit}/\mathcal{L}^3$ (Notice that $M_{\rm unit}$ also scales the accretion rate onto the black hole and strength of magnetic field at the same time, see Mościbrodzka et al. 2009 for details). It is therefore in principle necessary to redo the ray-tracing/radiative transfer computation whenever these parameters are varied, which is what we pursue here in contrast to an approximate scaling approximation designed to avoid the additional computational cost. The free parameters used to model the differences between electron and proton temperatures in various regions of different magnetization are $R_{\rm low}$ and $R_{\rm high}$ (motivated by Mościbrodzka et al. 2016 and Ressler et al. 2015, see also Event Horizon Telescope Collaboration 2019e). Specifically, the proton to electron temperature ratio reads:

$$\frac{T_p}{T_e} = R_{\rm low}\frac{1}{1+\beta^2} + R_{\rm high}\frac{\beta^2}{1+\beta^2}. \tag{1}$$

Given $T_{\rm p}$, which is equal to the gas temperature in the GRMHD simulation, as was done in Event Horizon Telescope Collaboration (2019e, 2022e), we can compute $T_{\rm e}$ using equation (1) and the synchrotron emissivities thus depend on the assumed $R_{\rm high}$, $R_{\rm low}$ parameters. As $R_{\rm high}$ and $R_{\rm low}$ are not two independent parameters, $R_{\rm low}$ is usually fixed to be a constant ($R_{\rm low} = 1$ in Event Horizon Telescope Collaboration 2019e). The inclination angle, i, and position angle, PA (orientation of the image on the sky with respect to the celestial North pole), are two remaining parameters of the model describing the geometrical orientation of the system with respect to the observer's line of sight. Finally, the distance of the observer to the source has to be assumed.

For current models only black hole spin and magnetic flux require an independent GRMHD simulation. This leads to higher computational efficiencies especially for higher dimensions and the ability to estimate the posterior distribution for each parameter including black hole mass and spin. We also drop the assumption about optical depth made in Event Horizon Telescope Collaboration (2019e, 2022e) where a crude flux rescaling was applied, treating BH mass and total flux as scale-free parameters (to within a limited range). Such a scaling can at most be valid in a finite range and more specifically for matter that is sufficiently optically thin. Our method reperforms the ray-tracing and radiative transfer on every likelihood evaluation and hence drops these assumptions. This is in itself a significant improvement over the current method.

In this work we focus on parameter estimation by scaling the dimension-less GRMHD simulations to Sgr A∗ system but other black hole masses can be assumed. We therefore fix the mass of the black hole to $M_{\rm BH,Sgr\ A*} = 4.1 \times 10^6\,{\rm M}_\odot$ and distance $D_{\rm SgrA*} = 8.5$ kpc (Gravity Collaboration 2018). Other model parameters i, PA, $M_{\rm unit}$, $R_{\rm low}$, $R_{\rm high}$ are allowed to float. This list can easily be generalized to include any parameter in the ray-tracing and radiative transfer code used.

In Fig. 1 (left-most upper panel) we show an example of an arbitrarily chosen sets of parameters of appearance of the 3D GRMHD simulation scaled to Sgr A∗ system parameters as seen by an observer on Earth. The model image is generated at an observational wavelength of $\lambda = 1.3$ mm ($\nu = 230$ GHz) at which EHT currently operates (in the future EHT will also operate at 0.87 mm/345 GHz).

2.2 VLBI data products and synthetic data generation

EHT is an interferometer which detects the sparsely sampled Fourier components of the image of the source on the sky, called visibilities. The visibility $V(u, v)$ is the 2D Fourier transformed complex function of intensity distribution $I(x, y)$ defined by e.g. Thompson, Moran & Swenson (2017) as:

$$V(u, v) = \iint I(x, y) e^{-2\pi i(ux+vy)} {\rm d}x{\rm d}y. \tag{2}$$

The visibility is by definition a complex function with amplitude A and phase ϕ: $V(u, v) = Ae^{-i\phi}$. In Fig. 1 (middle and right-most upper panels) we show the visibility amplitude and phase computed based on the GRMHD model image.

The visibility amplitudes are subject of a future work, so they are not discussed here. In the present we utilize closure phases, the sum of the complex visibility phases along a closed triangle baseline, which is:

$$\Phi_{i,j,k} = \arg(V_{ij} V_{jk} V_{ki}), \tag{3}$$

[1] The black hole spin is usually given in dimension-less units where $a_* = 0$ describes Schwarzschild black hole and $a_* = 1$ is maximally rotating Kerr black hole

[2] https://github.com/moscibrodzka/ipole

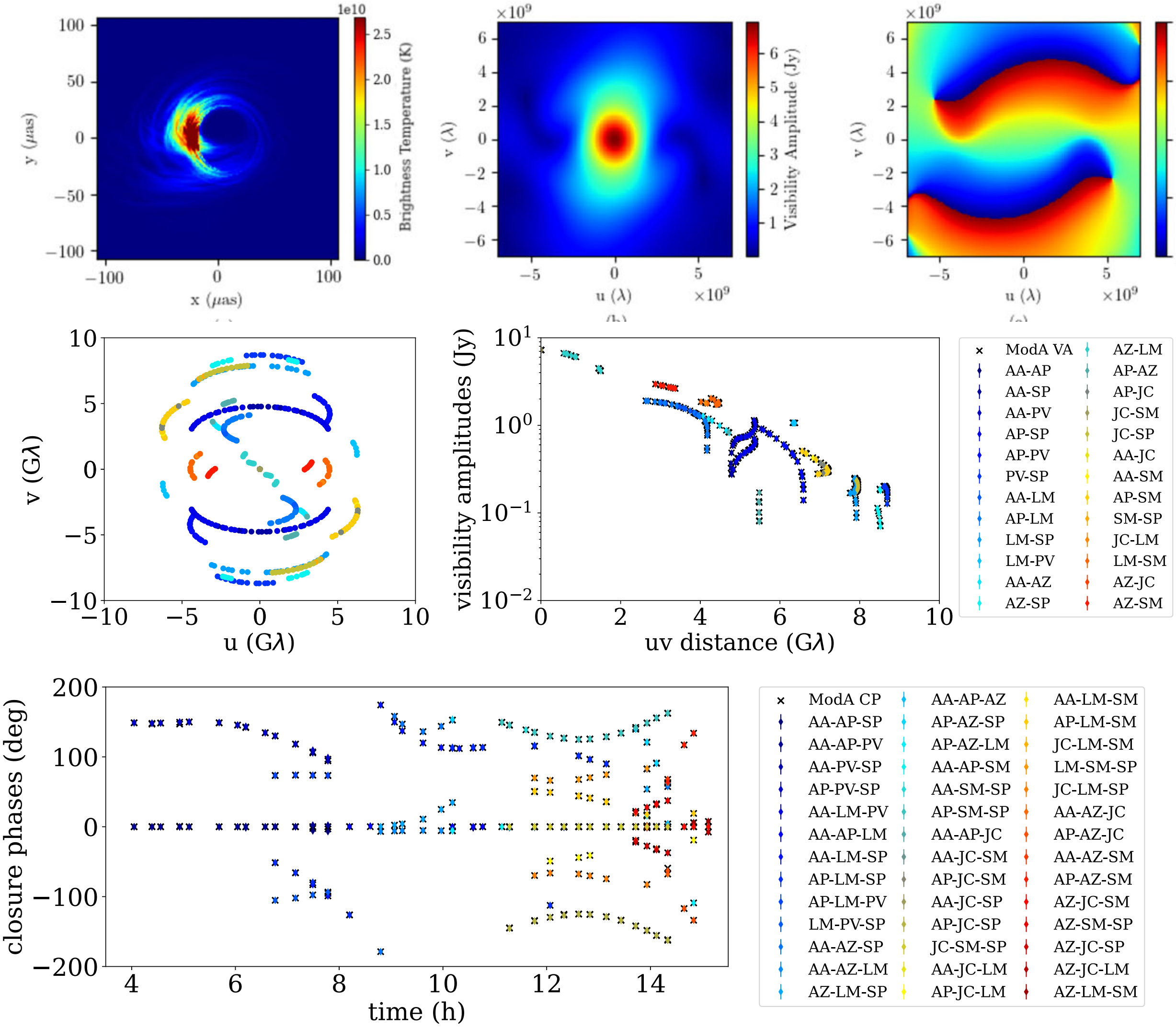

Figure 1. Top panels: 230 GHz image of GRMHD model of Sgr A∗. Physical parameters used to generate image are: $M_{\mathrm{unit}} = 3 \times 10^{18}$ gram, $R_{\mathrm{high}} = 3$, $R_{\mathrm{low}} = 3$, $i = 60°$, PA = 0°. Here, the model image has high resolution of 256 × 256 pixels. The colour shows the emission intensity (Stokes $\mathcal{I}$). The middle and right panels show the amplitudes and phases of the complex visibility function which is generated by the Fourier transformation (FT) of the model image. Bottom panels: Synthetic EHT 2017 data generated using `EHT-IMAGING` and `THEMIS`. The middle left panel shows the baseline (u, v) coverage of the observation and the middle right panels displays the visibility amplitudes as a function of (u,v) distance. The colours code different baselines. The bottom panel shows the synthetic closure phases. The different colours refer to different EHT station triangles. Black crosses denote the data modelled within `THEMIS` which are in excellent agreement with those from `EHT-IMAGING`.

where V_{ij}, V_{jk}, and V_{ki} are the visibility of baseline *ij*, *jk*, and *ki*. Due to degeneracy amongst possible triangles, an array with N antennas has $(N-1)(N-2)/2$*independent* closure phases (Thompson et al. 2017, Blackburn et al. 2020). For the 2017 EHT observations $N = 8$ which gives 21 independent closure phases but in general one can form up to 56 closure phase triangles $(N(N-1)(N-2)/3!)$ assuming that the source is visible at all sites during an observing window. In practice, using 2017 EHT (u, v) coverage, we generate 41 closure phases. The simulated observation is roughly 11 h long. The main advantage of the closure quantities is that they are insensitive to station-based errors (Chael et al. 2018).

In Fig. 1 (middle and bottom panels) we show the (u, v) coverage, synthetic visibility amplitudes and closure phases. Our example of synthetic VLBI data is generated based on GRMHD image assuming the following parameters: $i = 60°$, $M_{\mathrm{unit}} = 3 \times 10^{18}$ gram, $R_{\mathrm{high}} = 3$, $R_{\mathrm{low}} = 3$, and position angle PA = 0°. For the following M_{unit} will always be in (gram). The image has 128×128 pixels (see Appendix A for discussion of image resolution), and we assume that the source is on the celestial sphere where the ascension and the declination are same as Sgr A∗. The synthetic VLBI data are generated using the `EHT-IMAGING` library[3] (Chael et al. 2016, 2018). We simulate the EHT observation with EHT 2017 array configuration to observe Sgr A∗ with the baseline (u, v) coverage matching the EHT observation on 2017 April 7 (Event Horizon Telescope Collaboration 2022b). The centre of the observational frequency band is 229.1 GHz and the bandwidth is 1.8 GHz.

Our synthetic EHT data are time-averaged along (u, v) tracks into 10 min scans. Furthermore, the data have been treated to account for typical noises, such as thermal and systematic noises usually

[3] https://github.com/achael/eht-imaging

considered when analysing EHT data (see e.g. Event Horizon Telescope Collaboration 2019d, 2024a, b). The thermal noise (tn) is to account for fluctuations in the telescope during 'observation', while the systematic errors (syser), set at 1 per cent, 10 per cent, or 30 per cent, capture additional uncertainties from the instrument.

When fitting model to synthetic or real EHT data, the calculation of visibility amplitudes and closure phases from the given model image is the nested part of EHT data analyse framework THEMIS.[4]THEMIS is a massively parallel, modular, flexible, and extensible framework, containing all the utilities necessary to compare EHT data to a variety of model predictions for these data sets, including visibility amplitudes, closure phases and more. The FFTW3[5] is the fast Fourier transformation (FFT) library used to transform the intensity distribution on the celestial plane to the visibility of each uv data sets, which are read from input data files. In Fig. 1 we also show that the visibility amplitudes and the closure phases calculated with THEMIS match perfectly those produced by EHT-IMAGING library.

2.3 Model fitting

2.3.1 Likelihood and priors

THEMIS carries out Bayesian parameter estimation via MCMC sampling the log-likelihood. In the Bayesian statistics, if given the prior probability distribution of parameters to estimate, the posterior probability distribution is constrained by the likelihood function. The likelihood and the prior are defined by the user. The log-likelihood of closure phases is

$$\mathcal{L}(\vec{p}) = -\sum_j \frac{\Delta^2\left(\Phi_j - \hat{\Phi}_j(\vec{p})\right)}{2\sigma_j^2}, \tag{4}$$

where $\vec{p}$ is the vector of parameters to estimate, Φ_j and $\hat{\Phi}_j(\vec{p})$ are the observed and modelled closure phases, $\Delta(x)$ is the angular difference in the range $[-180^\circ, 180^\circ)$, and $\sigma_j = \sqrt{\sigma^2_{\mathrm{tn},j} + \sigma^2_{\mathrm{sys}} * \hat{\Phi}_j(\vec{p})^2}$. A link to the more traditional approach is the relation $\mathcal{L} = -\chi^2/2$, from where it follows that

$$\chi^2(\vec{p}) = \sum_j \frac{\Delta^2\left(\Phi_j - \hat{\Phi}_j(\vec{p})\right)}{\sigma_j^2} \;;\; \chi^2_{\mathrm{eff}} = \frac{\chi^2}{n_d - n_f}, \tag{5}$$

where $n_d = 358$ is the number of data points, affected by the data character, visibility amplitudes (VA), or closure phases (CP), and the observation specifics, number of telescopes, time, etc. The number of freedom is $n_f = 2$ for two-parameter fit and 4 for the multiparameter fit, equal to the number of parameters being fitted simultaneously.

In this likelihood definition, we assume that the errors in the closure phases have Gaussian distribution. However when signal-to-noise ratio (SNR) is low, the error distribution is more likely to be non-Gaussian (Thompson et al. 2017). How the error distribution and SNR affect the fitting accuracy is discussed in Broderick et al. (2020). Another problem of this likelihood is that it is unable to treat fitting multi-epoch data. Both, the non-Gaussian errors and the multi-epoch observation fitting are beyond the scope of this work.

Similarly to the phases, the log-likelihood for visibility amplitudes can be calculated as

$$\mathcal{L}(\vec{p}) = -\sum_j \frac{\left[|V|_j - |\hat{V}|_j(\vec{p})\right]^2}{2\sigma_j^2}. \tag{6}$$

We adopt prior for each parameter separately. Due to the lack of the knowledge of the true model parameters and for keeping the approach as agnostic as possible, we adopt flat prior for all parameters. It is worth mentioning that flat prior is not equivalent to non-informative prior, but it is sufficient for this fit.

2.3.2 Sampling parameters

MCMC methods are frequently used to sample the posterior from prior with defined likelihood. In order to efficiently sample the underlying parameter space and to faithfully infer model parameters. Special care is needed on top of standard MCMC to avoid trapping in local extrema. We adopt a (parallel) tempering technique (Swendsen & Wang 1986; Geyer 1991), in which high-tempered chains explore large regions in the parameter space with low precision while the low-tempered chains focus on small regions with high precision. The different tempered chains communicate and exchange their position information, which let the coldest chains (most accurate and used to be the final output) escape any local optimums. The scheme has been demonstrated to explore a variety of likelihood surfaces including multimodal distributions and is well described in Broderick et al. (2020). The sampler of choice for this project was the affine invariant (AI) method described in detail in Goodman & Weare (2010). More possibilities and reasoning behind this choice can be found in Appendix B.

Within THEMIS a variety of further options can be chosen, from VLBI data types, to sampling methods, number of walkers, temperatures, steps between communication of the walkers, number of processors per likelihood. Our final choice consists of fitting closure phase data (CP), with the addition of thermal noise and systematic errors (1 per cent, 10 per cent, 30 per cent), the affine invariant sampler, and the number of walkers and temperatures: $N_{\mathrm{W}} = 8$, $N_{\mathrm{T}} = 10$ for the two parameter fitting and $N_{\mathrm{W}} = 10$, $N_{\mathrm{T}} = 12$ for the multiparameter run (see the next section). We have calculated the effective sampling size for our two models (A and B, see Section 3) and we found minimum values of ~ 40 and ~ 67, respectively. We have not apply any thinning, i.e. retain only a subset of samples from the MCMC chains.

A summary of the physical and numerical parameters used in the calculations can be found in Table 1.

3 PARAMETER ESTIMATION PIPELINE: VALIDATION USING SYNTHETIC VLBI DATA

In the following, first we perform two parameter fitting, using fitting parameter pairs of PA together with one of the other parameters (while keeping the remaining two fixed), the details of which are presented in Section 3.1. Then we raise both the number of N_{T} and N_{W} by 2 and perform an all parameter fitting simultaneously, which is presented in Section 3.2 and finally we discuss variability of the source and the data and how to tackle it in Sections 3.3 and 3.4

3.1 Single snapshot, two parameter fitting

In the first test, model A (shown in the upper panels in Fig. 1) is used to generate the truth synthetic data via simulating the EHT 2017 observation.

[4] https://github.com/PerimeterInstitute/Themis

[5] FFTW is a publically available C subroutine library for computing the discrete Fourier transform in one or more dimensions, of arbitrary input size, and of both real and complex data. http://www.fftw.org/

Table 1. Physical and numerical model parameters explored in this work.

	Physical model parameters	
Name	Value/Range	Description
a_*	0.9375	Black hole spin parameter of a give GRMHD simulation snapshot; Fixed in this occasion.
i	(0, 180°)	Viewing angle (inclination) of the observer: $i = 0°$ is face-on, $i = 90°$ is edge-on.
PA	$[-\pi, \pi]$	Position angle of the black hole spin on the sky, measured east of north.
$M_{\rm unit}$	$[10^{17}, 6 \times 10^{19}]$	Mass unit parameter scales the density of the plasma, hence $\mathcal{M}_{\rm unit} \sim \dot{M}(g/s)$.
$R_{\rm high}$	[1,90]	Describes coupling of $T_{\rm e}$ w/ $T_{\rm p}$ in regions of weak magnetization (high plasma β region).
$R_{\rm low}$	3	Describes coupling of $T_{\rm e}$ w/ $T_{\rm p}$ in regions of strong magnetization (low plasma β region).
	Fitting numerical parameters	
	Two parameter fit	
Name	Value	Description
image res.	128^2	The size of model image, unit in pixels.
$N_{\rm W}$	8	The number of chains in MCMC sampler.
$N_{\rm T}$	10	The temperature level of the parallel tempering in MCMC sampler.
comm. freq.	2	The interval MCMC steps between communication events of different tempered chains.
burn-in	500	The number of initial MCMC steps removed from chains when building posterior.
end step	5000	The final MCMC step of the simulation
	Four parameter fit	
image res.	128^2	The size of model image, unit in pixels.
$N_{\rm W}$	10	The number of chains in MCMC sampler.
$N_{\rm T}$	12	The temperature level of the parallel tempering in MCMC sampler.
comm. freq.	2	The interval MCMC steps between communication events of different tempered chains.
burn-in	1000–3500	The number of initial MCMC steps removed from chains when building posterior.
end step	10 000	The final MCMC step of the simulation

The first pipeline test validates the scheme while fitting two parameters only: PA in combination with one of the three remaining parameters i, $M_{\rm unit}$ or $R_{\rm high}$. PA sampling alone is done more efficiently through an analytically marginalized likelihood however sampling i, $M_{\rm unit}$, and $R_{\rm high}$ requires repeating the GRRT simulation in every step. To initialize the MCMC chains we have chosen values in the middle of the range that we wish to explore (Table 1), using flat priors for all parameters.

The test results are summarized in Fig. 2 where we show the posterior probability distribution (PD) for all parameters and the trace plot and the log-likelihood for $M_{\rm unit}$ as an example. For all runs $\chi^2_{\rm eff} \leq 1$, as shown in Table 2. We achieve effective sampling sizes larger than 600 for both parameters and a median split $\hat{R} = (0.9996, 1.0017)$ in the two parameter fit.

The posterior densities and the Gaussian fits of them have been made with a burn-in of 500 MCMC steps. For all parameters except $R_{\rm high}$ the truth values are covered by the posterior (at 3 σ in the case of inclination). By contrast, the PD of $R_{\rm high}$ has a shifted peak revealing a smaller than 1 per cent bias (from the median value) and misses the truth value altogether.

To investigate $R_{\rm high}$ offset we carry out several tests with three improvements, such as (i) adding more stations in the telescope array (from 2021, 2022), (ii) using CP and VA data, (iii) using a cut in the data where we only use CP points with S/N above 4, and lastly (iv) inflating the systematic errors at 10 per cent. The posterior can be seen in Fig. 3. Improvements (ii), (iii), and (iv) all resolve the parameter offset problem separately. This points to a bias caused by the closure phases, which is fixed either by the SNR cut or its effects are weakened when also VA data are included. The offset problem is also resolved when the systematic errors are inflated and the posterior covers a wider range of values.

Two parameter fitting tests demonstrate that it is possible to fit other parameters apart from PA, in an adaptive way, and that they converge to the truth values in a fast and stable fashion. The $\chi^2_{\rm eff}$ values for all runs below are close to 1 which further strengthen these claims. The two parameter fits for $R_{\rm high}$ or i already have advantage over the standard EHT image library which samples these variables rather sparsely (usually $R_{\rm high} = 1, 10, 40, 160$, and $i = 10, 30, 50...$).

3.2 Single snapshot, four parameter fitting

Next, we simultaneously sample four parameters (PA, i, $M_{\rm unit}$, and $R_{\rm high}$). We still fit only the CP data, using the same initial values and ranges of the parameters as in the previous two parameter fitting tests.

Fig. 4 shows a triangle plot of the posteriors and the joint probability densities of parameters given two different truths, models A and B. The burn-in window has been set to 3500 MCMC steps. Our pipeline recovers both, significantly different, truth parameters. As visible in Table 3, the numerical parameters of the four parameter fits together with 10 000 MCMC steps result in parameter estimation accuracy and precision comparable to those of the two parameter fits.

3.3 Effects of time variability on the parameter estimation

Here we examine the behaviour of the pipeline when taking into account a possibility that the source may be changing in time (which is certainly true for Sgr A∗ over a single night and for M87∗ over time-scales of a week). In fact, our model image does not change with time yet (although this can be naturally incorporated in the future). Instead we examine the effects if the realization of variability in the model is different from the one in the data. We do so by fitting synthetic data from a model in a certain point in the simulation to synthetic EHT data created using a different time moment of the same simulation. model B + 100 M, has been created from the same GRMHD simulation and the same radiative transfer parameters, but 100 GM/c^3 later than model B, and similarly model B-500 M, 500 GM/c^3 earlier than model B. Note that GRMHD snapshots are known to become sufficiently uncorrelated when separated by 20 −

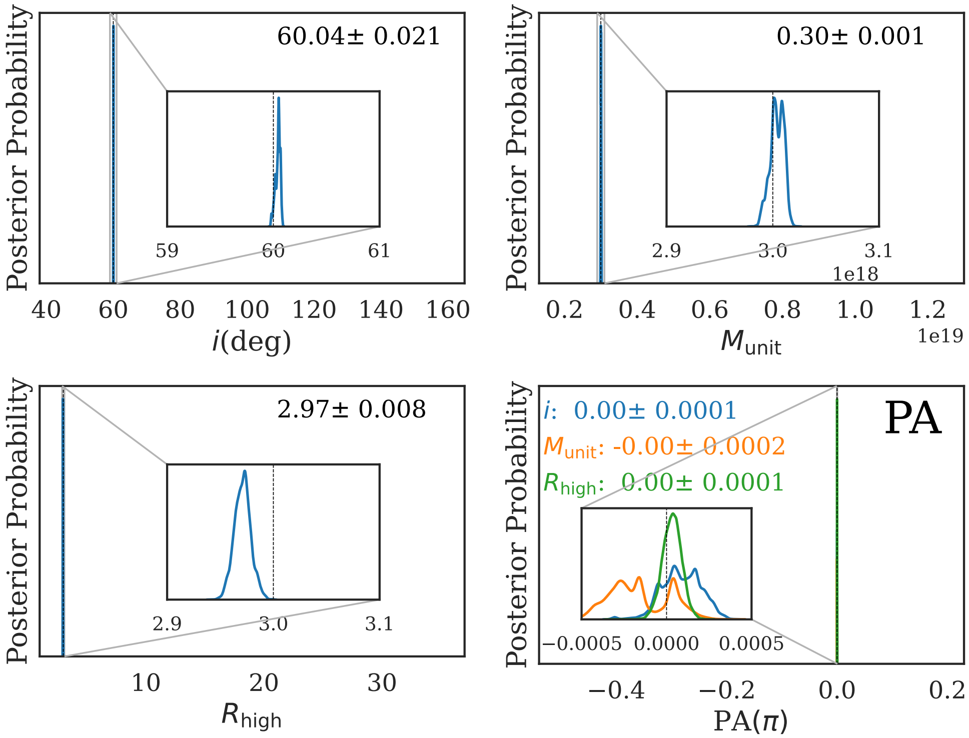

Figure 2. Results of the two parameters fits, showing the posterior probability distribution for all parameters.

Table 2. Truth parameters and the pipeline performance for two parameter fitting, for a burn-in step of 500 MCMC. For each listed run, the PA and one other parameter is varied.

	Truth	Estimated	$\chi^2_{\rm eff}$
PA(π)	0	$0 + O^{-4}$	–
$i(^{\circ})$	60	60.04 ± 0.021	0.66
$M_{\rm unit}(10^{18})$	3	3 ± 0.001	0.69
$R_{\rm high}$	3	2.97 ± 0.008	0.89

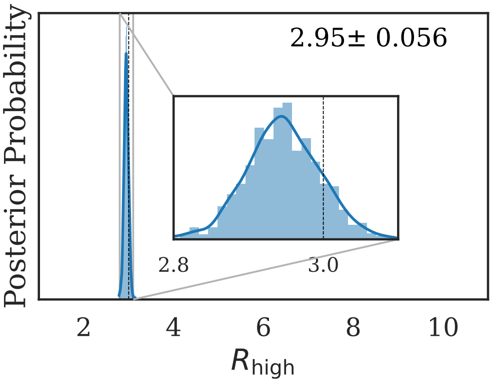

Figure 3. The posterior probability distribution for two parameters ($R_{\rm high}$ and PA) fits when assuming systematic errors of 10 per cent. Since we already have information about the truth, this test is limited to 2000 MCMC steps and a narrower prior for $R_{\rm high} \in (1, 10)$. When fitting real VLBI data we would use a wider prior and run the pipeline longer.

30 M and so the adopted time offsets can be considered significant (Georgiev et al. 2022). The snapshots of the GRMHD model for the same radiative transfer parameters at different time moments are shown in Fig. 5. We still use model B for synthetic data, and B + 100 M or B − 500 M as a template for fitting. Our goal is to assess: (i) how poor the fit quality gets for a given error budget and (ii) how large the bias can be.

Fig. 6 shows parameter estimation for both snapshots. Regarding B + 100 M, from the triangle plot it is clear that 2 parameters (i, $M_{\rm unit}$) have distributions with peaks shifted from the truth (7 per cent and 23 per cent, or 87σ and 6σ away from the truth, respectively), but in a coherent manner in a sense that the more distant snapshot is further away from the truth. Notice that in some cases the 3σ contours are on the edge of the truth (intersection of dashed lines). As for the two remaining parameters ($R_{\rm high}$ and PA), the distributions are able to cover the truth, perhaps a coincidence or an effect of time correlation given the poor fit quality (as expected in absence of inflated error budgets); in this test the $\chi^2_{\rm eff} = 150.4$.

In the same figure the orange line shows the parameter estimation for snapshot B − 500 M. It is evident that in this test the fit is unable to cover the truth in all four parameters, including PA. The log-likelihood is larger compared to that when fitting model B + 100 M, but $\chi^2_{\rm eff} = 181$ stays roughly at the same level.

The two tests above illustrate that on top of the expected poor-fit quality a large bias is typically introduced into parameter estimation due to the intrinsic source variability. Table 4 collects the best-fitting parameters for the two runs. In addition, Fig. 5 shows the best-fitting model images for the two tests above. The B + 100 M and B − 500 M best-fitting images look somewhat different compared to model B (shown in the left panel) but overall crescent shape of the emission region is preserved.

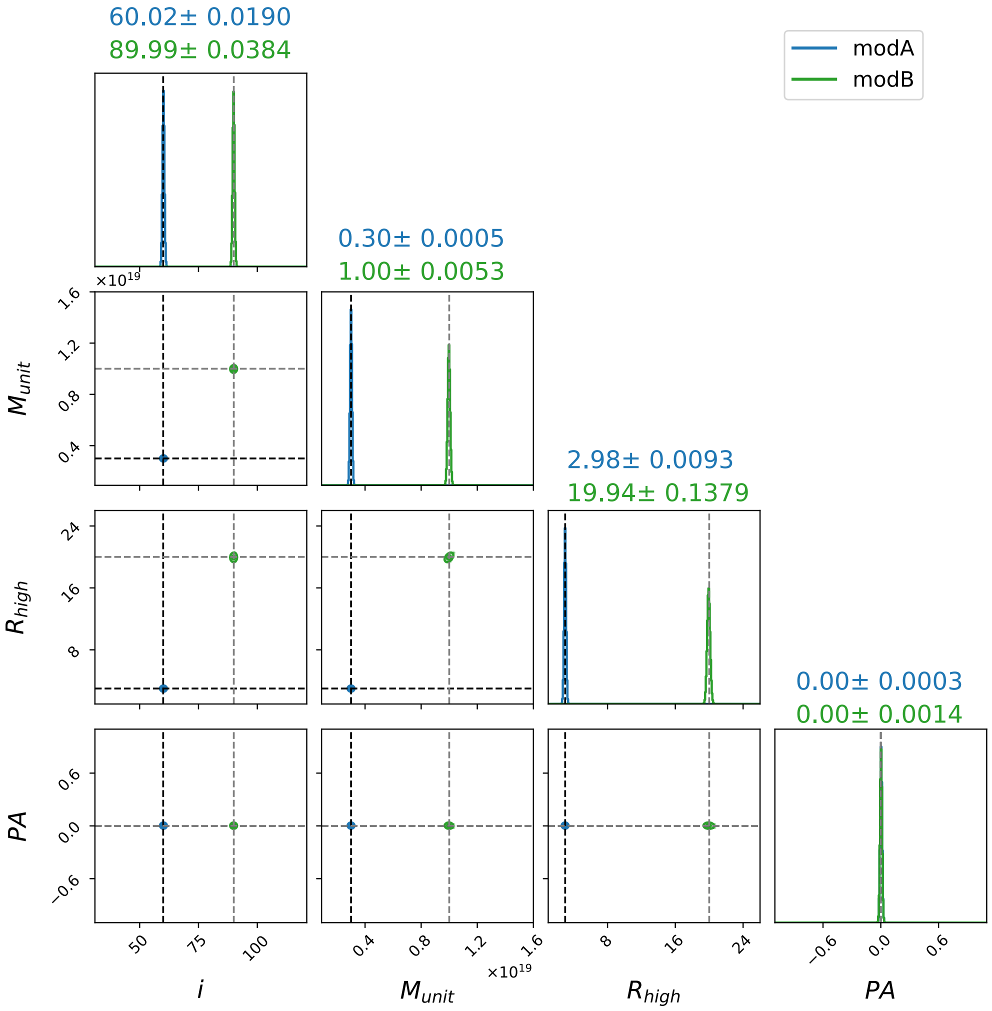

Figure 4. The triangle plot of four parameters estimation by fitting with the synthetic data based on models A and B. The main diagonal shows marginalized posterior distributions of all four parameters. The six plots in the lower left triangle show the joint densities for all the parameter combinations. The contours have been set to represent confidence of 0.68, and 0.9973 (1 and 3σ). Dashed lines denote the truth (black for model A, grey for B). The lowest χ^2_{eff} was 0.89 for both models.

Table 3. Truth parameters and the pipeline performance for four parameter fitting with 1 per cent systematic error added to the simulated data, for a burn-in step of 3500 MCMC. All parameters are sampled simultaneously, so there is only one χ^2_{eff} value.

	Truth	Estimated	χ^2_{eff}
		Model A	
PA(π)	0	0 ± 0.0003	0.88
i(°)	60	60.02±0.02	0.88
$M_{\mathrm{unit}}(10^{18})$	3	3 ± 0.0005	0.88
R_{high}	3	2.98 ± 0.01	0.88
		Model B	
PA(π)	0	0 ± 0.0014	0.82
i(°)	90	89.99 ±0.0378	0.82
$M_{\mathrm{unit}}(10^{19})$	1	1 ± 0.0052	0.82
R_{high}	20	19.93 ± 0.1345	0.82

3.4 Tackling variability: inflated errorbars and snapshot averaging

In this section, we carry out two additional tests which may be useful when designing strategies on how to tackle the variability issues.

The first obvious step to address variability impact on parameter estimation is to simply inflate the systematic errors of the data points to enable the analysis to be more permissive and fitting to be easier. Fig. 7 demonstrates model B fitting using snapshot B − 500M with three different choices of systematic error, namely: 1 per cent, 10 per cent, and 30 per cent. Already with systematic errors of 10 per cent, the fitting of snapshots separated by $\Delta t = 500\,\mathrm{M}$ recovers the truth. This test validates that the pipeline can find the truth even with snapshot misspecification, a necessary feature for real data fitting, where most certainly all our models will be (at best) only approximations to the real image. At 30 per cent errors, the posterior distributions of all free parameters widen (as expected), notice that $\chi^2_{\mathrm{eff}} = 179.5, 0.71, 0.41$ for increasing error budgets, while still covering the truths at 1σ confidence. For this specific example, the 30 per cent and arguably 10 per cent cases are slightly overfitted, presumably due to an overestimated systematic error to capture the intrinsic variability. Note that for general and realistic cases this may be slightly different, but this needs to be investigated more thoroughly in a dedicated future work. The produced Themis-fit images of model B-500M, for different error budgets are visible in Fig. 8.

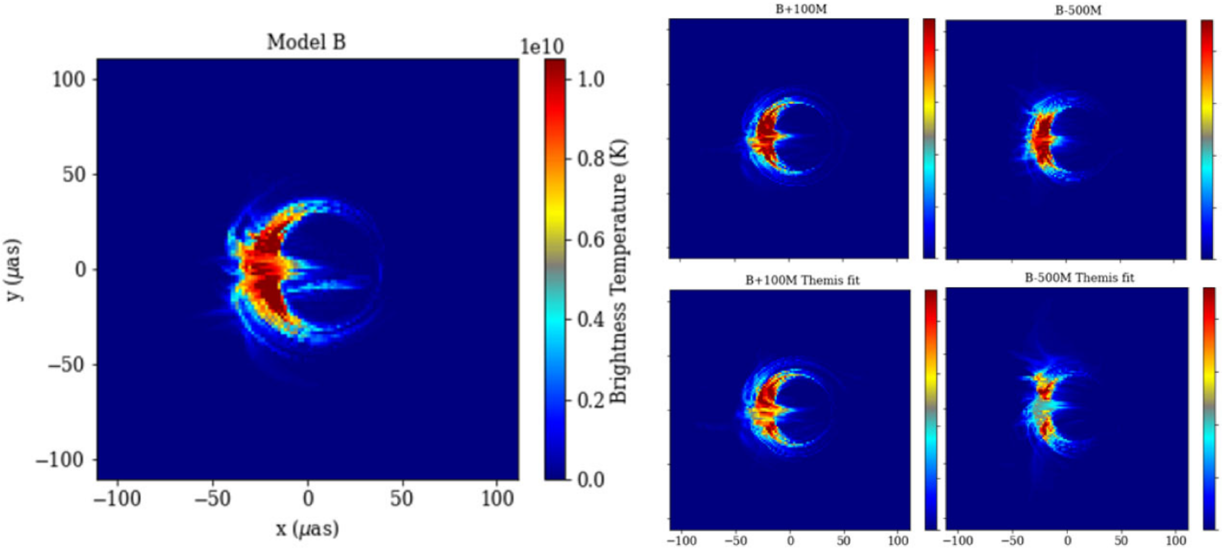

Figure 5. 230 GHz images of GRMHD model of Sgr A∗ from different snapshots with the same (left and right top panels) and fitted (right bottom panels) GRRT parameters. The colour codes the emission intensity (Stokes $\mathcal{I}$). A comparison between physical parameters and snapshots is presented in Table 4.

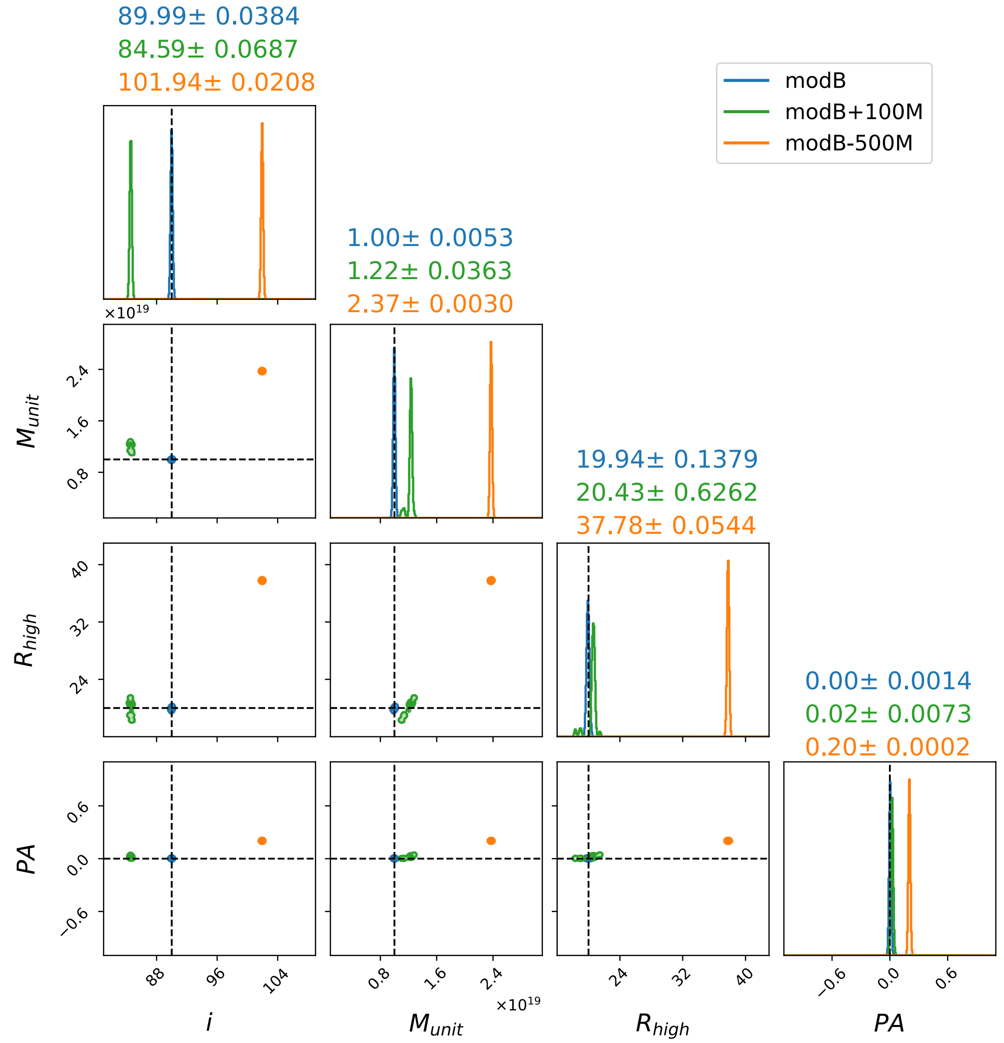

Figure 6. Same as Fig. 4 for modB, modB + 100M and modB − 500M (all with 1 per cent systematic error). The corresponding $\chi^2_{\rm eff} = 0.89$, 150, 181 for the three models, respectively.

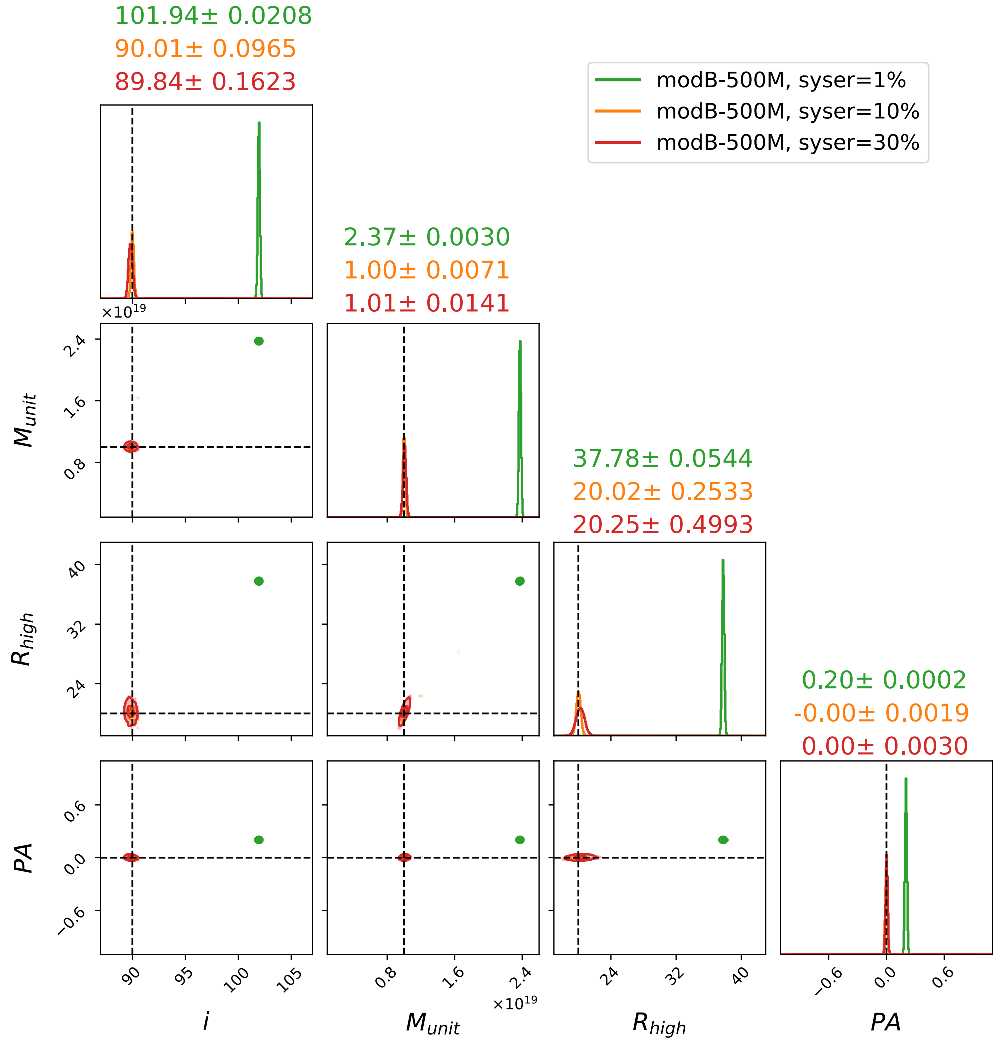

Figure 7. Same as Fig. 4 for modB − 500M (1 per cent systematic error), modB − 500M (10 per cent systematic error) and modB − 500M (30 per cent systematic error). The corresponding $\chi^2_{\rm eff}$ = 179.5, 0.71, 0.41 for increasing error budgets (1 per cent, 10 per cent, 30 per cent). The snapshots from all models (plus model B) are visible bellow, in Fig. 8.

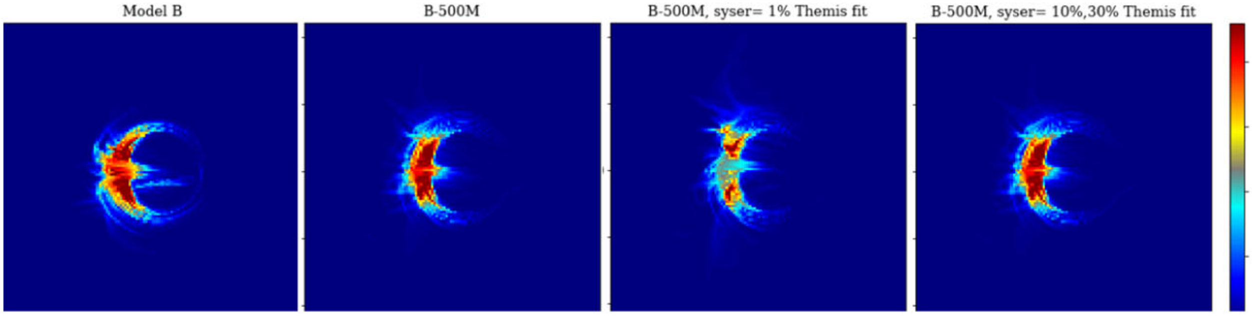

Figure 8. 230 GHz images of GRMHD models from Fig. 7. The figure aims to underline the improvement on the last two panels, going from 1 per cent errors, to 10 per cent, and 30 per cent.

Sagittarius A∗ is changing on time-scales that are short compared to a full night of EHT observation. To be precise 500 M for Sgr A∗ is equivalent to 165 min. In this case VLBI data collected over one full night represent a smoothed-out image of a varying accretion flow. To emulate such smoothing effect in our pipeline we create the truth synthetic data by averaging three snapshots (models: B − 500M, B, and B + 100M), called model 'avg'. We then fit a single snapshot to the 'averaged' truth (shown in Fig. 9, left panel). We consider two cases: first where the fitted snapshot is a part of the averaged image (model B, shown in Fig. 9) and second where the fitted snapshot is approximately 3500 M away from the average image (Fig. 9). These two fits are called 'avg $T_{\rm ref}$ = 0' and 'avg $T_{\rm ref}$ = 3500M'

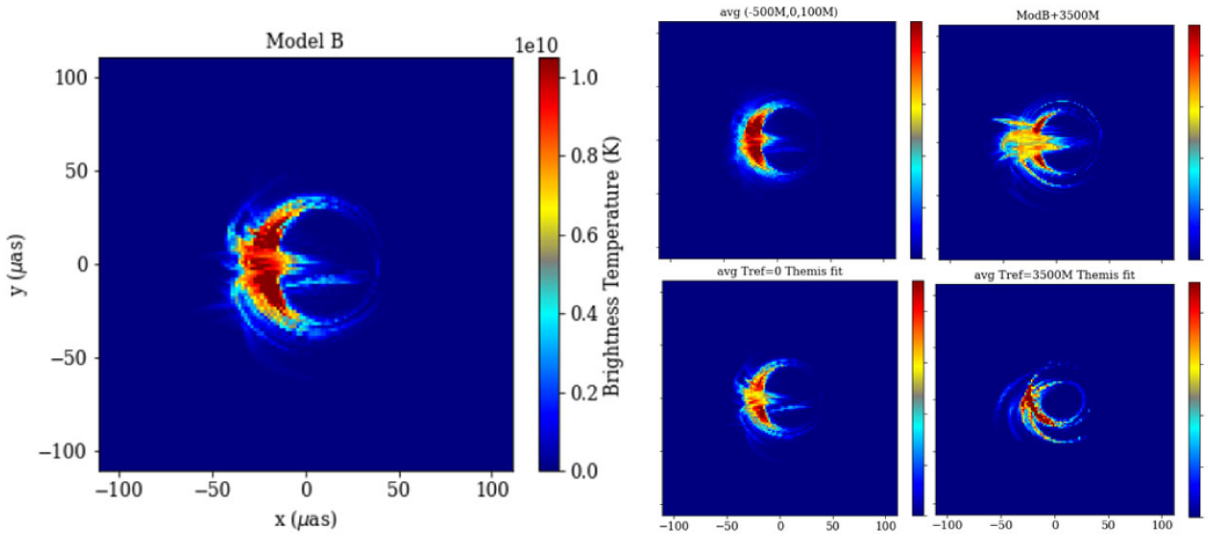

Figure 9. 230 GHz images of GRMHD model of Sgr A$*$ from different snapshots and using the GRRT parameters dictated by the fitting algorithm. For B + 100 M and B − 500 M the sampler was fitting B with the aforementioned templates. For avg models, $T_{\rm ref}$ corresponds to the template used by the sampler. The colour codes the emission intensity (Stokes $\mathcal{I}$). A direct comparison between physical parameters and snapshots can be seen in Table 5. Model avg $T_{\rm ref} =$ 3500 M Themis fit introduces the strongest biases. The most striking ones to the eye are $M_{\rm unit}$ and PA. The smaller $M_{\rm unit}$ is a reaction to the extended nature of the snapshot, while the PA = 36 deg result of the different symmetry of the source. Note, that we fit CP which is particularly sensitive to symmetries on the sky.

respectively. As a sanity check we also perform a fit with model avg (−500 M,100 M), that excludes the snapshot $T_{\rm ref} = 0$ from the truth. The systematic error added to the synthetic data was 10 per cent for all cases.

The results of this fitting exercise are shown in Fig. 10 and their parameters are listed in Table 5. The model avg $T_{\rm ref} = 0$ (orange line) converged to values close to the truth with the largest deviation for $R_{\rm high} = 26.9$ and total $\chi^2_{\rm eff} = 10.2$. The produced image avg $T_{\rm ref} = 0$ Themis fit is visible in Fig. 9 (bottom right panel) closely resembles the averaged image (the cross-correlation between them is 0.972 using the NXCORR tool within EHT-IMAGING software). The fit from model avg (−500M,100M), $T_{\rm ref} = 0$, is visible in Fig. 10. The results are of matching quality with similar biases and $\chi^2_{\rm eff} = 4.3$, ensuring that the smoothing of the truth is more significant than the existence of $T_{\rm ref} = 0$ in the averaged truth image.

In the second case of model avg $T_{\rm ref} = +3500$ M the biases are significantly larger, placing an upper limit in snapshot misspecification at roughly 500 M. The fitted image (avg $T_{\rm ref} = 3500$ M Themis fit) is visibly different from the averaged image, but explains some of the posterior values, such as the low $M_{\rm unit}$ to make emission narrower, and negative PA to roughly match the emission region asymmetry on the sky.

To sum up, both these tests suggest that variability will play a detrimental role in parameter estimation when moving to real data fitting. However, a more sophisticated implementation of the noise models arising from variability studies (Broderick et al. 2022; Georgiev et al. 2022), both on baseline and time domain, could prove extremely fruitful, and we plan to examine this in a future study.

4 DISCUSSION AND CONCLUSIONS

We created a pipeline towards Bayesian inference by fitting the GRMHD models to EHT observations and estimate model parameters. Similar efforts have been previously made by Kim et al. (2016) where sampling of GRMHD images was done only using two parameters: PA and total flux normalization. Also in Psaltis et al. (2022) a new MCMC algorithm was introduced for sampling of geometric, crescent models for image features (such as the shadow radius, the width of the ring etc.). In Medeiros et al. (2023) a PCA-based image reconstruction was developed, using an ensemble of simulated GRMHD images for fitting VLBI data. Lastly, in Jimenez-Rosales et al. (2023) image moments were used to characterize GRMHD snapshots as a means for model discrimination. Here we sample multiple parameters which require radiative transfer calculations in every MCMC-step which is a significant leap compared to previous work. We tested the pipeline over two distinct models (A, B) with differing inclination angle and $R_{\rm high}$, but more importantly, we made first steps towards tackling the time variability issue of such systems. The main results of this work lies in Fig. 6, where we showed that with the correct consideration of error budgets the pipeline is capable of retrieving correct parameters even for mis-specified snapshots. In Fig. 10 we showed that the miss-specification can work even for an averaged truth from multiple snapshots. Of course, that does not come without limitations as for the same averaged snapshot with a fitting template 3500 M away, the pipeline misses significantly the truth in all parameters. Template spacing of 1000 M, or 500 M to be more moderate, could potentially solve that and it would decrease the necessary snapshots by an order of 100, from 500 (cadence 10 M) used in AIS to 5–10 (cadence 1000 M–500 M) with our scheme.

In this stage we focused on fitting models to observed closure phases constructed from interferometric visibility phases. If one chooses to also fit visibility amplitudes, another thing that should be taken into account is that in case of Sgr A$*$ (but not M87$*$), the visibility function should be additionally modified to include smearing effects caused by refractive scattering of radio waves by free electrons in the Galaxy (Johnson & Gwinn 2015), causing artificial small-scale substructure in the image.

Despite being an improvement on static libraries, the adaptive parameter estimation is still time-consuming. In particular, to run the multiparameter fit on a university cluster, with 60 CPU cores, 1 GB per core, for 10 000 MCMC, takes approximately 170 wall-clock hours. In the same cluster running the snapshot scoring (part of AIS) takes 20 wall-clock hours with 24 cores, for a certain parameter

Table 4. The difference in GRMHD snapshots and physical parameters between the models presented in fitting test 3.3. The labels −100 M, +500 M, refer to how far a snapshot of the GRMHD simulation is with respect to model B ($T = 0$). Themis fit, refers to the best parameters from the MCMC sampler. For B + 100 M and B − 500 M the sampler was fitting B with the aforementioned templates. The percentile in the fitted models note systematic error level.

Model	i	$M_{\rm unit}$	$R_{\rm high}$	PA $[\pi]$	$\chi^2_{\rm eff}$
B + 100M	90	10^{19}	20	0	–
B − 500M	90	10^{19}	20	0	–
B + 100M Themis fit (1 per cent)	84.5	1.23×10^{19}	20.5	0.02	150
B − 500M Themis fit (1 per cent)	101.9	2.37×10^{19}	37.8	0.2	181
B − 500M Themis fit (10 per cent)	90.0	1.0×10^{19}	20.0	0.0	0.71
B − 500M Themis fit (30 per cent)	89.8	1.0×10^{19}	20.3	0.0	0.41

combination (so using only five inclination values and five $R_{\rm high}$ values takes 500 h). For a usage of ∼ 10 snapshots our method is already faster, plus the added value of not having to create, save, and manage the millions of snapshots required for AIS.

Another time-consuming part (for both the static library scoring and our approach) is the GRMHD simulation itself. At present we do not consider different GRMHD simulations and only call the ray-tracing code to generate different model images from the same simulation. The parameters which we are interested in the GRMHD, such as the spin of the black hole, could not be estimated under the current settings. Applying the current method to multiple simulations is a first, direct way forward leaving a model selection problem that could be tackled with Bayesian evidence or information criteria. How to simplify the model and generate the model image faster is a big challenge and that is the reason why fitting simple phenomenological models to observations is another practical way to compromise at present, such as Narayan & Yi (1995) and Broderick & Loeb (2006) or the more modern approaches of Palumbo et al. (2022), Chang et al. (2024), and Yfantis et al. (2024).

The MCMC sampling part is fast due to the highly parallel development. The computing performance could be improved by carefully choosing the numerical parameters (e.g. the temperature, the number of walkers) to be better adapted to the computer. Another bottleneck arises from load imbalance on the radiative

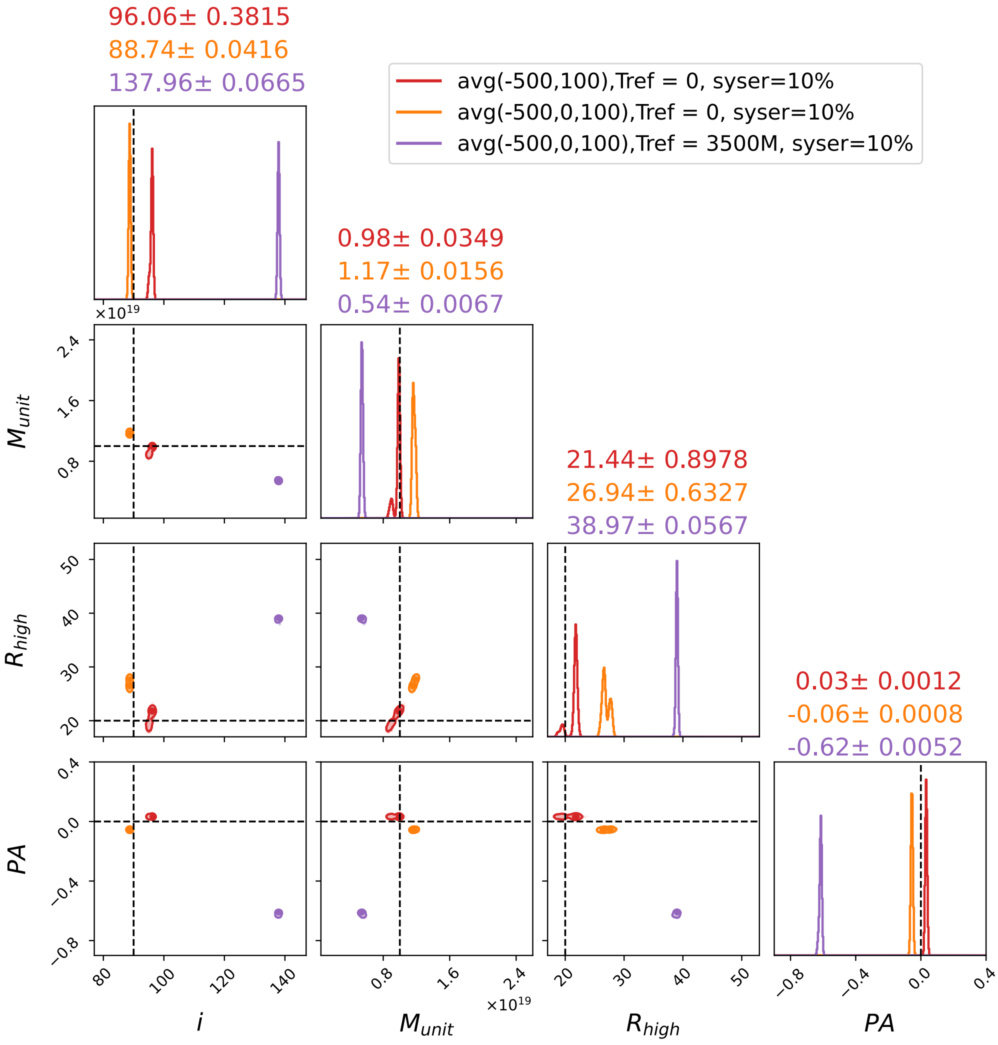

Figure 10. Same as Fig. 4 for models avg(−500,100), $T_{\rm ref} = 0$, avg(−500, 0, 100), $T_{\rm ref} = 0$ and avg(−500, 0, 100), $T_{\rm ref} =$ +3500 M, all with 10 per cent systematic error. The corresponding $\chi^2_{\rm eff} = 4.3$, 10.2, 30.0 for the three models, respectively..

Table 5. The difference in GRMHD snapshots and physical parameters for fitting test 3.4. Model avg was made using snapshots −500 M, 0 and +100 M. Themis fit, refers to the best parameters from the MCMC sampler. $T_{\rm ref}$ corresponds to the template used by the sampler. The percentile in the fitted models note systematic error level.

Model	i	$M_{\rm unit}$	$R_{\rm high}$	PA [π]	$\chi^2_{\rm eff}$
B + 3500 M	90	10^{19}	20	0	-
avg (−500 M, 0,100 M)	90	10^{19}	20	0	-
avg (−500 M,100 M) Tref = 0 Themis fit (10 per cent)	96.04	0.98×10^{19}	21.4	0.03	4.3
avg $T_{\rm ref} = 0$ Themis fit (10 per cent)	88.74	1.17×10^{19}	26.9	−0.06	10.2
avg $T_{\rm ref} = 3500$ M Themis fit (10 per cent)	137.96	0.54×10^{19}	40.0	−0.6	30.0

transfer side. At each tempering level independent model images are being generated for different parameter values, some of which such as higher $\dot{M}$ (i.e. higher density and opacity) will take longer to compute than others. On this front, recent developments in ray-tracing optimization, such as the GPU version of `IPOLE` presented in Moscibrodzka & Yfantis (2023) can be proven useful for the speed-up of the pipeline.

We have presented and validated the first Bayesian scheme to infer properties from GRMHD simulations from their simulated model images and visibility data as measured by an EHT-like VLBI configuration. This is a major step in fully utilizing the predictive power from GRMHD simulations of accreting black holes which previously have only been compared to VLBI data in more indirect ways and by using a-priori fixed parameter surveys. The work presented here eliminates simplifying scaling assumptions with total flux and BH mass in previous EHT VLBI analysis using GRMHD models Event Horizon Telescope Collaboration (2019e). It further allows improved conclusions from GRMHD models given a VLBI data set: (i) a refined inference in a continuous posterior distribution instead of discrete apriori chosen values (ii) efficient extension of the probed prior range (for instance beyond $R_{\rm high} = 160$) of the explored model parameters, which would otherwise become prohibitive with current strategies and (iii) a more efficient pathway to expand the parameter space to include any additional parameter that `IPOLE` can vary. More advanced samplers can easily handle much higher dimensional likelihood surfaces than will ever be explored with such models. Instead, the key improvement will be to speed up the evaluation of a single likelihood for instance by speeding up disc I/O.

Much work is still needed to get the most out of such inferences, but the next steps (sampler improvements, better likelihood approaches using complex visibilities, GRMHD-informed error budget etc.; Event Horizon Telescope Collaboration 2022d; Broderick et al. 2022; Georgiev et al. 2022) are both clear and already implemented in other analyses in `THEMIS`. Furthermore, following the results of Event Horizon Telescope Collaboration (Akiyama et al. 2021a), where polarization is resolved from M87, the so-called closure traces[6] can be included in the pipeline, making the fitting of the polarization to full visibilities possible. We envision that improved analysis schemes will greatly benefit the GRMHD community and theoretical interpretation of accreting black holes in the near future.

ACKNOWLEDGEMENTS

We thank Paul Tiede, George Wong, Lia Medeiros, and Michi Bauböck for fruitful discussions. This publication is part of the project the Dutch Black Hole Consortium (with project number NWA 1292.19.202) of the research programme the National Science Agenda which is financed by the Dutch Research Council (NWO).

DATA AVAILABILITY

Software used in the paper: `IPOLE`, `THEMIS`, `EHT-IMAGING`, Python. The scripts can be shared on reasonable request to the corresponding author.

REFERENCES

Balbus S. A., Hawley J. F., 1998, Rev. Mod. Phys., 70, 1
Blackburn L., Pesce D. W., Johnson M. D., Wielgus M., Chael A. A., Christian P., Doeleman S. S., 2020, ApJ, 894, 31
Blecher T., Deane R., Bernardi G., Smirnov O., 2017, MNRAS, 464, 143
Broderick A. E. et al., 2020, ApJ, 897, 139
Broderick A. E. et al., 2022, ApJ, 930, L21
Broderick A. E., Loeb A., 2006, MNRAS, 367, 905
Broderick A. E., Pesce D. W., 2020, ApJ, 904, 126
Chael A. A., Johnson M. D., Bouman K. L., Blackburn L. L., Akiyama K., Narayan R., 2018, ApJ, 857, 23
Chael A. A., Johnson M. D., Narayan R., Doeleman S. S., Wardle J. F. C., Bouman K. L., 2016, ApJ, 829, 11
Chael A., Narayan R., Johnson M. D., 2019, MNRAS, 486, 2873
Chang D. O., Johnson M. D., Tiede P., Palumbo D. C. M., 2024, ApJ, 974, 143
Dexter J., Agol E., Fragile P. C., 2009, ApJ, 703, L142
Doeleman S. S. et al., 2008, Nature, 455, 78
Doeleman S. S. et al., 2012, Science, 338, 355
EHT MWL Science Working Group 2021, ApJ, 911, L11
Event Horizon Telescope Collaboration 2019b, ApJ, 875, L2
Event Horizon Telescope Collaboration 2019c, ApJ, 875, L3
Event Horizon Telescope Collaboration 2019d, ApJ, 875, L4
Event Horizon Telescope Collaboration 2019e, ApJ, 875, L5
Event Horizon Telescope Collaboration 2019f, ApJ, 875, L6
Event Horizon Telescope Collaboration 2021b, ApJ, 910, L13
Event Horizon Telescope Collaboration 2022a, ApJ, 930, L12
Event Horizon Telescope Collaboration 2022b, ApJ, 930, L13
Event Horizon Telescope Collaboration 2022c, ApJ, 930, L14
Event Horizon Telescope Collaboration 2022d, ApJ, 930, L15
Event Horizon Telescope Collaboration 2022e, ApJ, 930, L16
Event Horizon Telescope Collaboration 2022f, ApJ, 930, L17
Event Horizon Telescope Collaboration 2024a, ApJ, 964, L25
Event Horizon Telescope Collaboration 2024b, ApJ, 964, L26
Event Horizon Telescope Collaboration Akiyama K. et al., 2021a, ApJ, 910, L12
Event Horizon Telescope Collaboration et al., 2019a, ApJ, 875, L1
Fish V. L. et al., 2011, ApJ, 727, L36
Fishbone L. G., Moncrief V., 1976, ApJ, 207, 962
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gammie C. F., McKinney J. C., Tóth G., 2003, ApJ, 589, 444
Georgiev B. et al., 2022, ApJ, 930, L20

[6] Closure traces, introduced by Broderick & Pesce (2020), is a data product constructed from polarimetric VLBI visibilities, that are immune to both station gains and polarization leakage (encoded in the so-called 'D-term').

Geyer C. J., 1991, in Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Interface Foundation of North America, p. 156

Gold R. et al., 2020, ApJ, 897, 148

Gold R., McKinney J. C., Johnson M. D., Doeleman S. S., 2017, ApJ, 837, 180

Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Gravity Collaboration 2018, A&A, 618, L10

Hada K. et al., 2013, ApJ, 775, 70

Hawley J. F., Balbus S. A., 1991, ApJ, 376, 223

Issaoun S. et al., 2019, ApJ, 871, 30

Issaoun S. et al., 2021, ApJ, 915, 99

Janssen M. et al. (2019) A&A, 626, A75

Jiménez-Rosales A., Dexter J., 2018, MNRAS, 478, 1875

Jimenez-Rosales A., Yfantis A., Mościbrodzka M., Dexter J., 2023, MNRAS, 527, 1847

Johnson M. D. et al., 2015, Science, 350, 1242

Johnson M. D. et al., 2018, ApJ, 865, 104

Johnson M. D., Gwinn C. R., 2015, ApJ, 805, 180

Kim J. Y. et al., 2018, A&A, 616, A188

Kim J., Marrone D. P., Chan C.-K., Medeiros L., Özel F., Psaltis D., 2016, ApJ, 832, 156

Lu R.-S. et al., 2018, ApJ, 859, 60

Medeiros L., Psaltis D., Lauer T. R., Özel F., 2023, ApJ, 943, 144

Mościbrodzka M., Falcke H., Shiokawa H., 2016, A&A, 586, A38

Mościbrodzka M., Falcke H., Shiokawa H., Gammie C. F., 2014, A&A, 570, A7

Mościbrodzka M., Gammie C. F., 2018, MNRAS, 475, 43

Mościbrodzka M., Gammie C. F., Dolence J. C., Shiokawa H., Leung P. K., 2009, ApJ, 706, 497

Moscibrodzka M., Yfantis A., 2023, ApJS Ser., 265, 22

Narayan R., Yi I., 1995, ApJ, 444, 231

Nelson B., Ford E. B., Payne M. J., 2014, ApJS, 210, 11

Palumbo D. C. M., Gelles Z., Tiede P., Chang D. O., Pesce D. W., Chael A., Johnson M. D., 2022, ApJ, 939, 107

Porth O. et al., 2019, ApJS Ser., 243, 26

Prather B. S. et al., 2023, ApJ, 950, 35

Psaltis D. et al., 2022, ApJ, 928, 55

Ressler S. M., Tchekhovskoy A., Quataert E., Chandra M., Gammie C. F., 2015, MNRAS, 454, 1848

Shiokawa H., 2013, PhD thesis, University of Illinois at Urbana-Champaign

Swendsen R. H., Wang J.-S., 1986, Phys. Rev. Lett., 57, 2607

Tchekhovskoy A., Narayan R., McKinney J. C., 2011, MNRAS, 418, L79

Ter Braak C. J. F., 2006, Stat. Comput., 16, 239

Thompson A. R., Moran J. M., Swenson Jr G. W., 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition. Springer, Cham

Tibbits M. M., Groendyke C., Haran M., Liechty J. C., 2014, J. Comput. Graph. Stat., 23, 543

Yfantis A. I., Mościbrodzka M. A., Wielgus M., Vos J. T., Jimenez-Rosales A., 2024, A&A, 685, A142

APPENDIX A: OPTIMAL IMAGE RESOLUTION

To reduce the computational cost of the pipeline we look for a minimum model image resolution for which the synthetic VLBI data are converged. Given the fiducial image of model A (with parameters $i = 60^\circ$, $M_{\rm unit} = 3 \times 10^{18}$, $R_{\rm high} = 3$, $R_{\rm low} = 3$, PA $= 0$), shown in Fig. 1, we use `EHT-IMAGING` to produce the visibility amplitudes and closure phases for four different image resolution cases: 32×32 pixels, 64×64 pixels, 128×128, and 256×256 pixels.

We compare all the observables generated by `EHT-IMAGING`, the results of which are plotted in Fig. A1. In the top row we can see the natural improvement of images with resolution. In the bottom one we show the VA and CP data for all resolutions. What we want to convey is that even though the data get better with higher resolutions, the step from 128 to 256 is small, while the steps before that large. Indeed, we see that points of green and orange show sizable divergence, but the blue points are often 'hidden' in the plots, suggesting that they match very well with the points from 256^2 (purple).

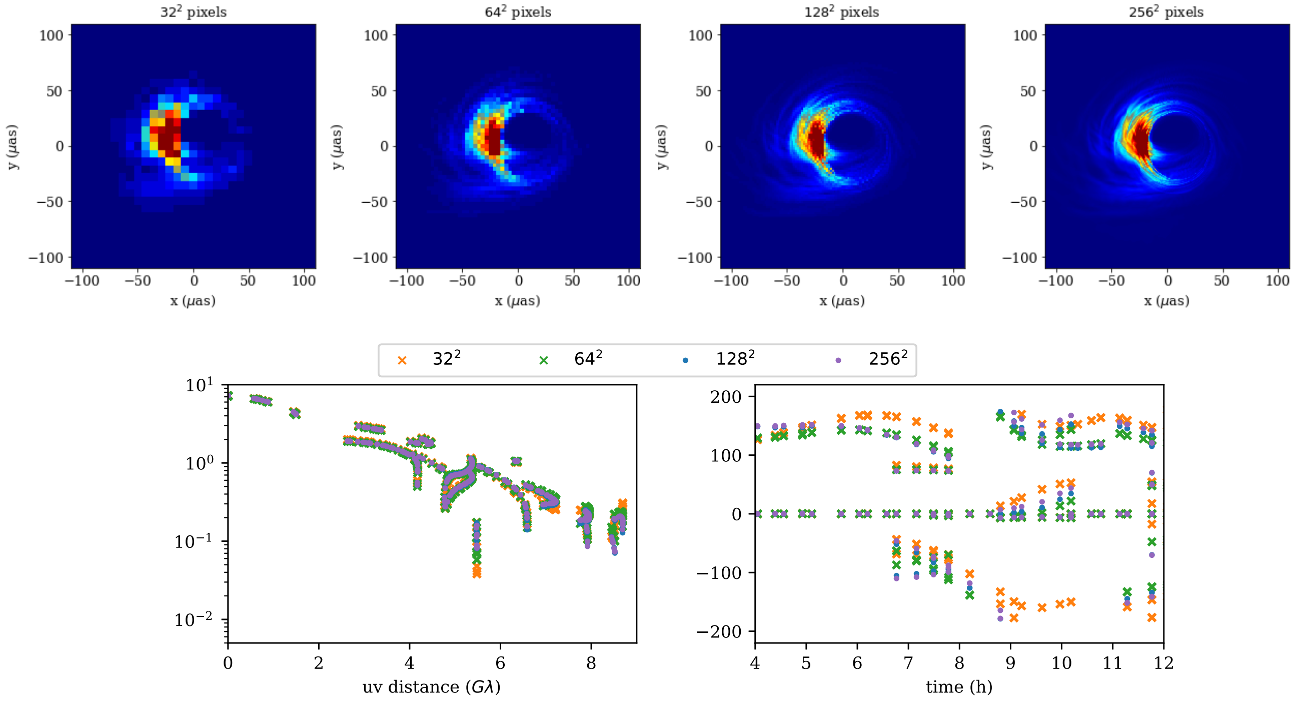

Figure A1. Top: the modelled images for 32×32 pixels, 64×64 pixels, 128×128 pixels, 256×256 pixels. Bottom:t he left plot shows the visibility amplitudes for all different resolutions; the right one all the closure phases. Crosses represent 32^2 and 64^2, while dots represent 128^2 and 256^2.

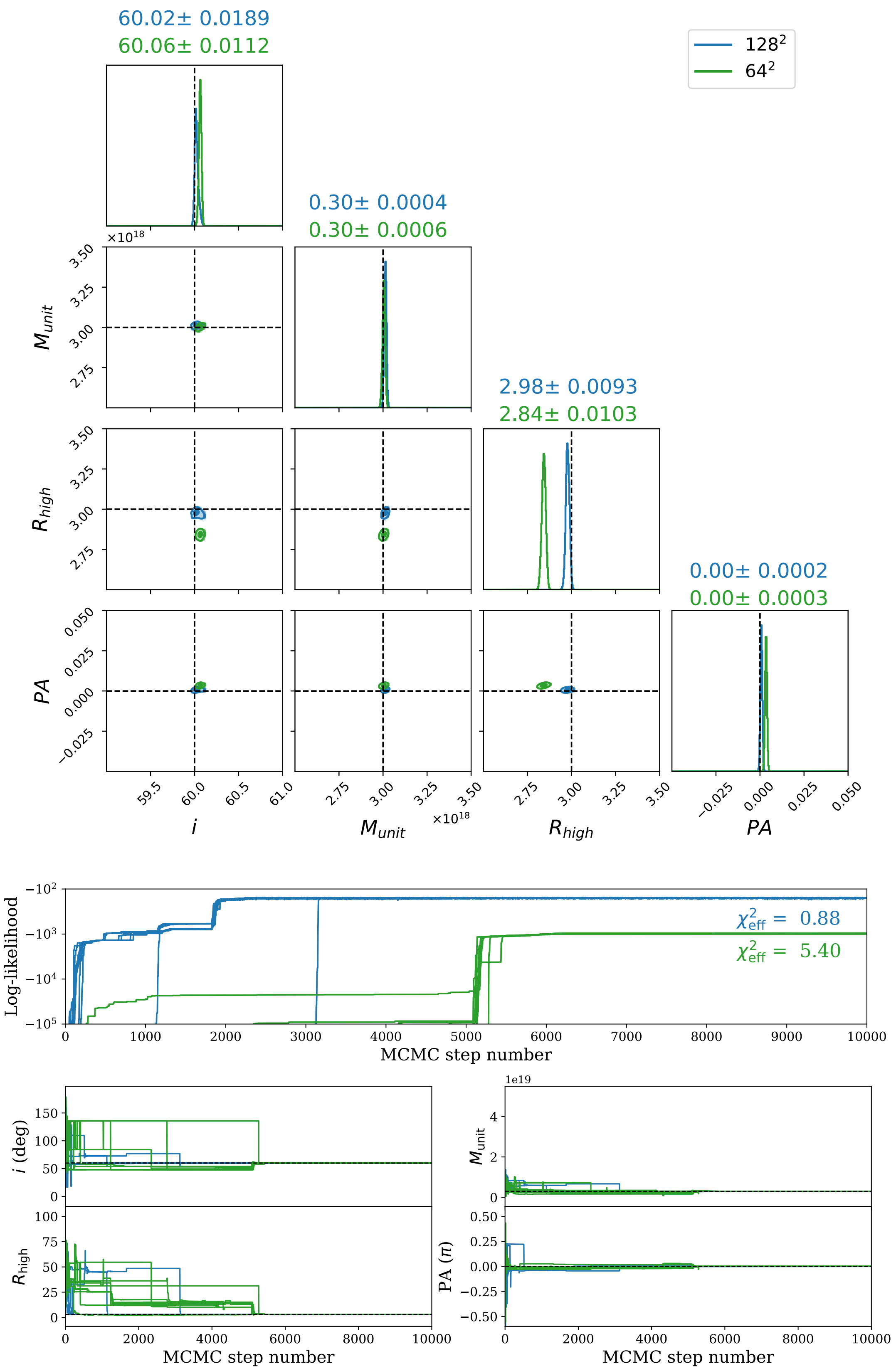

Figure A2. Top: Same as Fig. 4 for model A using different resolution (64^2, 128^2). Middle: The log-likelihood evolution through out the run. Bottom: The chains for the whole duration of the MCMC run. The burn-in for the PDs and the Gaussian fits has been set to 6000 and the total run time was 10 000 MCMC. The reported $\chi^2_{\rm eff}$ are the smallest values after the burn-in. In short: low resolution introduces poorer fit quality, but the posteriors are largely un-biased with the exception of $R_{\rm high}$. The latter bias in $R_{\rm high}$ is subtle in magnitude compared to the coarse spacing in the EHT libraries, but should be further investigated in subsequent work.

To further strengthen our thesis and demonstrate the benefit of using data with resolution 128^2, in Fig. A2 we compare the results from fitting model A, for all four parameters, using two different resolutions (64^2, 128^2). The higher resolution outperforms the lower one in a number of fronts. The posteriors for 64^2 are not covering the truth for all parameters (R_{high} is shifted). This is reflected in the likelihood graph as well, where in lower resolution the quality of the fit is not as good ($\chi^2_{\mathrm{eff}} = 5.37$). Lastly, the lower resolution exhibits a slower convergence, that can be seen both in the chains and the likelihood development.

Overall, it seems that a resolution of 128^2 is necessary for sufficient fit quality while resolution of 256^2 does not provide further improvement at higher computational expenses.

APPENDIX B: DEPENDENCE ON CHOICE OF SAMPLERS, WALKERS, TEMPERATURES, AND INITIALIZATION

In this appendix, we search for the optimal sampling method and chains initialization. We do this by various fitting procedures and direct comparisons for model A.

Within the parallel tempering sampler, an algorithm should be chosen to decide on the next step parameters. There is a handful of options implemented in THEMIS, from which we consider three for this project. The first one is the affine invariant (AI) method described in detail in Goodman & Weare (2010). Under the affine transformation the ill-shaped density probability will not bring any extra difficulties as the simple single-variable MCMC samplers (e.g. Gibbs sampler) do. This method is widely used and well tested in the famous MCMC python package EMCEE (Foreman-Mackey et al. 2013). The second possible algorithm is the differential evolution (DE) method introduced by Ter Braak (2006) and developed further in Nelson, Ford & Payne (2014). Similarly to the affine invariant method each chain draws new values by using positions from other chains in the parameter space. The difference is that DE uses all chains per draw while AI only one. This can lead to discrepancies in their performance, depending on the problem. Due to the pipeline constructions, the number of processors should be at least N_{T} and be integer times of $N_{\mathrm{T}} \times N_{\mathrm{W}}/2$ for both samplers. The last option is the automated factor slice (AFS) algorithm (Tibbits et al. 2014). It was implemented more recently than the others within THEMIS and it has a key difference since it operates only with one chain, multiplied by the temperatures. It improves from the simple Metropolis–Hastings (MH) sampler by using an in-between step of redefining the sampling pool on every step.

Fig. B1 shows the chains and the corresponding likelihood for all three samplers fitting two parameters (M_{unit} and PA). The top has been made with an initial value of 8×10^{18} while the bottom with 3×10^{19} (truth is at 3×10^{18}). In the top panel, all samplers behave similarly, they converge towards the truth fast (with AFS being the fastest ~ 10 steps) and with similar χ^2_{eff} values (AFS has a higher value that persists till after 5000 MCMC steps). Already we can notice that DE in contrast to AI is more jittery, probably to the fact that the number of chains and temperatures is too high for the level of communication DE sampler imposes.

In the bottom panel the picture is different. AFS which was the faster sampler in the previous case, now converges to the truth last (at 530 MCMC steps). Furthermore, even though it seems like DE is faster than AI, that is actually not the case, since almost all chains are close to the truth but not exactly and vary significantly for a long period. All this is reflected in the likelihood plot, where the oscillation of DE chains is more evident.

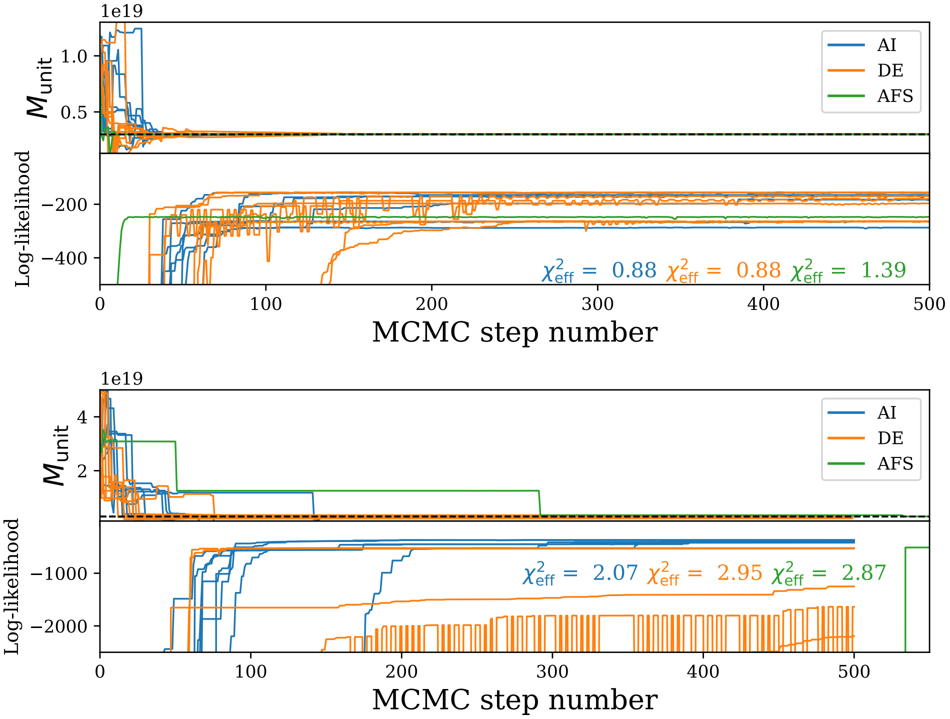

Figure B1. Trace of the chains fitting M_{unit} for different samplers (AI: affine invariant, DE: differential evolution, AFS: automated factor slice). The run shown in the top panel was initialized at 8×10^{18}, while the bottom at 3×10^{19}. Note that the last MCMC step in the bottom is 550. AI and DE were run with 40 cores and their computational efficiency was 1.5 steps per core per hour. AFS was run with 10 cores and the efficiency was higher at 2.5 steps per core per hour.

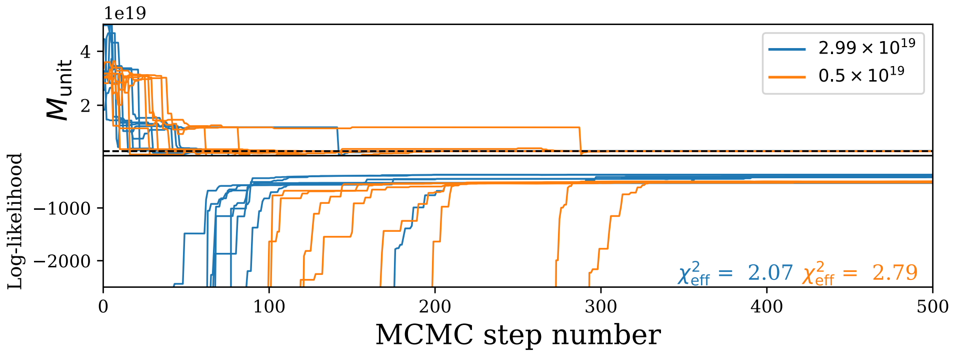

Figure B2. Trace of the chains fitting M_{unit} for different initial ranges using the AI sampler. The initial value is 3×10^{19}.

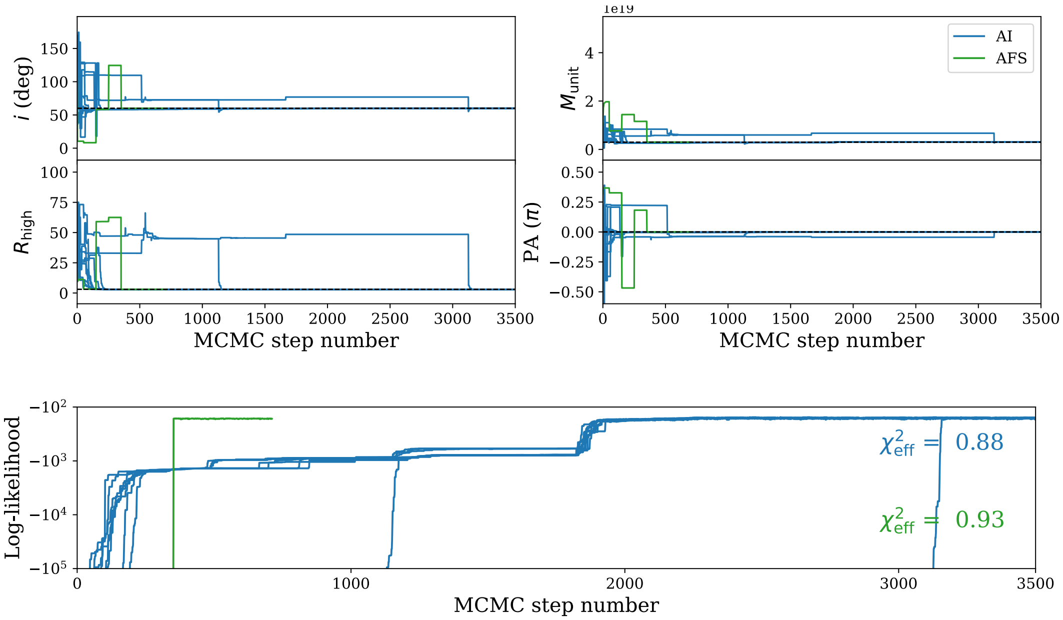

Figure B3. Trace of the chains fitting all four parameters for different samplers (AI: affine invariant, AFS: automated factor slice). The chains were initialized in the same way as for model A. AI was run with 40 cores and the computational efficiency was 1 step per core per hour. AFS was run with 12 cores and the efficiency was the same, 1 step per core per hour, with a much faster convergence though.

As a last remark the computational efficiency of AI and DE is 60 and 70 steps per hour, respectively, using 40 cores (8 chains and 10 temperatures, divided by 2), so ∼ 1.5 steps per core per hour while AFS is 25 steps per hour, using 10 cores (1 chain and 10 temperatures), so 2.5 steps per core per hour.

Fig. B3 shows the fitting procedure for all parameters using AI and AFS. For initial values we chose the same as in the main results, that is $i_0 = 90$, $M_{\text{unit},0} = 8 \times 10^{18}$, $R_{\text{high},0} = 30$, $\text{PA}_0 = 0$. It is evident that AFS converges faster in this example. The fact that some chains from AI get stuck for a significant amount of time can be a problem for more advance fitting procedures. This is a well known problem in certain fits with AI, reported already in Foreman-Mackey et al. (2013).

In Fig. B2 we show the chain evolution for the AI sampler for different initial range, using the same initial value. The convergence is faster with a wide range if the initial value is far from the truth, but with better informed prior a smaller range is to be preferred, as stated in Foreman-Mackey et al. (2013).

The combination of these tests, taking into account the computational time, makes AI our sampler of choice for this work. Furthermore, we choose to initialize our chains near the middle of the desired area of exploration with an initial range that covers it all to be as agnostic as possible. In the future a hybrid approach, using AI for a wide survey and AFS for a more detailed exploration, with priors obtained from AI, could prove optimal.

This paper has been typeset from a TEX/LATEX file prepared by the author.