S2 Advanced Statistical Methods - Coursework
Antikythera mechanism

Zihan Xu (zx282)

Lent Term 2025

Word Count: 2383

1 (a)

The measured hole location of can be plotted using the csv file in the data:

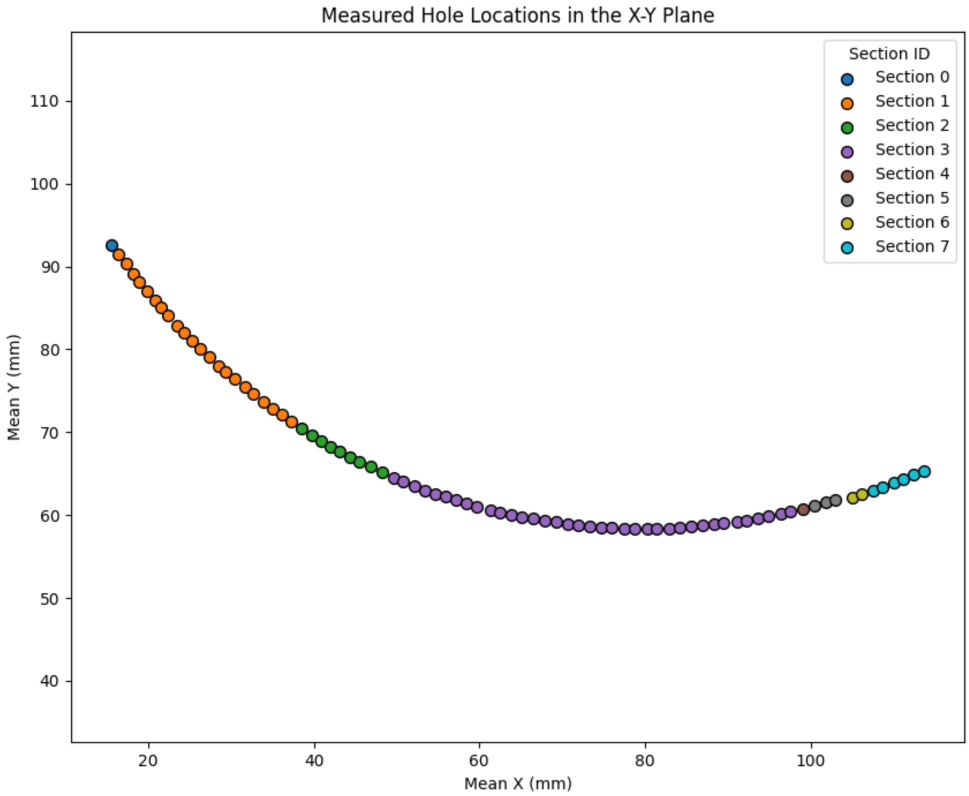

Figure 1: Measured Hole Positions

The holes are colored according to the section they belong to. The code that computes the figure is `code/visualize.py`.

2 (b)

The model for the hole location in Ref 2 first consist of angular positions specified for i^{th} hole in the j^{th} section [1]:

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

where the unknown parameters for each section are $(x_{0j}, y_{0j}, \alpha_j)$. The positions of i^{th} hole in the j^{th}, relative to the arc center is given by:

$$r_{ij} = (\cos \phi_{ij}, \sin \phi_{ij})$$

The Tangential component is:

$$t_{ij} = (\sin\phi_{ij}, \cos\phi_{ij})$$

The error between the predicted location of the model and the measured positions are assumed to be Gaussian distributed, with two types of covariance matrices, one being Isotropic and the other one being Radial/Tangential covariance matrix.

The code that implements the model described above is inside `code/antikythera_model.py`, as the `AntikytheraModel` class.

A figure with hole positions generated using the model above is plotted. The x_0, y_0 and α_0 are initialized to be:

$$x_{0j} = 80 + j \quad y_{0j} = 136 + 0.5j \quad \alpha_j = -145.0 - 0.5 * j$$

This choice of parameter is inspired by the statistics provided in reference [2].

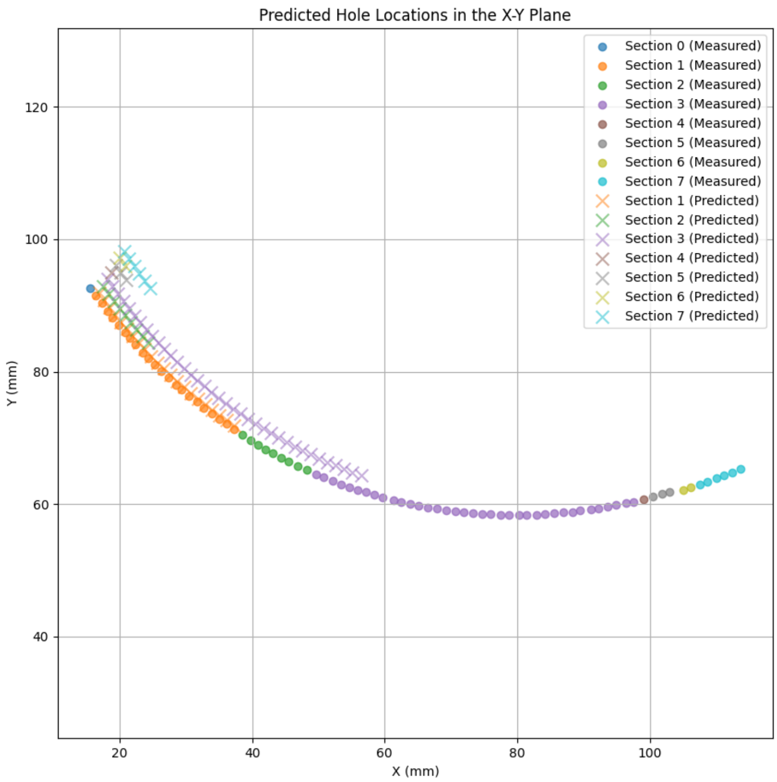

Figure 2: Predicted Hole Positions

The predicted hole locations closely follow the curvature of the measured hole locations, demonstrating strong agreement and thus validating the accuracy and reliability of the model.

3 (c)

The Gaussian likelihood function is defined as:

$$L(D|\theta) = \frac{\exp(-\frac{1}{2}[d_i - m_i]^T \Sigma^{-1} [d_i - m_i])}{2\pi\sqrt{\Sigma}}$$

Here, d_i represents the measured data and m_i denote the model prediction for hole i. Σ is the covariance matrix.

There are two options for covariance matrices:

$$\Sigma_1 = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} \sigma_r & 0 \\ 0 & \sigma_t \end{pmatrix}$$

Σ_1 represents the covariance matrix of Isotropic Gaussian distribution and Σ_2 represents that of a 2-dimensional Gaussian distribution with principle axes aligned with the Radial and Tangential directions.

Since the measured data are in x-y coordinate and the model defined previously is defined on principal axis aligned with the Radial-Tangential direction, m_i must be transformed into x-y coordinate:

$$x_{ij} = x_{0j} + r\cos\phi_{ij}$$

$$y_{ij} = y_{0j} + r\sin\phi_{ij}$$

where $m_i = (x_{ij}, y_{ij})$ if m_i belongs to section j.

Similarly, transformation of coordinate system must be performed on the covariance matrices. Let S be the change of coordinate system matrix:

$$\Sigma_1' = S\Sigma_1 S^{-1} = \Sigma_1 \qquad \Sigma_2' = S\Sigma_2 S^{-1}$$

The code implementation calculates S individually for each section. The details are in `code/antikythera_model.py`, under the `AntikytheraModel` class.

4 (d)

The gradient of the log likelihood, $\frac{\partial}{\partial\theta^\mu}\log L(D|\theta)$, is calculated using the PyTorch library, which facilitates automatic differentiation. In PyTorch, this is achieved by tracing the operations performed on tensors. As we perform operations such as addition, multiplication, or exponentiation, PyTorch automatically tracks these operations, enabling the computation of gradients with respect to the model parameters.

For my model, I ensure that it supports gradient tracing by using PyTorch's built-in functions. The `AntikytheraModel` class inherits from `nn.Module` to make it compartible with other functions in the PyTorch libeary. All relevant operations are performed using PyTorch tensors, which are capable of storing gradients. Specifically, for each parameter, I ensure that the tensor associated with it has the `requires_grad=True` flag set, indicating that PyTorch should track its gradient. The parameters are initialized as `nn.Parameter` to make optimizations happening in later questions easier.

To test the implementation, I ran several test cases where the gradients were computed for known simple functions such as multiplication and exponentiation, and verified that the results matched the analytically computed gradients. This step ensures that the automatic differentiation mechanism works correctly and the derivative is behaving as expected. The test is inside `diff_test`.

5 (e)

To compute the maximum likelihood parameters, a minus sign is applied to the log-likelihood function, transforming the problem into minimizing the negative log-likelihood. The resulting objective is the negative log-likelihood function.

To efficiently minimize this function, the Adam optimizer is employed. It adaptively adjusts the learning rate for each parameter by utilizing the first and second moment of the gradient. By iteratively minimizing the loss function, the Adam optimizer updates the parameters, eventually producing the maximum likelihood estimates.

In addition to the parameter initialization mentioned in section (b), N is initialized to be 354 and r initialized to be 77. For the Isotropic model, $\sigma = 0.1$ and for the Radial-Tangential model, $\sigma_r = 0.025$, $\sigma_t = 0.1$.

Using the initial parameters and the Adam optimizer with learning rate sent to 0.0001, an empirically effective rate, the maximum likelihood parameters found are the following:

Section	x_0 (mm)	y_0 (mm)	α (degrees)
0	79.53	136.19	-145.36
1	80.46	136.04	-145.21
2	82.51	136.49	-145.49
3	83.51	136.99	-145.99
4	84.52	137.48	-146.48
5	85.52	137.98	-146.98
6	86.52	138.48	-147.48
7	87.52	138.99	-147.98

Table 1: Isotropic Optimized Parameters

$N = 354.00$, $r = 76.49\ mm$ and $\sigma = 0.3306$

Section	x_0 (mm)	y_0 (mm)	α (degrees)
0	79.12	135.10	-145.81
1	80.05	135.56	-145.18
2	82.88	136.98	-145.01
3	83.94	138.40	-145.51
4	84.98	138.99	-146.00
5	85.98	139.49	-146.50
6	86.98	139.99	-147.00
7	87.98	140.49	-147.50

Table 2: Radial-Tangential Optimized Parameters

$N = 354.00$, $r = 76.04$ mm, $\sigma_r = 0.0250$ and $\sigma_t = 0.1000$

The values above are determined through 1000 iterations using the Adam optimizer. The value of negative log likelihood is investigated from the convergence trajectory:

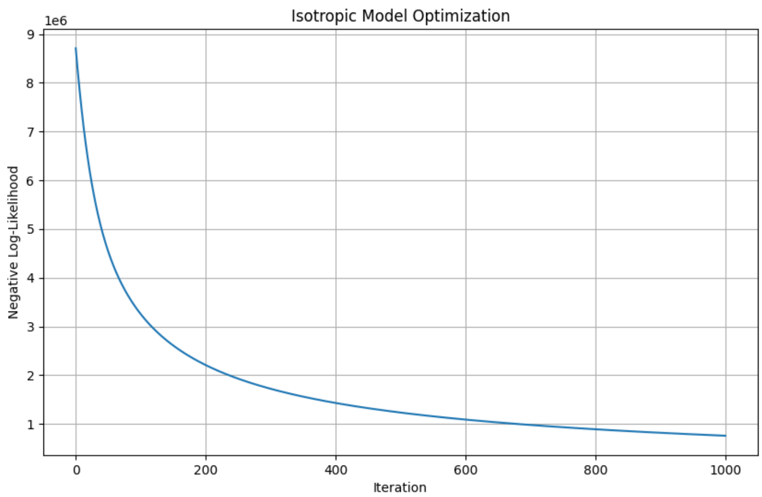

Figure 3: Isotropic Model Optimization Trajectory

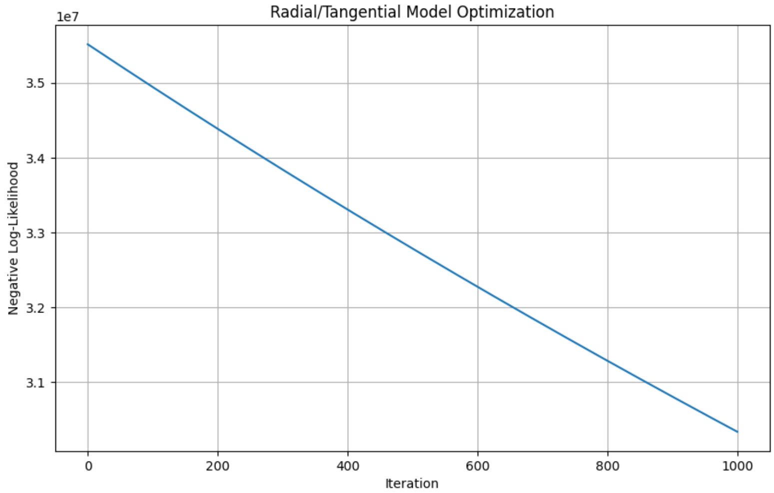

Figure 4: Radial-Tangential Model Optimization Trajectory

The Isotropic model demonstrates a significant reduction in the gradient, suggesting convergence toward an optimal solution. In contrast, the Radial-Tangential model exhibits a nearly linear trajectory, implying that additional iterations could potentially lead to a smaller likelihood value. However, the parameter values remain relatively stable even as more iterations are performed. Given this stability and considering computational efficiency, the results after 1000 iterations are presented as the final outcome.

Note that the likelihood function appears to be yielding extremely large values, which is likely a consequence of the large number of measured data points and the relatively small variances observed in both models. When the variance is small, the likelihood function tends to become very sensitive, amplifying small differences in the model's predictions and leading to larger likelihood values. The sheer volume of data points further compounds this effect, contributing to very high likelihood values as the model attempts to fit the data across many observations.

In addition, this phenomenon also influences the gradient of key parameters such as r, σ, σ_t and σ_t, causing them to exhibit significantly large gradients. Large gradients can pose challenges in optimization algorithms, as they may lead to instability in the update steps, particularly when these updates are large and potentially overshoot the optimal solution. This sensitivity in the gradient values could be another factor to why the Adam Optimizer is not performing ideally in locating the global minimum, as reflected in the convergence trajectory plot. The optimizer may struggle with the large gradients, leading to slow convergence, where the model fails to settle into a stable minimum despite multiple iterations. Even though the learning rate is set to be 0.0001, the small learning rate is still unable to counter the effect of huge gradient. Further exploration is needed for better determination of the parameters.

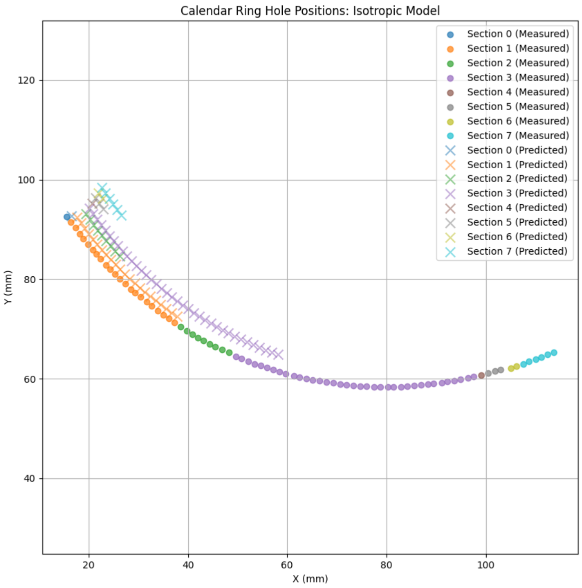

Figure 5: Isotropic Maximum Likelihood Parameter Hole Positions

The x-axis represents the horizontal positions (mm), and the y-axis represents the vertical positions (mm). Each colored point corresponds to a specific section. The Isotropic nature of this model indicates that the hole positions are evenly distributed in a more uniform manner across the calendar ring.

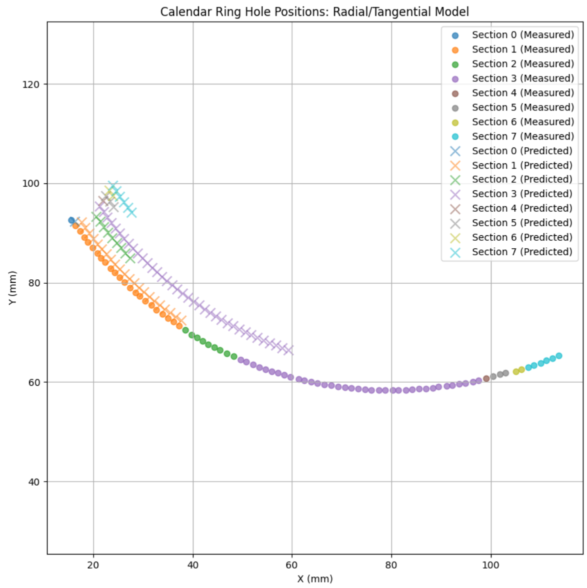

Figure 6: Radial-Tangential Maximum Likelihood Parameter Hole Positions

This plot displays the positions of the parameter holes for the Radial-Tangential model. Similar to the Isotropic plot, the x-axis and y-axis represent the horizontal and vertical positions, respectively. However, this model takes into account both Radial and Tangential components, leading to a slightly different distribution of hole positions.

Despite the differences in the models used, the hole positions from both the Isotropic and Radial-Tangential maximum likelihood models exhibit high similarities. In both cases, the parameter holes are distributed in a curved pattern with similar curvature observed across sections. The overall shape of the distributions is consistent between the two models, suggesting that despite the differing underlying frameworks of the Isotropic and Radial-Tangential models, both methods ultimately produce comparable estimates for the hole locations.

The program used to conduct the computations and generate the plots above is `code/optimize_antikythera.py`

6 (f)

A variation of the Hamiltonian Monte Carlo (HMC) method, the No-U-Turn Sampler, is used to sample the posterior distribution. This method is designed to improve sampling efficiency by automatically adjusting the trajectory length during sampling, which reduce the risk of inefficient sampling.

Empirically, this method performs at least as efficient as a standard HMC method with no additional computation cost for tuning runs [2]. The number of samples generated is chosen to be 100 with 200 steps for warm-up. The warm-up phase in NUTS is critical. During this phase, parameters such as the step size of the HMC algorithm are tuned. This phase allows the sampler to adapt to the specific geometry of the target distribution and to optimize its exploration strategy.

The prior distribution of σ, σ_r and σ_t is chosen to be half normal distribution, which is a folded version of the standard normal distribution, which only assigns positive probability to positive values of the variable. The mean is set to 0, and the standard deviation is 1. This choice of standard deviation allows the sampler to explore a broader range of values for the standard deviation, facilitating more flexible sampling. By setting the mean to 0, this accounts for the belief that, while the parameters should be positive, there is no strong prior knowledge favoring larger or smaller values within the positive domain. It allows the model to adjust and learn from the data without overly constraining the parameter space.

The first hole in the first section is chosen to be sampled:

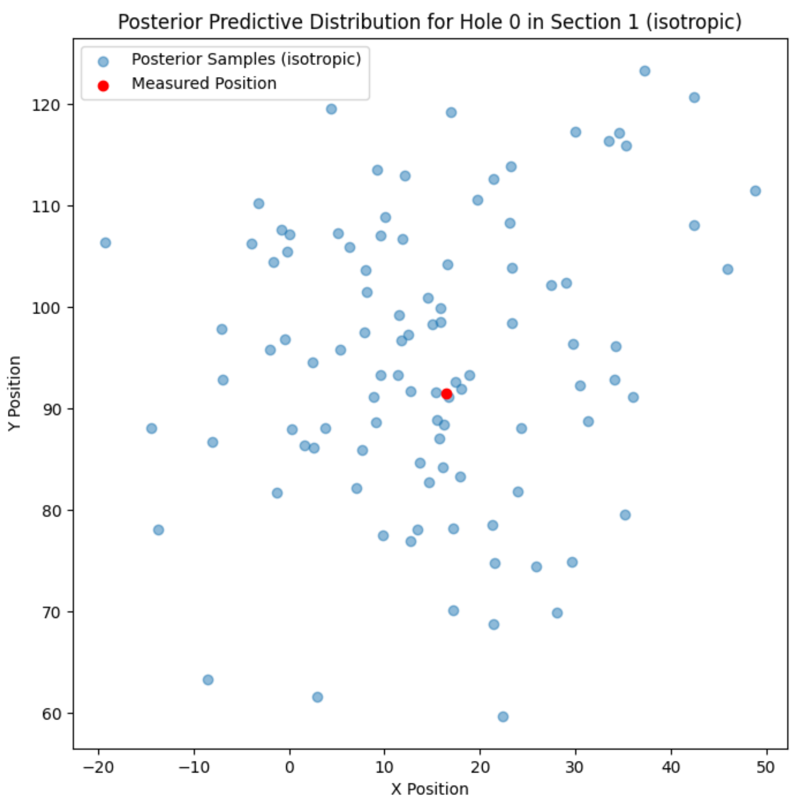

Figure 7: Isotropic Model - Posterior Predictive Distribution for Hole 0

The distribution of the posterior samples around the measured value indicates that the Isotropic model is effective in capturing the general pattern of the data. The samples are symmetrically scattered around the measured point, reflecting the assumption of the Isotropic model that variability is consistent in all directions. The relatively large spread of the points suggests that the posterior distribution assigns a high probability to values across a broad range. This is expected, as the first hole in each section can vary flexibly, provided that the subsequent holes remain reasonably aligned with each other.

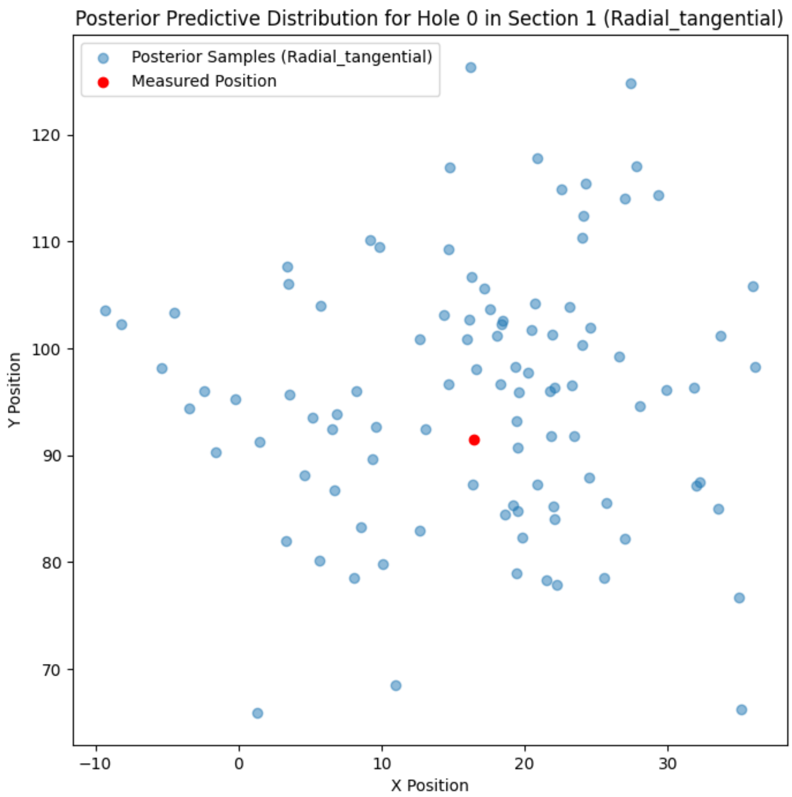

Figure 8: Radial Tangential Model - Posterior Predictive Distribution for Hole 0

The Radial Tangential model accounts for variability along specific Radial directions, and the samples are likely to follow patterns that align with this assumption. The HMC algorithm has explored the parameter space, allowing the model to capture the data's underlying structure more accurately than the Isotropic model. The samples are more compact and closely aligned with the measured position. The more compact distribution of samples suggests that the posterior distribution places higher probability on values that are close to the measured data point. This results in a posterior that better matches the observed data.

7 (g)

The covariance matrix Σ plays a crucial role in modeling the uncertainty of hole positions on the Antikythera calendar ring. It defines the relationship between errors in the x and y coordinates of the measured positions, and its structure affects the accuracy of model predictions. Two covariance models are considered: the Isotropic covariance model and the Radial/Tangential covariance model.

The Radial/Tangential model is a priori considered particularly appropriate for this problem because it aligns with the natural geometry of the calendar ring. Since the holes are arranged along a circle, the errors in their placement are more likely to follow the curvature of the ring. For example, holes placed near the outer parts of the ring may exhibit different error characteristics compared to holes closer to the center. The Isotropic model, by assuming uniform error distribution in all directions, might not capture these nuances effectively due to its simplicity.

Below is an example for the predicted position for section 2 using both models:

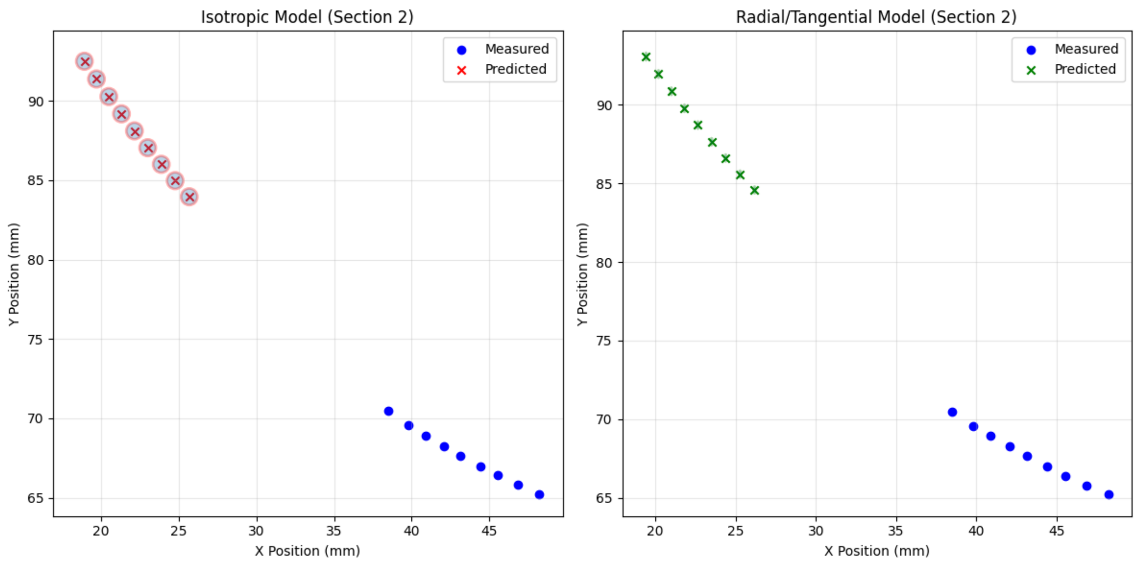

Figure 9: Covariance Model Comparison: Predicted Position for Section 2

The circles around the predicted points serve as a visual representation of the uncertainty associated with each prediction. The size of the circles corresponds to the standard deviation of the predicted positions, with each circle representing a one standard deviation away from the predicted point. It could be seen that the Radial-Tangential model has a much smaller circle, indicating a higher certainty for the predicted position.

To further explore the differences between the two models, more metrics are explored, namely, log likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC). AIC and BIC are defined as the following:

$$AIC = 2k - 2\ln(\hat{L})$$

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

Here $\hat{L}$ denotes the maximum likelihood, k represents the number of estimated parameters and n is the number of data points.

If the true model is included among the candidate models, then the Bayesian Information Criterion (BIC) is the preferred metric for selecting the best model. However, if the true model is not present within the candidate models, the Akaike Information Criterion (AIC) is more suitable for identifying the model that best approximates the true model [3]. Also, both metrics are independent of the prior, which reduces the risk of invalid comparison due to inappropriate belief on the parameters. Lower the value of AIC and BIC, the better the model is in describing the data. Below is a plot showing the relevant metrics for the two models:

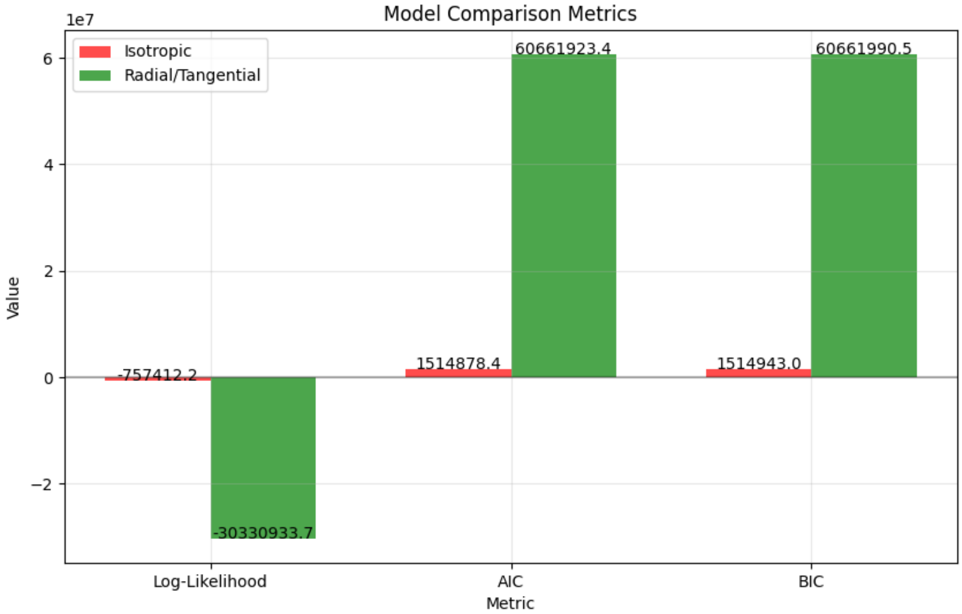

Figure 10: Metrics for Covariance Model Comparison

Despite analysis on the uncertainty previously, the metrics shows strong preference for the Isotropic model.

The Log-Likelihood for both models shows that the Isotropic model (-757412.2) has a significantly better fit compared to the Radial/Tangential model (-30330933.7). Based on Log-Likelihood alone, the Isotropic model would be favored since it yields a higher value.

For the AIC values, the Isotropic model (1514878.4) again performs better than the Radial/Tangential model (60661923.4). Since AIC penalizes complexity with a term of $2k$ (k is the number of parameters), the Isotropic model is slightly favored. However, since there is only a difference of 1 free parameter between the two models, the favor for perplexity should not greatly affect the outcome of the AIC. This result suggests that the Isotropic model offers a better balance between fit and complexity of the model compared to the Radial/Tangential assumption.

Finally, in terms of BIC, the Isotropic model (1514943.0) still outperforms the Radial/Tangential model (60661990.5). The results are consistent with those from the AIC. BIC penalizes complexity more heavily than AIC, particularly as sample size increases, making it more conservative in favoring simpler models. The fact that the Isotropic model remains clearly favored, reinforces the conclusion that the added complexity of the Radial/Tangential model is not supported by the data.

Taken together, these three metrics, log-likelihood, AIC and BIC, consistently point toward the Isotropic model as the better choice. This consistent advantage across all criteria suggests that the data do not support the added assumptions introduced by the Radial/Tangential model. While more complex models can sometimes capture nuanced behavior in the data, in this case, the additional parameter and complexity fail to offer practical improvement in fit.

8 Acknowledgment

This report was developed with the assistance from Anthropic's Claude 3.5 Sonnet. Specifically, Claude was used to improve the clarity of written content and assist with debugging Python code implementations. All AI-generated suggestions were manually reviewed.

References

[1] G. Woan & J. Bayley, "An Improved Calendar Ring Hole-Count for the Antikythera Mechanism: A Fresh Analysis", The Horological Journal, July 2024, https://arxiv.org/abs/2403.00040

[2] Devlin, L., Horridge, P., Green, P. L., & Maskell, S. (2021). The No-U-Turn Sampler as a Proposal Distribution in a Sequential Monte Carlo Sampler with a Near-Optimal L-Kernel. arXiv preprint arXiv:2108.02498.

[3] Wikipedia contributors. (n.d.). Akaike information criterion: Comparison with BIC. Wikipedia. Retrieved March 28, 2025, from https://en.wikipedia.org/wiki/Akaike_information_criterion#Comparison_with_BIC