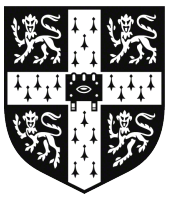


UNIVERSITY OF CAMBRIDGE

S2 Advanced Statistics - Coursework Report

Jacob Tutt (JLT67)

Department of Physics, University of Cambridge

Coursework Documentation

April 4, 2025

Word Count: 2987

1 Introduction

As one of civilisation's earliest natural science, astronomy has documented records of systematic observations dating back as far as 1000 BCE, with the Assyro-Babylonians' study of the periodic motion of celestial bodies [4]. A key item in the portfolio of archaeological artifacts is the 2,000-year-old Antikythera Mechanism, which for a century after its discovery in 1901, was widely believed to represent a 365-day Egyptian calendar [8]. However accurate measurements provided through CT imaging in Ramsey [9] revealed a 354-day lunar calendar [3], providing evidence for the Ancient Greeks' strong understanding of the near universe and lunar phases [5].

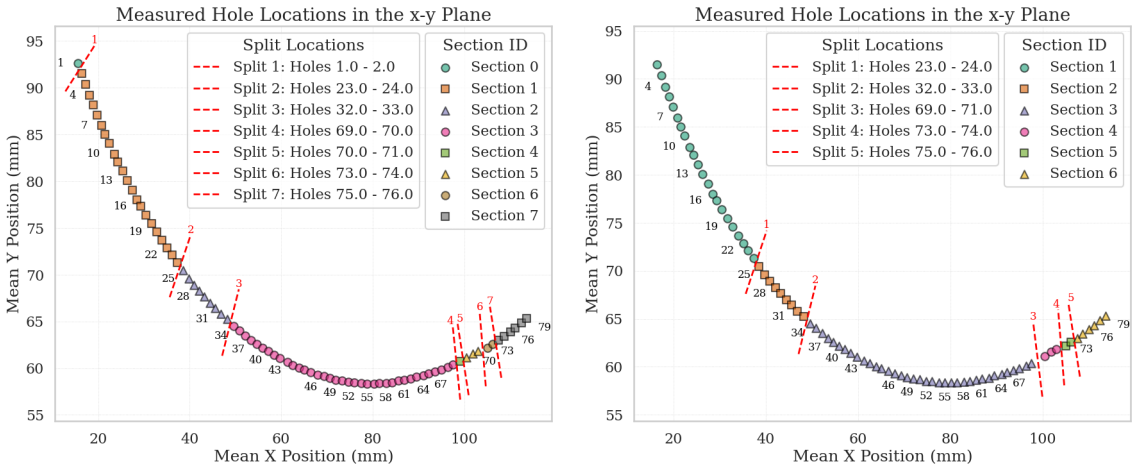
This project aims to apply a Bayesian Hamiltonian Monte Carlo posterior analysis to reevaluate the approximately 25% of the ring that remains, following a similar approach to that presented in Woan and Bayley [11]. Through this analysis, we seek to infer the most probable number of holes in the overall mechanism thus uncovering its original purpose.

2 Analysis Pipeline

As outlined in section 3 and section 4, this project considers three distinct (sub-)datasets, each evaluated under both an isotropic and anisotropic Gaussian likelihood function. To automate the different analysis approaches, the `Calender_Analysis` pipeline was built which handles all permutations through the initialisation parameters `model_type` and `filtering`. Full documentation for which is available here.

3 Data

To account for the holes being distributed across eight fragmented sections, we define two parameter types: global parameters describing the full ring and fragment-specific ones accounting for the relative shifts and rotations (see section 4). As the goal is to constrain the global parameters, we must consider the information each fragment provides about the overall structure relative to the additional dimensionality. Originally presented by Budiselic et al. [3], this report considers three filtering schemes: using all data (Figure 1a), excluding sections with only one hole (Figure 1b), and removing fragments with fewer than three holes and edge points. Below, the resultant datasets for the first two are shown with the third visualised in the repository. Throughout this report, we predominantly discuss the results using the moderate filtering criteria, however the results for all three are presented at the end.



(a) Unfiltered

(b) Lightly Filtered

Figure 1: Visualisation of Datasets

DATA SUMMARY	
Total Sections	8
Total Holes	81
Section ID	**Hole Range**
1	1–1
2	2–23
3	24–32
4	33–69
5	70–70
6	71–73
7	74–75
8	76–81

Table 1: Full Dataset Summary

DATA SUMMARY	
Total Sections	6
Total Holes	79
Section ID	**Hole Range**
1	2–23
2	24–32
3	33–69
4	71–73
5	74–75
6	76–81

Table 2: Filtered Dataset Summary

4 Probabilistic Model

This section presents the parameters used to model the hole positions, following Woan and Bayley [11]'s framework, along with the two likelihood functions employed. We present this in a general form for the i-th hole in the j-th section, assuming s total sections ($j \in \{0, 1, \ldots, s-1\}$).

4.1 Hole Positions

The global parameters of the ring, previously mentioned, are the total number of holes, N, and the fixed radius r. Each section j then has its own set of parameters, accounting for translational offsets with unique origins for the center of each arc, $\vec{r}_0^{\,j} = (x_0^j, y_0^j)$, and rotational offsets through similar angular offset parameters, α_j.

Overall, for a given total number of holes, N, the angular position of the i-th hole in the j-th section can be defined as:

$$\phi_{ij} = 2\pi \frac{(i-1)}{N} + \alpha_j, \tag{1}$$

The Cartesian coordinates of the hole can then be given by:

$$x_i = x_0^j + r\cos(\phi_{ij}) \tag{2}$$

$$y_i = y_0^j + r\sin(\phi_{ij}) \tag{3}$$

The notable assumptions are that the holes are evenly spaced around the circular ring and that all holes lie in the two-dimensional (x–y) plane. Furthermore, we assume all parameters to be continuous. This is trivial for all but N, which can be interpreted by allowing a single anomaly for the hole spacings.

4.2 Likelihood

4.2.1 Mahalanobis Distance

We employ the Mahalanobis distance to quantify the discrepancy between observed $\vec{d}_{obs,i}$ and predicted hole positions $\vec{d}_{model,i}$, which generalises the Euclidean distance to account for measurement uncertainty during construction through the covariance matrix Σ:

$$D_M^2 = (\vec{d}_{\text{obs},i} - \vec{d}_{\text{model},i})^\top \Sigma^{-1} (\vec{d}_{\text{obs},i} - \vec{d}_{\text{model},i}), \tag{4}$$

In evaluating the overall dataset's agreement with the model, we assume that the covariance matrix is constant across all holes and the errors are independent and identically distributed. This allows us to adopt a Gaussian error model, which is applied using two approaches:

4.2.2 Isotropic Model

We first quantify the discrepancies using an isotropic covariance, in which we assume the measurement errors in the orthogonal dimensions of the plane to be equal, σ and uncorrelated. Note that in this case the choice of orthogonal directions is arbitrary.

$$\Sigma_{iso} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \tag{5}$$

In defining the cartesian residuals in the x and y dimensions as:

$$e_{i,x} = x_{\text{obs},i} - x_{\text{model},i}, \tag{6a}$$
$$e_{i,y} = y_{\text{obs},i} - y_{\text{model},i}. \tag{6b}$$

The overall Gaussian log-likelihood function can be written as:

$$\log L(D|\theta) = -\frac{1}{2} \sum_i \left(\frac{(e_{i,x})^2 + (e_{i,y})^2}{\sigma^2} \right) - n \log(2\pi\sigma^2) \tag{7}$$

4.2.3 Anistropic Model

We next consider a model that accounts for directionally dependent uncertainties by treating the radial (σ_r) and tangential (σ_t) errors (with respect to the ring's center) separately, resulting in an anisotropic covariance matrix. The a priori reasoning for this model choice is expanded upon in section 8.

$$\Sigma_{aniso} = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_t^2 \end{bmatrix}, \tag{8}$$

The cartesian residuals (Equation 6) are transformed to radial and tangential error components using the angle of the hole with respect to the ring, ϕ_i (Equation 1):

$$e_{i,r} = e_{i,x} \cos\phi_i + e_{i,y} \sin\phi_i, \tag{9}$$
$$e_{i,t} = -e_{i,x} \sin\phi_i + e_{i,y} \cos\phi_i, \tag{10}$$

Our resultant anisotropic Gaussian log-likelihood function can be written as:

$$\log L(D|\theta) = -\frac{1}{2} \sum_i \left(\frac{(e_{i,r})^2}{\sigma_r^2} + \frac{(e_{i,t})^2}{\sigma_t^2} \right) - n \log(2\pi\sigma_r\sigma_t) \tag{11}$$

4.3 Implementation

Along with N and r, these uncertainty terms are treated as global parameters. The model thus has between 15 and 28 parameters depending on the model type and filtering scheme employed. The `likelihood` function within the pipeline supports both models and can optionally provide the logarithmic and/or negative values. Due to the high dimensionality of the parameter space, likelihood values are typically extreme and thus the logarithmic form is typically more numerically stable.

5 Likelihood Derivatives

To support gradient-based optimisation and later Hamiltonian Monte Carlo, the pipeline provides the models derivatives with respect to all parameters.

$$\frac{\partial}{\partial \theta^{\mu}} \log \mathcal{L}(D \mid \boldsymbol{\theta}), \tag{12}$$

This is implemented in `grad_likelihood` using the automatic differentiation functionality of the JAX library [2]. Its validity was verified by deriving the analytical derivates for the log-likelihood using the chain rule, which is presented in `analytic_grad_loglikelihood`. Below, we present an example comparing the derivative values for the global parameters in the anisotropic model, with the full set of results for both models reported in the repository.

Parameter	Auto-Diff	Analytic-Diff	Difference
N	0.308698	0.308698	3.552714e-15
r	-253.295342	-253.295342	8.526513e-14
σ_r	-391.598526	-391.598526	0.00000e+00
σ_t	-579.331346	-579.331346	2.273737e-13

Table 3: Comparison of Automatic and Analytic Differentiation

To check both implementations' robustness this was run for 100 different initialisations of the parameter spaces which showed a maximum deviation of 2.3×10^{-12}. Noteably, the analytical method is significantly faster (66.5%) and more memory efficient (65.1%).

Method:	Auto-Diff	Analytic-Diff
Avg Execution Time (s)	0.053956 ± 0.109242	0.018073 ± 0.078173
Avg Peak Memory (KB)	269.99 ± 105.16	94.21 ± 78.67
Gradient Agreement	MATCH	
Max Deviation	2.274e−12	

Table 4: Comparison of automatic vs manual differentiation

6 Maximium Likelihood Estimates

To obtain an initial estimate of the parameters, we attempt to determine the Maximum Likelihood Estimate (MLE) by minimising the negative log-likelihood. This is implemented in the `max_likelihood_est` function which support various gradient-based optimisation algorithms, including Adam and SciPy's BFGS minimisation [10]. Given the high dimensionality of the parameter space, we introduce optional support for performing each update based on sub-batches of the observed data. The introduction of stochasticity is intended to improve robustness by helping the optimiser escape local minima.

We first verified the ability of gradient-based methods to effectively learn and converge to a minimum by plotting their performance over time, for a single run from a randomly chosen initialisation. An example of this is shown in Figure 2, using the anisotropic model.

While results for the Adam optimiser are included in the repository, we focus on BFGS, which demonstrated better and more consistent performance. To obtain robust estimates, the minimisation was repeated for 20 randomly sampled initialisations from the parameter space. These are taken from the broad priors defined in subsection 7.2 to ensure sensible values. Below we report the top three log-likelihood values achieved by each optimiser across those runs.

No.	log(L)	AIC
1	219.34	-394.68
2	188.09	-332.18
3	-30.85	105.70

Table 5: Best MLE fits for Anisotropic Model

No.	log(L)	AIC
1	140.00	-238.00
2	134.37	-266.74
3	121.64	-201.28

Table 6: Best MLE fits for Isotropic Model

Two notable results are seen, while accounting for differences in model complexity using the Akaike Information Criterion (AIC), the anisotropic model achieves a lower score providing initial but inconclusive evidence that it is more successful at describing the observed data.

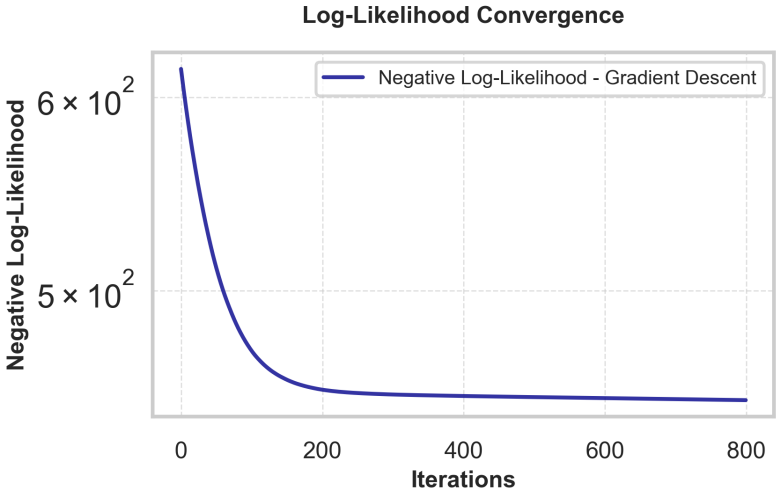
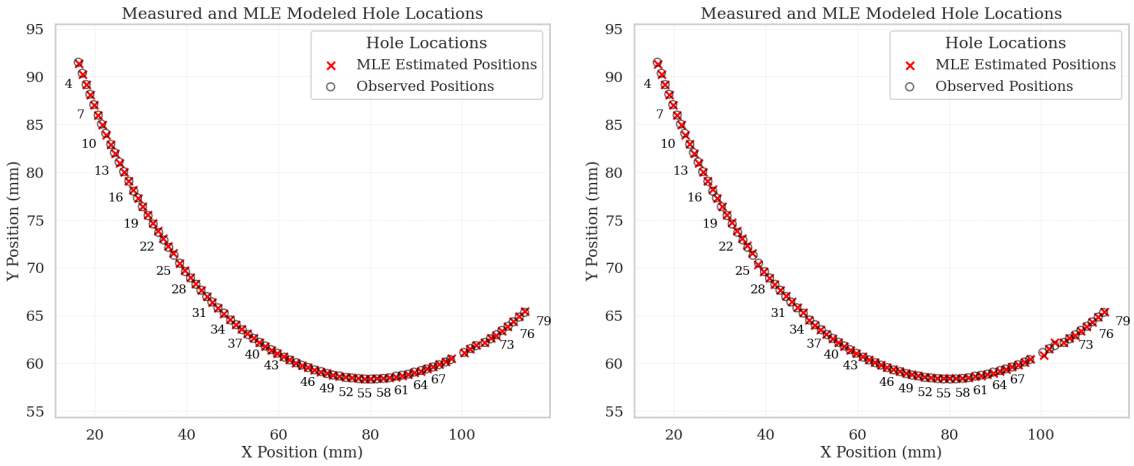


Figure 2: Convergence of Log-likelihood in Gradient Descent (Anisotropic Model)

Secondly, the isotropic model yields much more consistent results across different initialisations, suggesting that its reduced dimensionality (21 versus 22) allows the minimisation to converge more accurately and efficiently.



(a) Ansitropic

(b) Isotropic

Figure 3: MLE Hole Positions

The achieved parameters (Table 7 and Table 8) appear sensible through a visual comparison of the modelled and observed hole positions in Figure 5. Notably, the only significant deviation is section 4 of the isotropic model, where the estimated values for x_0, y_0, and α_0 are relatively offset at 83.48, 131.11, and $-131.21°$, respectively. This offset is attributed to the fragment containing only three holes, resulting in its associated parameters being weakly constrained. This observation supports the arguments for filtering out smaller fragments, as discussed in section 3. Particularly interesting from this preliminary analysis is the suggestion that the isotropic model supports an Egyptian calendar with N = 364.84, while the anisotropic model yields N = 355.11, more consistent with a lunar calendar. Although likely the result of unsuccessful convergence in a local minimum, it highlights the subtlety and sensitivity of the parameter N with no obvious visual discrepancies despite the deviation.

Parameter	Value
N	355.11
r	77.39
σ_r	0.03
σ_t	0.12
$x_{0,1}$	79.68
$x_{0,2}$	80.36
$x_{0,3}$	79.88
$x_{0,4}$	79.12
$x_{0,5}$	79.17
$x_{0,6}$	82.13
$y_{0,1}$	136.11
$y_{0,2}$	135.54
$y_{0,3}$	135.76
$y_{0,4}$	135.49
$y_{0,5}$	136.1
$y_{0,6}$	136.06
α_1	-145.56
α_2	-151.87
α_3	-151.31
α_4	-145.00
α_5	-150.08
α_6	-153.13
Log-Likelihood	219.34

Table 7: Best MLE fits for Anisotropic Model

Parameter	Value
N	364.84
r	79.45
σ	0.10
$x_{0,1}$	80.96
$x_{0,2}$	79.37
$x_{0,3}$	80.02
$x_{0,4}$	83.48
$x_{0,5}$	79.76
$x_{0,6}$	81.96
$y_{0,1}$	137.79
$y_{0,2}$	138.38
$y_{0,3}$	137.84
$y_{0,4}$	131.11
$y_{0,5}$	137.52
$y_{0,6}$	138.17
α_1	-145.00
α_2	-143.85
α_3	-144.43
α_4	-131.21
α_5	-143.85
α_6	-145.56
Log-Likelihood	134.00

Table 8: MLE Parameters for Isotropic Model

7 Bayesian Posterior Estimates

This paper now moves on to greater understand the uncertainties in our measurements through posterior sample estimates, providing greater insight into the space's curvature and structure. This is achieved using a Hamiltonian Monte Carlo (HMC) algorithm.

7.1 Hamiltonian Monte Carlo (NUTS)

Due to the parameter space's high dimensionality and thus a need to outperform traditional random-walk MCMC methods, Hamiltonian Monte Carlo (HMC) [1] was employed. Efficient exploration is achieved by creating an analogy of physical phase space, thus allowing the use of Hamilton dynamics to evolve the system in 'time'. To do so, the parameter space $\vec{\theta}$ is expanded with momentum variables $\vec{p}$ which allows the algorithm to make informed, long-distance proposals while maintaining a high acceptance probability.

$$\mathcal{H}(\vec{\theta}, \vec{p}) = U(\vec{\theta}) + K(\vec{p}), \tag{13}$$

where the potential energy is defined as:

$$U(\vec{\theta}) = -\log p(\vec{\theta} \mid \mathcal{D}), \tag{14}$$

and the kinetic energy, with mass matrix, M:

$$K(\vec{p}) = \frac{1}{2}\vec{p}^{\top} M^{-1} \vec{p}, \tag{15}$$

Overall the system is governed by Hamilton's equations:

$$\frac{d\theta^k}{dt} = \frac{\partial \mathcal{H}}{\partial p^k}, \tag{16a}$$

$$\frac{dp^k}{dt} = -\frac{\partial \mathcal{H}}{\partial \theta^k}, \tag{16b}$$

HMC uses the leapfrog integrator to numerically evolve steps forwards due to its 'sympletic' properties. In this work, we exploit the Numpryo library's implementation of a No-U-Turn Sampler

(NUTS), a variant of HMC that builts in 'adaptive path length and mass matrix adaptation' [7] to avoid redundant computation.

7.2 Prior Choices

Our initial prior distributions were chosen to be largely uninformative and hence broad. For all parameters where the order of magnitude was intuitively constant, we applied uniform priors to reflect scale invariance in the fragment origins and rotations. However, the uncertainty parameter, σ, could not be trivially constrained to a given order of magnitude, and thus a log-uniform prior (Jeffreys) was more appropriate.

To inform these broad prior ranges, initial estimates were calculated via a simple least squares fitting, shown in Figure 4. The error parameter σ was loosely estimated by the standard deviation of the inter-hole distances, approximately 0.127. Notably, the prior on N allowed the chains to explore both Egyptian and lunar calendar hypotheses.

Parameter	Description	Distribution
N	Total number of holes	Uniform(330, 370)
r	Ring radius	Uniform(65, 90)
x_0	Section offsets (x-dimension)	Uniform(60, 100)
y_0	Section offsets (y-dimension)	Uniform(120, 160)
α	Angular offsets (degrees)	Uniform(-160, -120)
Errors	Measurement uncertainties (σ)	LogUniform(10^{-5}, 5)

Table 9: Prior distributions

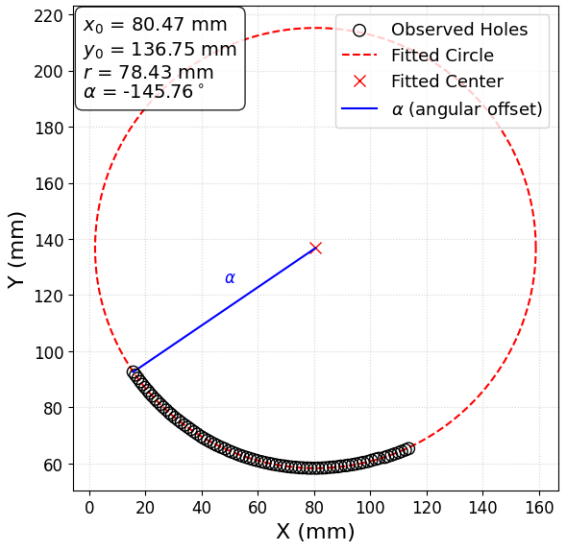


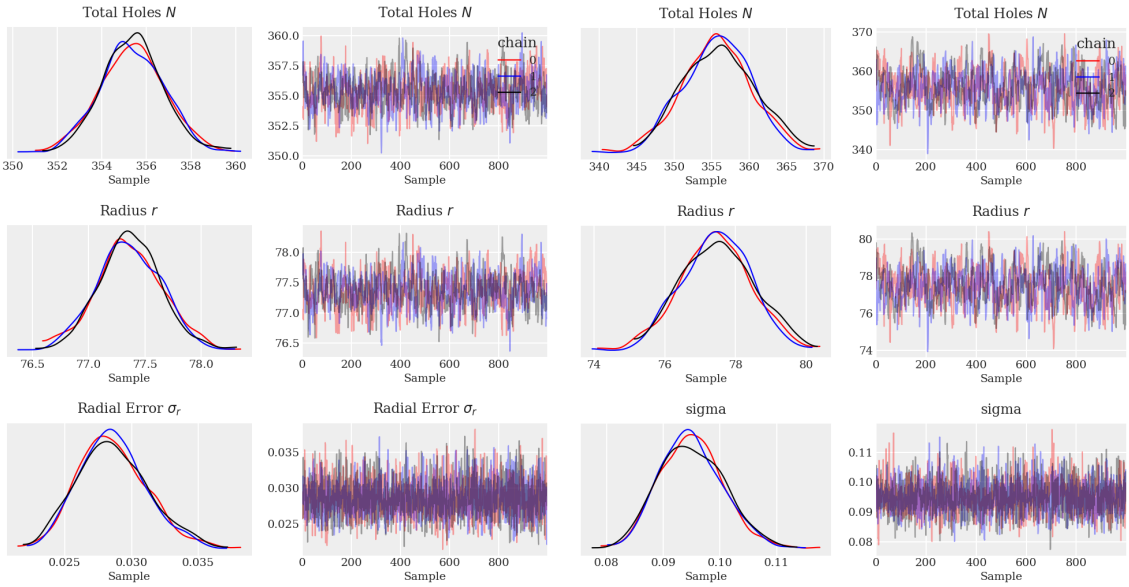
Figure 4: Least Squares Fit

7.3 Burn-In Period

The burn-in period is particularly relevant for NUTS samplers as it is used to automatically fine-tune both the step size and mass matrix. Helping the chains reach the target distribution before analysing samples as well as optimising the exploration of the parameter space. We found a burn-in period of 600 steps is sufficient for both models, based on visual inspection of the sampled trajectories across multiple chains. However, given the low computational cost, a more conservative burn-in of 2000 steps is used in the final analysis. More quantitative measures for convergence, including the Gelman–Rubin statistic and autocorrelation length, are discussed in subsection 7.4.

7.4 Optimising Initialisation

A grid search was conducted to identify the most computationally efficient initialisations for both the isotropic and anisotropic models (`run_hmc_optimisation`). This investigated step sizes (0.05,



(a) Ansitropic Model

(b) Isotropic Model

Figure 5: Demonstration of convergence of chains after burn-in of 600

0.1, 0.5, 1, 5), acceptance probabilities (0.65, 0.75, 0.85, 0.90), and mass matrix types (dense vs diagonal). The efficiency was optimised based on the 'time per effective samples', accounting for both the autocorrelation length and the iteration's computational cost. This used the number of effective samples calculated by `arviz` [6], taking the minimum across parameters for robustness. Convergence was further verified using the Gelman–Rubin statistic.

Notably, both models favored dense mass matrices, likely due to the strong parameter correlations between N and r. Beyond this, they showed differing preferences for step size and acceptance rate. Overall, the anisotropic model achieves more efficient sampling, achieving a minimum time per effective sample of 0.001435 seconds.

One caveat is ArviZ's tendency to suggest a number of effective samples that exceed the total number of draws for highly uncorrelated chains, leading to autocorrelation lengths below one. To ensure uncorrelated results, all subsequent thinning is applied by using autocorrelation lengths rounded up to the nearest integer. The best five results for each model are presented below.

Step Size	Acceptance Probability	Dense Matrix	Autocorr Length	Effective Samples	GR Stat.	Time/Iter (s)	Time/Eff Sample (s)
0.50	0.75	True	2.6555	753.17	1.0048	0.002176	0.005779
5.00	0.75	True	2.9802	671.09	1.0043	0.002012	0.005997
5.00	0.85	True	2.0157	992.28	1.0019	0.003420	0.006893
0.10	0.90	True	2.4076	836.69	1.0014	0.003383	0.008145
0.05	0.90	True	2.0792	961.93	1.0030	0.004041	0.008402

Table 10: Isotropic HMC optimisation

Step Size	Acceptance Probability	Dense Matrix	Autocorr Length	Effective Samples	GR Stat.	Time/Iter (s)	Time/Eff Sample (s)
0.05	0.65	True	0.8762	2282.63	1.0009	0.001637	0.001435
0.10	0.75	True	0.9740	2053.42	1.0011	0.001782	0.001735
0.50	0.65	True	1.1107	1800.60	1.0007	0.001592	0.001768
0.50	0.75	True	1.0111	1977.95	1.0007	0.001852	0.001873
1.00	0.65	True	1.1437	1748.65	1.0007	0.001780	0.002036

Table 11: Anisotropic HMC optimisation

7.5 Results

Using these optimised initialisations, the `run_hmc_optimised` and `thinned_hmc_analysis` functions automate the sampling process by calculating the autocorrelation and applying thinning. We

generate results for all three filtering schemes across both models. While aiming for 12,000 effective samples, some runs yield slightly fewer due to varying autocorrelation lengths. Below the full results for the mid-level filtering scheme are presented, along with the results of N across all models and filtering schemes. Full results for all can be found within the repository.

Parameter	**Median**	**68%**	**95%**	**99%**
N	355.28	$^{+1.37}_{-1.37}$	$^{+2.75}_{-2.72}$	$^{+3.54}_{-3.64}$
r	77.34	$^{+0.29}_{-0.28}$	$^{+0.56}_{-0.56}$	$^{+0.75}_{-0.74}$
σ_r	0.028	$^{+0.003}_{-0.002}$	$^{+0.006}_{-0.004}$	$^{+0.008}_{-0.005}$
σ_t	0.129	$^{+0.012}_{-0.010}$	$^{+0.024}_{-0.019}$	$^{+0.033}_{-0.024}$
$x_{0,1}$	79.69	$^{+0.20}_{-0.20}$	$^{+0.33}_{-0.33}$	$^{+0.40}_{-0.53}$
$x_{0,2}$	79.91	$^{+0.22}_{-0.22}$	$^{+0.37}_{-0.38}$	$^{+0.60}_{-0.61}$
$x_{0,3}$	79.86	$^{+0.04}_{-0.04}$	$^{+0.06}_{-0.06}$	$^{+0.09}_{-0.09}$
$x_{0,4}$	81.44	$^{+1.09}_{-1.12}$	$^{+1.798}_{-1.95}$	$^{+2.85}_{-3.05}$
$x_{0,5}$	81.49	$^{+2.44}_{-2.37}$	$^{+4.50}_{-3.95}$	$^{+5.80}_{-5.94}$
$x_{0,6}$	83.24	$^{+0.40}_{-0.39}$	$^{+0.66}_{-0.65}$	$^{+1.06}_{-1.04}$
$y_{0,1}$	136.03	$^{+0.20}_{-0.21}$	$^{+0.35}_{-0.34}$	$^{+0.41}_{-0.55}$
$y_{0,2}$	135.72	$^{+0.26}_{-0.27}$	$^{+0.44}_{-0.45}$	$^{+0.53}_{-0.70}$
$y_{0,3}$	135.71	$^{+0.28}_{-0.29}$	$^{+0.47}_{-0.48}$	$^{+0.57}_{-0.76}$
$y_{0,4}$	135.84	$^{+0.83}_{-0.81}$	$^{+1.30}_{-1.46}$	$^{+1.47}_{-2.32}$
$y_{0,5}$	135.85	$^{+0.44}_{-0.47}$	$^{+0.74}_{-0.83}$	$^{+0.86}_{-1.10}$
$y_{0,6}$	136.43	$^{+0.29}_{-0.29}$	$^{+0.49}_{-0.50}$	$^{+0.59}_{-0.78}$
α_1	-145.70	$^{+0.06}_{-0.06}$	$^{+0.11}_{-0.11}$	$^{+0.11}_{-0.17}$
α_2	-145.65	$^{+0.17}_{-0.23}$	$^{+0.29}_{-0.34}$	$^{+0.46}_{-0.52}$
α_3	-145.54	$^{+0.17}_{-0.17}$	$^{+0.34}_{-0.34}$	$^{+0.52}_{-0.52}$
α_4	-146.68	$^{+0.92}_{-0.92}$	$^{+1.49}_{-1.49}$	$^{+2.46}_{-2.29}$
α_5	-146.28	$^{+1.83}_{-1.94}$	$^{+3.09}_{-3.50}$	$^{+4.70}_{-4.61}$
α_6	-147.82	$^{+0.46}_{-0.46}$	$^{+0.74}_{-0.69}$	$^{+1.09}_{-1.09}$

Table 12: Anisotropic: Posterior Estimates (middle filtering)

Parameter	**Median**	**68%**	**95%**	**99%**
N	355.31	$^{+4.20}_{-4.16}$	$^{+8.32}_{-8.11}$	$^{+9.99}_{-10.90}$
r	77.35	$^{+0.91}_{-0.90}$	$^{+1.80}_{-1.74}$	$^{+2.17}_{-2.34}$
σ	0.095	$^{+0.006}_{-0.005}$	$^{+0.013}_{-0.010}$	$^{+0.017}_{-0.013}$
$x_{0,1}$	79.69	$^{+0.65}_{-0.63}$	$^{+1.06}_{-1.05}$	$^{+1.63}_{-1.66}$
$x_{0,2}$	79.69	$^{+0.74}_{-0.75}$	$^{+1.26}_{-1.23}$	$^{+1.96}_{-1.99}$
$x_{0,3}$	79.87	$^{+0.11}_{-0.11}$	$^{+0.24}_{-0.18}$	$^{+0.30}_{-0.19}$
$x_{0,4}$	81.65	$^{+3.51}_{-3.57}$	$^{+5.89}_{-5.96}$	$^{+9.42}_{-8.93}$
$x_{0,5}$	81.92	$^{+3.47}_{-3.69}$	$^{+12.09}_{-9.24}$	$^{+16.45}_{-13.20}$
$x_{0,6}$	83.18	$^{+1.43}_{-1.39}$	$^{+2.81}_{-2.43}$	$^{+3.64}_{-1.71}$
$y_{0,1}$	136.03	$^{+0.66}_{-0.67}$	$^{+1.34}_{-1.14}$	$^{+1.76}_{-2.26}$
$y_{0,2}$	135.74	$^{+0.87}_{-0.85}$	$^{+1.69}_{-1.77}$	$^{+2.17}_{-2.26}$
$y_{0,3}$	135.71	$^{+0.93}_{-0.91}$	$^{+1.82}_{-1.72}$	$^{+2.21}_{-2.35}$
$y_{0,4}$	136.08	$^{+1.30}_{-1.33}$	$^{+2.04}_{-2.25}$	$^{+3.19}_{-3.77}$
$y_{0,5}$	135.97	$^{+2.20}_{-2.55}$	$^{+3.80}_{-5.23}$	$^{+4.55}_{-6.21}$
$y_{0,6}$	136.42	$^{+1.05}_{-1.03}$	$^{+1.89}_{-1.60}$	$^{+2.33}_{-2.48}$
α_1	-145.70	$^{+0.17}_{-0.17}$	$^{+0.29}_{-0.29}$	$^{+0.52}_{-0.46}$
α_2	-145.65	$^{+0.57}_{-0.63}$	$^{+1.15}_{-1.03}$	$^{+1.55}_{-1.67}$
α_3	-145.53	$^{+0.57}_{-0.63}$	$^{+1.15}_{-1.03}$	$^{+1.38}_{-1.60}$
α_4	-146.79	$^{+2.92}_{-2.75}$	$^{+4.83}_{-4.64}$	$^{+7.51}_{-7.10}$
α_5	-146.62	$^{+5.04}_{-5.73}$	$^{+9.80}_{-10.54}$	$^{+10.54}_{-12.83}$
α_6	-147.77	$^{+1.32}_{-1.38}$	$^{+2.58}_{-2.69}$	$^{+3.38}_{-3.44}$

Table 13: Isotropic: Posterior Estimates (middle filtering)

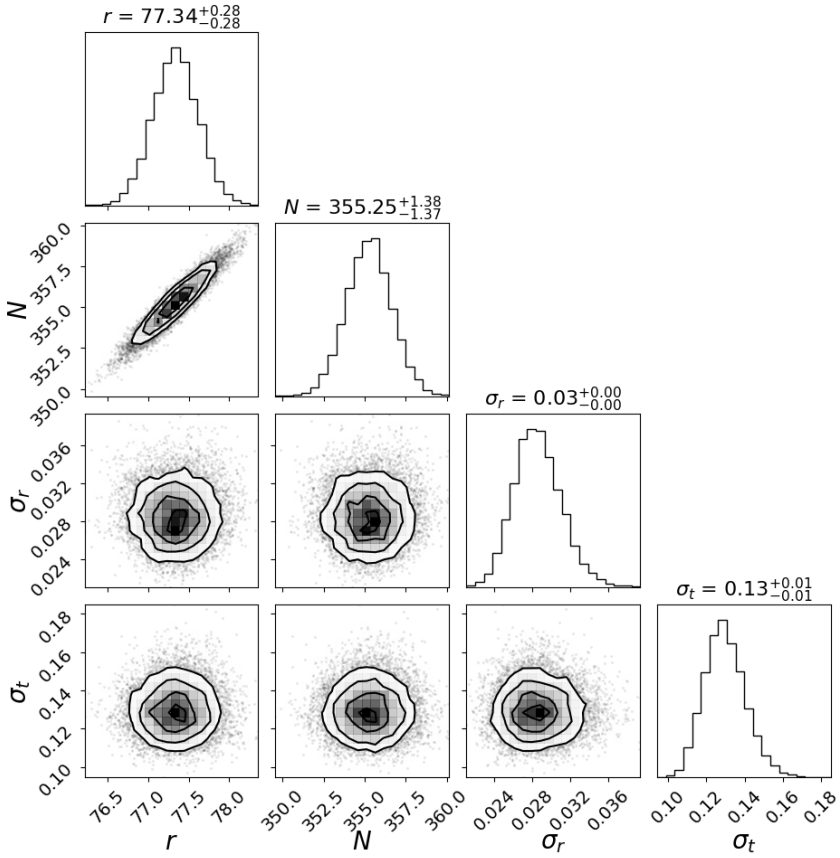


Figure 6: Anisotropic: Corner Plots (Mean)

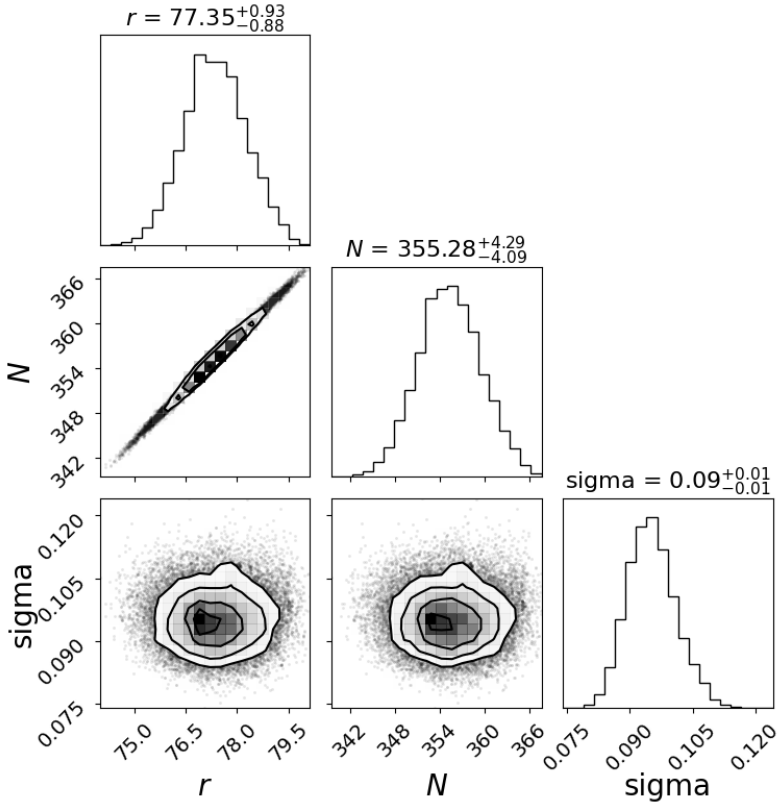


Figure 7: Isotropic: Corner Plot (Mean)

Filtering	Anisotropic N	Isotropic N
None	$355.182^{+1.32}_{-1.39}$	$354.791^{+4.58}_{-5.44}$
Basic	$355.281^{+1.37}_{-1.37}$	$355.308^{+4.20}_{-4.16}$
Full	$354.090^{+1.44}_{-1.48}$	$353.806^{+4.10}_{-4.72}$

Table 14: Posterior medians for N across filtering schemes.

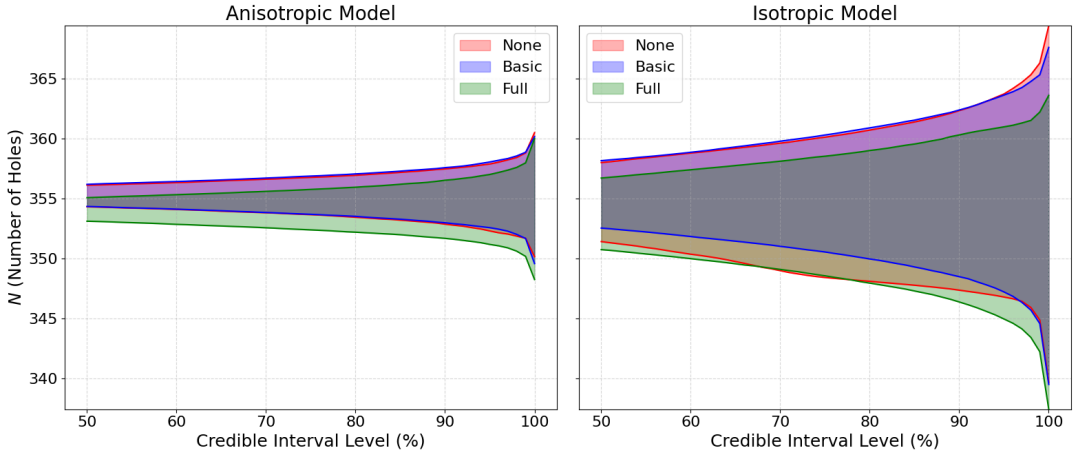


Figure 8: Credible Intervals for N under different filtering schemes.

7.6 Discussion

Results in Table 12 were first cross-validated against those published in Woan and Bayley [11], showing near-perfect agreement and falling well within the respective 68% uncertainty. Both corner plots demonstrate the expected strong correlation between parameters N and r, which are inherently linked within the model.

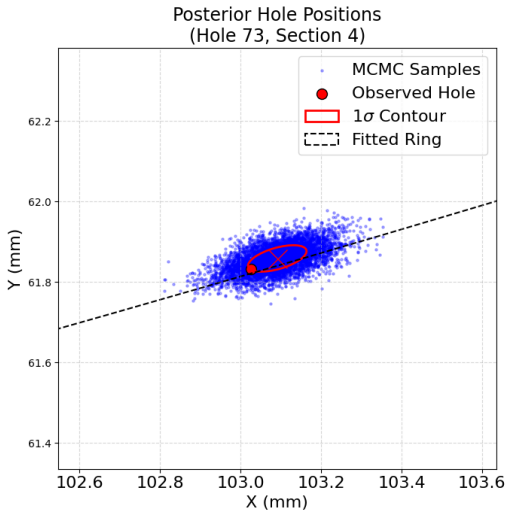
Best illustrated in Figure 8, we see the posterior medians appear mostly consistent between the isotropic and anisotropic models. However, the isotropic model consistently achieves credible intervals typically 2–3 times wider than those of the anisotropic model, indicating a weaker ability to constrain the parameter's estimates. This provides further evidence that the isotropic model is less successful in describing the observed data.

Focusing on the parameter N, we assess the effect of the different data filtering criteria. As shown in Table 14 and Figure 8, and similarly reported in Woan and Bayley [11], the harshest filtering leads to lower estimates of N, more consistent with the lunar calendar. This outcome is reasoned by the fact it removes the 4th and 5th fragments, which are shown by Table 12 and Table 13 to have the most poorly constrained individual parameters (x_0, y_0 and α_0).

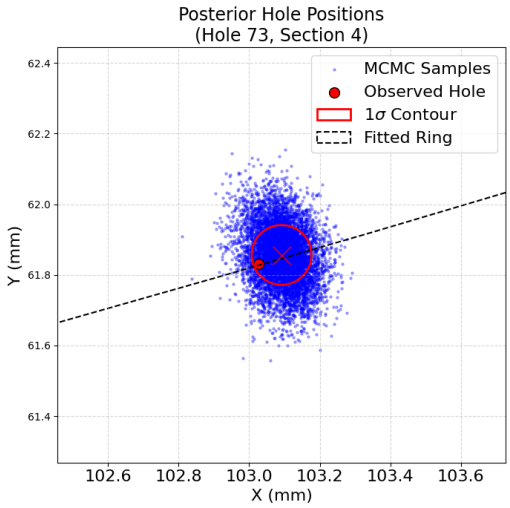
As is the overall focus of this paper, we find that all model variants and filtering schemes confidently exclude the Egyptian calendar value of $N = 365$ with 95% confidence. The anisotropic model can more confidently exclude it, with 99% certainty. Overall we conclude this analysis provides overwhelming statistical support for a lunar calender, as originally proposed in Budiselic et al. [3].

7.7 Visualisation of Posterior Samples

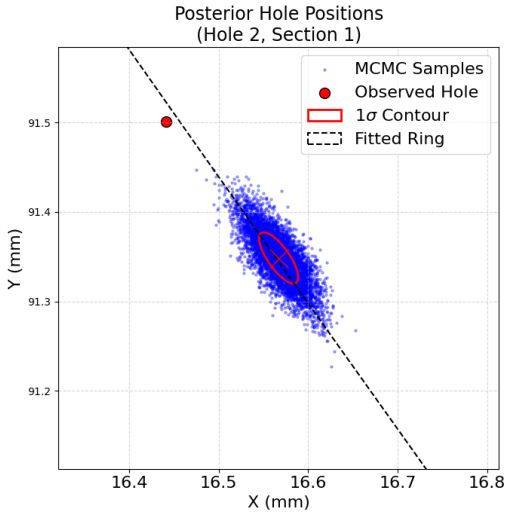
In Figure 9, we provide further evidence for the advantages of the anisotropic model in its ability to account for the distance-dependent uncertainties. We show two cases: a typical example (Hole 73) and a more extreme example (Hole 2). Both clearly show that the discrepancies between the modelled and observed hole positions are much greater in the tangential direction than in the radial direction. In allowing the anisotropic model to capture this asymmetry, it is able to provide a more accurate and realistic representation of the data.



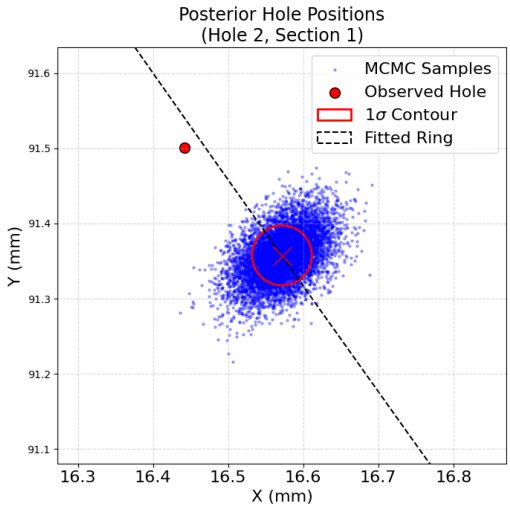
(a) Anisotropic Hole 73



(b) Isotropic Hole 73



(c) Anisotropic Hole 2



(d) Isotropic Hole 2

Figure 9: Posterior Hole Distributions

8 Model Comparison

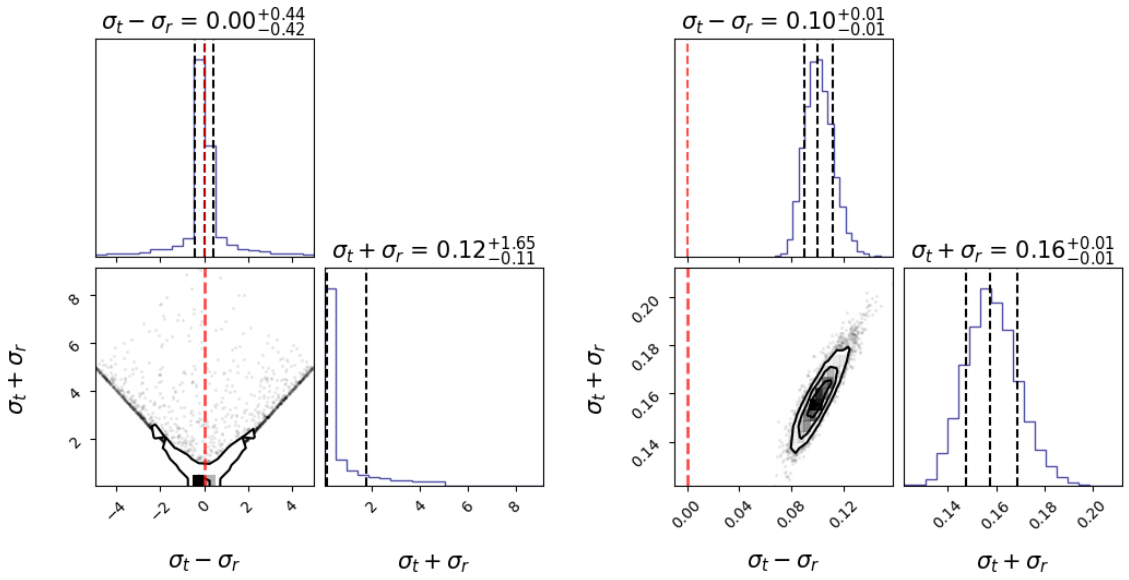
As introduced in section 4, this paper explores the advantages of an anisotropic covariance matrix, Σ_{aniso} compared with an isotropic matrix representing a univariate scalar uncertainty σ. This additional flexibility allows the model to account for directionally dependent uncertainties, which in this project, are rotated to be the radial and tangential directions relative to the ring. This is considered beneficial a priori, as the original construction of the ring likely relied on an accurate circular stencil. The placement of holes is therefore likely more precise in the radial direction than along the circumference's arc. Throughout this paper, we have provided indications of the anisotropic model's success through its ability to greater constrain the global parameters in subsection 7.6 and visual inspection of errors alignment in subsection 7.7. This section now aims to apply some more quantitative model comparisons.

8.1 Savage-Dickey

By considering the isotropic model as a nested generational of the anisotropic model, with $\sigma_r = \sigma_t$, we are able to apply the Savage-Dickey Density Ratio. This allows us to determine the Bayes factor by simply evaluating the posterior and prior densities of the anisotropic model at the condition, $\sigma_r = \sigma_t$. This is achieved by defining the transformed parameter $X = \sigma_t - \sigma_r$.

$$R_{\text{SD}} = \frac{P_{\text{posterior}}(\sigma_t = \sigma_r)}{P_{\text{prior}}(\sigma_t = \sigma_r)} = \frac{P_{\text{posterior}}(X = 0)}{P_{\text{prior}}(X = 0)} \tag{17}$$

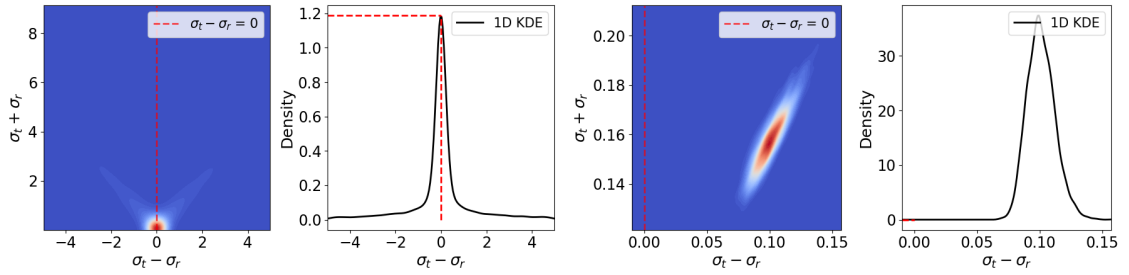
These probabilities are estimated through a one-dimensional kernel density estimation (KDE) on X, using samples drawn from the HMC posterior (as described above) and directly from the prior using `NumPyro`'s built-in sampling, overall this is automated through `savage_dickey_comparison`. This report presents the results based on 10,000 samples being drawn from each.



(a) Transformed Prior

(b) Transformed Posterior

Figure 10: Savage Dickey Sample Distributions



(a) Prior

(b) Posterior

Figure 11: Kernel Density Estimates

For the prior probability, $P_{\text{prior}}(\sigma_t = \sigma_r)$, we obtain a value of 1.18749. In contrast, the anisotropic posterior samples (Figure 12b) show minimal support for the nested isotropic condition, $X = \sigma_t - \sigma_r = 0$, and the kernel density estimate approaches zero. While not sufficient to definitively quote a Bayes factor of zero, we can confidently conclude that the anisotropic model has significantly more support from the observed data.

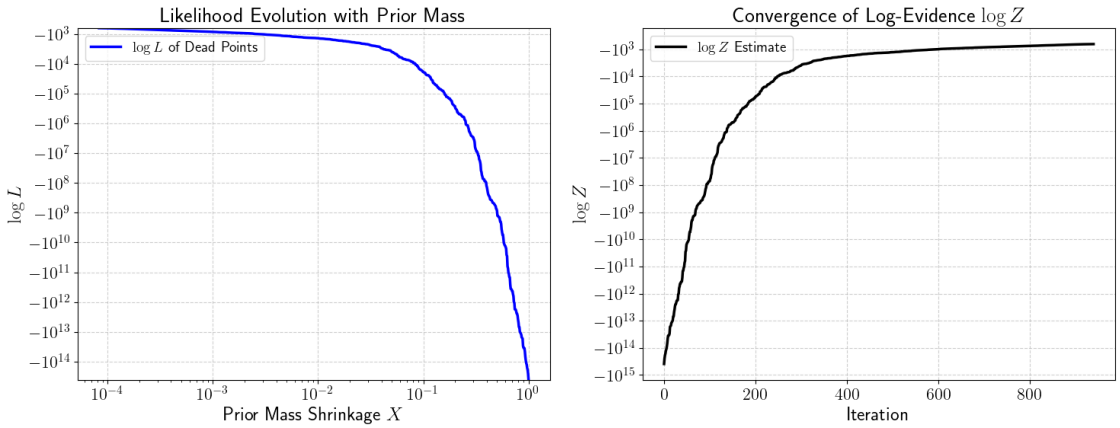
8.2 Nested Sampling

To provide an additional method for estimating bayes factor, this paper aimed to build its own custom Nested Sampling implementation (`run_nested_sampling`). This was optimised for flexibility and numerical stability, with `ns_prior_transform` supporting unit-cube reparameterisation compatible with a wealth of Numpyro's prior distributions for robust use in the future. Additionally, the algorithm exploits numpy's `logaddexp` to allow all updates and uncertainty quantification to be done in log-space, with full implementation outlined in the documentation.

$$\log Z = \log \sum_i \exp(\log L_i + \log w_i) \tag{18}$$

where L_i is the i-th dead point's likelihood and w_i, is its associated weight (change in prior mass).

In preliminary tests the isotropic model successfully converged to a Bayesian evidence estimate of $log(Z) = -641$, using 100 live points and a tolerance of 0.20. However, reliable convergence could not be achieved under the same conditions for the anisotropic model due to the increased dimensionality, with an additional Jeffreys prior, and thus no conclusive results were achieved. Example plots for the isotropic model are included below for reference.



(a) Evidence Evolution

(b) Prior Mass Shrinkage

Figure 12: Nested Sampling for Isotropic model

9 Conclusion

Through probabilistic reanalysis of the Antikythera Mechanism, this report shows strong support for the hypothesised lunar calendar, originally proposed by Budiselic et al. [3]. Across all models and filtering schemes, we are able to exclude the 365-hole (Egyptian calendar) hypothesis with at least 95% confidence. We further show visually (Figure 9) and quantitatively (section 8) that the observed data strongly favors an anisotropic model which is thus able to provide more tightly constrained parameters which can exclude the Egyptian calendar with 99% certainty. Across the three filtering schemes, the anisotropic model estimates N as $355.182^{+1.32}_{-1.39}$, $355.281^{+1.37}_{-1.37}$, and $354.090^{+1.44}_{-1.48}$ respectively, all of which are consistent with a lunar-based calendar.

9.1 Declaration of Use of Autogeneration Tools

This report made use of Large Language Models (LLMs), to assist in the development of the project. These tools have been employed to assist:

- Formatting plots to enhance presentation quality.
- Generating docstrings for the repository's documentation.
- Performing iterative changes to already defined code.
- Debugging code and identifying issues in implementation.
- Latex formatting for the report.
- Identifying spelling and punctuation inconsistencies within the report.
- Suggesting more concise phrasing to reduce the word count.

References

[1] Michael Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. 2018. arXiv: 1701.02434 [stat.ME]. URL: https://arxiv.org/abs/1701.02434.

[2] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. URL: http://github.com/jax-ml/jax.

[3] C. Budiselic et al. "Antikythera Mechanism Shows Evidence of Lunar Calendar". In: *SocArXiv* (2020). Revised February 2021. DOI: 10.31235/osf.io/fzp8u. URL: https://osf.io/preprints/socarxiv/fzp8u_v1.

[4] European Space Agency. *A History of Astrometry - Part I: Mapping the Sky from Ancient to Pre-Modern Times*. https://sci.esa.int/web/gaia/-/53196-the-oldest-sky-maps. Last updated: 1 September 2019. 2019. (Visited on 04/02/2025).

[5] Tony Freeth and Alexander Jones. "The Cosmos in the Antikythera Mechanism". In: *ISAW Papers* 4 (2012). ISAW Papers 4. URL: https://dlib.nyu.edu/awdl/isaw/isaw-papers/4/.

[6] Ravin Kumar et al. "ArviZ a unified library for exploratory analysis of Bayesian models in Python". In: *Journal of Open Source Software* 4.33 (2019), p. 1143. DOI: 10.21105/joss.01143. URL: https://doi.org/10.21105/joss.01143.

[7] Du Phan, Neeraj Pradhan, and Martin Jankowiak. "Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro". In: *arXiv preprint arXiv:1912.11554* (2019).

[8] Derek de Solla Price. "Gears from the Greeks: The Antikythera Mechanism—A Calendar Computer from ca. 80 BC". In: *Transactions of the American Philosophical Society* 64.7 (1974), pp. 1–70.

[9] Andrew Ramsey. "The latest techniques reveal the earliest technology - A new inspection of the Antikythera Mechanism". In: June 2007.

[10] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[11] Graham Woan and Joseph Bayley. *An improved calendar ring hole-count for the Antikythera mechanism*. 2024. arXiv: 2403.00040 [physics.hist-ph]. URL: https://arxiv.org/abs/2403.00040.