

Exoplanet Astrometry for Gaia and GaiaNIR

Hugo Elmhäll

Division of Astrophysics
Department of Physics



LUND
UNIVERSITY

2026–EXA258

Degree project of 15 higher education credits
June 2026

Supervisor: David Hobbs

Division of Astrophysics
Department of Physics
Box 118
SE-221 00 Lund
Sweden

Abstract

The Gaia space observatory has produced astrometric measurements of extraordinary precision, enabling detection of exoplanets through the reflex motion of their host stars. With Gaia Data Release 4 (DR4) scheduled for December 2026, which will include full epoch astrometry, methods for orbital parameter recovery are needed. This work presents a Bayesian Markov Chain Monte Carlo (MCMC) framework for fitting astrometric orbits to Gaia epoch data, applied to five Gaia DR3 candidates with distinct orbital configurations. With real epoch astrometry data not yet publicly available, observations are simulated based on published Gaia orbital solutions, using conservative per-epoch noise of $\sigma_{AL} = 0.1$ mas, for both five-year and ten-year mission scenarios.

The framework successfully recovers orbital parameters for all five candidates, with the ten-year mission generally providing improved precision over the five-year mission, particularly for lower signal-to-noise systems. Injection-recovery tests demonstrate reliable period and eccentricity recovery across wide parameter ranges. An overestimation of inclination is identified for nearly face-on orbital configurations, consistent with a known mathematical consequence of the back-conversion from Thiele-Innes to Campbell elements. Period perturbation sweeps reveal the presence of period aliases at the 2:1 and 3:2 mean-motion resonance periods, highlighting the need for careful treatment when interpreting real DR4 detections. Additionally, one candidate has independently been identified as a false positive, demonstrating the importance of follow-up observations for solely astrometrically detected companions. The method will be directly applicable to real DR4 epoch data with no fundamental modifications, and the results support the scientific case for GaiaNIR as a potential successor mission with an extended observational baseline.

Populärvetenskaplig beskrivning

Finns det liv på andra planeter? Det är en fråga som forskare världen över har ägnat sina liv åt att besvara. En del menar att universum är så ofattbart stort att det är en självklarhet att liv existerar någon annanstans. Andra hävdar att sannolikheten för att de omständigheter som gav upphov till liv på jorden skulle upprepas är så oerhört liten att det är närmast omöjligt. Trots dessa skilda åsikter är jakten på utomjordiskt liv mer aktiv än någonsin, men för att kunna hitta liv på planeter måste planeterna själva först upptäckas.

För ungefär tretton år sedan sköts en rymdfarkost känd som Gaia-satellitten upp i rymden, till en plats drygt 1.5 miljoner kilometer från Jorden. Utrustad med högkvalitativa kameror och vetenskapliga instrument hade den ett enda syfte: att skapa den mest precisa kartan över vår galax Vintergatan som någonsin framställts. Under drygt elva år observerade Gaia tålmodigt nästan två miljarder stjärnor och bestämde varje stjärnas position med en precision liknande att urskilja en golfboll på månens yta. Det är just denna extrema noggrannhet som gör Gaia perfekt för att inte bara kartlägga stjärnor, utan även för att upptäcka planeter som kretsar kring dem. Dessa planeter kallas extrasolära planeter, eller exoplaneter, vilket betyder ”planeter som inte kretsar kring solen”.

Men hur upptäcker vi dessa exoplaneter? Stjärnor kan observeras eftersom de lyser med ett starkt och detekterbart ljus, vilket planeter inte gör. Svaret ligger i den gravitationella dansen mellan värdstjärnan och dess planet. En planet kretsar inte ensamt kring sin värdstjärna, utan de kretsar båda kring sitt gemensamma masscentrum. Detta gör att värdstjärnan svajar något, och det är just detta svaj som med Gaias höga precision kan detekteras och från vilket en kretsande planet kan härledas.

Denna teknik kallas astrometri och har först på senare år börjat tillämpas för att hitta planeter, tack vare Gaias precision. Tidigare har andra metoder dominerat fältet, men astrometrin börjar nu ta mark. Gaia observerar varje stjärna ungefär 14 gånger per år och noterar dess position vid varje tillfälle. Genom att analysera data insamlad över fem år kan stjärnans rörelse till följd av en kretsande planet spåras. Utifrån denna rörelse bryts banan ned i sju parametrar, var och en innehållande specifik information om banan, såsom den tid det tar för planeten att fullfölja ett varv kring stjärnan och banans form. Dessa parametrar kan vara mycket svåra att lösa direkt, och därför används en metod känd som Markov Chain Monte Carlo (MCMC). Metoden använder sannolikheter för att hitta de korrekta parametervärdena, och fungerar på följande sätt: föreställ dig att du står med ögonbindel i ett bergigt landskap och försöker hitta den lägsta punkten i en dal. Genom att känna lutningen och ta upprepade små steg kan du ändå, utan att se, hitta botten av dalen. MCMC fungerar på samma sätt och navigerar i ett sannolikhetslandskap för att hitta det mest sannolika värdet.

Genom att kombinera astrometri och MCMC bidrar resultaten av detta arbete till sökandet efter exoplaneter i vår galax, till förståelsen av vilka slags banor de följer, och i förlängningen till frågan om vilka av dem som möjligen skulle kunna hysa liv. Gaia beräknas ha detekterat tiotusentals planetsystem när dess slutliga data har analyserats och publicerats omkring år 2030, och vem vet, kanske är ett av dem vårt framtida hem.

Acknowledgements

I would like to express my gratitude to Dr. David Hobbs for providing me this project and for his invaluable expertise and support throughout its completion. I would also like to acknowledge the European Space Agency and the scientists involved for designing, launching and operating the Gaia mission, without which this project would not have been possible. Finally, I thank my fellow students and professors at Fysicum for making my three years at the bachelor's programme both enriching and memorable.

Contents

1	Introduction	1
1.1	The Gaia satellite	1
1.2	Using Gaia to detect exoplanets	2
2	Theory	3
2.1	Orbital parameters	3
2.1.1	Thiele Innes parameters	4
2.2	Astrometric signature and S/N	4
2.2.1	Companion mass estimation	5
2.3	Bayesian inference	5
2.4	Markov Chain Monte Carlo	6
3	Method	8
3.1	Candidate selection	8
3.2	Astrometric and Orbital data	8
3.3	Simulated epoch data	8
3.4	Astrometric uncertainty propagation	9
3.5	MCMC implementation	9
3.5.1	Perturbations of values	9
3.6	Posterior analysis	10
3.7	Convergence	10
3.8	Testing	10
3.8.1	Five-year and ten-year comparison	10
3.8.2	Systematic injection-recovery grid	10
3.8.3	Period sweep	11
3.8.4	Convergence robustness	11
4	Results	12
4.1	Parameter recovery for selected candidates	12
4.1.1	Candidate 1386158505321789696	12
4.1.2	Candidate 4062446910648807168	13
4.1.3	Candidate 423297927866697088	14
4.1.4	Candidate 1878822452815621120	15
4.1.5	Candidate 4698424845771339520	16
4.2	Companion mass estimates	17
4.3	Injection-recovery	18
4.3.1	Period recovery	18

4.3.2	Eccentricity recovery	19
4.3.3	Inclination recovery	21
4.4	Period perturbation sweep	22
4.4.1	Candidate 1386158505321789696 sweep	22
4.4.2	Candidate 423297927866697088 sweep	22
4.4.3	Candidate 4062446910648807168 sweep	23
4.5	Convergence robustness	23
5	Discussion	24
5.1	Observational setup	24
5.2	Interpretation of results	25
5.3	Applicability to DR4 and GaiaNIR	27
6	Conclusion	28
A	Figures from 4.1	30
B	Figures from 4.4 and 4.5	56

Chapter 1

Introduction

1.1 The Gaia satellite

The Gaia satellite was a space observatory operating in the L2 Lagrange point at an approximate distance of 1.5 million km beyond Earth's orbit around the sun. It was launched by the European Space Agency (ESA) on 19 December 2013 with the objective of creating the most precise map of the Milky Way. The satellite is equipped with two telescopes, each scanning regions separated by 106.5 degrees as the satellite spins around its own axis, such that every star in its line of sight is measured twice per rotation. As it spins, it also precesses, allowing Gaia to observe new stars and survey the entirety of the sky repeatedly over time, observing each star an average of 70 times over a five year period. For data collection, three separate scientific instruments are used. The first is the astrometric instrument that measures the precise positions of objects in the sky. With a series of measurements, this is used to determine the parallax and proper motion of the observed objects. The second is the photometric instrument, which measures the intensity and wavelength of incoming light. This can then be used to determine the color, age, temperature, and chemical compositions of stars and objects. The third is the radial velocity spectrometer, which measures the radial velocity of observed objects by considering the Doppler shift of the spectral lines.

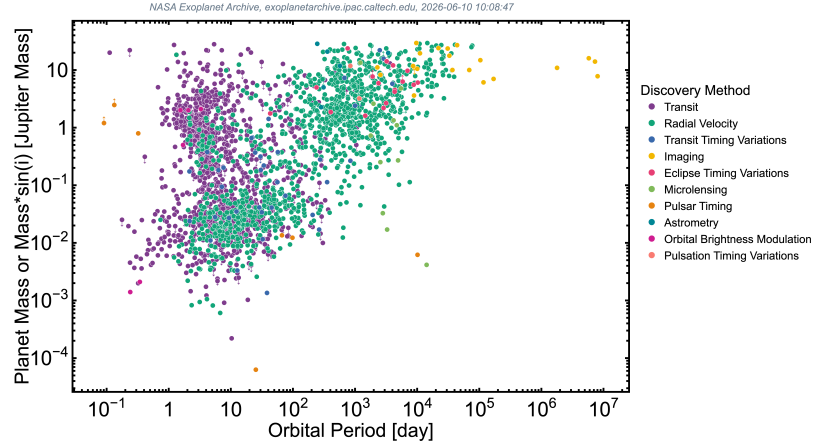
As of 15 January 2025, Gaia had reached its end of observations due to it running out of cold-gas propellant used for small orientation adjustments. Without it, Gaia could no longer maintain its precision for measurements, and in March 2025 it was sent by ESA into a retirement orbit around the sun. A potential successor known as GaiaNIR (Gaia Near Infra-Red), proposed by Hobbs et al. (2016), is being actively studied within ESA. This satellite would carry on Gaia's legacy and improve and extend the mapping of the galaxy by instead absorbing wavelengths in the near infra-red spectrum. This would allow for observation of fainter stars that generally are more red, as well as stars hidden behind interstellar dust, through which visible light does not pass, due to extinction.

Although Gaia is no longer active, future releases based on years of collected data are still planned. The next one is Data Release 4 (DR4) set to be released in December 2026, which will be based on 66 months of observational data. Beyond its improved values compared to earlier data releases, DR4 will also release a full time-series, enabling detailed studies of the motion of stars. Following this data release, sometime before the end of 2030 Data Release 5 (DR5) will be released, which will be based on all mission data collected by the Gaia satellite since its launch. This is expected to yield further precision and improvements, and eventually represent the most comprehensive and detailed stellar catalogue ever assembled.

1.2 Using Gaia to detect exoplanets

To date, over 6000 exoplanets have been confirmed (Christiansen et al. 2025), with the largest contribution from the transit method and radial velocity method. The data collected by Gaia contribute to both of these techniques, as well as the astrometry method. Each method has its own benefits, and is sensitive to specific orbital configurations.

The transit method uses photometry and is applied to near edge-on orbits when the star’s companion, relative to the observer, passes in front of the star and creates a momentary decrease in the star’s measured flux. From this decrease, the existence of the companion is derived (Perryman 2018). This method works best for large mass and/or short period companions and has been used to discover over 4600 exoplanets (Christiansen et al. 2025).



The radial velocity and astrometry methods both exploit a common orbital feature. A star and an exoplanet orbit each other around their shared system barycentre (center of mass), which is placed in a focus of their re-

spective Keplerian elliptical orbits. As the star traces this orbit it undergoes a small but detectable reflex motion, which with the two different methods is detected and utilized in their own ways.

Figure 1.1: Plot of currently known exoplanets, with Orbital Period on the x-axis and Planet Mass on the y-axis, and color indicating discovery method. <https://exoplanetarchive.ipac.caltech.edu/>

The radial velocity method measures the projected motion along the observer’s line of sight by examining the Doppler shift in wavelengths of the star’s emitted light, where an observed receding motion yields longer stretched wavelengths (redder light), and an observed approaching motion yields shorter compressed wavelengths (bluer light) (Perryman 2018). This method is used most effectively when the star-exoplanet orbit is oriented edge-on relative to the observer, which results in a larger radial velocity. However, with a perfectly face-on orbit, the radial velocity is zero and the method is futile. This method has contributed to discovering over 1100 exoplanets (Christiansen et al. 2025), and is sensitive to longer orbital periods compared to the transit method.

The astrometry method relies on extremely high positional precision to measure the motion of a star on the sky. By tracking the star’s position over an extended period, small deviations of its trajectory are detected. In the presence of a companion, these deviations arise from the star’s reflex motion around the system’s center of mass, producing a characteristic astrometric orbit (Perryman 2018). This method is best utilized for face-on orbits, but has so far only contributed to a handful of exoplanet discoveries (Christiansen et al. 2025). This is the method employed in this thesis using the high-precision measurements provided by Gaia Data Release 3 (DR3) (Gaia Collaboration et al. 2023).

Chapter 2

Theory

2.1 Orbital parameters

All observed orbits follow the laws of a Keplerian elliptic orbit. To describe the orbital motion of a host star induced by an orbiting companion, the orbit must be parameterized in a way that connects the physical orbit to the observed astrometric signal. This is done by using a set of orbital parameters, each describing certain aspects of the orbit, such as size, shape, orientation, and timing.

To break the orbital motion down into parameters, we first need a reference point from which the parameters are then described. When observing a star, a reference plane is considered that is perpendicular to the viewing direction and passes through the focal point of the elliptic orbit. The reference uses the viewing direction, the plane, the northern direction projected onto the plane, and the ascending node, illustrated in Figure 2.1. The orbital motion is then decomposed into seven orbital parameters: $(a, e, P, T_0, i, \Omega, \omega)$.

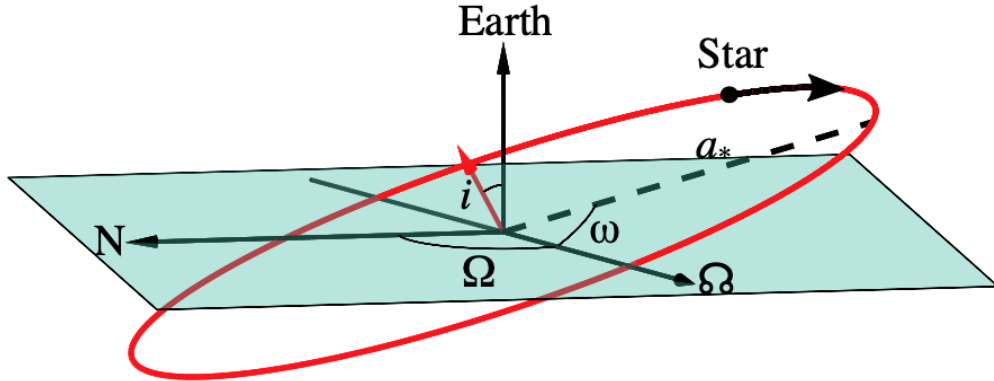


Figure 2.1: Visual representative example of an observed orbit, along with the reference plane, northern direction and ascending node, and orbital parameters a , ω , Ω and i (Ranalli et al. 2018).

The semi-major axis (a) and the eccentricity (e) together describe the size and shape of the orbit respectively, where (a) is half of the major axis and (e) describes the ellipticity of the orbit. The period (P) is the time it takes for the planet to complete the orbit, and the time of periastron passage (T_0) gives the time of the planet at a specific passage point, relative to a reference epoch. The three remaining parameters describe the orientation of the orbit relative to the plane of reference. The longitude of ascending node (Ω) is the angle between the north-pointing axis and the intersection

of the orbit with the plane. The argument of pericentre (ω) is the angle between the orbit's intersection with the reference plane and the "tip" of the orbit (pericentre). The inclination (i) is the angle between the orbital plane and the reference plane, which describes the tilt of the orbit relative to the observer (Perryman 2018).

2.1.1 Thiele Innes parameters

In the astrometric orbit model used by Gaia, the orbital motion of stars is described in terms of the Thiele-Innes parameters, which provide an alternative representation of orbital elements. The Thiele-Innes parameters (A, B, F, G) are linear combinations of the orbital elements a, i, ω, Ω and are formulated in the following way (Perryman 2018):

$$\begin{aligned} A &= a(\cos \omega \cos \Omega - \sin \omega \sin \Omega \cos i) \\ B &= a(\cos \omega \sin \Omega + \sin \omega \cos \Omega \cos i) \\ F &= a(-\sin \omega \cos \Omega - \cos \omega \sin \Omega \cos i) \\ G &= a(-\sin \omega \sin \Omega + \cos \omega \cos \Omega \cos i). \end{aligned}$$

Using these parameters, the apparent motion of the star on the plane of the sky can be written in tangential coordinates as

$$\begin{aligned} x &= AX + FY \\ y &= BX + GY \end{aligned}$$

where the variables X and Y describe the motion in the orbital plane and are defined as

$$\begin{aligned} X &= \cos(E) - e \\ Y &= \sqrt{1 - e^2} \sin E. \end{aligned}$$

Here, e is the eccentricity and E is the eccentric anomaly, being an angle that describes the position of the star along its orbit.

In this project, the Thiele-Innes parameters are directly extracted from the Gaia data and used instead of ω, Ω and i . This choice is motivated by the linear properties of the parameters. When using sampling methods, such as Markov Chain Monte Carlo, the angular parameters introduce issues related to angular wrapping, due to their periodicity. This is avoided when using Thiele-Innes parameters, and once they have been determined, the angular elements can be inversely recovered analytically.

2.2 Astrometric signature and S/N

The detectability of a companion using Gaia astrometry is governed by the astrometric signature (Perryman et al. 2014)

$$\alpha = \frac{a_\star}{d} = \frac{M_p}{M_\star} \cdot \frac{a_p}{d}$$

which describes the angular size of the host star's reflex motion on the sky, where $a_\star = \frac{M_p}{M_\star} a_p$ is

the semi-major axis of the host star’s orbit around the system barycentre and d is distance from observer to the system. The signal-to-noise ratio per transit is then defined as

$$S/N = \frac{\alpha}{\sigma_{AL}}$$

where σ_{AL} is the per-epoch along-scan measurement uncertainty. Since the photocentric semi-major axis a_0 expressed in angular units derived from the Thiele-Innes constants is exactly α , the signal-to-noise ratio is calculated as a_0/σ_{AL} .

2.2.1 Companion mass estimation

Once the astrometric signature α and orbital period P are determined, a rough order-of-magnitude estimate of the companion mass can be derived (Perryman 2018). The physical semi-major axis of the host star’s reflex motion is obtained with

$$a_\star = a_0 \cdot d$$

where a_0 is in mas and d is the distance in parsecs, giving $a_\star$ in astronomical units. The semi-major axis of the companion orbit a_p is then derived from Kepler’s third law:

$$a_p = (P^2 \cdot M_\star)^{1/3}$$

where P is expressed in years, $M_\star$ is the stellar mass in solar masses, and a_p is returned in astronomical units. By conservation of momentum, the star and companion orbit their common barycentre such that $M_\star a_\star = m_p a_p$, giving the companion mass as

$$m_p = M_\star \cdot \frac{a_\star}{a_p}$$

This estimate requires an independent determination of the stellar mass $M_\star$, obtained from spectral type or stellar models. In this work, stellar masses are taken from literature values where available. If not available, they are instead estimated from spectral type, following Pecaut & Mamajek (2013).

2.3 Bayesian inference

Bayesian inference provides a method for estimating model parameters by combining prior knowledge with observational data. This is done through Bayes’ theorem, which relates the posterior probability distribution of some parameters to the likelihood of the data and the prior distribution as

$$P(\Theta|D) = \frac{P(D|\Theta)P(\Theta)}{P(D)}.$$

Here, Θ represents the set of parameters to be explored, and D denotes the observed data. The term $P(\Theta|D)$ is the posterior distribution, $P(D|\Theta)$ is the likelihood, $P(\Theta)$ is the prior, and $P(D)$ is the evidence, which acts as a normalization constant. In practice, the evidence is often difficult to compute, and therefore parameter estimation focuses on sampling from the posterior distribution up to a proportionality constant (Speagle 2019).

In complex cases, the posterior distribution can be impossible to compute analytically. This problem can be overcome by the use of numerical methods, such as Markov Chain Monte Carlo, to generate samples from the posterior distribution.

2.4 Markov Chain Monte Carlo

In this project, Markov Chain Monte Carlo (MCMC) methods are employed using the python package `emcee` (Goodman & Weare 2010; Foreman-Mackey et al. 2013). MCMC methods are widely used in Bayesian statistics when sampling from complex probability distributions. The goal is to, using Bayes theorem, determine the posterior probability distribution of a set of variables, which in this case are the orbital parameters. MCMC combines two methods; Monte Carlo sampling and Markov chains.

Monte Carlo sampling

Monte Carlo methods are numerical techniques that use random sampling in order to estimate certain quantities or parameters that are difficult to compute analytically. Instead of solving a problem through deterministic calculations, the method is used to instead approximate the desired quantity through generation of random samples and averaging the result. This method is particularly useful for approximating integrals, sampling from probability distributions, or exploring complex parameter spaces (Katzgraber 2009).

The basic idea is rooted in the law of large numbers, where if provided with enough samples from a probability distribution, the median of the samples converges to the true value. For example, say we intend to estimate the integral of a function $f(x)$ with respect to a probability distribution $p(x)$

$$I = \int_a^b f(x)p(x)dx.$$

With sufficiently large N this can be estimated as

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i)$$

as the average of the samples converge to the true value. The accuracy of this convergence improves roughly as $1/\sqrt{N}$, so an increasing sample size reduces the statistical uncertainty of the result (Katzgraber 2009).

Markov chains

A Markov chain is a stochastic process in which the probability of moving from one state to another depends solely on the current state of the system and not on the history of previous steps. This property, known as the Markov property, implies that the future evolution of states is independent of the past state once the current state is known (Katzgraber 2009).

In a Markov chain, the system steps through a sequence of states, often denoted $x_1, x_2, \dots, x_n$, where each transition between states is controlled by a probability distribution. The probability of transitioning from state x to a new state x' is described by a transition probability $P(x \rightarrow x')$.

The collection of these transition probabilities determines the dynamics of the chain and how the exploration of the state space is conducted (Katzgraber 2009).

Emcee

As briefly stated, we make use of the python package `emcee`, in which an affine-invariant ensemble sampler is implemented. A standard MCMC method, such as the Metropolis algorithm, uses an individual walker to explore the parameter space. The `emcee` package utilizes an ensemble of walkers on which the Markov chain is defined, instead of a single state. This algorithm evolves a set of samples simultaneously, where each walker is updated based on the position of the others. New positions are proposed using the so called stretch move, where for a walker X_k a new position is proposed using another walker X_j according to

$$Y = X_j + Z(X_k - X_j)$$

where Z is a random scaling variable. The proposed move is then accepted or rejected based on its probability using the Metropolis-Hastings rule (Goodman & Weare 2010).

Despite each walker being based on the position of others, the Markov property is still satisfied, as each state of the ensemble depends only on its current configuration and not its past.

Chapter 3

Method

3.1 Candidate selection

The stellar systems analyzed in this project are all documented in the ESA Gaia DR3 Non-Single Star (NSS) catalogue, specifically in the `nss_two_body_orbit` table. This table contains stars for which a two-body orbital solution has been fitted, as described in Holl et al. (2023). The chosen systems were individually tested and selected by hand. Four systems were picked from an exoplanet candidate list published on the ESA Gaia cosmos portal, derived from Marcussen & Albrecht (2023), which compiles systems for which the observed wobble detected by Gaia is consistent with a planetary companion. A fifth candidate was also picked and is a well-documented binary star system with known Gaia DR3 NSS orbital solutions with epoch astrometry.

The selection was motivated to contain a range of varying orbital geometries, including both near edge-on and near face-on configurations, combined with other varying parameters. This is to explore the effect on the accuracy of the fitting as a result of distinct mixes of parameters and dissimilar orbits.

3.2 Astrometric and Orbital data

For each target, the five-parameter astrometric solution $(\alpha, \delta, \mu_{\alpha*}, \mu_{\delta*}, \varpi)$ was retrieved from `gaiadr3.gaia_source` using the Gaia archive query interface, together with the full 5 by 5 covariance matrix constructed from the reported uncertainties and correlation coefficients. Orbital parameters (P, e, T_0, A, B, F, G) were retrieved from `gaiadr3.nss_two_body_orbit` and served as the true parameters to which the recovered parameters were compared. This paper expresses T_0 in offset in days from the reference epoch Julian Date (JD) 2457388.5 (1 January 2016).

3.3 Simulated epoch data

Rather than using real Gaia epoch astrometry, which will not be publicly available until DR4 in December of 2026, the epoch data were instead simulated based on the orbital solution for each target. Two sets of simulations are run: a five-year mission with $N = 70$ simulated transits, and a ten-year mission with $N = 140$ simulated transits. The specific time of simulated transits was uniformly randomized over the duration of the mission, and the scan angles were drawn uniformly over the interval $[0, \pi)$. The true along-scan orbital signal was computed at each epoch using the Gaia-published Thiele-Innes constants, and Gaussian noise of $\sigma_{AL} = 0.1$ mas per epoch was added,

representative of the astrometric precision of moderately bright stars. Ranalli et al. (2018) report $\sigma_{AL} = 0.034$ mas for $V \leq 14$; the adopted value is therefore a conservative approximation for the targets analyzed in this paper, which lie in range $G \approx 4 - 15$.

Given the value of $\sigma_{AL} = 0.1$ mas, the signal-to-noise ratios are calculated, following section 2.2, and presented in section 4.1.

3.4 Astrometric uncertainty propagation

The uncertainty in the Gaia five-parameter astrometric solution $(\alpha, \delta, \mu_\alpha, \mu_\delta, \varpi)$ contributes to the along-scan measurements at each epoch. This contribution was computed by projecting the 5×5 astrometric covariance matrix $\mathbf{C}$ onto the along-scan direction via the design vector $\mathbf{h}_i$, following Brandt (2021)

$$\sigma_{astro,i}^2 = \mathbf{h}_i^\top \mathbf{C} \mathbf{h}_i$$

with $\mathbf{h}_i$ constructed as

$$\mathbf{h}_i = [\sin \theta_i, \cos \theta_i, \Delta t_i \sin \theta_i, \Delta t_i \cos \theta_i, \Pi_{\alpha,i} \sin \theta_i + \Pi_{\delta,i} \cos \theta_i].$$

Here, θ_i is the scan angle at epoch i , Δt_i is the time in years passed since the reference epoch, and $\Pi_{\alpha,i}, \Pi_{\delta,i}$ are parallax factors. The parallax factors describe the displacement along the sky due to the Earth changing position in its orbit around the sun, computed from the barycentric position of the Earth at each observation epoch, following Lindegren et al. (2021).

3.5 MCMC implementation

The seven orbital parameters were inferred using the Bayesian- and MCMC methods described in sections 2.3 and 2.4. The likelihood is defined as Gaussian, as the residuals between the observed and modelled along-scan measurements follow a normal distribution with zero mean and standard deviation, defined as $\sigma_{total,i} = \sqrt{\sigma_{AL,i}^2 + \sigma_{astro,i}^2}$. The Gaussian log-likelihood was

$$\ln P(D|\Theta) = -\frac{1}{2} \sum_i \left[\left(\frac{\xi_i - \xi_{orb,i}}{\sigma_{total,i}} \right)^2 + 2 \ln \sigma_{total,i} \right]$$

where $D = \{\xi_i\}$ is the set of along-scan observations. A log-uniform prior (Jeffreys prior) was set on P , and a uniform prior in $[0, 1)$ was set on e , with the other parameters unconstrained. The sampler was run with 64 walkers. An initial burn-in phase was run to locate the high probability region, after which the walkers were then again started from the best-fit position to initialize the production phase.

3.5.1 Perturbations of values

To guarantee that the MCMC had a genuine search to do for parameters, the initial starting point for the MCMC were in some cases perturbed from the true values, which tests the samplers ability to recover the true parameter. With the sampler searching for 7 variables, there can be a multitude of combinations of perturbations. This project focuses mainly on period and eccentricity perturbations.

3.6 Posterior analysis

A posterior analysis of the fitted parameters was performed by computing the 16th, 50th and 84th percentiles and comparing these to the true parameters. The median (50th percentile) serves as the point estimate, and the 16th and 84th percentiles mark the credible interval of values based on the central 68% of the posterior probability, matching a 1σ uncertainty for a Gaussian distribution. The Thiele-Innes parameters were then also converted back into the Campbell orbital elements (a , ω , Ω , i) to allow for their posterior analysis and comparison to the true values.

Using the recovered a_0 values, system distances, and stellar mass estimations, the companion mass estimates are derived following the theory of subsection 2.2.1.

3.7 Convergence

The convergence of the MCMC chains is assessed using the integrated autocorrelation time τ (Goodman & Weare 2010; Foreman-Mackey et al. 2013), which measures how many steps are required for consecutive samples in a chain to become statistically independent. For a chain of N_{prod} production steps, convergence is considered to be achieved when

$$N_{\text{prod}} > 50 \tau_{\text{max}}, \quad (3.1)$$

where τ_{max} is the largest autocorrelation time across all parameters. This ensures that the chain contains at least 50 independent samples per parameter, making the posterior statistically reliable.

3.8 Testing

In order to evaluate the precision of the method when extracting exoplanet orbital information using Gaia astrometric data combined with the MCMC implementation, several tests are performed. Each test isolates a specific aspect of the fitting procedure’s behavior, allowing the quality of the recovery of the orbital parameters to be determined and any systematic limitations of the method to be identified.

3.8.1 Five-year and ten-year comparison

Using the five selected candidates mentioned in section 3.1, the fitting procedure was applied to both the five-year and ten-year simulated missions for each system. This allows for a direct comparison of the precision and accuracy of the recovered parameters between the two mission lengths. Therefore, it evaluates the impact of the use of a larger number of transits provided by the ten-year mission, noting improved precision and reduced bias in the recovered parameters.

3.8.2 Systematic injection-recovery grid

To address how well the system recovers certain orbital parameters, one-dimensional injection-recovery sweeps were performed on the host star HD164604, which has defined Gaia- and `nss_two_body_orbit` data, meaning it has a known exoplanet. However, instead of using the already defined exoplanet data, a planetary system with fabricated orbital parameters was simulated. The parameters recovered in three separate tests are the period, eccentricity, and inclination. For each test, the

investigated parameter was varied in an interval, while the other parameters were kept fixed, with details shown in Table 3.1.

	P	e	i	a_0	ω	Ω	T_0
Test 1	100 – 1000d	0.3	45°	1mas	0°	0°	0d
Test 2	500d	0 – 1	45°	1mas	0°	0°	0d
Test 3	500d	0.3	0 – 90°	1mas	0°	0°	0d

Table 3.1: The values of parameters used in the injection-recovery testing.

The period and eccentricity were also perturbed in two separate runs, one initializing the MCMC at period $1.1P_{\text{true}}$, and one initializing at eccentricity $e = 0.01$. For each procedure, the MCMC was run with 64 walkers, 2000 burn-in steps, and 4000 production steps. The recovered parameters were then compared to the true ones, similar to the approach of Ranalli et al. (2018), and the bias in recovered parameters was noted.

3.8.3 Period sweep

To investigate the ability of the MCMC to recover the correct orbital period of a system, the initial starting position of the MCMC away from the true value was perturbed. Rather than testing a single perturbation, a sweep was performed over a range of starting periods, from which the precision of the recovery can be evaluated as a function of the initial offset from the true period. This procedure was applied to three systems, defined in Table 3.2 along with the swept-over period intervals and step size used.

Source id	Start	End	Increments
1386158505321789696	$0.5P$	$2.0P$	$0.02P$
4062446910648807168	$0.5P$	$2.0P$	$0.02P$
423297927866697088	$0.5P$	$2.0P$	$0.02P$

Table 3.2: The systems used in the period perturbation sweep with corresponding source id, and the intervals and step-increments used in the sweep.

Each procedure used a five-year mission with 64 walkers, 2000 burn-in steps, and 4000 production steps.

3.8.4 Convergence robustness

With the MCMC initialized using a fixed random seed and a set number of walkers, the results could in theory depend on these choices rather than reflecting the true posterior. To verify the robustness of these settings, tests were performed on a well-behaved system, candidate 1878822452815621120. The MCMC was run using three different random seeds (42, 121, 739) and three different number of walkers (32, 64, 128), with all other settings fixed, combined creating nine individual runs. For each run, the results were compared in a plot of recovered parameters and error margins were tabulated. Each run used a five-year mission with 3000 burn-in steps and 5000 production steps.

Chapter 4

Results

4.1 Parameter recovery for selected candidates

The five candidates selected in section 3.1 are briefly described individually in the following subsections, with Gaia catalogue parameters and MCMC recovery results presented for both the five-year and ten-year simulated missions, with the MCMC using 3000 burn-in steps and 5000 production steps in all cases. For each candidate; burn-in traces, production traces, corner plots, residuals plots, and sky-orbit projections were plotted for the five-year missions. To conserve space these are presented in Appendix A.

4.1.1 Candidate 1386158505321789696

Candidate **1386158505321789696**, also known as **Phi Herculis** (ϕ Her) is a well-known binary star system located approximately 74 parsecs from Earth, with a Gaia G-band magnitude of 4.19 making it the brightest of the five candidates. The system has a nearly face-on orbital configuration with a large photocentric semi-major axis, yielding a large signal-to-noise ratio of $S/N = 82.6$. The Gaia catalogue orbital parameters are presented in Table 4.1, together with the MCMC-recovered parameters.

Parameter	Gaia catalogue	5-year	10-year
P (days)	554.4	$554.7^{+2.2}_{-2.2}$	$554.3^{+1.5}_{-1.4}$
e	0.474	$0.474^{+0.014}_{-0.013}$	$0.474^{+0.013}_{-0.013}$
T_0 (days)	78.5	$78.2^{+2.1}_{-2.2}$	$78.7^{+3.7}_{-3.8}$
A (mas)	-8.23	$-8.22^{+0.20}_{-0.19}$	$-8.23^{+0.21}_{-0.21}$
B (mas)	-0.63	$-0.63^{+0.21}_{-0.21}$	$-0.66^{+0.22}_{-0.21}$
F (mas)	0.66	$0.67^{+0.39}_{-0.38}$	$0.65^{+0.44}_{-0.47}$
G (mas)	-7.61	$-7.63^{+0.28}_{-0.28}$	$-7.64^{+0.25}_{-0.26}$
Derived Campbell elements			
a_0 (mas)	8.26	$8.31^{+0.20}_{-0.19}$	$8.33^{+0.21}_{-0.20}$
i (deg)	4.5	$5.1^{+2.3}_{-2.2}$	$5.2^{+2.3}_{-2.2}$
ω (deg)	184.1	$209.1^{+142.7}_{-21.8}$	$210.6^{+140.4}_{-23.5}$
Ω (deg)	0.6	$54.4^{+118.5}_{-47.6}$	$37.3^{+135.7}_{-29.4}$
Converged ($N_{\text{prod}} > 50 \tau_{\text{max}}$)		✓	✓

Table 4.1: Gaia catalogue values and MCMC recovered parameters (median $^{+84\text{th}}_{-16\text{th}}$ percentile) for candidate 1386158505321789696 for the five-year and ten-year simulated missions. Parameters a_0 , i , ω and Ω are derived from the Thiele-Innes constants and are not directly fitted. Convergence criterion follows section 3.7.

The MCMC recovers the orbital parameters and Thiele-Innes constants with great precision for both mission lengths. The period is recovered within 0.3 days of the true value in both cases, the eccentricity value agrees with the median of both missions to three decimal places with small uncertainties, and time of periastron passage stays within 0.3 days of the true value. The Thiele-Innes constants A and G dominate the signal and are well constrained, while F shows larger uncertainties. The derived inclination of $i \approx 5^\circ$ is accurate, although with notable uncertainty of ± 2.2 . The angles ω and Ω are essentially unconstrained, with uncertainties spanning nearly the full range of possible values. The ten-year mission here offers no notable improvement over the five-year mission.

4.1.2 Candidate 4062446910648807168

Candidate **4062446910648807168**, also known as **HD164604** is a star located approximately 40 parsecs from Earth, having a Gaia G-band magnitude of 9.33 with a confirmed exoplanet companion (**HD164604b**). The system has $S/N = 5.7$, and a nearly face-on orbital configuration with an eccentricity of 0.615, being the most eccentric of the candidates. The Gaia catalogue values and MCMC recovered parameters are presented in Table 4.2.

Parameter	Gaia catalogue	5-year	10-year
P (days)	615.5	$617.5^{+5.7}_{-5.7}$	$615.2^{+2.2}_{-2.2}$
e	0.615	$0.645^{+0.047}_{-0.041}$	$0.620^{+0.034}_{-0.032}$
T_0 (days)	-193.6	$-204.7^{+14.7}_{-14.3}$	$-195.6^{+7.4}_{-7.4}$
A (mas)	-0.49	$-0.49^{+0.03}_{-0.03}$	$-0.48^{+0.02}_{-0.02}$
B (mas)	-0.26	$-0.23^{+0.03}_{-0.03}$	$-0.25^{+0.02}_{-0.02}$
F (mas)	0.19	$0.14^{+0.07}_{-0.07}$	$0.16^{+0.05}_{-0.05}$
G (mas)	-0.47	$-0.51^{+0.05}_{-0.06}$	$-0.49^{+0.04}_{-0.04}$
Derived Campbell elements			
a_0 (mas)	0.565	$0.591^{+0.037}_{-0.032}$	$0.580^{+0.027}_{-0.026}$
i (deg)	7.7	$12.2^{+6.3}_{-5.7}$	$11.5^{+4.8}_{-4.6}$
ω (deg)	337.8	$314.7^{+22.9}_{-24.7}$	$321.1^{+16.3}_{-15.8}$
Ω (deg)	47.5	$59.6^{+42.1}_{-24.3}$	$60.0^{+13.9}_{-12.9}$
Converged ($N_{\text{prod}} > 50 \tau_{\text{max}}$)		✓	✓

Table 4.2: Gaia catalogue values and MCMC recovered parameters (median $^{+84\text{th}}_{-16\text{th}}$ percentile) for candidate 4062446910648807168 (HD164604b).

The MCMC recovers the parameters and Thiele-Innes constants well in both missions, with slight improvement using the ten-year mission for most values. The period is recovered within 2 days and 0.3 days for the five-year and ten-year missions respectively. The eccentricity is slightly over-estimated in both cases but within good range of the true value, similar to the recovered time of periastron passage. The Thiele-Innes constants A , B and G are consistently recovered, with a marginally worse value and error for F . The derived inclination in both missions with $i \approx 12^\circ$ is somewhat larger than the Gaia catalogue value of $i = 7.7^\circ$, with uncertainties $\pm 5 - 6^\circ$. ω and Ω both yield somewhat accurate values, with larger uncertainties for the five-year mission compared to the ten-year mission.

4.1.3 Candidate 423297927866697088

Candidate **423297927866697088** is a star located approximately 79 parsecs from Earth, with no common catalogue designation beyond the Gaia DR3 source-id. It has a Gaia G-band magnitude of 14.77 and an unconfirmed exoplanet companion. The system has S/N = 5.7, is nearly edge-on with $i = 89.0^\circ$ and has a very low eccentricity of 0.095, being a nearly circular orbit. The Gaia catalogue values and MCMC recovered parameters are shown in Table 4.3.

Parameter	Gaia catalogue	5-year	10-year
P (days)	491.9	$491.1^{+4.4}_{-4.2}$	$491.3^{+2.3}_{-2.2}$
e	0.095	$0.038^{+0.059}_{-0.025}$	$0.070^{+0.095}_{-0.051}$
T_0 (days)	55.0	$38.3^{+44.3}_{-78.8}$	$59.4^{+29.8}_{-28.2}$
A (mas)	0.02	$-0.08^{+0.24}_{-0.28}$	$0.02^{+0.16}_{-0.14}$
B (mas)	0.01	$-0.05^{+0.17}_{-0.21}$	$0.01^{+0.15}_{-0.11}$
F (mas)	0.44	$0.39^{+0.07}_{-0.31}$	$0.40^{+0.04}_{-0.08}$
G (mas)	0.37	$0.28^{+0.06}_{-0.22}$	$0.33^{+0.04}_{-0.06}$
Derived Campbell elements			
a_0 (mas)	0.574	$0.562^{+0.033}_{-0.034}$	$0.545^{+0.026}_{-0.026}$
i (deg)	89.0	$91.4^{+5.7}_{-5.7}$	$89.2^{+5.9}_{-6.4}$
ω (deg)	271.8	$261.5^{+38.4}_{-122.7}$	$271.9^{+23.9}_{-20.7}$
Ω (deg)	40.2	$35.8^{+3.9}_{-3.7}$	$39.9^{+3.6}_{-3.0}$
Converged ($N_{\text{prod}} > 50 \tau_{\text{max}}$)		×	×

Table 4.3: Gaia catalogue values and MCMC recovered parameters (median $^{+84\text{th}}_{-16\text{th}}$ percentile) for candidate 423297927866697088.

The period is recovered with great accuracy, within 0.8 days and 0.6 days of the true value, marking slight improvement using the ten-year mission. For the ten-year mission the eccentricity is recovered with greater accuracy, but with larger uncertainty, and the time of periastron passage value is better, with its uncertainty shrunk by 53%. For the Thiele-Innes constants, we see a poor recovery with the five-year mission. The constants are all ill-recovered with large uncertainties, whereas with the ten-year mission they are rather well recovered, although still with significant but improved error in A and B . Semi-major axis and inclination both show acceptable recovery with uncertainty $\pm \approx 0.03$ and $\pm \approx 6^\circ$ respectively. ω has large error with the five-year mission, and shows improvement with the ten-year mission. Ω is accurately recovered in both cases with minor uncertainties. Neither mission meets the convergence criterion.

4.1.4 Candidate 1878822452815621120

Candidate **1878822452815621120**, with alternate name **HIP 110894** is a star with unconfirmed exoplanet companion, listed as a spectroscopic binary by Simbad. It is located approximately 62 parsecs from Earth and has a Gaia G-band magnitude of 9.76, with S/N = 5.5. The system is rather edge-on, inclined by 77.9° , with a large period of 1009 days. The full Gaia catalogue values and MCMC recovered parameters are presented in Table 4.4.

Parameter	Gaia catalogue	5-year	10-year
P (days)	1009.3	$1005.7^{+19.5}_{-19.6}$	$1010.8^{+9.4}_{-9.0}$
e	0.446	$0.448^{+0.035}_{-0.033}$	$0.472^{+0.043}_{-0.036}$
T_0 (days)	459.4	$473.3^{+31.7}_{-32.0}$	$427.8^{+27.6}_{-27.1}$
A (mas)	0.55	$0.53^{+0.04}_{-0.04}$	$0.53^{+0.04}_{-0.05}$
B (mas)	-0.02	$-0.02^{+0.02}_{-0.02}$	$-0.01^{+0.02}_{-0.02}$
F (mas)	0.08	$0.00^{+0.13}_{-0.13}$	$0.21^{+0.10}_{-0.11}$
G (mas)	0.06	$0.04^{+0.03}_{-0.03}$	$0.03^{+0.03}_{-0.03}$
Derived Campbell elements			
a_0 (mas)	0.552	$0.542^{+0.033}_{-0.033}$	$0.576^{+0.027}_{-0.026}$
i (deg)	77.9	$82.2^{+7.1}_{-7.0}$	$83.3^{+5.0}_{-5.5}$
ω (deg)	352.2	$40.6^{+314.5}_{-40.7}$	$337.1^{+10.9}_{-12.8}$
Ω (deg)	178.4	$176.3^{+1.1}_{-175.8}$	$175.9^{+1.9}_{-175.7}$
Converged ($N_{\text{prod}} > 50 \tau_{\text{max}}$)		✓	✓

Table 4.4: Gaia catalogue values and MCMC recovered parameters (median $^{+84\text{th}}_{-16\text{th}}$ percentile) for candidate 1878822452815621120.

The period is well recovered with median within 4 days and 2 days for the five-year and ten-year missions respectively, with larger uncertainty using the five-year mission. Eccentricity shows good recovery with uncertainty $\pm \approx 0.04$ for both missions, with the five-year mission having more accurate median, and time of periastron passage has uncertainties of $\pm \approx 30$ days, where the true value falls within these ranges for both missions. The Thiele-Innes constants A , B and G are well recovered for both missions, whereas F has large uncertainty with disagreeing medians for both missions, where the true value falls outside the recovered range for the ten-year mission. The semi-major axis and inclination are well recovered, with uncertainties $\pm \approx 0.03$ mas and $\pm 5 - 7^\circ$ respectively. ω is poorly recovered and unconstrained for the five-year mission, but is rather well recovered for the ten-year mission, and Ω shows great recovery in median but is unconstrained in both missions, spanning nearly the full range of angles.

4.1.5 Candidate 4698424845771339520

Candidate **4698424845771339520**, also known as **L88-59** is a white dwarf with unconfirmed exoplanet companion. It is located at a distance approximately 10 parsecs from Earth with a Gaia G-band magnitude of 13.70, and S/N = 2.3. The system is nearly edge-on with $i = 83.9^\circ$ and has an extremely low orbital period of 33.65 days. All Gaia catalogue values and MCMC recovered parameters are shown in Table 4.5.

Parameter	Gaia catalogue	5-year	10-year
P (days)	33.65	$33.67^{+0.04}_{-0.04}$	$33.63^{+0.02}_{-0.02}$
e	0.198	$0.194^{+0.172}_{-0.097}$	$0.182^{+0.115}_{-0.080}$
T_0 (days)	6.0	$5.9^{+2.2}_{-4.6}$	$6.8^{+2.0}_{-4.5}$
A (mas)	-0.04	$-0.06^{+0.03}_{-0.03}$	$-0.05^{+0.03}_{-0.03}$
B (mas)	0.13	$0.13^{+0.10}_{-0.08}$	$0.12^{+0.07}_{-0.07}$
F (mas)	0.04	$0.04^{+0.04}_{-0.07}$	$0.04^{+0.03}_{-0.05}$
G (mas)	-0.18	$-0.23^{+0.19}_{-0.06}$	$-0.17^{+0.14}_{-0.05}$
Derived Campbell elements			
a_0 (mas)	0.234	$0.291^{+0.029}_{-0.026}$	$0.233^{+0.022}_{-0.027}$
i (deg)	83.9	$79.0^{+9.5}_{-9.0}$	$79.0^{+11.4}_{-10.5}$
ω (deg)	52.7	$67.6^{+203.3}_{-32.8}$	$60.0^{+200.5}_{-31.2}$
Ω (deg)	106.0	$106.2^{+5.4}_{-5.4}$	$107.9^{+6.4}_{-5.8}$
Converged ($N_{\text{prod}} > 50 \tau_{\text{max}}$)		×	×

Table 4.5: Gaia catalogue values and MCMC recovered parameters (median $^{+84\text{th}}_{-16\text{th}}$ percentile) for candidate 4698424845771339520.

The recovered parameters are for the most part accurate throughout. The period is recovered with a remarkable precision of 0.02 days, or approximately 30 minutes, while eccentricity experiences a large uncertainty range of $\pm \approx 0.1$ for both missions. The Thiele-Innes constants A and F are well recovered with small uncertainties, while B and G show worse precision, with slight improvement for the ten-year mission. The semi-major axis is for the five-year recovery not within range, and sees great improvement in accuracy of the median for the ten-year mission. The angular Campbell elements stay consistent for the five-year and ten-year mission, with the inclination being rather accurate, ω showing acceptable median but with large uncertainty, and Ω being near spot-on with small uncertainties. Neither mission meets the convergence criterion.

4.2 Companion mass estimates

Using the method described in subsection 2.2.1, order-of-magnitude companion mass estimates are derived for each candidate using the recovered a_0 and P values from the ten-year mission, the system distances from the Gaia catalogue, and stellar masses from the following sources: literature values for candidates 1386158505321789696, 4062446910648807168 and 4698424845771339520 (Torres 2007; Arriagada et al. 2010; Subasavage et al. 2017), the Gaia DR3 FLAME module for candidate 1878822452815621120 (Gaia Collaboration et al. 2023), and for candidate 423297927866697088, estimated from its Gaia DR3 BP-RP color of 2.71, corresponding to spectral type M3–M3.5 with $M_\star \approx 0.30 M_\odot$ following Pecaut & Mamajek (2013) with related online table¹. The results are presented in Table 4.6, with companion masses expressed in solar masses ($M_\odot$) or Jupiter masses (M_J).

No uncertainties are reported as the estimates are order-of-magnitude, only intended to characterize the properties of each host star and companion rather than provide precise mass calculations.

¹http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt, version 2022.04.16

Candidate	d (pc)	$M_\star$ ($M_\odot$)	$a_\star$ (AU)	a_p (AU)	m_p
1386158505321789696	73.66	3.05^L	0.61	1.92	$0.97 (M_\odot)$
4062446910648807168	40.02	0.80^L	0.02	1.31	$14.80 (M_J)$
423297927866697088	79.18	0.30^C	0.04	0.82	$16.63 (M_J)$
1878822452815621120	62.29	0.80^F	0.04	1.83	$16.40 (M_J)$
4698424845771339520	9.71	0.57^L	0.002	0.17	$7.05 (M_J)$

Table 4.6: Order-of-magnitude companion mass estimates for the five candidates, derived from the ten-year recovered a_0 and P values following subsection 2.2.1.

^L literature value, ^F Gaia DR3 FLAME module, ^C estimated from Gaia DR3 BP-RP color. Note that candidate 1386158505321789696 is a stellar companion, and therefore expressed in solar masses.

4.3 Injection-recovery

This section presents the results of the injection-recovery tests described in subsection 3.8.2, using parameter values presented in Table 3.1. For each investigated orbital parameter, the recovery performance is evaluated by comparing the median of the posterior distribution to the true value, with bias defined as $\theta_{true} - \theta_{med}$ and scatter as $\pm(\theta_{84} - \theta_{16})/2$. All tests use the host star HD164604, with per-epoch measurement noise $\sigma_{AL} = 0.1$ mas and a photocentric semi-major axis $a_0 = 1.0$ mas, along with $\omega = \Omega = 0^\circ$ and $T_0 = 0$ days. The MCMC for each test uses 2000 burn-in steps and 4000 production steps. Throughout, two mission lengths are considered: the five-year mission ($N = 70$ transits) and the extended ten-year mission ($N = 140$ transits).

4.3.1 Period recovery

Here, four runs are performed: a standard five-year and ten-year run, and a five-year and ten-year run with walkers initialized at a period 10% above the true value ($P_{start} = 1.1 P_{true}$). The plots generated present the simulated "true" period vs the bias in recovered period ($\Delta P = P_{true} - P_{recovered}$), shown in Figure 4.1.

The period is well-recovered across the full range of 100-1000 days in all four considered cases, with the bias staying within acceptable range of the zero-bias. The average bias in all cases, along with worst recovered case are presented in Table 4.7. This shows that the ten-year mission recovers more precise periods both on average and in worst recovery.

	5-year		10-year	
	No pert.	+10% P	No pert.	+10% P
$\langle \Delta P \rangle$ (% P_{true})	0.282	0.760	0.188	0.521
$\max \Delta P $ (% P_{true})	0.818	14.029	0.417	12.232
Convergence (%)	98.9	92.3	96.7	94.5

Table 4.7: The average recovery bias and worst recovery bias for each of the four tests expressed as percentages of the true period, along with a convergence percentage of the 91 explored cases in each test. All runs, whether converged or not, are included in the bias-calculations.

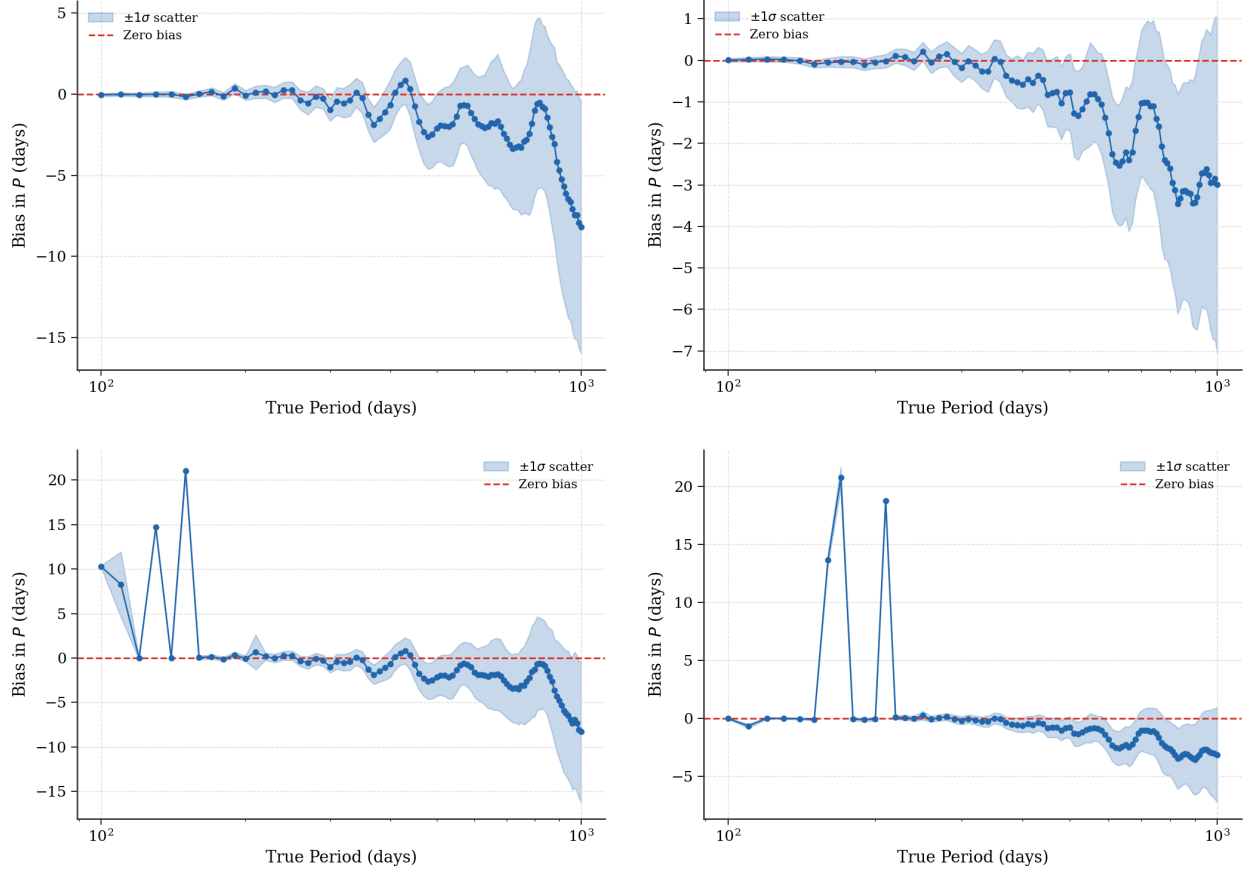


Figure 4.1: Simulated "true" period vs. bias in recovered period ($P_{\text{true}} - P_{\text{recovered}}$). Left and right plots show five-year and ten-year missions respectively. Upper and lower plots have different MCMC starting positions; the upper starting at the true period, and lower starting at $1.1P_{\text{true}}$. Each point represents an individual run with corresponding scatter error (16th-84th percentile).

In the majority of cases when recovering the period without perturbation, the MCMC converges, with only a few instances of non-convergence. There is also a higher percentage of converged recoveries without a perturbation compared to with a perturbation. This is in the worst cases due to the MCMC when mispositioned exploring other probability densities, not finding the largest probability density corresponding to the true value. In other cases it finds the correct value eventually, but then lacks enough steps to converge. The clearly poorly recovered periods in the plots are results of the former, and the lower convergence percentage of the ten-year compared to the five-year mission without perturbation is a result of the latter.

4.3.2 Eccentricity recovery

Four runs are performed: a standard five-year and ten-year run, and a five-year and ten-year run with walkers initialized at $e = 0.01$, assuming a nearly circular orbit. The plots generated show the simulated "true" eccentricity vs. the bias in recovered eccentricity ($\Delta e = e_{\text{true}} - e_{\text{recovered}}$), presented in Figure 4.2.

The eccentricity is accurately recovered with no large discrepancies in the standard five-year and

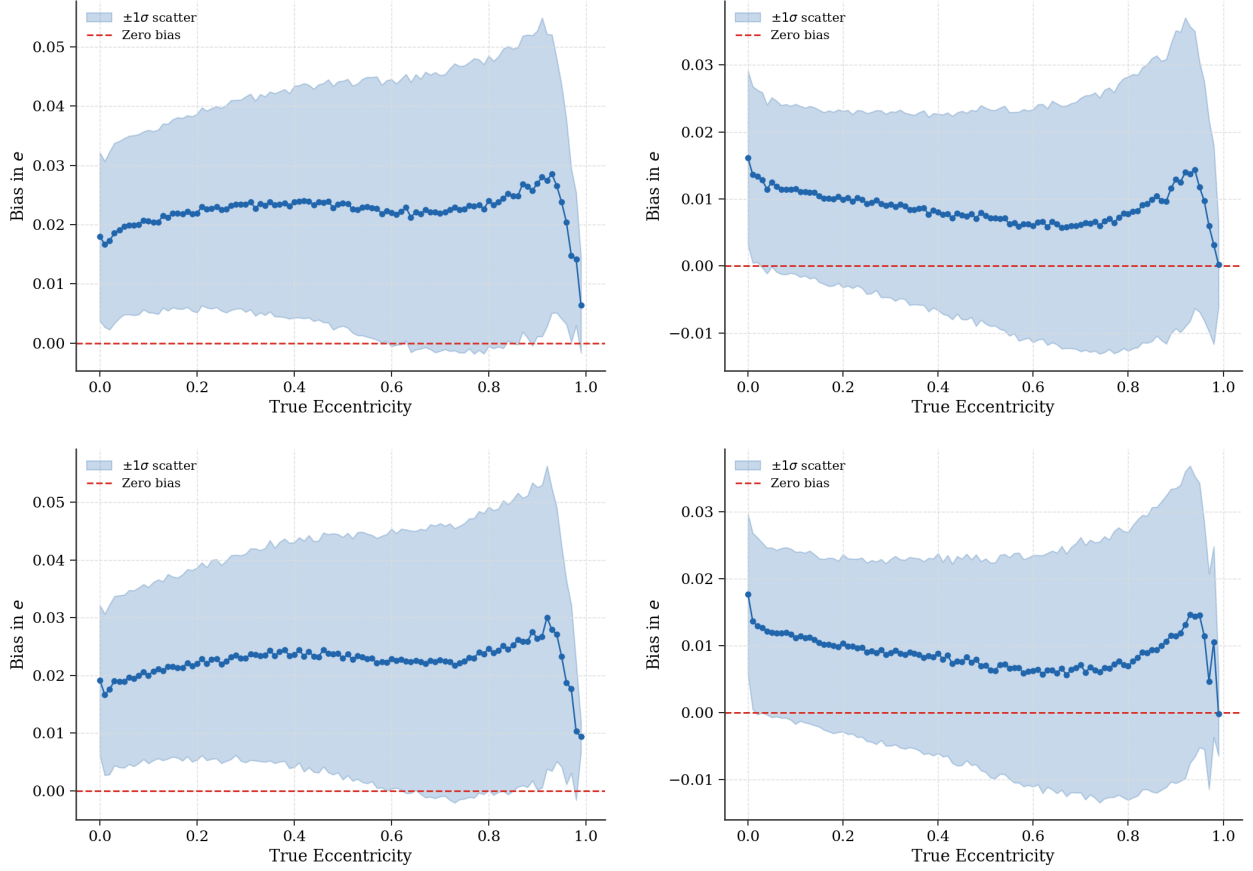


Figure 4.2: Simulated ”true” eccentricity vs. bias in recovered eccentricity ($e_{\text{true}} - e_{\text{recovered}}$). Left and right plots show five-year and ten-year missions respectively. Upper and lower plots have different MCMC starting positions; the upper starting at true eccentricity, and lower starting at $e = 0.01$, assuming circular orbit.

ten-year cases, although with a constant bias of $\Delta e \approx +0.01 - 0.02$ over all eccentricities. The average- and worst bias in the four tests are presented in Table 4.8, showing an improvement in the eccentricity recovery using the ten-year mission. The perturbed recoveries yield marginally worse values, but there is no readily apparent difference in the quality of recovery between the cases of no perturbation and the ones starting at $e = 0.01$.

	5-year		10-year	
	No pert.	$e_{\text{start}} = 0.01$	No pert.	$e_{\text{start}} = 0.01$
$\langle \Delta e \rangle$	0.0225	0.0226	0.0088	0.0089
$\max \Delta e $	0.0286	0.0301	0.0161	0.0177
Convergence (%)	76	77	78	79

Table 4.8: The average and worst recovery bias for each of the four tests, along with a convergence percentage of the 100 explored cases in each test. All runs, whether converged or not, are included in the bias-calculations.

In all cases considered, approximately 3/4 cases meet the convergence criterion, where a clear majority of non-converging cases occur at eccentricities in the ranges $e_{\text{true}} \approx 0 - 0.1$ and $e_{\text{true}} \approx 0.9 - 1$. This test demonstrates no clear improvement in convergence using the longer mission.

4.3.3 Inclination recovery

Two runs are performed: a standard five-year mission and a standard ten-year mission. The plots generated show true vs. recovered inclination, and true vs. bias in recovered inclination ($\Delta \cos i = \cos i_{\text{true}} - \cos i_{\text{recovered}}$), presented in Figure 4.3.

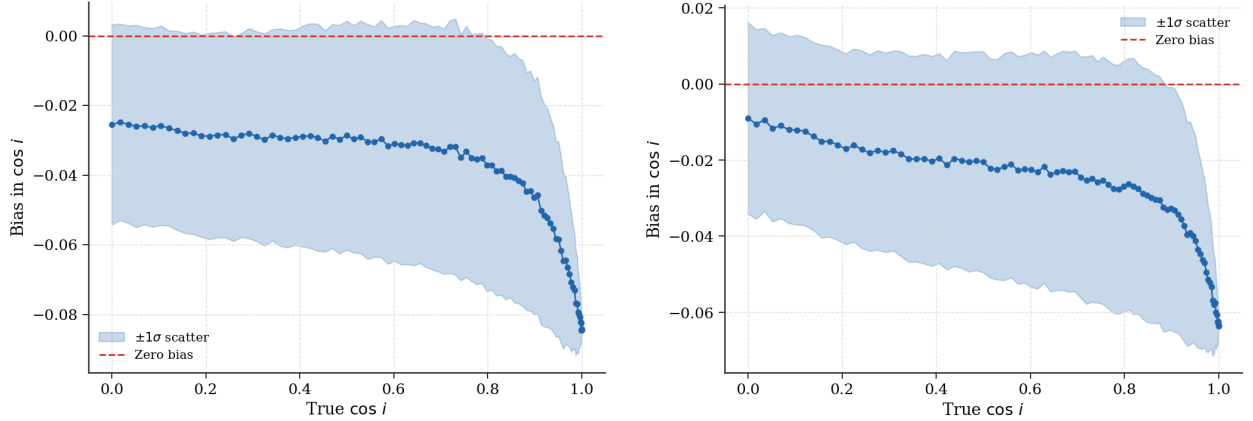


Figure 4.3: The true inclination ($\cos i_{\text{true}}$) plotted against inclination bias ($\Delta \cos i = \cos i_{\text{true}} - \cos i_{\text{recovered}}$). Left plot shows the five-year mission recovery and right plot the ten-year missions.

The plots illustrate an overestimation in the recovered inclinations compared to the true values, most apparent in smaller inclinations. For the five-year mission, the first inclination $i_{\text{true}} = 0^\circ$ ($\cos i = 1$) recovers $i_{\text{recovered}} = 23.726^\circ$ ($\cos i = 0.915$) and possesses the largest bias, with the same sample for the ten-year mission giving $i_{\text{recovered}} = 20.551^\circ$ ($\cos i = 0.936$). Continuing, with each increasing step in true inclination, the bias on average decreases for both missions. The overestimation in inclination is accompanied by an overestimation in the semi-major axis. Results for selected inclinations with respective recoveries, along with respective recovered semi-major axis, are shown in Table 4.9.

	5-year				10-year			
$i_{\text{true}} (^\circ)$	$i_{\text{rec}} (^\circ)$	$\Delta i (^\circ)$	$a_{0,\text{rec}} (\text{mas})$	$\Delta a_0 (\text{mas})$	$i_{\text{rec}} (^\circ)$	$\Delta i (^\circ)$	$a_{0,\text{rec}} (\text{mas})$	$\Delta a_0 (\text{mas})$
0	23.73	+23.73	1.036	+0.036	20.55	+20.55	1.030	+0.030
30	34.48	+4.48	1.012	+0.012	33.31	+3.31	1.013	+0.013
60	61.87	+1.87	1.004	+0.004	61.35	+1.35	1.008	+0.008
90	91.46	+1.46	1.003	+0.003	90.51	+0.51	1.006	+0.006

Table 4.9: Recovered inclination and photocentric semi-major axis at selected true inclinations for the five-year and ten-year simulated missions. Note that the large degree bias near $i = 0^\circ$ appears smaller when expressed in $\cos i$.

4.4 Period perturbation sweep

This section presents the results of the period perturbation sweep described in subsection 3.8.3. The sweep is performed over a five-year mission on three systems, one of which is a known binary star. The sweep only simulates a perturbation in the period, while all other parameters are initialized from the Gaia `nss_two_body_orbit` catalogue values. The recovery of the period is evaluated by generating plots of the perturbation factor $P_{\text{start}}/P_{\text{true}}$ against the resulting bias $\Delta P = \hat{P} - P_{\text{true}}$ in percent.

4.4.1 Candidate 1386158505321789696 sweep

This source is the binary star system. The recovery results are shown in Figure 4.4. A section of correct recovery is visible in the range $P_{\text{start}}/P_{\text{true}} = 0.68 - 1.58$, where the MCMC recovers the true period regardless of the initial perturbation. Outside of this range the period is incorrectly recovered. However, in the bottom range of $P_{\text{start}}/P_{\text{true}} = 0.52 - 0.66$ there is a consistency where the MCMC recovers $P \approx 0.62P_{\text{true}}$, and in the upper range of $P_{\text{start}}/P_{\text{true}} = 1.60 - 2.00$ the MCMC consistently recovers $P \approx 2P_{\text{true}}$. All cases, except $P_{\text{start}}/P_{\text{true}} = 0.50$, have an approximate MCMC acceptance fraction of 0.49, suggesting the respective posterior peak in which the walkers are located was adequately explored.

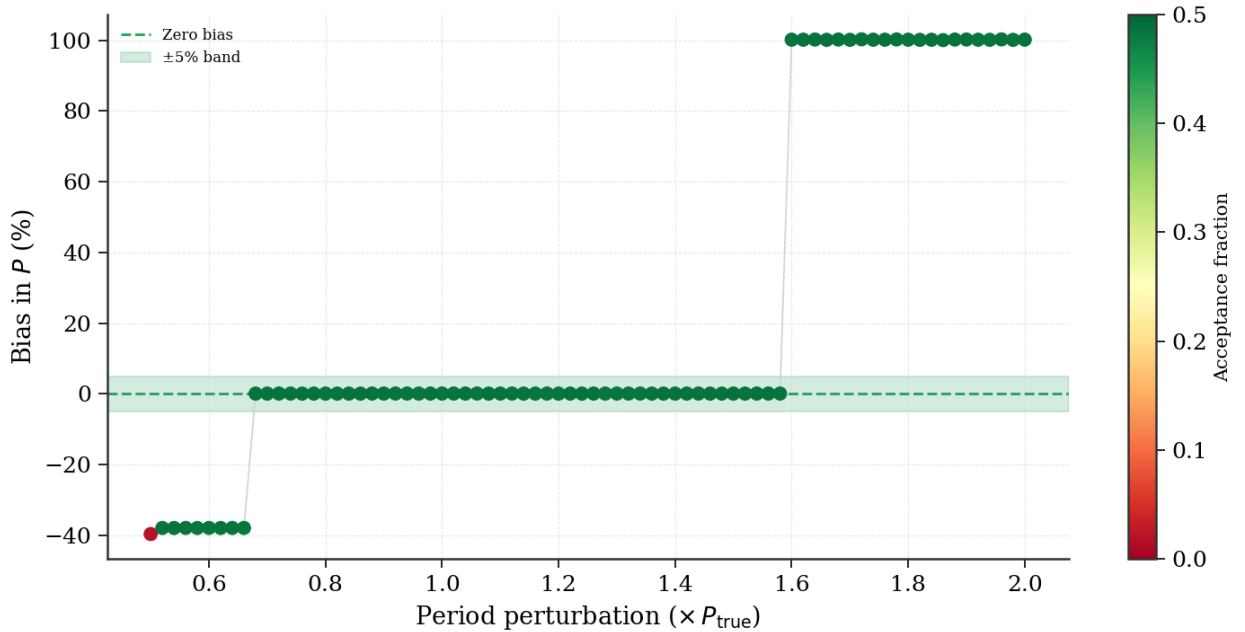


Figure 4.4: Period perturbation as a fraction of P_{true} plotted against the bias in percent in recovered period over a five-year mission, for source 1386158505321789696. The green horizontal dashed line marks zero bias, and marked green region indicates $\pm 5\%$ bias. The acceptance fraction is indicated by color.

4.4.2 Candidate 423297927866697088 sweep

The recovery results of this exoplanet system are shown in Figure B.1. A great majority of perturbations are correctly recovered, all in the region of $P_{\text{start}}/P_{\text{true}} = 0.66 - 2.00$. Each recovered

period in this range have corresponding acceptance fraction of approximately $0.22 - 0.33$ with a mean of 0.2816 , which again demonstrates the posterior being sufficiently explored. In range $P_{\text{start}}/P_{\text{true}} = 0.50 - 0.64$ all runs incorrectly recovered $P \approx 0.58P_{\text{true}}$. However, two of these incorrectly recovered runs have acceptance fraction in the same satisfactory range.

4.4.3 Candidate 4062446910648807168 sweep

The recovery results of this exoplanet system are shown in Figure B.2. Here there is a larger spread in recovered periods, with a majority of correctly recovered periods being in regions $P_{\text{start}}/P_{\text{true}} = 0.80 - 1.36$ and $1.86 - 2.00$ with large acceptance fraction. Although with lower acceptance fraction, there here also exists some consistency in incorrectly recovered parameters around $P \approx 0.53P_{\text{true}}$ and $P \approx 2P_{\text{true}}$.

4.5 Convergence robustness

This section presents the results of the convergence robustness tests described in subsection 3.8.4. For candidate 1878822452815621120, a total of nine runs are performed with random seeds 42, 121, and 739, together with number of walkers 32, 64, and 128. The recovered median of parameters for each individual run is shown in Figure B.3.

The figure demonstrates clear consistency in recovered parameters over all nine runs, with only minor individual deviations. The average and maximum bias in each recovered parameter from each run is tabulated in Table 4.10. Each parameter presents a precise recovery, with exception of F . However, this large error percentage is not due to bad recovery, but rather a consequence of recovered F having value near zero. The absolute deviations are small, and the true value falls within the error bounds.

Parameter	$\langle \Delta\theta/\bar{\theta} \rangle$ (%)	$\max \Delta\theta/\bar{\theta} $ (%)
P	0.0222	0.0438
e	0.1204	0.2556
T_0	0.0859	0.1842
A	0.1096	0.3182
B	2.0143	5.7380
F	81.1006	143.8607
G	1.2940	2.7520

Table 4.10: Average and maximum percentage deviation of the recovered posterior median across all nine runs for candidate 1878822452815621120. Deviations are computed relative to the mean across all runs.

Chapter 5

Discussion

5.1 Observational setup

For this project, the candidates used for the MCMC recovery tests were hand-picked from an exoplanet candidate list, along with one binary with known two-body orbital solution. The selection of the suspected exoplanets was rooted in finding candidates with specific and unique orbital configurations, with emphasis on period, eccentricity and inclination. Candidates from this list were individually tested and properties were noted. However, all candidates were not tested or considered in this process, meaning systems with parameter-configurations of greater interest or importance may have been overlooked. A more rigorous approach, instead of hand-picking candidates, would be to do a pre-planned methodical search through the candidates for properties of interest, using specified intervals of values for considered parameters. This would streamline the process, ensuring interesting candidates and saving time. Furthermore, the selected sample itself may also not be representative of the general Gaia exoplanet candidate population, as there exists a multitude of orbital parameter combinations and this project only focuses on a limited amount of exoplanets with parameter values of specific interest.

The per-epoch along-scan noise $\sigma_{AL} = 0.1$ was consistent for all candidates and was chosen to resemble the astrometric precision of moderately bright stars. The value was chosen as a conservative approximation as to not accidentally overestimate the precision of recovery by choosing a noise lower than the true noise for a specific candidate. This approach introduces incorrect noise estimations, although through underestimation. This shows that a system that is accurately recovered, would also be accurately recovered with a lower and correct along-scan noise, but conversely that a system that is not correctly recovered could possibly have been, with correct use of noise. For candidate 1386158505321789696 with $G_{\text{mag}} = 4.19$ the chosen noise is very pessimistic. A more delicate approach would be to assign a per-epoch along-scan noise relative to the magnitude of the candidate, as this assigns legitimate accuracy depending on the observed system.

A fundamental limitation of the method is that it is self-referential. The "true" orbital parameters are taken from the Gaia DR3 `nss_two_body_orbit` data, and the simulated epoch data are generated from these same solutions. This means that any biases present in the DR3 solutions that are due to the methods of Gaia themselves are invisible to this test, as they would be consistent in both the "true" data and simulated solutions. This is unfortunately unavoidable, but can be addressed with the release of the DR4 data.

5.2 Interpretation of results

The standard five-year and ten-year recovery missions covered in section 4.1 both do a great job of recovering parameters. The binary candidate 1386, being a high S/N system does not experience much improvement with the ten-year mission. The five-year mission is already sufficient in recovering the parameters. The ten-year mission becomes more of importance with the lower S/N ratio candidates, where the extra simulated transits are valuable in defining the orbital properties. This can be seen clearly in candidate 4062, where all parameters and Campbell elements show slight improvement in their values and uncertainties. Out of the five candidates, two do not meet the convergence criterion in either mission. A particularly interesting case is candidate 4698, with its short period of 33.7 days. The probability distribution of the period is extremely well defined and recovered with great accuracy, which is due to the short period being densely sampled over both missions. However, a handful of walkers become stuck at approximate periods 35 and 37.5 days, suggesting the presence of small local maximum in the otherwise flat probability landscape, probably due to noise originating from the low S/N ratio of 2.3. A similar situation appears for other parameters. With a flat probability landscape, the walkers wander almost aimlessly and therefore do not de-correlate to satisfy the autocorrelation time limit. Although finding the correct value, respective values of parameters are not well defined enough to converge. The largest driver of the non-convergence is T_0 with $\tau = 279$ compared to needed $N_{\text{prod}}/50 = 100$, tightly followed by G , F and B . For the other non-converging candidate, 4232, the same drivers appear.

A reoccurring issue in the standard testing is the imprecision in the derived Campbell elements ω and Ω . This is due to an issue in the back-conversion from the Thiele-Innes constants to the Campbell elements. One generic case is for face-on orbits with $i = 0^\circ$, where ω and Ω lose any meaningful definition. The recovery is reduced down to certain combinations of $\omega + \Omega$ (the longitude of periastron, ϖ) or $\omega - \Omega$, resulting in the elements becoming individually unrecoverable. The issue remains for nearly face-on orbitals and is clearly apparent in the uncertainties of the standard recovery of candidate 1386 with $i = 4.5^\circ$, but is rather resolved for candidate 4062 with $i = 7.7^\circ$. In edge-on cases with $i = 90^\circ$, Ω becomes well constrained, but ω and T_0 become degenerate, explaining why T_0 is the main driver for non-convergence from previous section, tightly followed by the ω -dependent TI-constants. This is seen clearly in the five-year recovery of the nearly edge-on candidate 4232, and both recoveries of nearly edge-on candidate 4698, where Ω is well defined and ω is essentially unconstrained.

The calculation of masses for host stars and companions, with respective semi-major axis, serve only as an estimated characterization of the systems and are not to be fully trusted. These calculations use rough estimates of the star masses, and the ten-year recovered a_0 and P without regard of uncertainties. The main focus of this paper was put into testing and the actual recovery of the orbital parameters.

Four of the five candidates were selected from the Gaia exoplanet candidate list and have already passed an initial filtering step, identifying the observed astrometric wobble as consistent with a planetary companion. Despite this, false positive contamination remains a concern. Ranalli et al. (2018) shows that the fraction of false detections depends strongly on S/N and the detection criterion applied, with false positive rates increasing significantly at low S/N. A particular source of contamination is unresolved binary stars, where the photocentric wobble of an unresolved stellar companion can mimic the astrometric signal of a planetary companion. Following Rogers et al. (2024), candidate 4698424845771339520 has been proven to be a false positive, confirming that the

suspected exoplanet does not exist. This shows that companions discovered through astrometric methods should be treated with caution, and it is beneficial to follow up and support the findings with other methods, for example through radial velocity measurements. This misinterpretation originates from the poor signal-to-noise ratio of 2.3 for this candidate, which falls below the limit of $S/N < 3$ needed for reliable detection of a system with period $P \leq 5$ yr (Casertano et al. 2008).

The injection-recovery of simulated period and eccentricity in section 4.3 is precise in most cases, with only a very few cases of ill-recovered parameters. A minor issue in the eccentricity recovery is the small constant offset, which unfortunately is of unknown origin. A known problem in eccentricity recovery using MCMC appears at or near boundaries $e = 0$ and $e = 1$ (Lucy & Sweeney 1971). At $e \approx 0$, the MCMC is forced to search within interval $[0, 1)$, which means values below 0 are rejected, causing some overestimation of eccentricities. Conversely, at $e \approx 1$ there is an underestimation visible as e approaches 1, where the bias deviates from previous offset and approaches 0. This is what causes the non-convergence in the cases near the boundaries. The difference in recovery due to the perturbation is minor, proving the MCMC identifies the correct eccentricity despite starting at $e = 0.01$. As for the comparison between mission lengths; the ten-year mission has a more precise recovery, demonstrating here an improvement using the extra number of transits. The inclination injection-recovery sweep reveals an overestimation in inclination at low inclinations, which decreases with each step toward higher inclinations. This is consistent with Makarov (2025), who links this to a mathematical consequence of the back-conversion from Thiele-Innes constants to the Campbell elements. Near $i = 0^\circ$, $\cos i$ is nearly flat and insensitive to changes in inclination, where small noise-driven perturbations result in large errors in recovered inclination. This is also accompanied by an overestimate of respective semi-major axis (a_0), although Makarov (2025) reports a larger overestimation of $\sim 24\%$, compared to our overestimate of $\sim 3.6\%$ for $i = 0^\circ$ using the five-year mission, with $S/N = 10$. Worth mentioning is that the injection-recovery tests use HD164604 exclusively, and results here depend on the specific properties of this star, such as distance, magnitude, and astrometric prior. This means that results using different stars with simulated systems may vary, which is not controlled and is a limitation of this test.

The period perturbation sweep presents for candidate 1386158505321789696 a persistent recovery in (almost) all perturbed initializations, despite converging toward different periods. The periods to which the MCMC converges are (apart from the true period) two period aliases, with values $\approx 0.62P_{\text{true}}$ and $\approx 2P_{\text{true}}$. These values reflect secondary maxima within the likelihood landscape explored by the MCMC, and as can be seen by the high acceptance fractions of ≈ 0.49 are very well defined. The reason for these specific values is physically motivated. Yahalomi et al. (2026) shows that an eccentric astrometric orbit produces a signal that is mathematically indistinguishable from a pair of planets in a mean-motion resonance. The alias at $\approx 2P_{\text{true}}$ corresponds to a 2:1 resonance and the alias at $\approx 0.62P_{\text{true}}$ likely corresponds to a 3:2 resonance, thereby mimicking the presence of a two-planet resonant system rather than a single eccentric planet. The other two candidates show a less structured alias behavior. Looking at candidate 423297927866697088, all perturbations above $0.64P_{\text{true}}$ converge to the true period, but there is a small secondary maximum at $\approx 0.58P_{\text{true}}$ to which some of the lower perturbations slightly converge. Due to the lower acceptance fraction, combined with its lower S/N ratio of 5.7, this is probably noise-driven as it has no clear resonance interpretation. Candidate 4062446910648807168 has a much larger spread of recovered periods, and shows some slightly higher acceptance fractions for two points near $2P_{\text{true}}$. The rest are most likely noise-driven local maxima, where the MCMC fails to locate the maximum corresponding to the true period.

The convergence robustness test demonstrate the results independence of random seed and number of walkers. All nine runs produce consistent results with only very minor variance, and no single combination of seed and walker count proved itself superior or inferior to another. This proves that the used settings of random seed 42 and 64 walkers throughout the testing are reliable, and that the recovered posteriors reflect the true probability distribution.

5.3 Applicability to DR4 and GaiaNIR

As has been stated, this project simulates the epoch astrometry data that will become publicly available with Gaia Data Release 4 in December 2026. The method developed is directly applicable to the real data, as the code used is already structured to import and accept real Gaia epoch astrometry through Gaia DataLink services. This means the transition to and use of the DR4 data requires no fundamental changes, allowing for continued and immediate application of the MCMC framework to real Gaia detections.

The use of real epoch data will also address one of the main limitations in this work, namely the fixed per-epoch noise assumption. With DR4, per-epoch uncertainties will be provided directly for each observation, which naturally reflects true measurement precision. This removes the need for an assumed along-scan measurement noise, which in this project was fixed to $\sigma_{AL} = 0.1$ mas independent of the magnitudes of observed systems, and recovered orbital parameters are therefore expected to be more precise.

The period alias identified in the perturbation sweep, corresponding to 2:1 and 3:2 mean-motion resonances, will also be relevant when interpreting the DR4 data. As discussed by Yahalomi et al. (2026), these degeneracies are expected to appear in the real DR4 data and will require careful treatment to distinguish single-planet solutions from two-planet mimics. This treatment could involve radial velocity follow-up, or closer analysis of the structure of the likelihood landscape that is explored.

Finally, the ten-year mission results demonstrate improved average parameter precision, supporting the scientific case for GaiaNIR as a successor mission. An extended observational period, combined with improved instruments and methods, would improve the parameter recovery, most notably for long period systems which have only been partially sampled within Gaia’s mission length. It would also help resolve period ambiguities and reduce inclination bias, as both these points benefit from a more extended baseline.

Chapter 6

Conclusion

This project investigated the applicability of Bayesian MCMC orbit fitting to Gaia astrometric data for the recovery of exoplanet orbital parameters, using simulated epoch data from five Gaia DR3 candidates across a range of orbital configurations.

The MCMC framework successfully recovers parameters for all five candidates, with the ten-year mission providing improved precision over the five-year mission, specifically for lower S/N systems. The period and eccentricity are generally well recovered across the injection-recovery tests, and the period perturbation sweep demonstrates that the period, and with it the full system, can be recovered across a wide range of initial period offsets, while also finding local maxima, revealing the presence of period aliases. The convergence robustness tests confirm that the results are genuine explorations of the posterior distribution, and not dependent on random seed or walker count.

Several limitations are encountered: the back-conversion from Thiele-Innes constants to Campbell elements introduces a systematic overestimation of inclination and semi-major axis for face-on configurations (Makarov 2025), along with a degeneracy in ω and Ω for face-on and edge-on configurations. Non-convergence at lower S/N reflects flat posterior landscapes rather than computational flaws, and the confirmed false positive in candidate 4698424845771339520 demonstrates the importance of follow-up observations for companions detected solely through astrometry.

With Gaia DR4 scheduled for release in December 2026, the method developed here is directly applicable to real epoch astrometry. The results of this work illustrate that MCMC-based orbit fitting is a viable tool for future Gaia exoplanet detections.

Bibliography

- Arriagada, P., Butler, R. P., Minniti, D., et al. 2010, *ApJ*, 711, 1229
- Brandt, T. D. 2021, *ApJS*, 254, 42
- Casertano, S., Lattanzi, M. G., Sozzetti, A., et al. 2008, *A&A*, 482, 699
- Christiansen, J. L., McElroy, D. L., Harbut, M., et al. 2025, , 6, 186
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, 674, A1
- Goodman, J. & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
- Hobbs, D., Høg, E., Mora, A., et al. 2016, arXiv e-prints, arXiv:1609.07325
- Holl, B., Sozzetti, A., Sahlmann, J., et al. 2023, *A&A*, 674, A10
- Katzgraber, H. G. 2009, arXiv e-prints, arXiv:0905.1629
- Lindgren, L., Klioner, S. A., Hernández, J., et al. 2021, *A&A*, 649, A2
- Lucy, L. B. & Sweeney, M. A. 1971, *AJ*, 76, 544
- Makarov, V. V. 2025, *The Astrophysical Journal Letters*, 981, L23
- Marcussen, M. L. & Albrecht, S. H. 2023, *AJ*, 165, 266
- Pecaut, M. J. & Mamajek, E. E. 2013, *ApJS*, 208, 9
- Perryman, M. 2018, *The Exoplanet Handbook*
- Perryman, M., Hartman, J., Bakos, G. Á., & Lindgren, L. 2014, *ApJ*, 797, 14
- Ranalli, P., Hobbs, D., & Lindgren, L. 2018, *A&A*, 614, A30
- Rogers, L. K., Debes, J., Anslow, R. J., et al. 2024, *MNRAS*, 527, 977
- Speagle, J. S. 2019, arXiv e-prints, arXiv:1909.12313
- Subasavage, J. P., Jao, W.-C., Henry, T. J., et al. 2017, *AJ*, 154, 32
- Torres, G. 2007, *AJ*, 133, 2684
- Yahalomi, D. A., Lu, T., Armitage, P. J., et al. 2026, , 999, L9

Appendix A

Figures from 4.1

All figures in this appendix correspond to the five-year simulated mission with 3000 burn-in steps and 5000 production steps in section 4.1. The captions for candidate 1386158505321789696 contain full descriptions applicable to all corresponding figures throughout the appendix.

1386158505321789696 plots

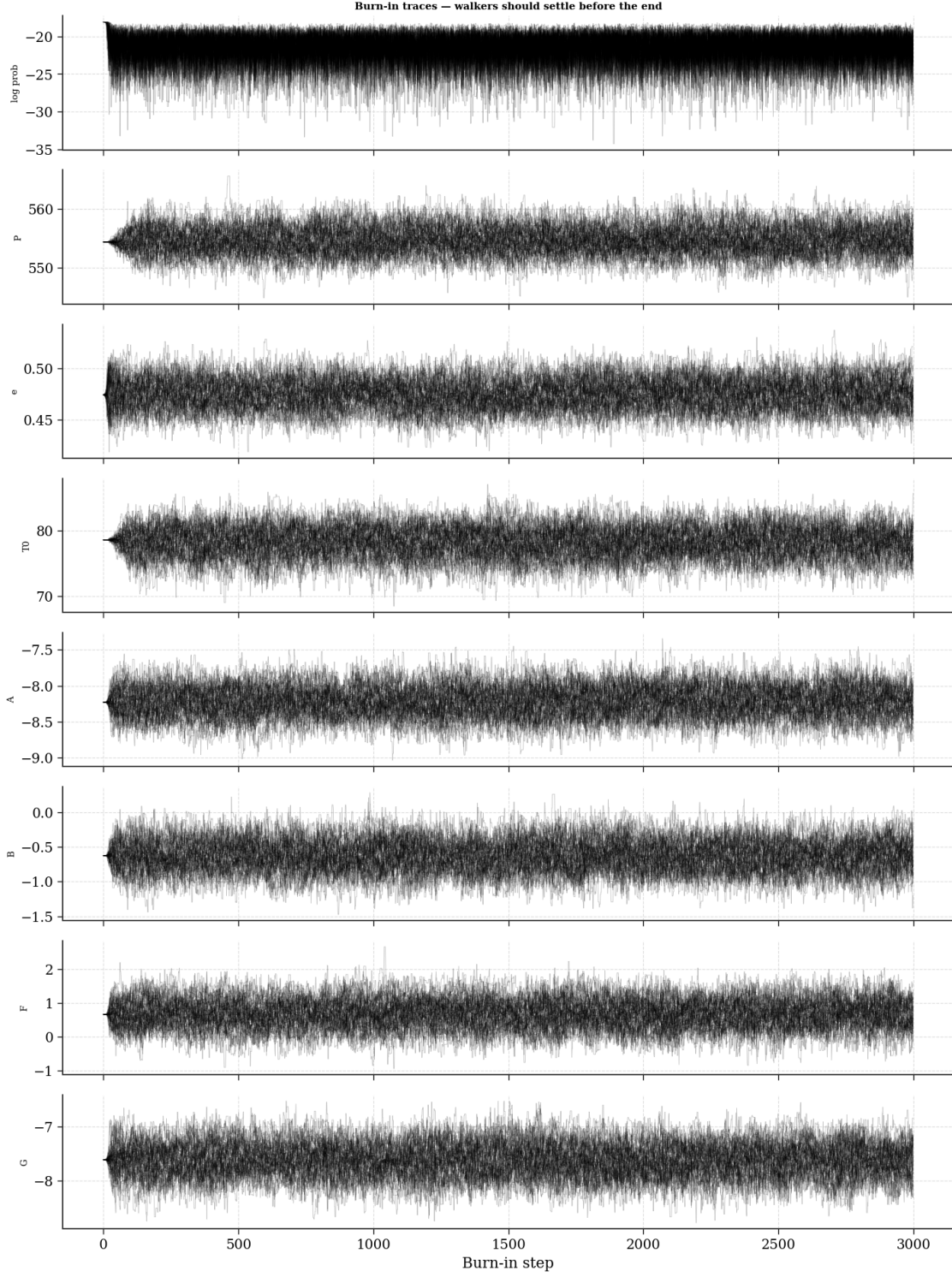


Figure A.1: MCMC burn-in traces for candidate 1386158505321789696 showing the log-posterior probability and the seven fitted orbital parameters $[P, e, T_0, A, B, F, G]$ over 3000 burn-in steps. Each line represents one walker.

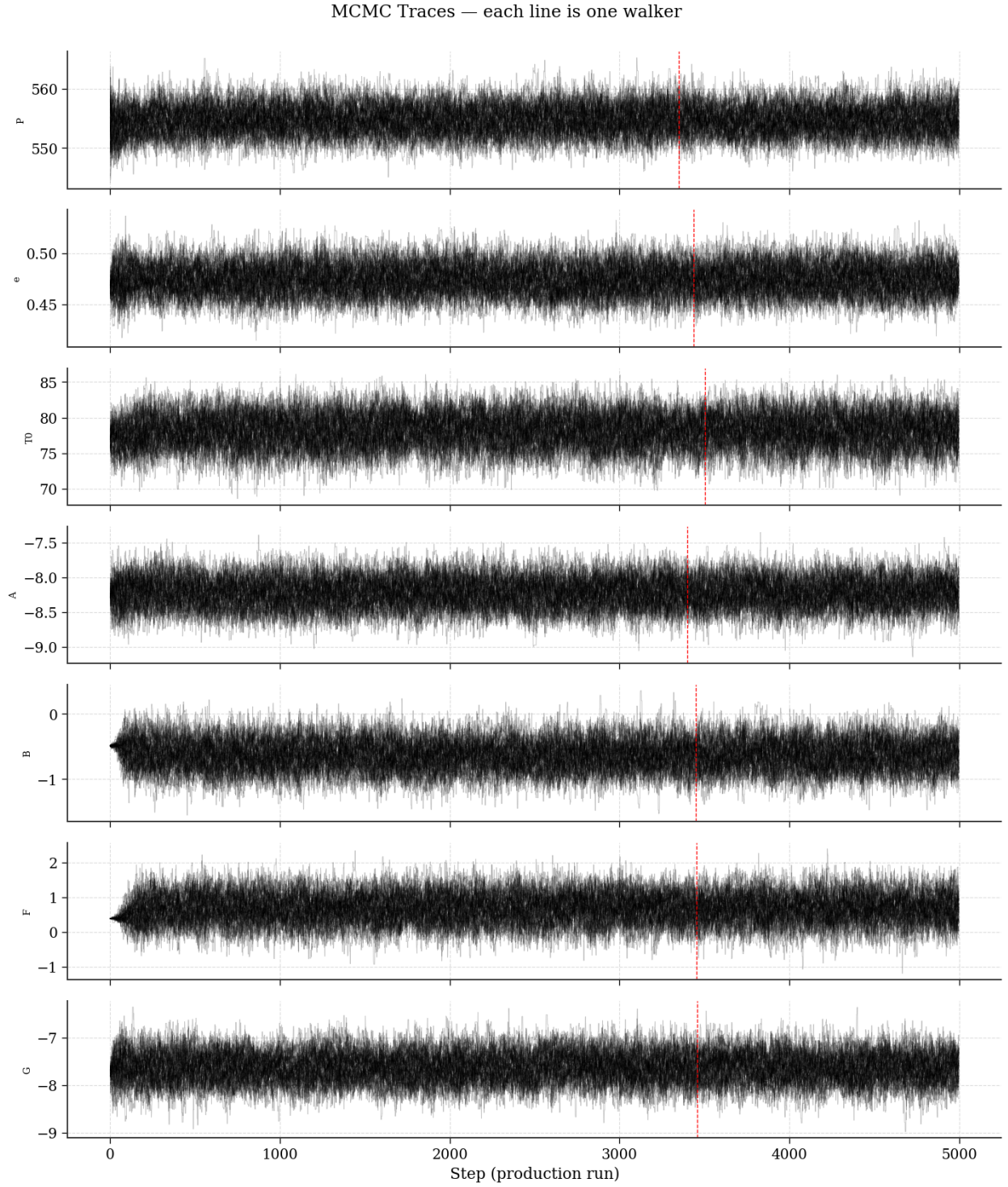


Figure A.2: MCMC production traces for candidate 1386158505321789696 showing the seven fitted orbital parameters $[P, e, T_0, A, B, F, G]$ over 5000 production steps. Each line represents one walker. Convergence per parameter is indicated with red vertical lines.

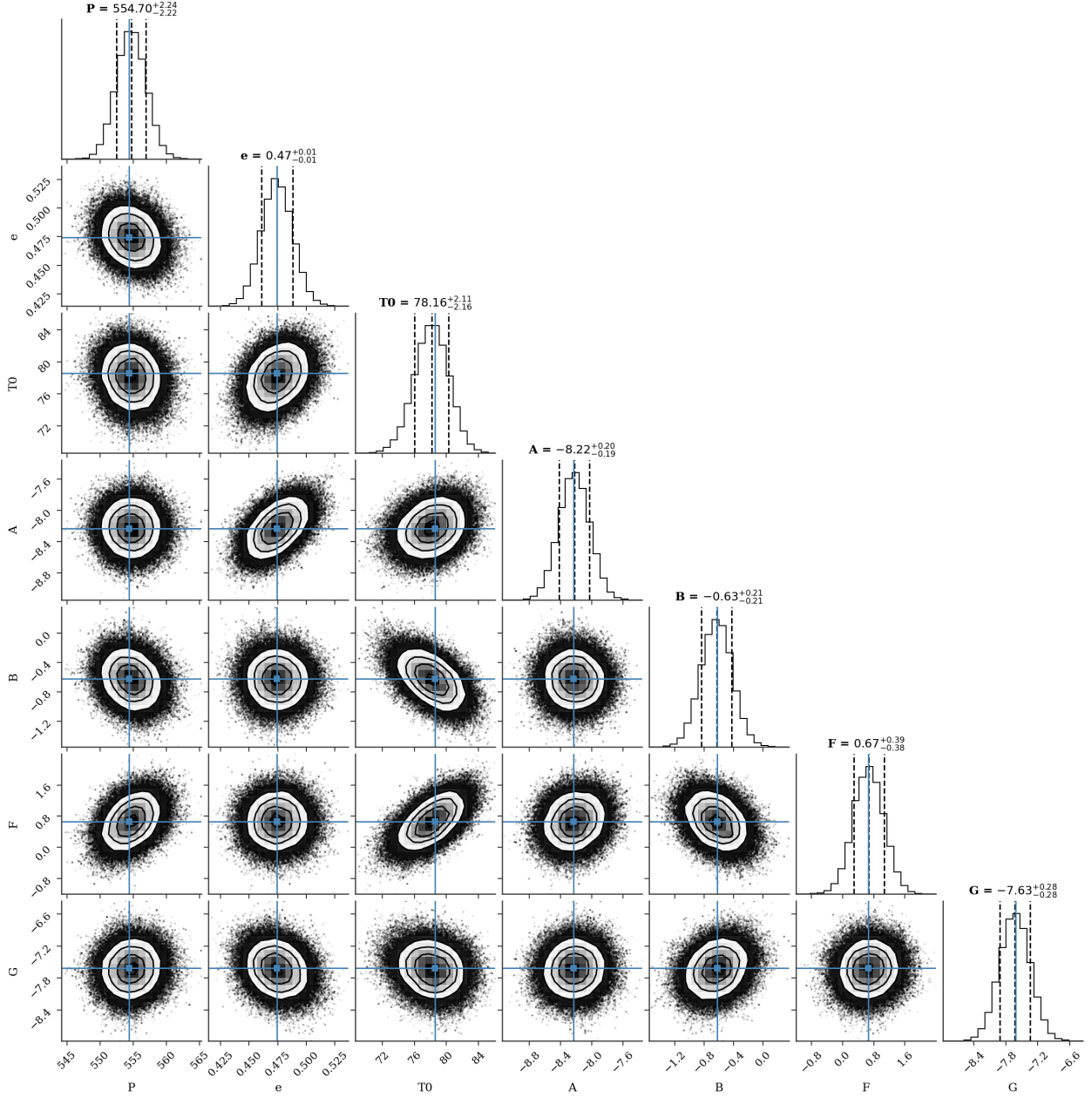


Figure A.3: Corner plot showing the posterior distribution for the seven fitted orbital parameters $[P, e, T_0, A, B, F, G]$ for candidate 1386158505321789696, using the five-year simulated mission with 3000 burn-in steps and 5000 production steps. Diagonal panels show the 1D marginalized posteriors with the 16th, 50th and 84th percentiles indicated with dashed lines. Other panels show the 2D joint posteriors. Blue lines indicate Gaia DR3 catalogue values.

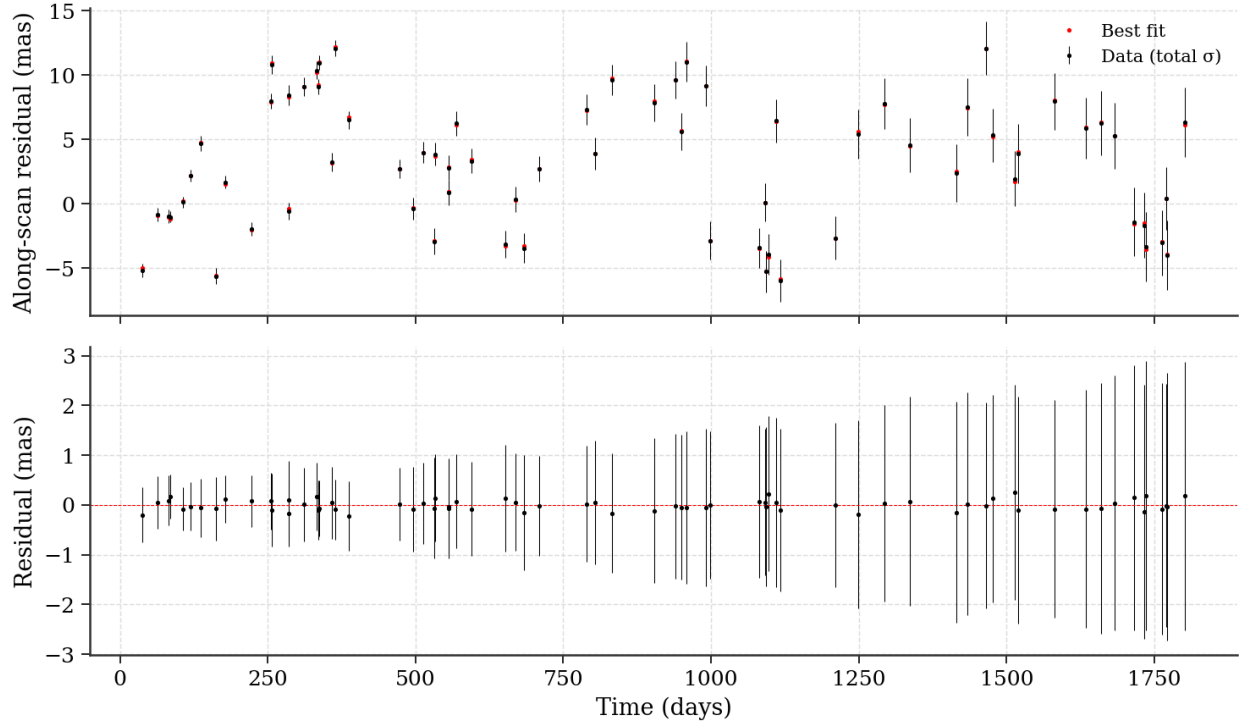


Figure A.4: Residuals plot for candidate 1386158505321789696, using the five-year simulated mission with 3000 burn-in steps and 5000 production steps. Top panel: simulated along-scan observations (black) and best-fit model (red) as a function of time, with total per-epoch uncertainties shown as error bars. Bottom panel: residuals after subtracting the best-fit model, with the dashed red line marking zero.

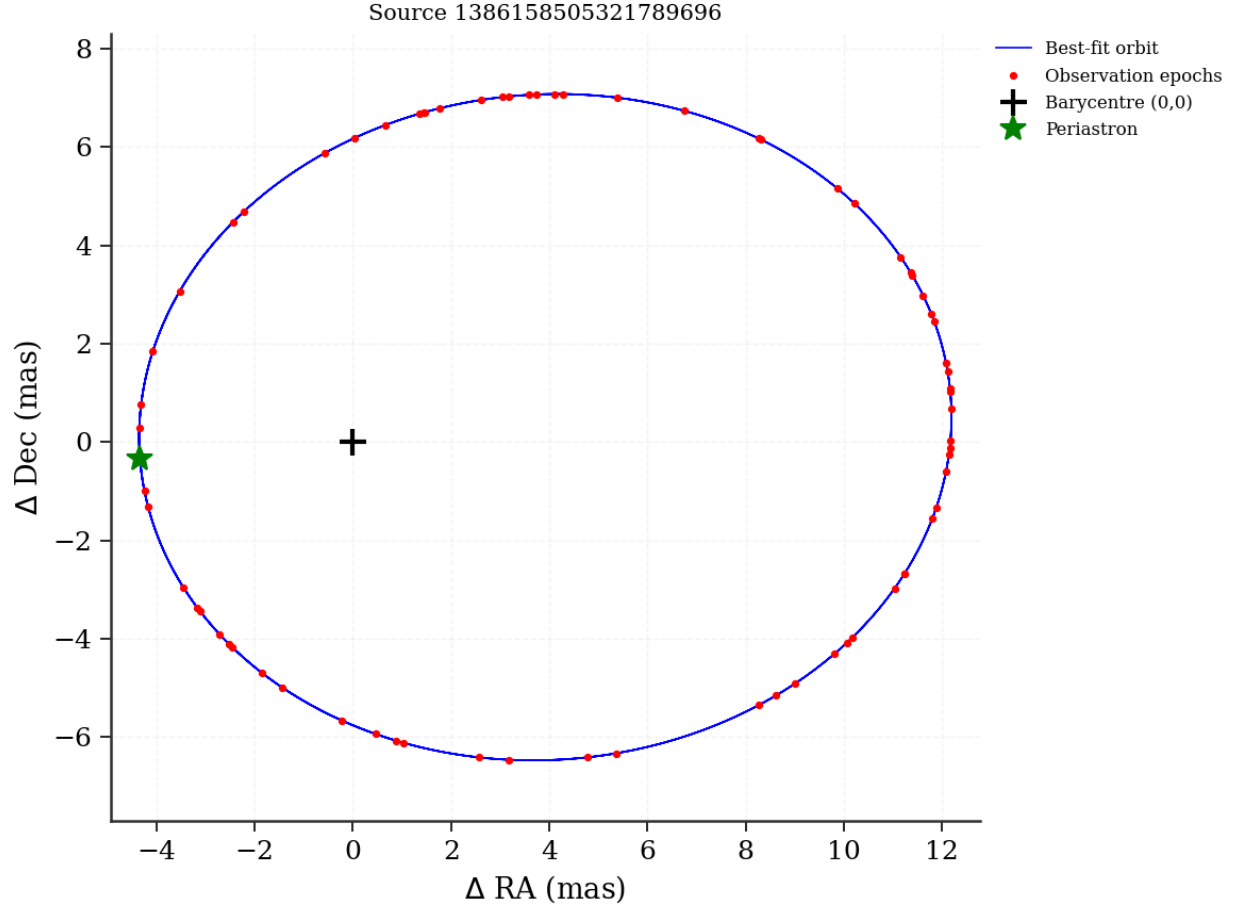


Figure A.5: Projected orbit of candidate 1386158505321789696 on the sky-plane, using the five-year simulated mission with 3000 burn-in steps and 5000 production steps. Red dots mark simulated observation epochs, blue line shows best-fit orbital, black cross indicates barycentre at the origin, and the green star marks the periastron passage point.

4062446910648807168 plots

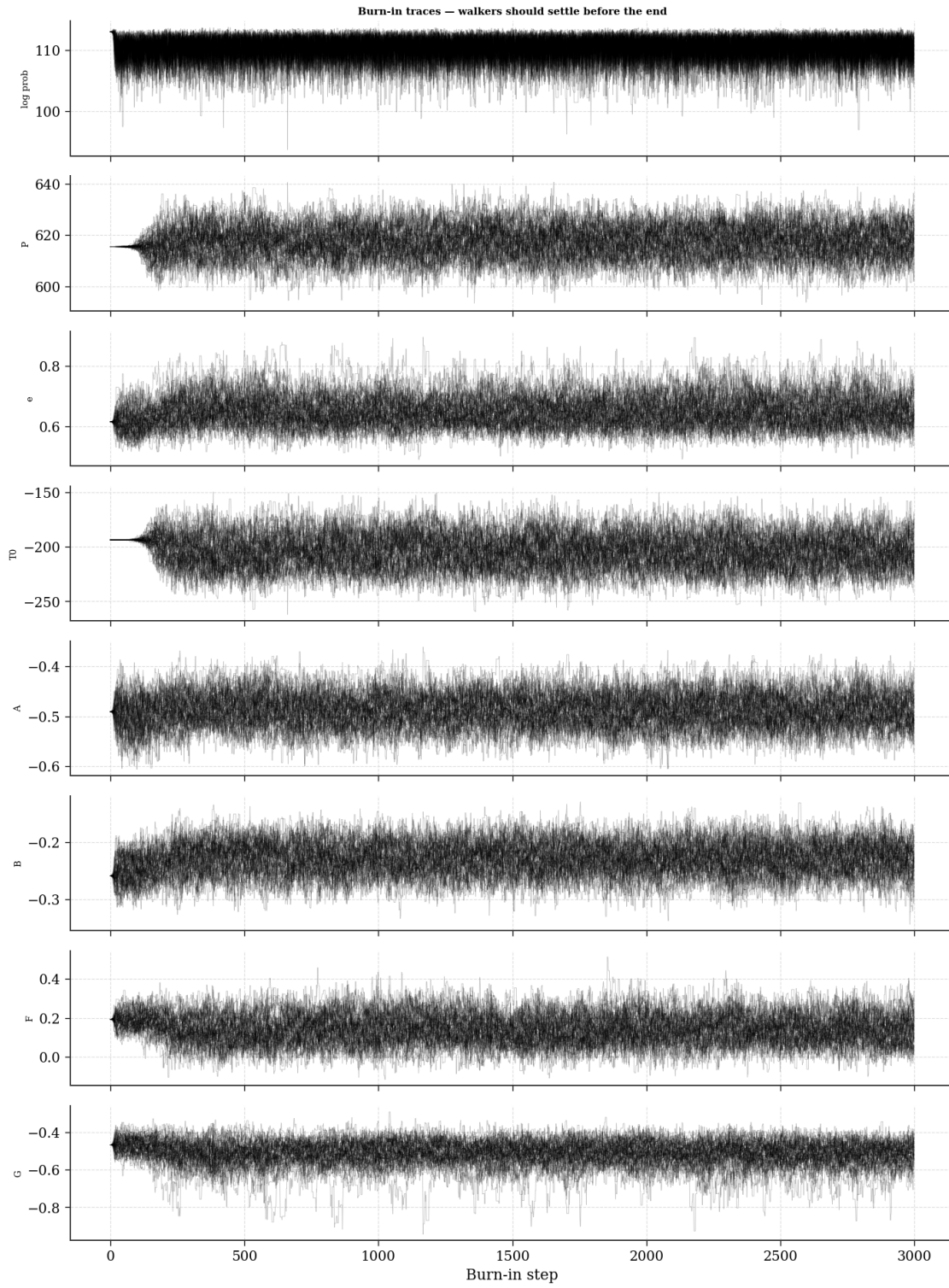


Figure A.6: MCMC burn-in traces for candidate 4062446910648807168.

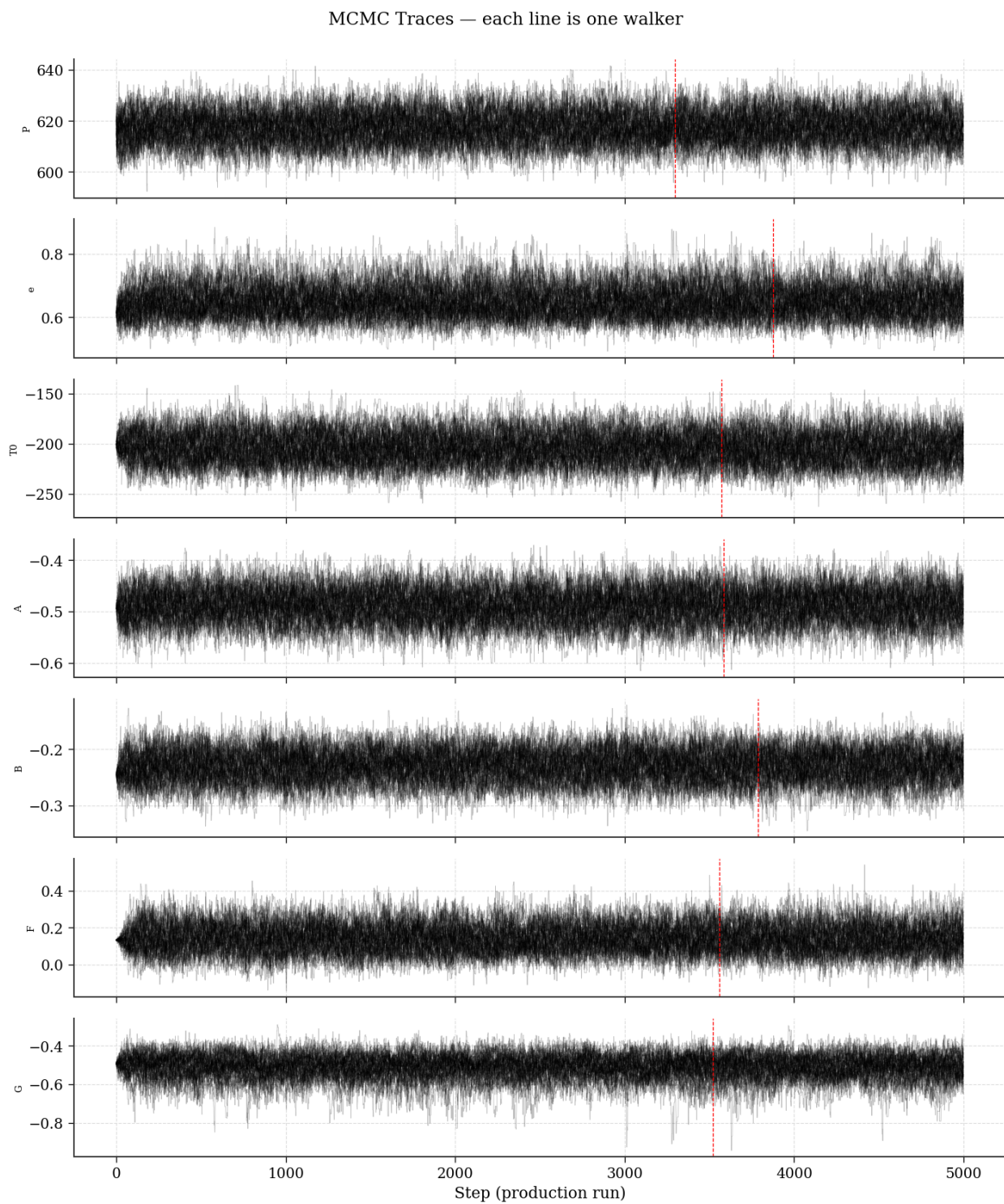


Figure A.7: MCMC production traces for candidate 4062446910648807168.

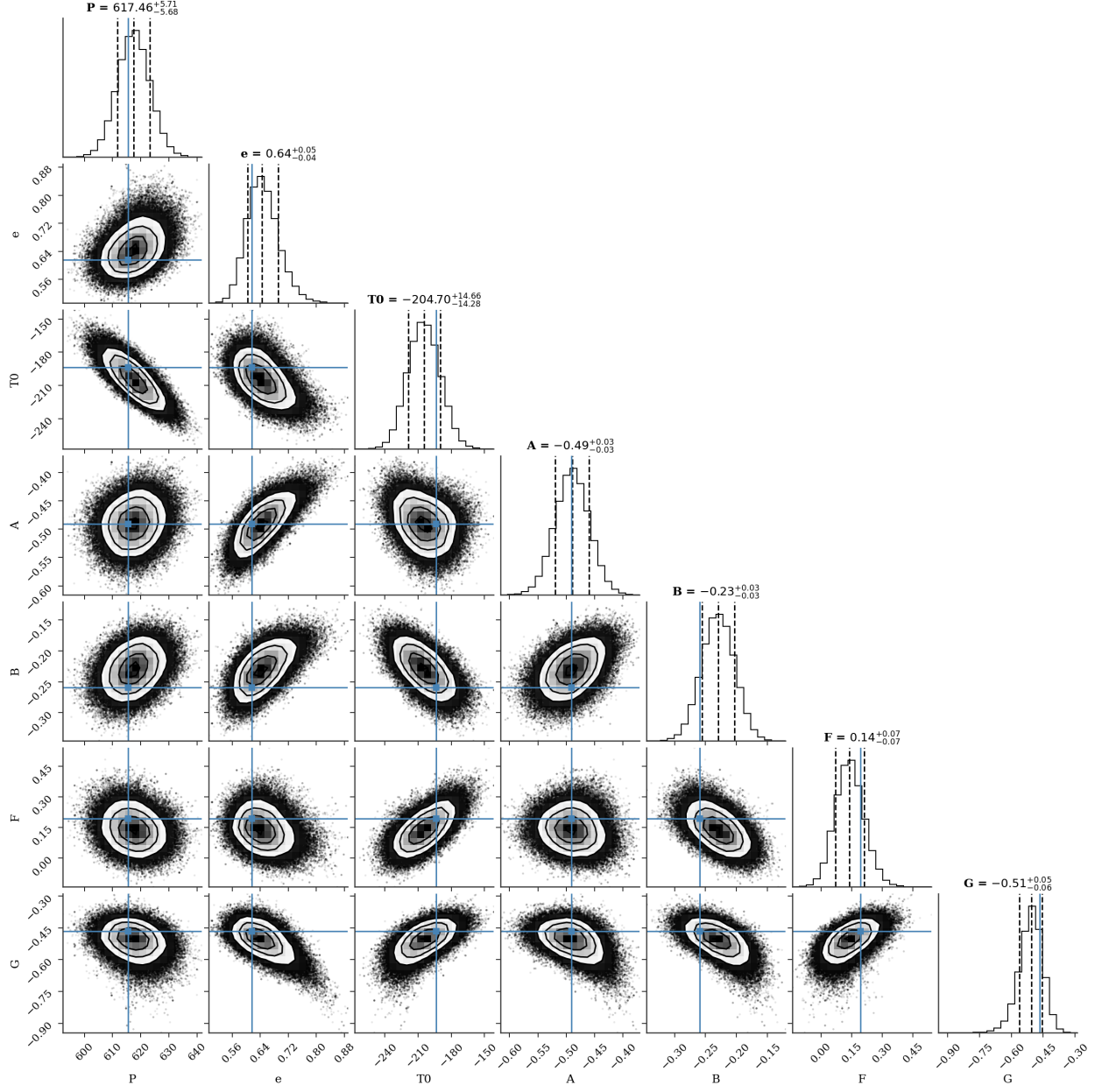


Figure A.8: Corner plot for candidate 4062446910648807168.

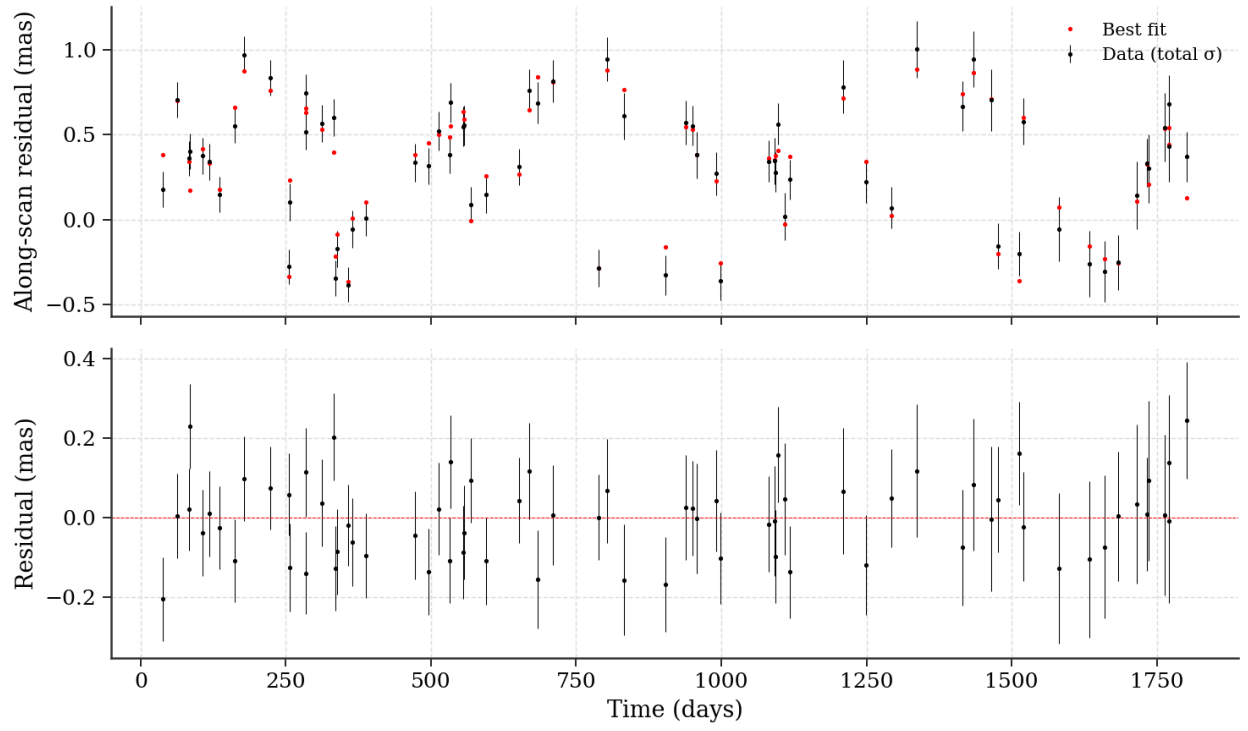


Figure A.9: Residuals plot for candidate 4062446910648807168.

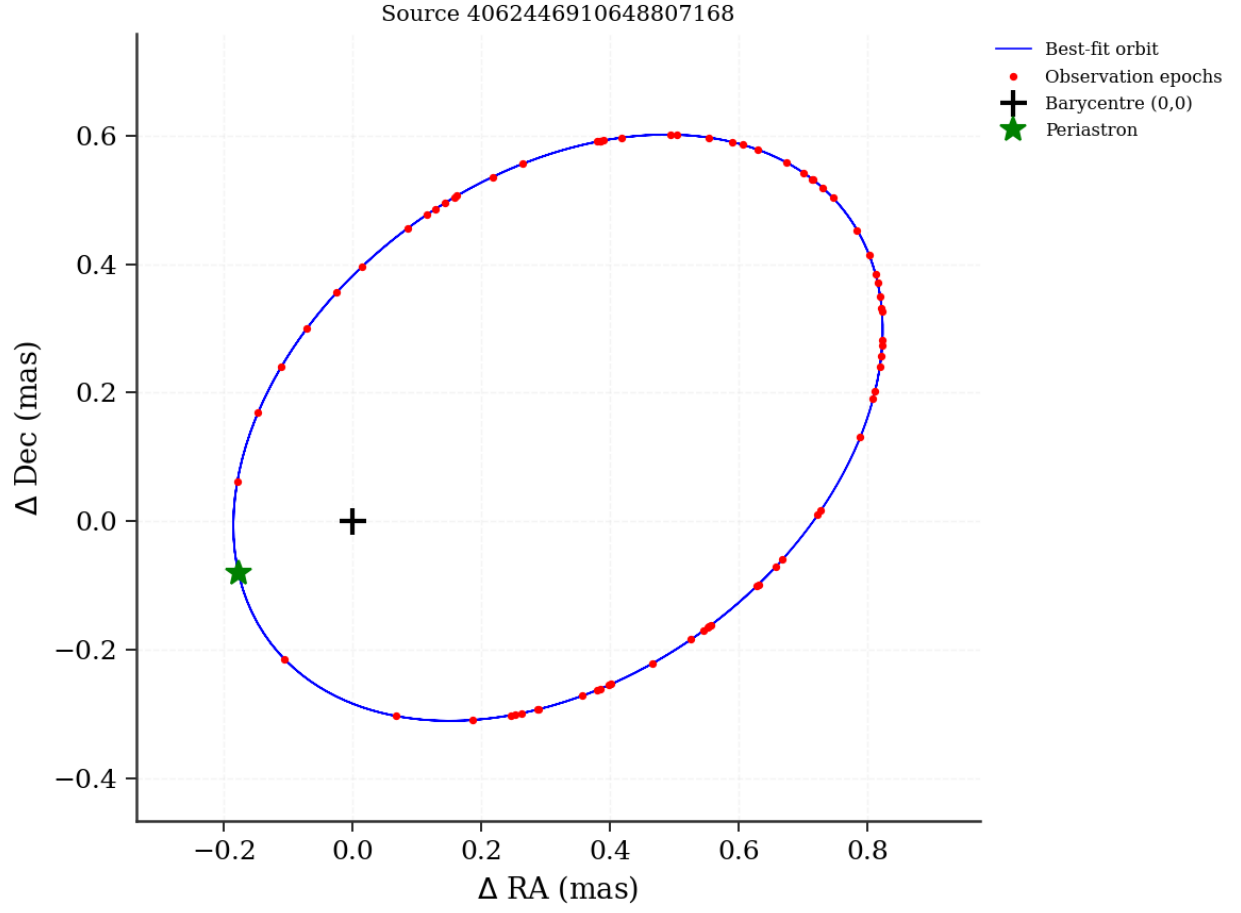


Figure A.10: Projected orbit of candidate 4062446910648807168 on the sky-plane.

423297927866697088 plots

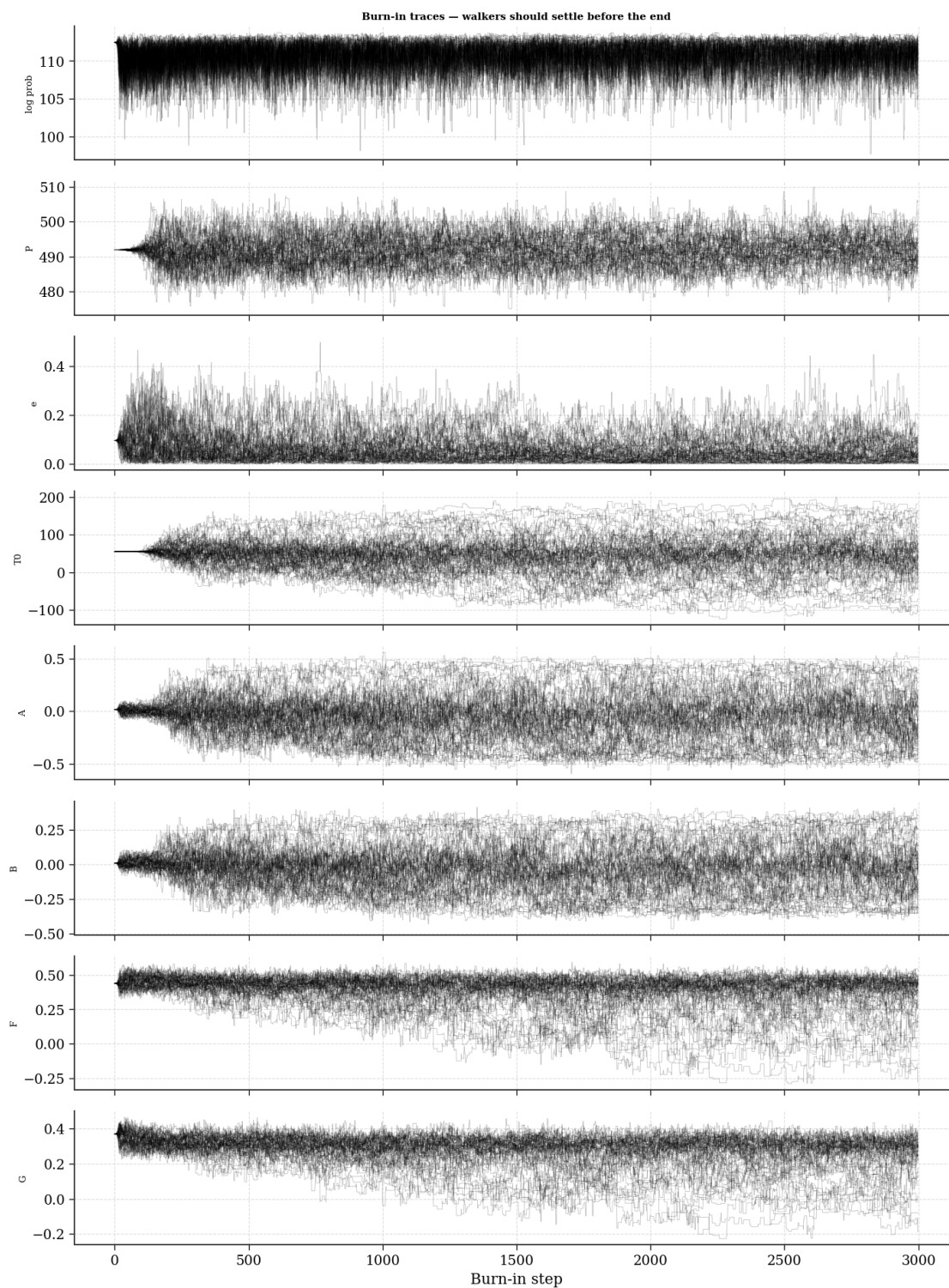


Figure A.11: MCMC burn-in traces for candidate 423297927866697088.

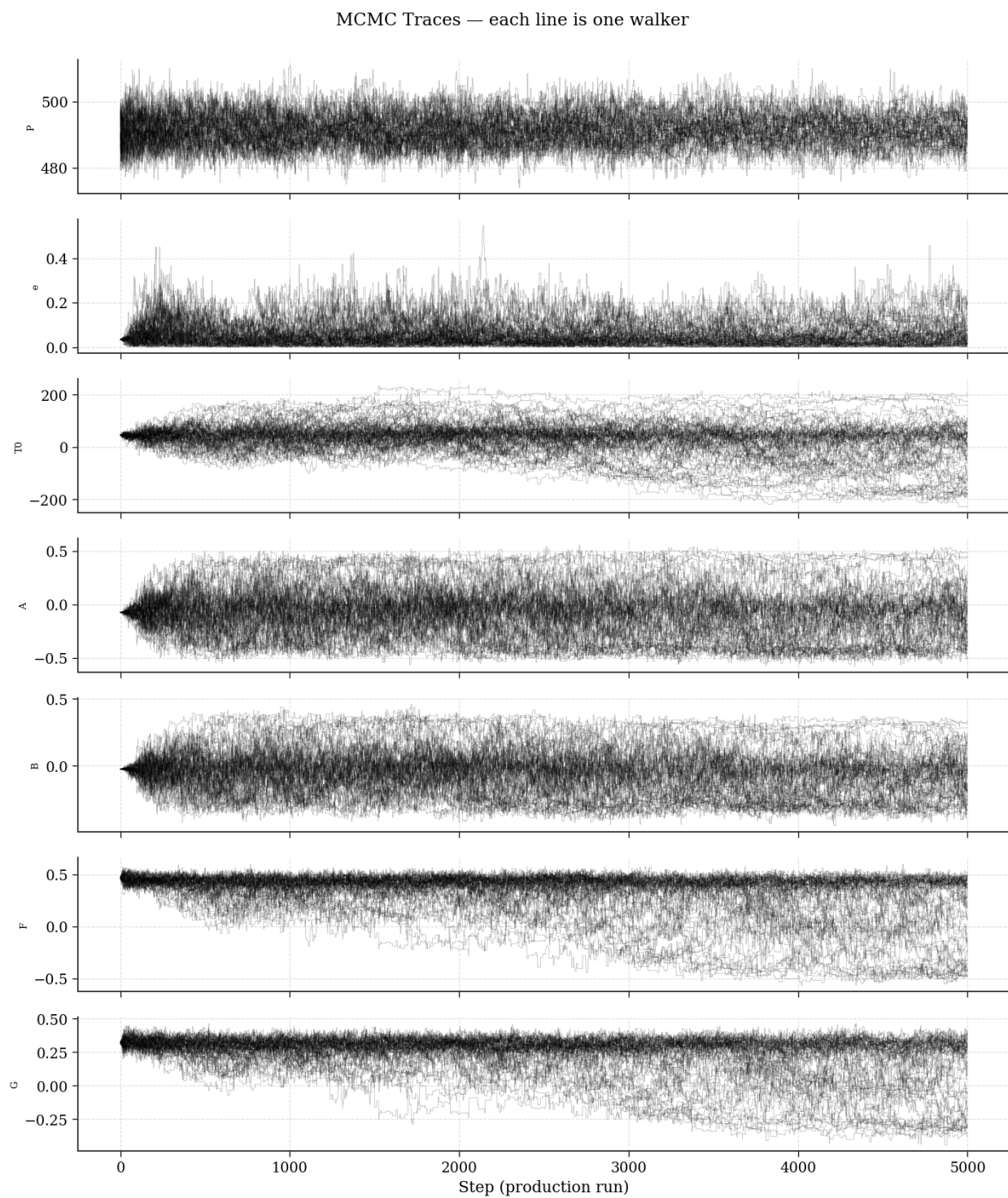


Figure A.12: MCMC production traces for candidate 423297927866697088.

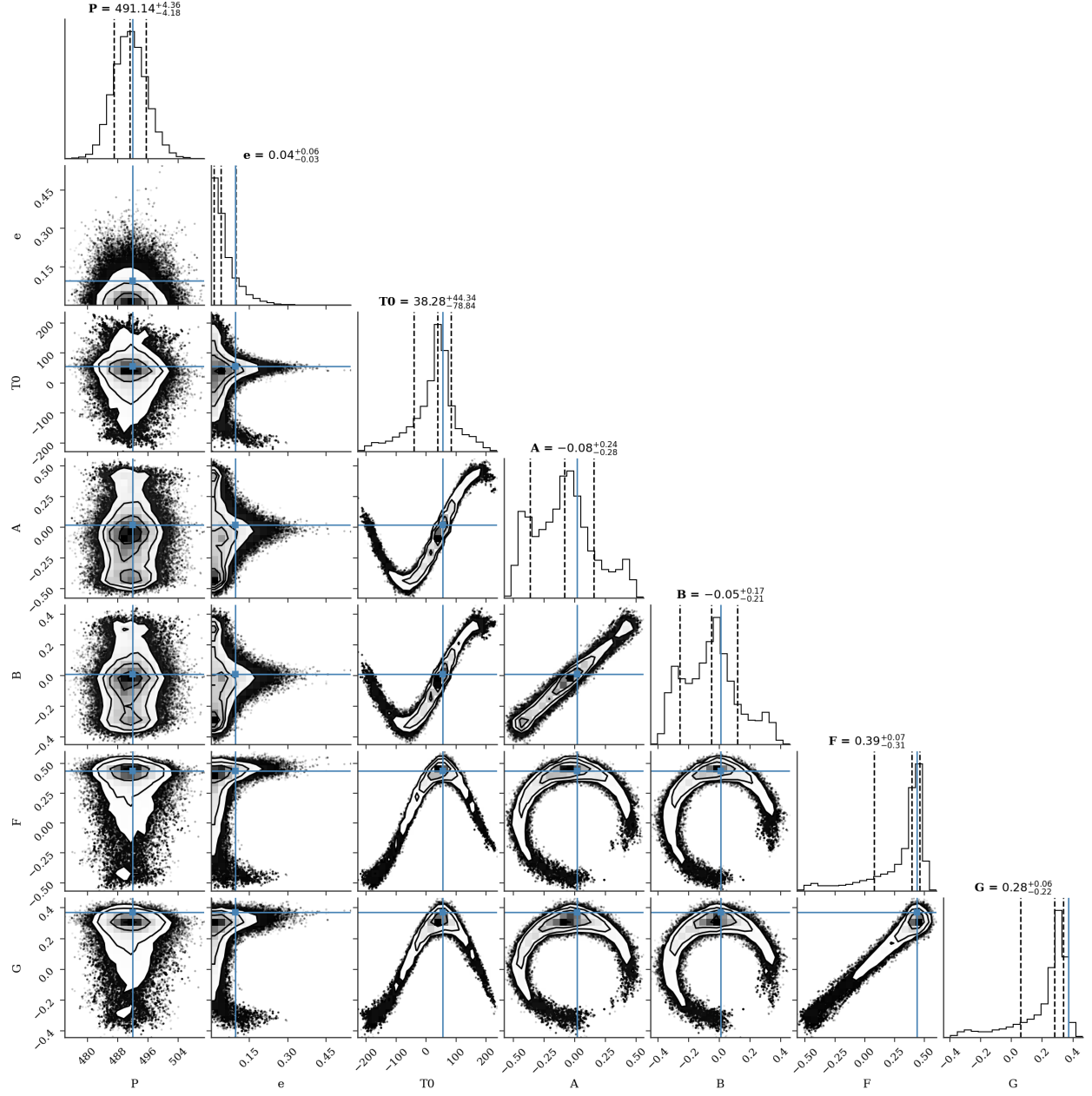


Figure A.13: Corner plot for candidate 423297927866697088.

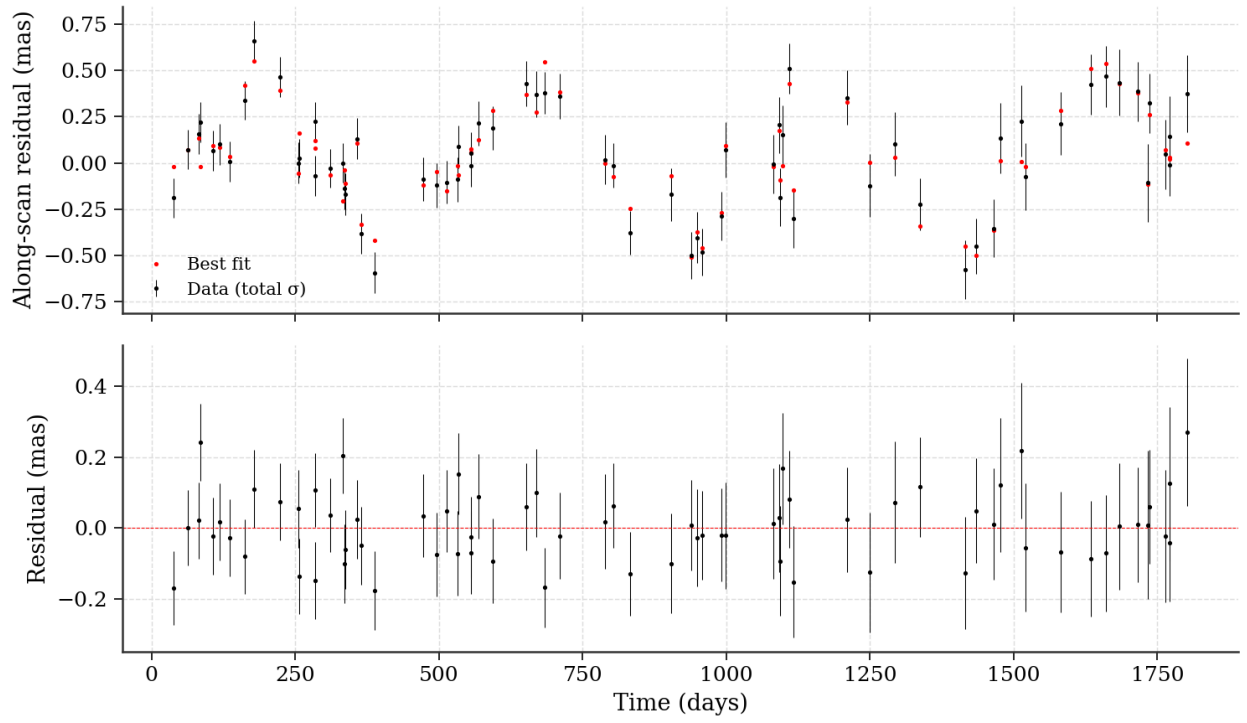


Figure A.14: Residuals plot for candidate 423297927866697088.

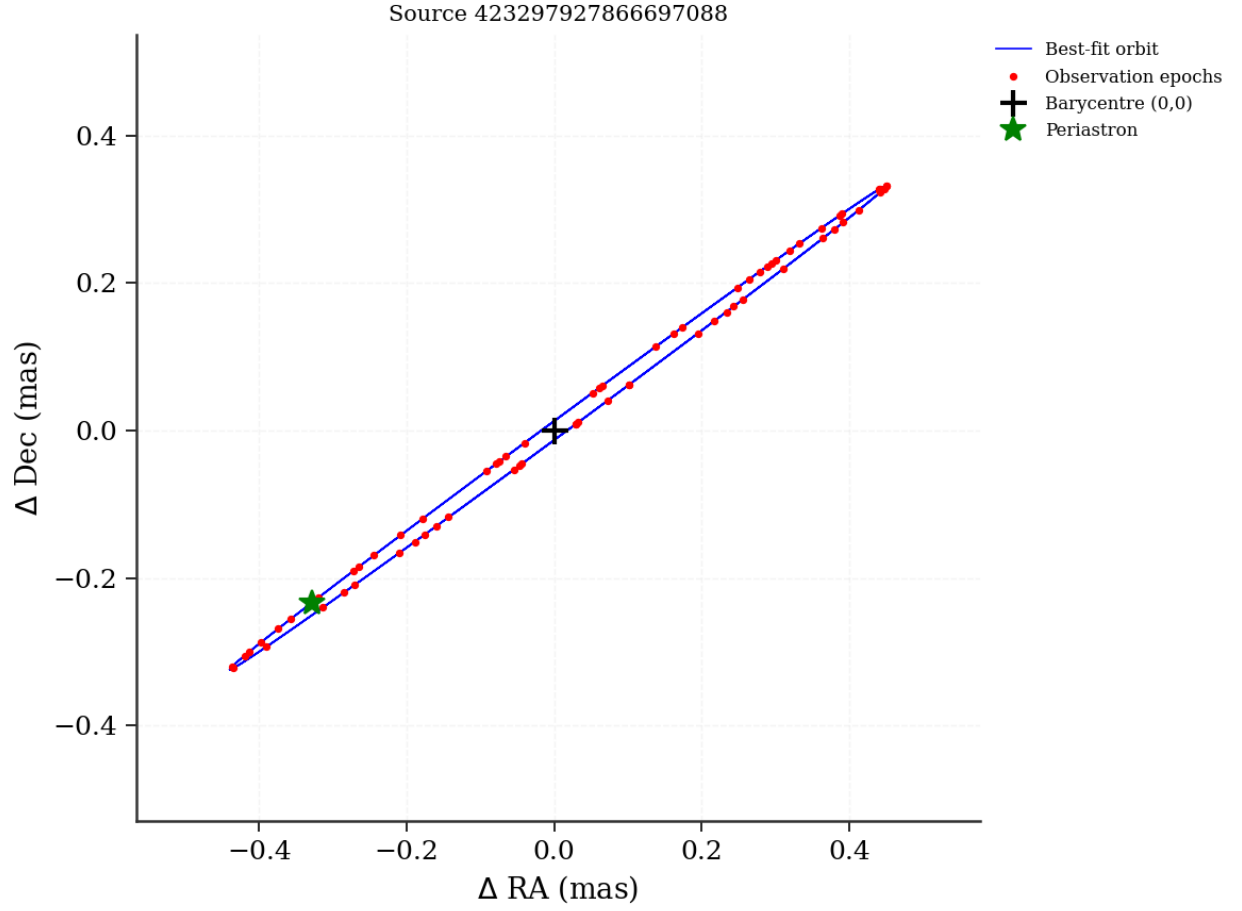


Figure A.15: Projected orbit of candidate 423297927866697088 on the sky-plane.

1878822452815621120 plots

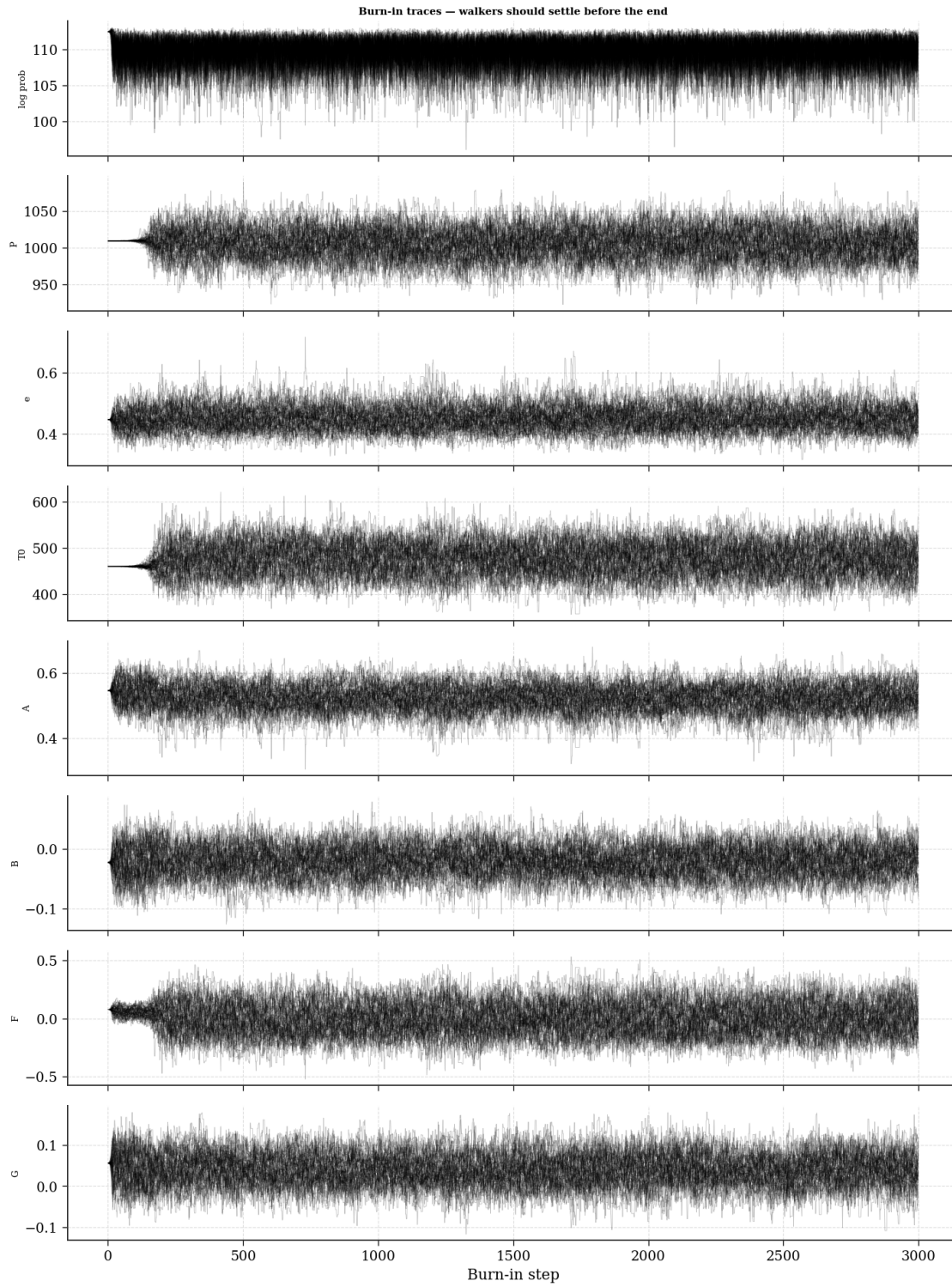


Figure A.16: MCMC burn-in traces for candidate 1878822452815621120.

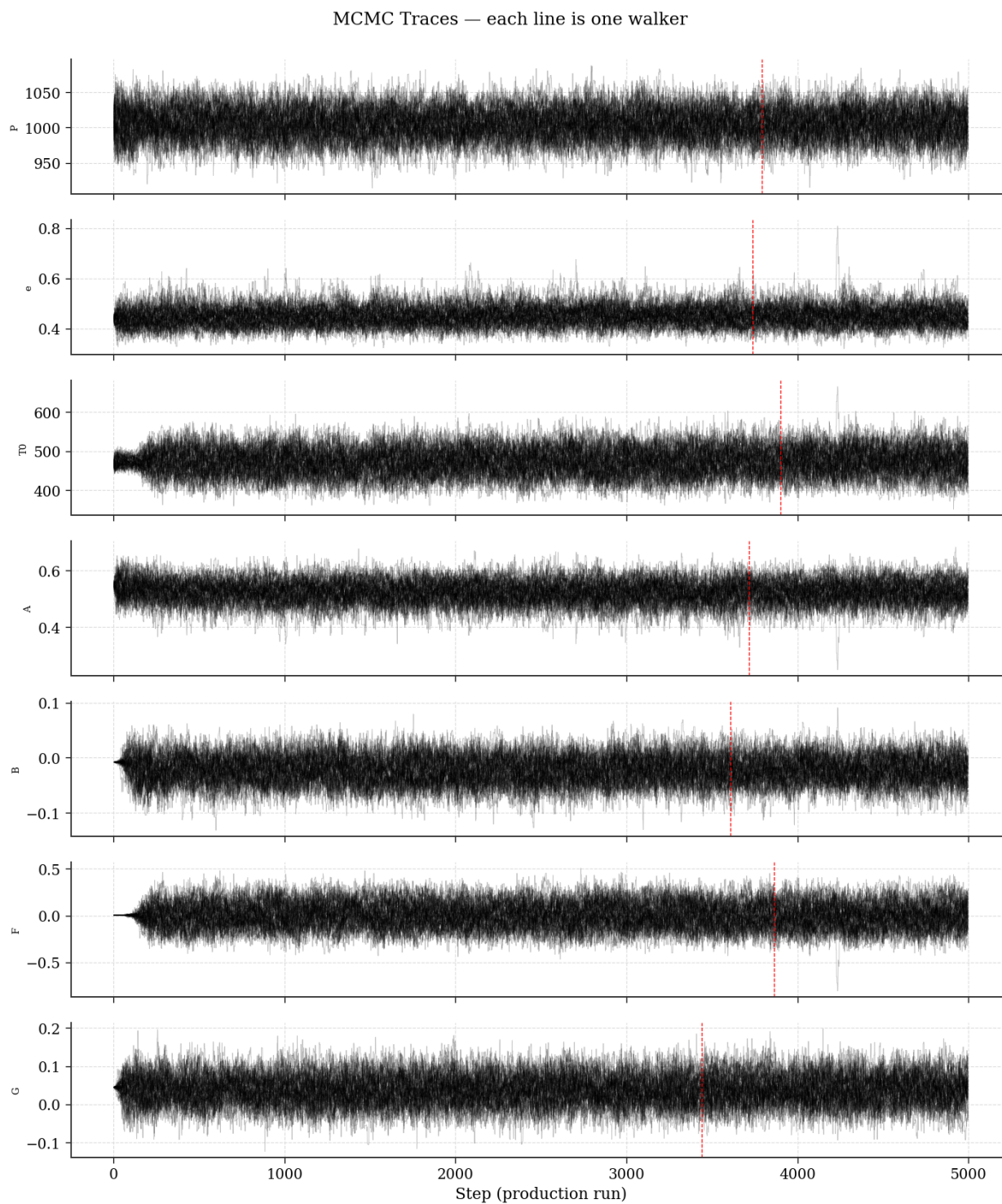


Figure A.17: MCMC production traces for candidate 1878822452815621120.

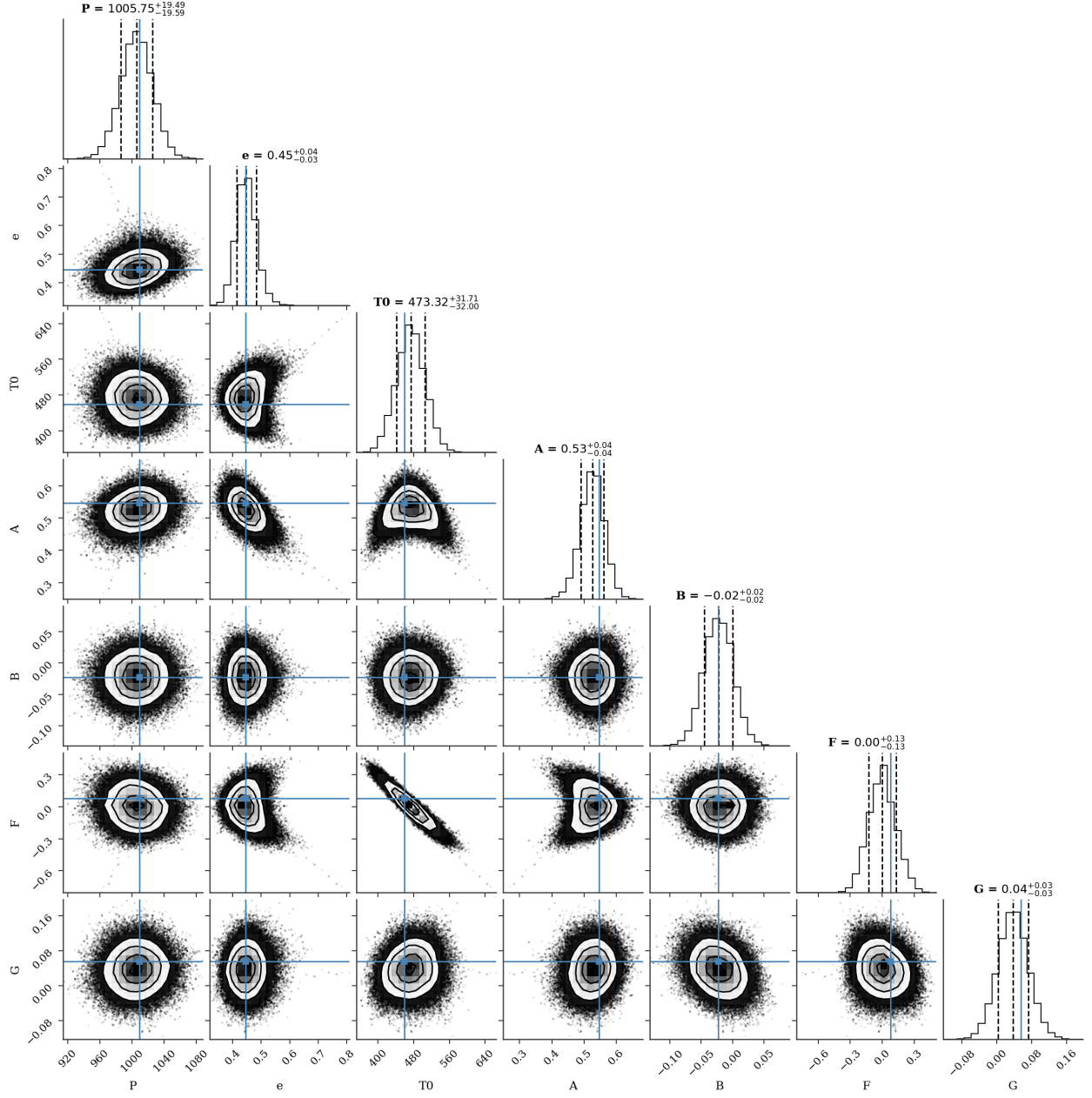


Figure A.18: Corner plot for candidate 1878822452815621120.

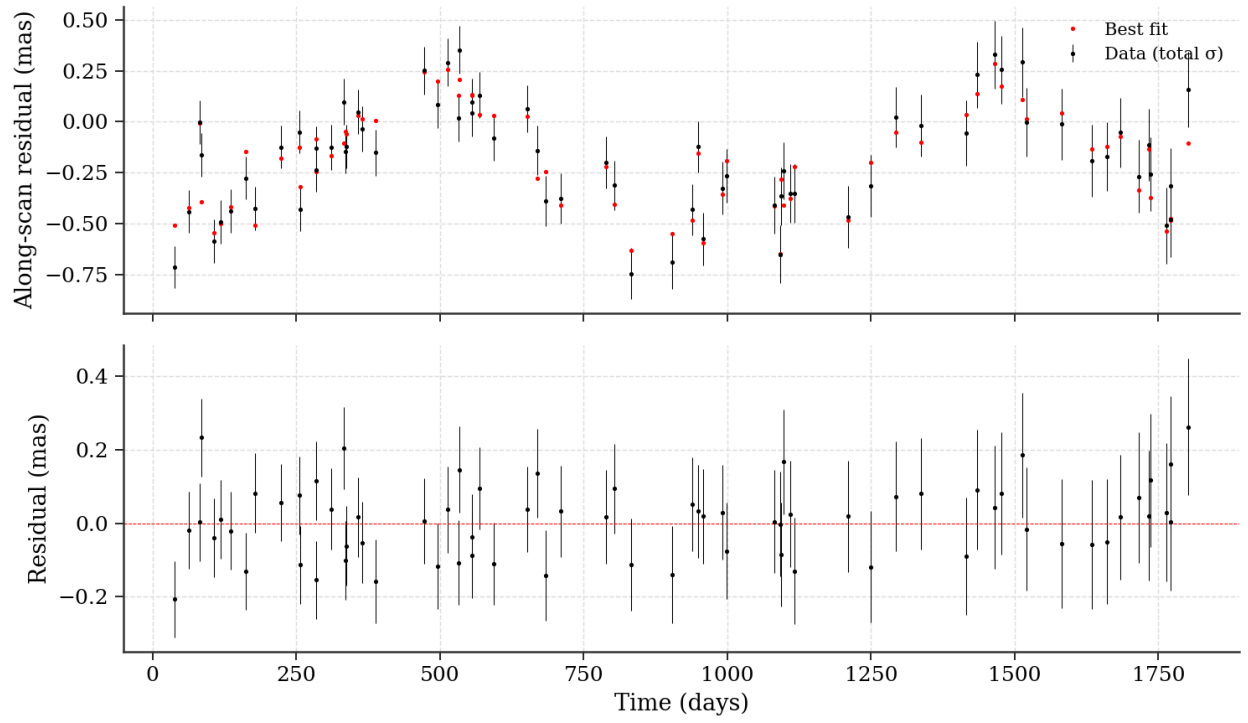


Figure A.19: Residuals plot for candidate 1878822452815621120.

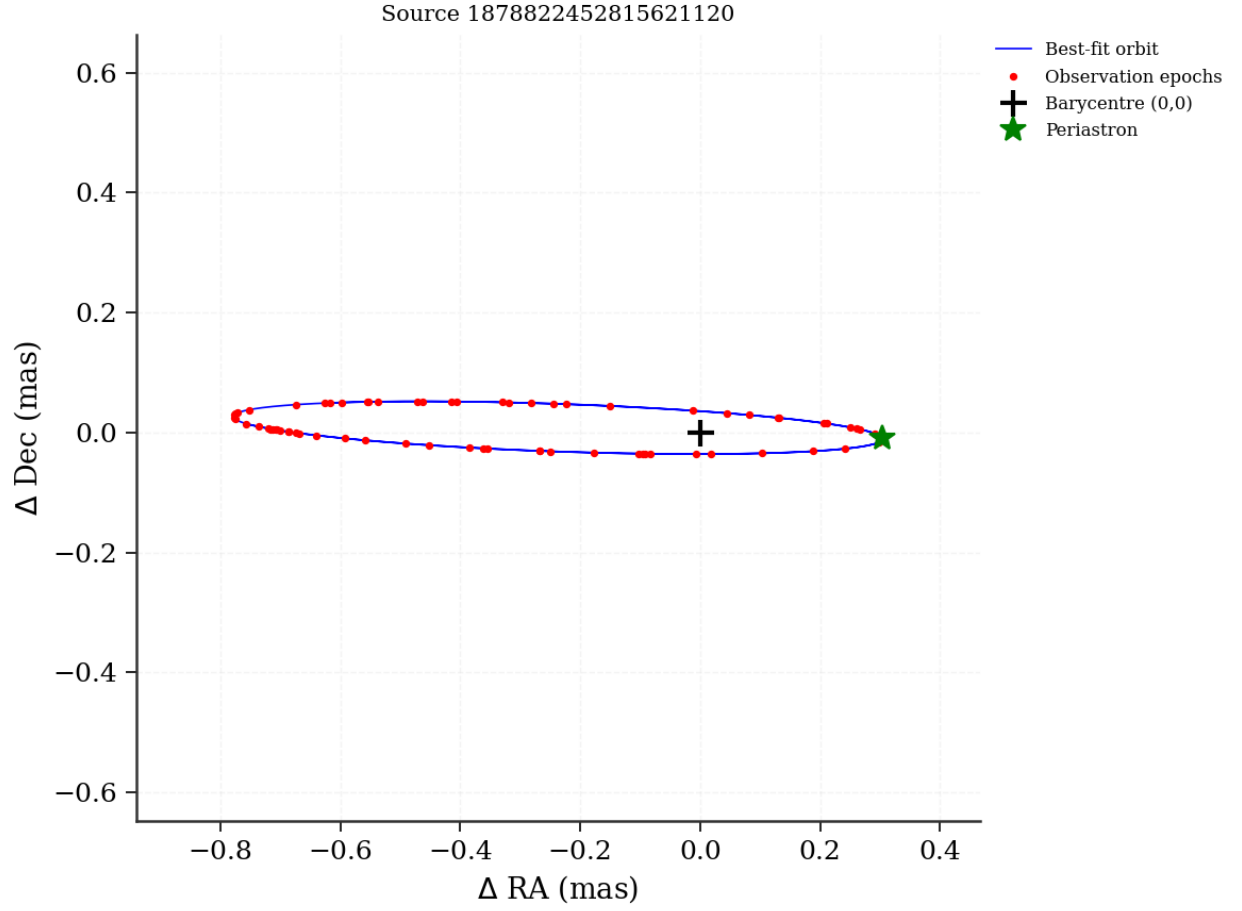


Figure A.20: Projected orbit of candidate 1878822452815621120 on the sky-plane.

4698424845771339520 plots

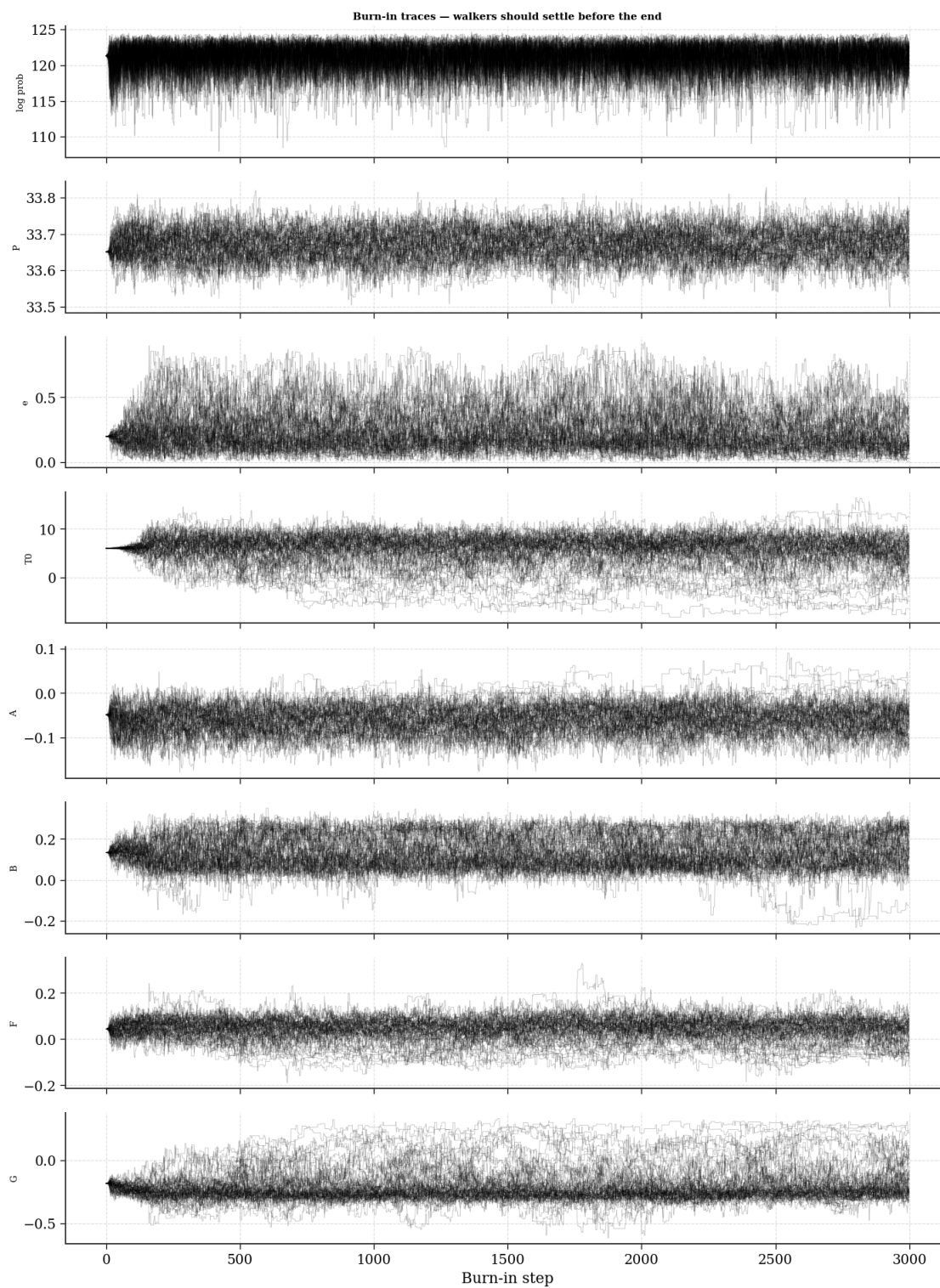


Figure A.21: MCMC burn-in traces for candidate 4698424845771339520.

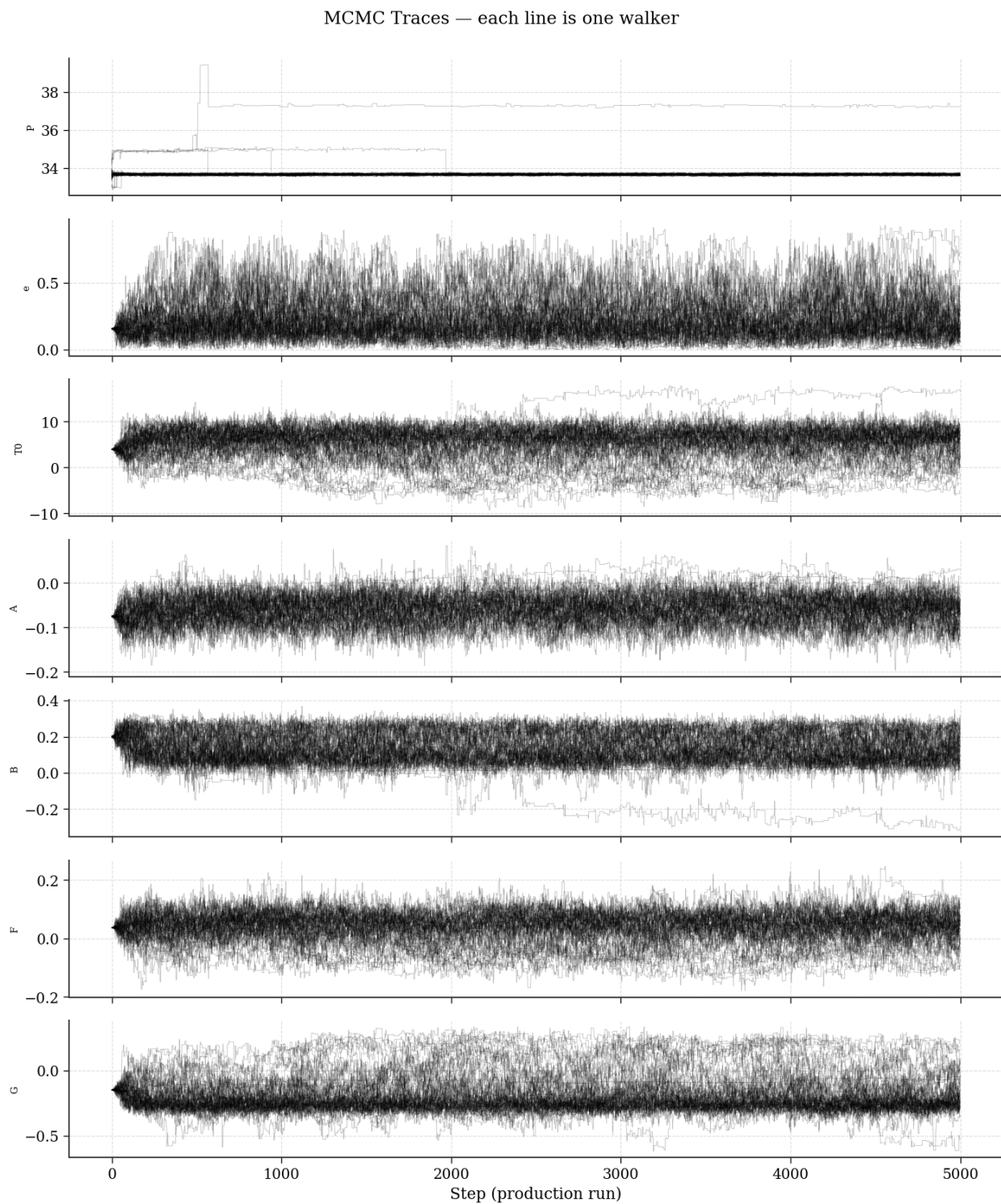


Figure A.22: MCMC production traces for candidate 4698424845771339520.

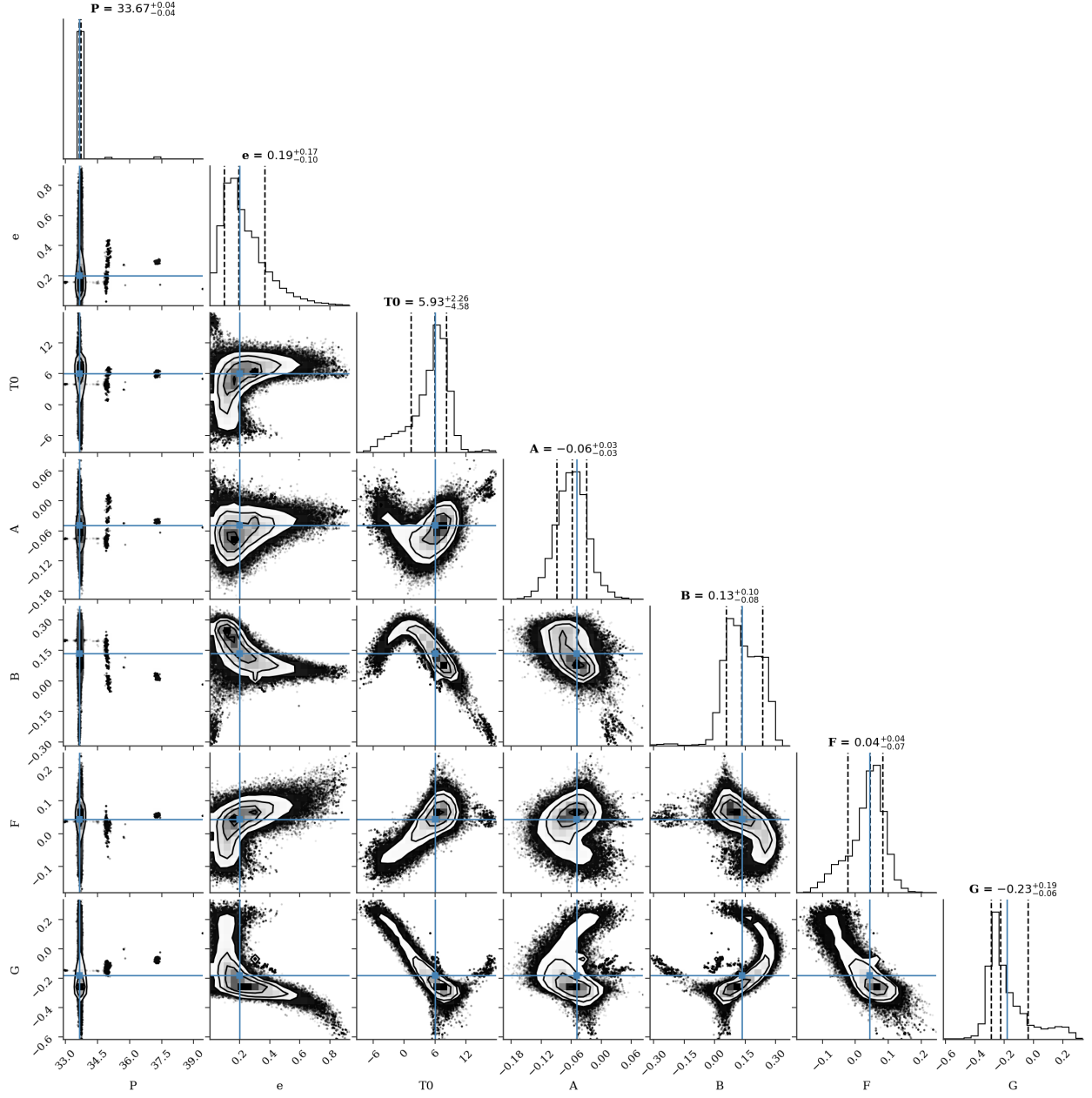


Figure A.23: Corner plot for candidate 4698424845771339520.

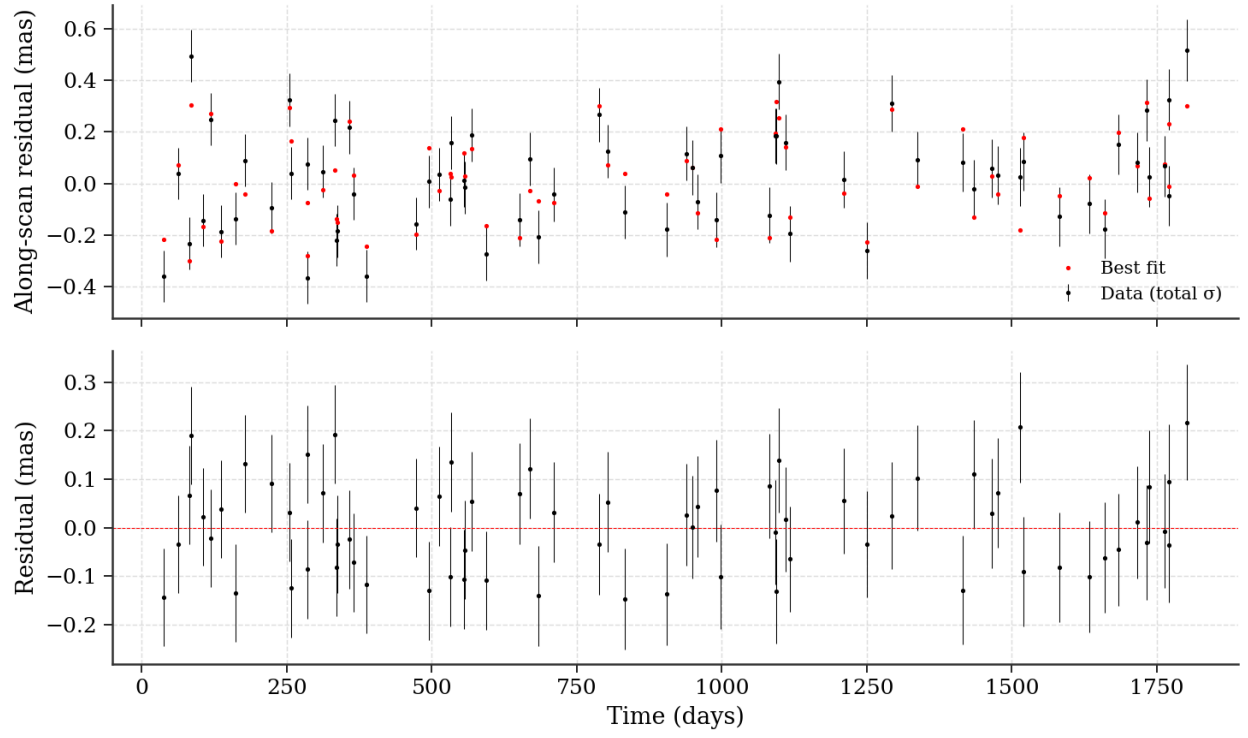


Figure A.24: Residuals plot for candidate 4698424845771339520.

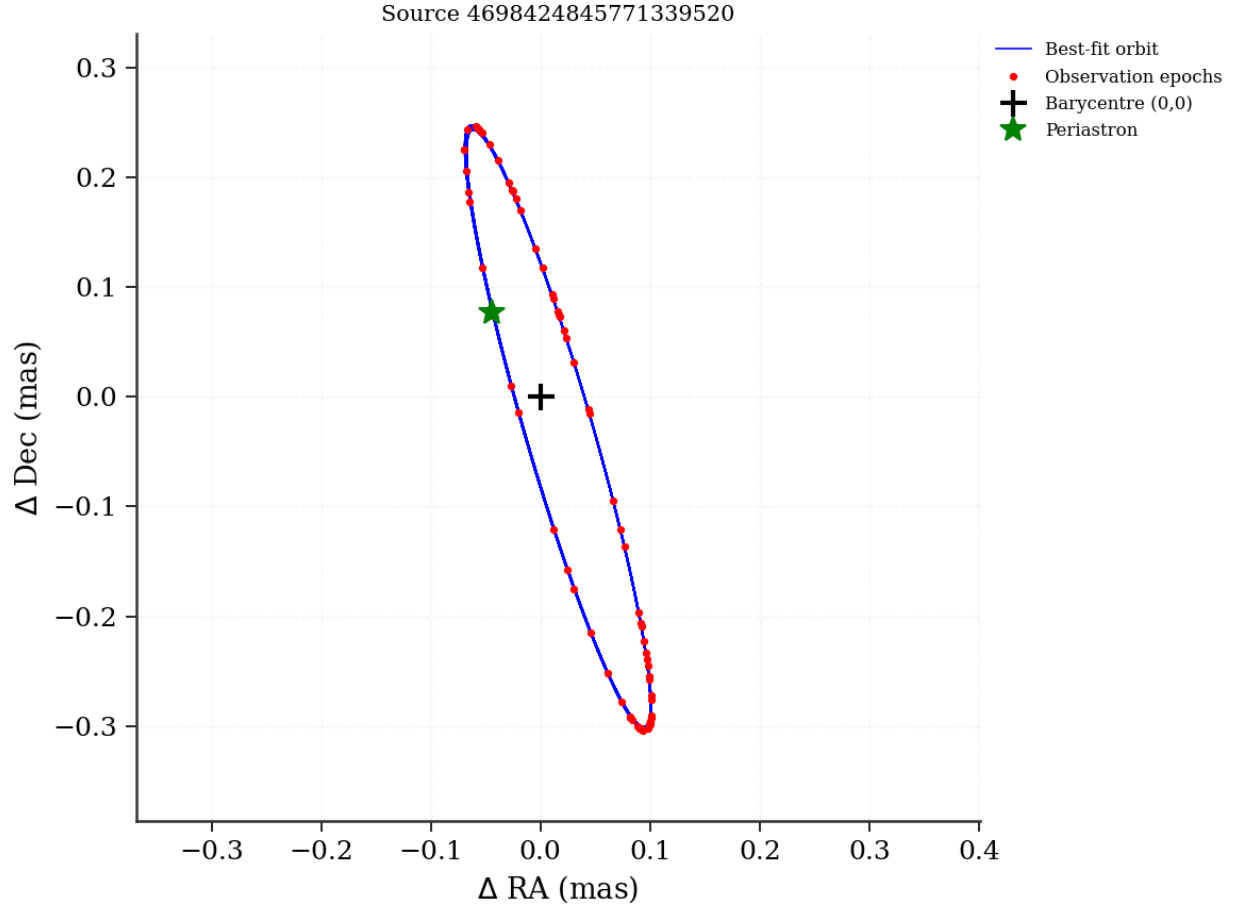


Figure A.25: Projected orbit of candidate 4698424845771339520 on the sky-plane.

Appendix B

Figures from 4.4 and 4.5

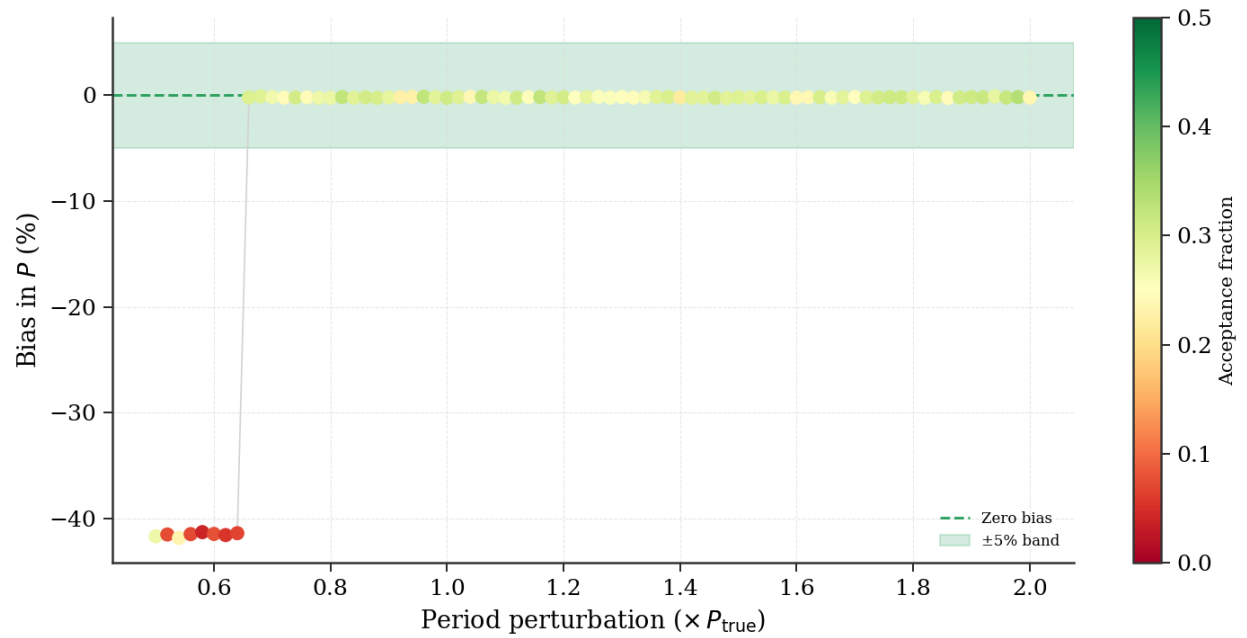


Figure B.1: Period perturbation as a fraction of P_{true} plotted against the bias in percent in recovered period over a five-year mission, for source 423297927866697088 subsection 4.4.2.

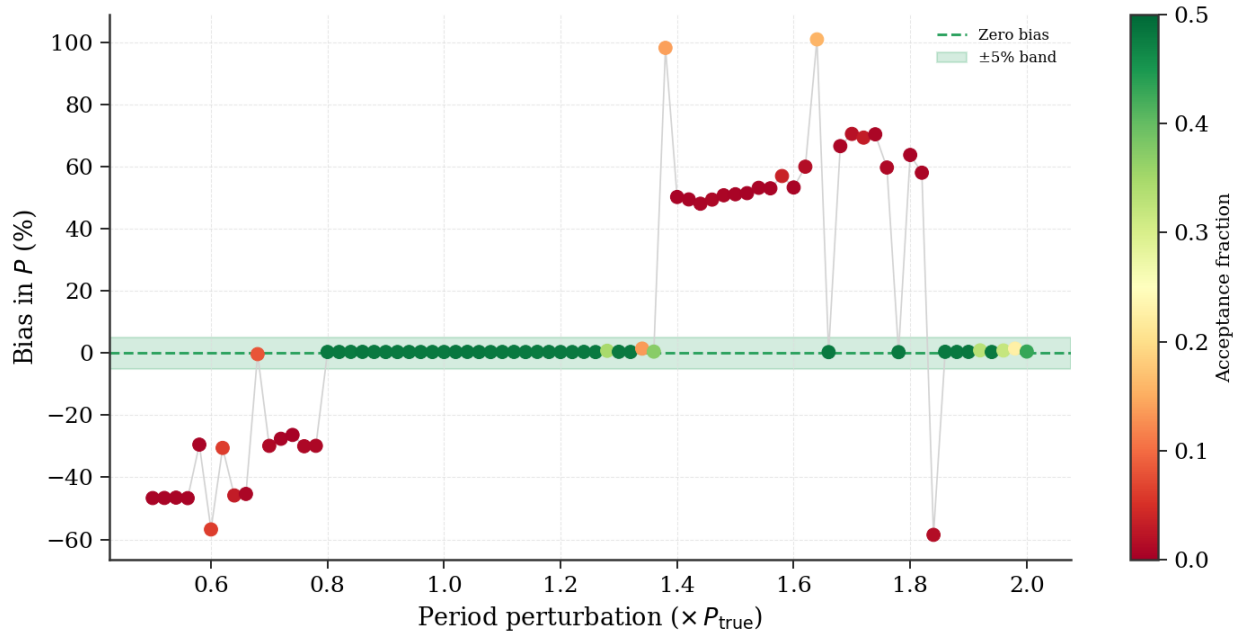


Figure B.2: Period perturbation as a fraction of P_{true} plotted against the bias in percent in recovered period over a five-year mission, for source 4062446910648807168 in subsection 4.4.3.

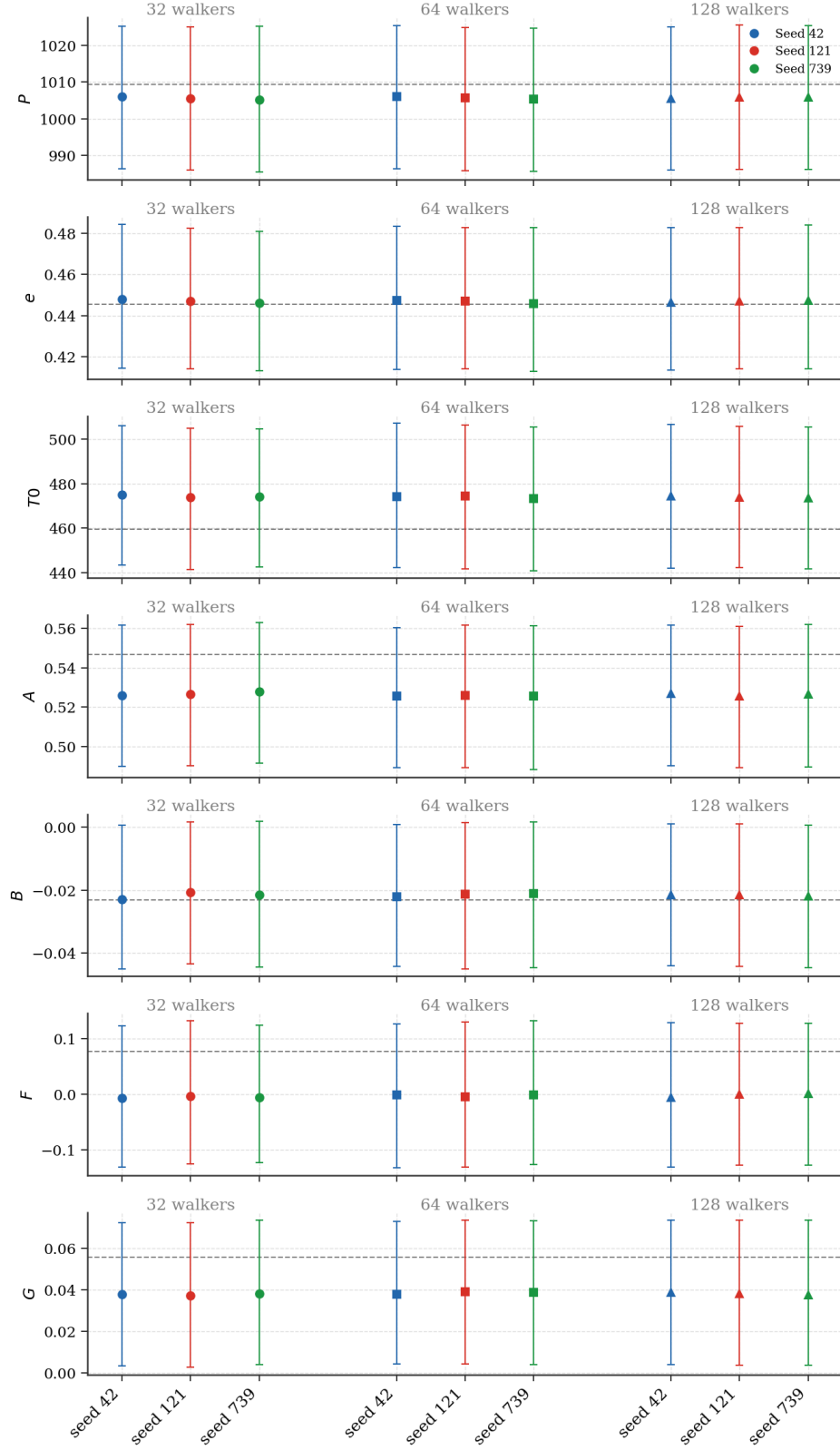


Figure B.3: Convergence robustness test for candidate 1878822452815621120 in section 4.5. Color denotes random seed, and columns are grouped by number of walkers. Plotted is the recovered posterior median with the 16th and 84th percentiles as error bars. The dashed black lines mark the true values.