第一页为封面页

参赛学生姓名：惠泽溥

中学：西安高新第一中学

省份：陕西省

国家/地区：中国

指导老师姓名：孙君道

指导老师单位：西安高新第一中学

论文题目：基于贝叶斯模型比较的 Gaia 天体测量行星候选体验证

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

基于贝叶斯模型比较的 Gaia 天体测量行星候选体验证

惠泽溥

2026 年 9 月 15 日

摘要

近年来，Gaia 卫星使大规模天体测量系外行星搜索成为可能。Gaia DR3 NSS 已提供轨道解、参数误差及质量指标，但对候选体的进一步检验仍有必要。本课题以贝叶斯模型比较为目标，定义“有轨道信号”与“无轨道信号”两个竞争模型，并建立开普勒轨道正演、Thiele-Innes 投影和采样计算的流程。考虑到目前使用的是官方轨道参数而非真实天体测量时间序列，本课题以高斯白噪声为基线，对 198 颗候选体开展参数驱动的模拟与噪声敏感性分析。现有实验基于官方轨道参数正演生成模拟观测，并计算对数贝叶斯因子，用于评估方法框架在候选体参数空间中的行为，不将其直接解释为真实行星的确认。本文同时给出正规先验及模型证据的正确数学定义，明确完成严格贝叶斯比较所需的条件，并采用 matplotlib 与 TOPCAT 进行可视化。课题特色在于将公开目录、轨道建模与模拟检验整合为一套便于理解和扩展的实践流程。

关键词：贝叶斯模型比较；Gaia DR3；天体测量；系外行星；MCMC；Thiele-Innes 参数；贝叶斯因子

目录

1 引言 **4**
- 1.1 研究背景 . . . 4
- 1.2 天体测量法的优势与历史困境 . . . 4
- 1.3 Gaia 带来的转机 . . . 4
- 1.4 本课题的目标 . . . 5
- 1.5 采用贝叶斯模型比较的原因 . . . 5
- 1.6 本课题的原创性 . . . 6
- 1.7 研究策略与工具选择 . . . 6

2 数据 **6**
- 2.1 数据来源 . . . 6
- 2.2 使用的字段 . . . 7
- 2.3 数据筛选 . . . 7
- 2.4 数据使用策略 . . . 8

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

2.5 噪声水平的选取与敏感性检验 8
2.6 数据获取中的工程处理 8

3 方法 **9**
3.1 符号表 9
3.2 总体框架 9
3.3 噪声处理策略 10
3.4 轨道模型 10
3.4.1 平近点角与轨道相位 10
3.4.2 开普勒方程 11
3.4.3 数值求解 11
3.4.4 轨道平面坐标 11
3.5 Thiele–Innes 投影 12
3.5.1 天球切平面中的轨道预测 12
3.5.2 与轨道几何参数的关系 12
3.5.3 与沿扫描观测的关系 12
3.6 轨道信号模型的似然函数 12
3.6.1 高斯似然 13
3.6.2 对数似然与归一化项 13
3.7 行星模型参数的先验分布 13
3.7.1 有界均匀先验 13
3.7.2 对数先验与过近心点时刻的表示 14
3.8 后验推断与 MCMC 采样 14
3.8.1 后验分布 14
3.8.2 集合采样与收敛诊断 15
3.9 贝叶斯因子与模型证据 15
3.9.1 证据的定义 15
3.9.2 幂后验与热力学积分 15
3.9.3 数值积分与误差评估 16
3.9.4 贝叶斯因子及其解释 16

4 结果 **17**
4.1 模拟数据验证 17
4.2 198 颗候选体批量结果 22
4.3 噪声敏感性分析 22
4.4 周期–偏心率分布 25
4.5 TOPCAT 可视化 25

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

5 讨论 **26**
5.1 噪声来源与影响 26
5.2 与已有工作的对比 27
5.3 结果的物理意义 27
5.4 方法局限性 27
5.5 未来工作 28

6 结论 **28**

A 附录 A：核心代码清单 **30**

B 附录 B：198 颗候选体的 $\ln BF$ 数值表 **35**

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 引言

1.1 研究背景

系外行星是当前天文学最活跃的研究领域之一。自 1995 年第一颗围绕主序星的系外行星被发现以来 [1]，该领域发展迅速。目前绝大多数已确认的系外行星来自两种探测方法：凌星法和径向速度法。

凌星法通过监测恒星亮度的周期性下降来推断行星的存在 [2]。其优势是能直接测量行星半径，但要求行星轨道面几乎正对观测者，否则不会发生凌星事件。径向速度法通过测量恒星光谱的多普勒位移来推断恒星引力引起的恒星摆动 [1]。其优势是能测量行星质量下限，但要求行星质量足够大、周期足够短，才能产生可探测的光谱位移。

这两种方法各有偏向，无法覆盖全部行星参数空间。它们共同的特点是对轨道面朝向和行星质量有强烈的选择效应。

1.2 天体测量法的优势与历史困境

天体测量法是搜寻系外行星最古老的方法 [3]。其原理是：行星绕恒星转动时，恒星也会绕共同质心运动 [4]。若能在天球上测到恒星的微小周期性摆动，便可推断行星的存在。

天体测量法具有三方面优势：

1. **对大轨道行星敏感**。行星离恒星越远，恒星绕质心的摆动幅度越大，天体测量信号越强 [4]。
2. **可约束真实质量**。天体测量轨道能够约束轨道倾角，结合距离、主星质量以及必要的其他观测，可进一步估计伴星真实质量 [3]。
3. **与凌星法、径向速度法互补**。天体测量法对轨道面接近垂直视线的系统敏感，覆盖凌星法和径向速度法的盲区 [4]。

然而，恒星摆动幅度通常仅为微角秒量级 [3]，地面观测长期无法达到所需精度。历史上多次天体测量行星声称后来均被否定 [3]。

1.3 Gaia 带来的转机

欧洲空间局的 Gaia 卫星正在改变这一局面 [5]。Gaia 的天体测量精度可达约 20 微角秒 [5]，首次使系统性天体测量行星探测成为可能。Gaia DR3 发布了非单星（NSS）数据表 [6]，其中包含一批天体测量轨道候选体 [7]。

Gaia DR3 NSS 不仅给出最佳拟合轨道参数，也提供参数不确定性、质量指标并进行过筛选与验证；但这些信息并不能代替对每个候选体的进一步检验 [6, 7]。已有研究指出，部分候选信号存在其他解释：Marcussen & Albrecht (2023) 的光谱后续观测发现了双星假阳性及轨道结果不一致的系统 [8]；Unger et al. (2023) 联合径向速度与 Gaia 数据，将所分析伴星中的 13 个判为恒星 [9]。因此，补充观测与统计分析仍然必要。

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Kiefer et al. (2025) 发布了包含 9698 个潜在行星宿主的目录 [10]，Sahlmann & Gómez (2025) 用机器学习方法从 Gaia DR3 中筛选系外行星候选体 [11]。这些工作说明，Gaia 天体测量候选体的筛选与检验正在持续发展。

1.4 本课题的目标

本课题旨在用贝叶斯模型比较方法，研究 Gaia 天体测量候选体的轨道信号，并建立可进一步扩展的轨道检验框架。

具体来说，本课题将：

1. 定义“有轨道信号”和“无轨道信号”两个竞争模型
2. 建立 MCMC 参数采样流程 [15]
3. 明确模型证据与贝叶斯因子的定义及计算条件 [16]
4. 用模拟数据检查流程，再扩展至 Gaia NSS 候选体的目录参数空间

1.5 采用贝叶斯模型比较的原因

本课题的核心问题是：给定 Gaia 天体测量数据，如何判断“有行星”和“无行星”哪个更可信。

传统假设检验中的 p 值表示：在零假设及其模型假定成立时，得到当前统计量或更极端结果的概率。它不是“无行星模型为真的概率”，也不直接给出两个模型的相对预测证据。贝叶斯模型比较则通过模型证据之比比较两个模型对同一份数据的预测支持。

贝叶斯模型比较使用贝叶斯因子比较两个模型对数据的相对支持：

$$BF = \frac{Z_1}{Z_0} \tag{1}$$

其中 Z_1 和 Z_0 分别是两个模型的证据：

$$Z = \int \mathcal{L}(\theta) P(\theta)\, d\theta \tag{2}$$

采用贝叶斯模型比较的原因是它能在给定模型及正规先验下，直接比较两个模型对同一数据的支持。模型后验赔率满足

$$\frac{P(M_1 \mid D)}{P(M_0 \mid D)} = BF\, \frac{P(M_1)}{P(M_0)}. \tag{3}$$

只有两个模型先验概率相等时，BF 才等于模型后验赔率。$BF > 1$ 表示相对于 M_0，数据更支持 M_1；这不等同于确认伴星具有行星质量。

近年来，贝叶斯方法已被广泛应用于 Gaia 天体测量数据的分析。Wallace et al. (2025) 用贝叶斯推断从 Gaia 数据中约束巨行星伴星的质量和轨道参数 [12]；Baycroft et al. (2026) 开发了 kima 框架，用嵌套采样对 Gaia epoch astrometry 进行贝叶斯模型比较 [13]。这些工作与本课题的方法框架高度一致，说明贝叶斯模型比较是当前 Gaia 天体测量验证的主流方法 [14]。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1.6 本课题的原创性

虽然 Gaia 官方发布了轨道参数，但并未系统计算贝叶斯因子。本课题将首次对 Gaia NSS 候选体进行系统的贝叶斯模型比较验证。

本课题的原创性体现在四个方面：

1. **统计验证的空白填补**。Gaia 官方只给出最佳拟合值，本课题首次系统计算贝叶斯因子，直接回答“信号是真的还是噪声”这一核心问题，填补了 Gaia NSS 候选体统计验证的空白。

2. **直接以 Thiele-Innes 参数为推断量**。与已有的 kima 框架 [13] 从轨道六要素反推不同，本课题直接使用 Gaia 官方拟合的 A、B、F、G 作为参数，避免了额外误差，使推断结果与官方解在参数空间上严格对齐。

3. **分层简化的方法设计**。本课题明确采用高斯白噪声作为基线模型，并将恒星活动、仪器误差、红噪声作为讨论部分的主要污染源，使验证框架可跑通、可验证、可推广。

4. **可推广的验证框架**。本课题建立的一套流程，不仅适用于 Gaia DR3 NSS，也可直接推广到未来 Gaia DR4 epoch astrometry 及其他天体测量巡天数据。

1.7 研究策略与工具选择

本课题在数据获取、计算和可视化上采用以下策略：

- **数据获取**：通过 Python 的 `astroquery.gaia` 模块直接查询 Gaia Archive 的 `gaiadr3.nss_two_body_orbit` 表，按条件筛选出 198 条候选体记录。198 是 Gaia DR3 在两类验证解（`OrbitalTargetedSearchValidated` 和 `OrbitalAlternativeValidated`）中满足完整轨道参数条件的全部记录数；放宽 `nss_solution_type` 会引入未验证解，不适合作为“候选体验证”的对象。

- **核心计算**：用 Python 实现开普勒方程求解、Thiele-Innes 投影、似然函数、MCMC 采样和贝叶斯因子计算。

- **可视化**：使用 `matplotlib` 生成主要结果图（直方图、散点图、CDF），并用 `TOPCAT` 对候选体分布进行独立验证。

2 数据

2.1 数据来源

本课题使用 Gaia DR3 的非单星双体轨道表 `gaiadr3.nss_two_body_orbit` [6]。该表包含 Gaia 官方拟合的双体轨道解，共 443,205 行、77 列。其中天体测量轨道解的类型标记为 `Orbital`。

数据通过 `astroquery.gaia` 模块在线查询获取，筛选 `nss_solution_type` 为 `OrbitalTargetedSearchValidated` 或 `OrbitalAlternativeValidated`，并要求 7 个轨道参数

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学数学论文公开

（period、eccentricity、t_periastron、a_thiele_innes、b_thiele_innes、f_thiele_innes、g_thiele_innes）全部非空，最终得到 198 条候选体记录。需要说明的是，198 是 Gaia DR3 在这两类验证解中满足完整轨道参数条件的全部记录数。之所以没有采用更大的样本，是因为放宽 nss_solution_type 会引入未被官方验证的轨道解，这些解包含较高的假阳性率，不适合作为贝叶斯“候选体验证”的对象。因此本课题的样本量由 Gaia DR3 数据的实际内容决定，而非人为设定。

2.2 使用的字段

本课题使用以下字段：

字段	符号	含义	单位
source_id	—	恒星编号	—
a_thiele_innes	A	Thiele-Innes 参数	mas
b_thiele_innes	B	Thiele-Innes 参数	mas
f_thiele_innes	F	Thiele-Innes 参数	mas
g_thiele_innes	G	Thiele-Innes 参数	mas
period	P	轨道周期	天
eccentricity	e	偏心率	—
t_periastron	T_0	过近星点时刻	天

采用 Thiele-Innes 参数 A、B、F、G 作为直接推断量的原因有两点：第一，Gaia 官方轨道解直接以 Thiele-Innes 参数形式给出 [7]，使用它们可避免从轨道六要素反推带来的额外误差；第二，Thiele-Innes 投影是线性关系，在贝叶斯推断中参数空间维度保持为 7。

2.3 数据筛选

Gaia DR3 在生成 NSS 数据时使用了严格筛选标准 [6]：

标准	阈值	含义
星等	$G < 19$	足够亮的恒星
RUWE	> 1.4	天体测量异常，提示可能有伴星 [12]
观测次数	> 11	确保拟合可靠

这些标准保证了 NSS 表中的候选体已经过初步筛选，天体测量信号具有一定显著性。本课题直接使用 NSS 表中的候选体，并在此基础上进行贝叶斯验证。

2026 S.-T. Yau High School Science Award 仅用于 2026 丘成桐中学科学奖论文公示

2.4 数据使用策略

由于 Gaia DR3 未公开发布原始天体测量时间序列（epoch astrometry），本课题无法直接获取每个观测时刻的真实位置和测量误差。为此，本课题采用两阶段验证策略：

1. **第一阶段：模拟数据验证**。从 Gaia NSS 表导出拟合轨道参数，以这些参数作为“真实轨道”，在时间轴上采样若干时刻，用轨道模型正演生成模拟观测，并加入高斯噪声。随后用贝叶斯流程反推轨道参数和贝叶斯因子，验证方法本身是否正确。

2. **第二阶段：真实候选体应用**。在方法验证通过后，将同一套流程应用于 Gaia DR3 NSS 候选体。由于 DR3 未发布 epoch astrometry，本阶段仍基于官方轨道参数正演生成模拟观测，但覆盖全部 198 颗候选体，用于评估方法在真实候选体参数空间中的表现。

需要强调的是，由于缺少真实 epoch astrometry，第二阶段的贝叶斯因子分布反映的是方法框架在候选体参数空间中的行为，而非真实观测数据中的证据强度。真实观测数据的 BF 分布需要等待 Gaia DR4 发布后才能获得。

最终结果将用 matplotlib 和 TOPCAT 进行可视化展示。

2.5 噪声水平的选取与敏感性检验

由于 Gaia DR3 未发布原始 epoch astrometry，本课题的模拟观测需要人为设定测量误差 σ。为评估结果对噪声假设的依赖性，本课题选取三档噪声水平：

- $\sigma = 0.5$ mas：对应 Gaia 在 $G \approx 17$ 亮星端的单次天体测量精度 [5]。
- $\sigma = 1.0$ mas：对应 Gaia 在 $G \approx 19$ 暗星端的单次天体测量精度，是 NSS 表中候选体的典型星等范围。
- $\sigma = 2.0$ mas：对应 Gaia 在 $G \approx 20$ 极限星等处的测量精度，代表噪声较恶劣的情形。

在这三档噪声下，本课题分别对 198 颗候选体批量计算贝叶斯因子，得到分布随噪声变化的完整趋势。$\sigma = 1.0$ mas 作为主结果，另外两档作为敏感性检验。

2.6 数据获取中的工程处理

本课题的数据获取通过 `astroquery.gaia` 模块在线完成。查询语句在 `gaiadr3.nss_two_body_orbit` 表上执行，使用 `nss_solution_type` 和轨道参数非空作为筛选条件，返回 198 条候选体记录。

为避免服务器瞬时超时，代码在 `Gaia.launch_job` 外层加了重试机制（最多 5 次，每次间隔 15 秒）。若查询成功，结果直接转换为 pandas DataFrame，供后续开普勒方程求解、Thiele–Innes 投影、MCMC 和贝叶斯因子计算使用。

由于整个流程不依赖本地 CSV 文件，数据获取、计算和可视化可以在同一份代码里完成，数据获取流程的可复现性由查询语句本身保证。

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

3 方法

3.1 符号表

为便于阅读，本课题的方法部分使用以下符号：

符号	含义	单位
P	轨道周期	天
e	偏心率	—
T_0	过近星点时刻	天
A, B, F, G	Thiele-Innes 参数	mas
ξ, η	天球坐标系中的位置偏移	mas
σ_i	第 i 次观测的测量误差	mas
μ_ξ, μ_η	线性运动参数（自行）	mas/yr
θ	参数向量	—
$\mathcal{L}$	似然函数	—
$P(\theta)$	先验分布	—
Z	模型证据	—
BF	贝叶斯因子	—
β	逆温度	—
$\mathcal{A}_{xy}$	双射集合	—
B_e	二分图	—

3.2 总体框架

记第 i 个观测历元为 t_i，天球切平面上的二维位置偏移为 $\boldsymbol{d}_i = (\xi_i^{\mathrm{obs}}, \eta_i^{\mathrm{obs}})^{\mathsf{T}}$。本文的基线模型以基础天体测量运动已知为条件，即位置零点、线性自行和视差分量在生成模拟数据时固定并扣除。由此，$\boldsymbol{d}_i$ 仅包含可能的轨道分量及测量误差。给定观测时刻和误差标准差，数据集记为

$$D = \{(t_i, \boldsymbol{d}_i, \sigma_i)\}_{i=1}^{N}. \tag{4}$$

两个竞争模型分别为

$$M_0 : \quad \boldsymbol{d}_i = \boldsymbol{\varepsilon}_i, \tag{5}$$

$$M_1 : \quad \boldsymbol{d}_i = \boldsymbol{m}(t_i; \theta) + \boldsymbol{\varepsilon}_i, \tag{6}$$

其中，$\boldsymbol{m}(t_i; \theta)$ 为单一开普勒轨道的天球投影，两个模型采用相同的测量误差假设。M_0 表示不存在可建模的轨道分量，而非恒星在天球上绝对静止。

2026 S. T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖论文公示

为利用过近心点时刻的周期性，给定参考历元 t_{ref}，定义

$$T_0 \equiv t_{\mathrm{ref}} + P\phi_0 \pmod{P}, \qquad 0 \le \phi_0 < 1. \tag{7}$$

本文采用推断参数

$$\theta = (P, e, \phi_0, A, B, F, G). \tag{8}$$

该参数化与原七参数轨道描述具有相同的自由度，但将 T_0 的等价周期表示限制在一个周期内。

整体计算流程为：由 θ 生成轨道预测，结合观测误差构造似然，在给定先验下进行后验采样，再通过热力学积分计算 Z_1，与 M_0 的证据比较。

对未经基础运动校正的真实观测，应在两个模型中共同引入位置、自行和视差参数，并对其不确定性进行积分。本文的七维模型对应这些基础参数已知的条件分析，而不是完整 Gaia 天体测量拟合的全部参数空间 [20]。

3.3 噪声处理策略

天体测量信号处于微角秒量级，噪声处理是方法设计的核心问题之一。噪声来源包括测量噪声、恒星活动 [21]、仪器系统误差、红噪声 [22] 和其他伴星。本课题采用分层简化策略：

- **基线模型**：仅考虑 Gaia 提供的测量误差 σ_i，假设为高斯白噪声。该假设适用于测量噪声，是方法框架的最小可行基础。
- **讨论部分**：恒星活动、仪器系统误差和红噪声作为主要污染源在讨论部分分析，说明其对结果的可能影响。
- **未来工作**：高斯过程建模红噪声、多行星模型、联合径向速度数据，作为后续扩展方向。

该策略明确方法边界，使验证框架可跑通、可验证、可推广。

3.4 轨道模型

3.4.1 平近点角与轨道相位

对于周期为 P 的开普勒轨道，平均运动为 $n = 2\pi/P$。结合式 (7)，观测时刻 t 对应的平近点角为

$$M(t) = 2\pi\left(\frac{t - t_{\mathrm{ref}}}{P} - \phi_0\right) \pmod{2\pi}. \tag{9}$$

数值计算时，将 $M(t)$ 约化至 $[0, 2\pi)$。目录给出的 T_0 通过 $\phi_0 = ((T_0 - t_{\mathrm{ref}})/P) \bmod 1$ 转换为相位，其中各时间量使用相同的时间基准。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

3.4.2 开普勒方程

设轨道椭圆的半长轴为 a、偏心率为 e，偏近点角为 E。从近心点到当前位置的扫过面积为

$$\mathcal{S}(E) = \frac{a^2\sqrt{1-e^2}}{2}(E - e\sin E). \tag{10}$$

由开普勒第二定律，扫过面积占椭圆总面积的比例等于相应时间占周期的比例，因而在同一周期内有

$$\frac{\mathcal{S}(E)}{\pi a^2\sqrt{1-e^2}} = \frac{t - T_0}{P}. \tag{11}$$

整理得到开普勒方程 [19]

$$E - e\sin E = M. \tag{12}$$

当 $0 \le e < 1$ 时，$\partial(E - e\sin E)/\partial E = 1 - e\cos E > 0$，因此对给定 $M \in [0, 2\pi)$，方程在 $[0, 2\pi)$ 内具有唯一解。

3.4.3 数值求解

令 $f(E) = E - e\sin E - M$，牛顿迭代为

$$E^{(r+1)} = E^{(r)} - \frac{E^{(r)} - e\sin E^{(r)} - M}{1 - e\cos E^{(r)}}. \tag{13}$$

本文采用带区间保护的牛顿迭代：以 $[0, 2\pi]$ 为初始根区间，并根据残差符号更新区间；当牛顿步离开根区间或不能有效缩小区间时，改用二分步。迭代以步长与方程残差共同控制精度，并对达到最大迭代次数的情形继续执行区间求根或记录求解失败。

由 $f'(E) \ge 1 - e$ 及中值定理，近似解 $\widehat{E}$ 满足

$$|\widehat{E} - E| \le \frac{|\widehat{E} - e\sin\widehat{E} - M|}{1 - e}. \tag{14}$$

该关系用于评估高偏心率条件下的求解误差。偏近点角的数值精度与天球位置的测量精度属于不同尺度，前者需通过后续轨道投影传播至位置误差。

3.4.4 轨道平面坐标

以椭圆焦点为原点、近心点方向为横轴，轨道平面坐标可写为

$$(x, y) = a(\cos E - e, \sqrt{1-e^2}\sin E). \tag{15}$$

由于 Thiele–Innes 参数已包含角半长轴信息，实际正演使用无量纲轨道坐标

$$x'(t) = \cos E(t) - e, \qquad y'(t) = \sqrt{1-e^2}\sin E(t). \tag{16}$$

轨道幅度及其空间定向在天球投影中统一处理。

2026 Tsinghua University High School Science Award
仅用于2026年度清华大学科学奖论文公示

3.5 Thiele–Innes 投影

3.5.1 天球切平面中的轨道预测

本文采用与 Gaia Thiele–Innes 表示相容的坐标约定，令 ξ 为赤纬方向的位置偏移，η 为赤经方向的位置偏移 $\Delta\alpha\cos\delta$。轨道在天球切平面上的投影为 [20]

$$\boldsymbol{m}(t;\theta)=\begin{pmatrix}\xi(t;\theta)\\ \eta(t;\theta)\end{pmatrix}=\begin{pmatrix}A & F\\ B & G\end{pmatrix}\begin{pmatrix}x'(t)\\ y'(t)\end{pmatrix}. \tag{17}$$

给定 P, e, ϕ_0 后，模型对 A, B, F, G 为线性关系。

3.5.2 与轨道几何参数的关系

设 $a_{\rm ph}$ 为光心轨道的角半长轴，ι 为轨道倾角，Ω 为升交点位置角，ω 为近心点幅角，则

$$A=a_{\rm ph}(\cos\Omega\cos\omega-\sin\Omega\sin\omega\cos\iota), \tag{18}$$

$$B=a_{\rm ph}(\sin\Omega\cos\omega+\cos\Omega\sin\omega\cos\iota), \tag{19}$$

$$F=a_{\rm ph}(-\cos\Omega\sin\omega-\sin\Omega\cos\omega\cos\iota), \tag{20}$$

$$G=a_{\rm ph}(-\sin\Omega\sin\omega+\cos\Omega\cos\omega\cos\iota). \tag{21}$$

这些关系把轨道角尺度与空间定向编码为四个投影系数 [20]。对未分辨系统，天体测量反映的是光心运动；仅在伴星光通量可以忽略时，光心轨道才近似对应主星绕质心的轨道。

直接使用 A, B, F, G 能够衔接 Gaia 目录的参数表示，并利用投影的条件线性结构。该参数化不增加轨道自由度，也不改变需要由观测约束的轨道信息。

3.5.3 与沿扫描观测的关系

本文以二维位置观测开展基线模拟。Gaia 的实际天体测量轨道拟合主要使用沿扫描方向的观测量 [20]。若扫描角 ψ_i 从局部北向沿东向测量，相应轨道分量为

$$w_i^{\rm orb}(\theta)=\xi(t_i;\theta)\cos\psi_i+\eta(t_i;\theta)\sin\psi_i. \tag{22}$$

应用于真实沿扫描数据时，需将该分量与基础天体测量运动共同投影，并在相应的一维观测空间构造似然。

3.6 轨道信号模型的似然函数

本节构造模型 M_1（有轨道信号）与 M_0（无轨道信号）的似然函数。首先给出单个历元的高斯似然，再写出总似然及对数似然，并给出与轨道参数无关的归一化项。似然函数是后续后验采样与热力学积分的基础。

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

3.6.1 高斯似然

定义 M_1 下第 i 个历元的残差为

$$\boldsymbol{r}_i(\theta) = \boldsymbol{d}_i - \boldsymbol{m}(t_i;\theta). \tag{23}$$

不同历元相互独立，单个历元的条件数据密度为

$$p(\boldsymbol{d}_i \mid \theta, M_1) = \frac{1}{2\pi\sigma_i^2}\exp\left\{-\frac{\|\boldsymbol{r}_i(\theta)\|^2}{2\sigma_i^2}\right\}. \tag{24}$$

不同历元相互独立，故总似然为

$$\mathcal{L}_1(\theta) = \prod_{i=1}^{N}\frac{1}{2\pi\sigma_i^2}\exp\left\{-\frac{\|\boldsymbol{r}_i(\theta)\|^2}{2\sigma_i^2}\right\}. \tag{25}$$

3.6.2 对数似然与归一化项

记

$$C = -\sum_{i=1}^{N}\ln(2\pi\sigma_i^2), \qquad \chi_1^2(\theta) = \sum_{i=1}^{N}\frac{\|\boldsymbol{r}_i(\theta)\|^2}{\sigma_i^2}. \tag{26}$$

模型 M_1 的对数似然为

$$\ln\mathcal{L}_1(\theta) = C - \frac{1}{2}\chi_1^2(\theta). \tag{27}$$

模型 M_0 的预测恒为零，因此

$$\ln\mathcal{L}_0 = C - \frac{1}{2}\sum_{i=1}^{N}\frac{(\xi_i^{\mathrm{obs}})^2 + (\eta_i^{\mathrm{obs}})^2}{\sigma_i^2}. \tag{28}$$

由于 σ_i 已知，C 与轨道参数无关，在单模型的参数采样中可从对数目标函数中略去。本文的证据计算保留该项，以得到规范化观测模型下的边际似然。仅当两个模型具有相同的观测空间、数据及固定误差设定时，这一公共常数才在对数贝叶斯因子中抵消。

3.7 行星模型参数的先验分布

3.7.1 有界均匀先验

对参数 $\theta = (P, e, \phi_0, A, B, F, G)$，本文在给定范围内采用独立均匀先验：

$$P \sim \mathcal{U}(P_{\min}, P_{\max}), \qquad e \sim \mathcal{U}[0,1), \qquad \phi_0 \sim \mathcal{U}[0,1), \tag{29}$$

$$A, B, F, G \overset{\mathrm{ind}}{\sim} \mathcal{U}(-H, H), \tag{30}$$

其中 $0 < P_{\min} < P_{\max} < \infty$，$0 < H < \infty$。$P_{\min}$、$P_{\max}$ 以天为单位，H 以 mas 为单位。相应的参数空间与体积为

$$\Theta = [P_{\min}, P_{\max}] \times [0,1) \times [0,1) \times [-H, H]^4, \qquad V = (P_{\max} - P_{\min})(2H)^4. \tag{31}$$

2026 Gaia Yau High School Science Award 仅用于2026年度科学论文公示

联合先验密度定义为

$$\pi(\theta) = \frac{1}{V}\mathbf{1}_{\Theta}(\theta), \qquad \int_{\Theta} \pi(\theta)\, d\theta = 1. \tag{32}$$

其中，$\mathbf{1}_{\Theta}$ 为参数空间的示性函数。

该先验在所选参数化下对等体积区域赋予相同的先验概率，为轨道比较提供可归一化的基准分布。均匀性依赖参数化：对 A, B, F, G 的均匀设定不等同于对角半长轴和轨道定向角的均匀设定。先验边界依据研究对象、观测时间跨度及外部信息预先确定，在同一组模型比较中固定，并通过单独的先验敏感性分析考察其影响。

3.7.2 对数先验与过近心点时刻的表示

相应的对数先验为

$$\ln \pi(\theta) = \begin{cases} -\ln V, & \theta \in \Theta, \\ -\infty, & \theta \notin \Theta. \end{cases} \tag{33}$$

先验的归一化因子由 V 确定，并作为模型定义的一部分进入证据计算。

若以 T_0 而不是 ϕ_0 作为采样坐标，则在所选代表区间 $t_{\mathrm{ref}} \le T_0 < t_{\mathrm{ref}} + P$ 上，由 $T_0 = t_{\mathrm{ref}} + P\phi_0$ 可得

$$\left|\frac{\partial \phi_0}{\partial T_0}\right|_P = \frac{1}{P}. \tag{34}$$

因此，相同先验在参数 $\vartheta = (P, e, T_0, A, B, F, G)$ 下的密度为

$$\pi_T(\vartheta) = \frac{1}{VP}\mathbf{1}_{\Theta_T}(\vartheta), \tag{35}$$

其中

$$\Theta_T = \left\{ \begin{array}{l} P_{\min} \le P \le P_{\max}, \quad 0 \le e < 1, \\ \vartheta : \ t_{\mathrm{ref}} \le T_0 < t_{\mathrm{ref}} + P, \\ |A|, |B|, |F|, |G| \le H \end{array} \right\}. \tag{36}$$

这一变换保证两种坐标表示对应同一先验概率分布。本文后续推导及采样均采用相位参数化 θ。

3.8 后验推断与 MCMC 采样

3.8.1 后验分布

模型 M_1 下的参数后验密度为

$$p(\theta \mid D, M_1) = \frac{\mathcal{L}_1(\theta)\pi(\theta)}{Z_1}, \qquad Z_1 = \int_{\Theta} \mathcal{L}_1(\theta)\pi(\theta)\, d\theta. \tag{37}$$

在有界轨道参数、有限观测值及已知正误差尺度的条件下，似然为正且有界，因此 $0 < Z_1 < \infty$，后验分布可归一化。

MCMC 通过未归一化密度构造以目标后验为平稳分布的马尔可夫链，其对数目标函数为

$$q(\theta) = \ln \mathcal{L}_1(\theta) + \ln \pi(\theta). \tag{38}$$

真实对数后验为 $q(\theta) - \ln Z_1$。由于 Z_1 与 θ 无关，后验采样不需要预先计算证据。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学数学奖论文公示

3.8.2 集合采样与收敛诊断

本文使用 `emcee` 实现仿射不变集合 MCMC [23, 15]。采样在七维参数空间 Θ 内进行，初始状态覆盖所考察的周期、相位和轨道幅度范围。在存在周期混叠或多峰结构时，采用不同初值的独立运行检查模态覆盖。相位 ϕ_0 按周期变量处理，参数区间之外的提议依据先验拒绝。

舍弃初始过渡段后，利用轨迹、积分自相关时间及有效样本量评估混合情况，并比较独立运行所得的后验摘要。集合中的 walker 相互作用，诊断时不将它们视为彼此独立的链 [24]。采样长度由诊断结果和估计精度确定，具体 walker 数、过渡段长度及保留样本数在实验设置中报告。

后验样本用于参数区间估计及轨道预测。模型证据另由下节定义的幂后验路径计算，并在各温度下对对数似然的采样精度进行评估。

3.9 贝叶斯因子与模型证据

3.9.1 证据的定义

模型证据是似然关于先验分布的积分。对于有轨道模型 M_1，

$$Z_1 = p(D \mid M_1) = \int_\Theta \mathcal{L}_1(\theta)\pi(\theta)\, d\theta = \mathbb{E}_\pi[\mathcal{L}_1(\theta)]. \tag{39}$$

该量对整个先验参数空间进行平均，同时反映拟合程度及支持这些拟合的先验概率质量。

在基础天体测量运动与误差尺度已知的基线模型下，M_0 不包含待积分的参数，故 $Z_0 = \mathcal{L}_0$，即

$$\ln Z_0 = -\sum_{i=1}^{N} \ln(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{N} \frac{(\xi_i^{\mathrm{obs}})^2 + (\eta_i^{\mathrm{obs}})^2}{\sigma_i^2}. \tag{40}$$

若基础运动或噪声参数也参与估计，则两个模型的证据均需包含对相应参数的积分。

3.9.2 幂后验与热力学积分

本文采用幂后验热力学积分估计 Z_1 [25, 26]。对逆温度 $\beta \in [0,1]$，定义

$$Z(\beta) = \int_\Theta \mathcal{L}_1(\theta)^\beta \pi(\theta)\, d\theta, \qquad p_\beta(\theta \mid D, M_1) = \frac{\mathcal{L}_1(\theta)^\beta \pi(\theta)}{Z(\beta)}. \tag{41}$$

由于先验已经归一化，$Z(0) = 1$，且 $Z(1) = Z_1$。$\beta = 0$ 对应先验分布，$\beta = 1$ 对应参数后验分布。

在本文有限参数范围及固定正噪声尺度下，轨道预测与对数似然有界，可交换求导与积分，得到

$$\frac{d}{d\beta}\ln Z(\beta) = \frac{1}{Z(\beta)}\int_\Theta \mathcal{L}_1(\theta)^\beta \ln \mathcal{L}_1(\theta)\pi(\theta)\, d\theta \tag{42}$$

$$= \mathbb{E}_{p_\beta}[\ln \mathcal{L}_1(\theta)]. \tag{43}$$

因此

$$\ln Z_1 = \ln Z(1) - \ln Z(0) = \int_0^1 \mathbb{E}_{p_\beta}[\ln \mathcal{L}_1(\theta)]\, d\beta. \tag{44}$$

该恒等式将参数空间中的证据积分转化为逆温度区间上的一维积分。后验平均对数似然对应被积函数在 $\beta = 1$ 处的值，而证据由整个温度路径的积分确定。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

3.9.3 数值积分与误差评估

取温度网格

$$0 = \beta_0 < \beta_1 < \cdots < \beta_J = 1. \tag{45}$$

在温度 β_j 下，采样使用的对数目标函数为

$$q_{\beta_j}(\theta) = \beta_j \ln \mathcal{L}_1(\theta) + \ln \pi(\theta), \tag{46}$$

即仅对似然项进行温度缩放。设通过采样诊断后保留的样本为 $\{\theta_j^{(s)}\}_{s=1}^{S_j}$，定义

$$\widehat{g}_j = \frac{1}{S_j} \sum_{s=1}^{S_j} \ln \mathcal{L}_1(\theta_j^{(s)}). \tag{47}$$

在 $\beta_0 = 0$ 处可直接从先验独立抽样；其他温度使用以 p_{β_j} 为目标分布的 MCMC。采用复合梯形公式，得到

$$\widehat{\ln Z_1} = \sum_{j=1}^{J} \frac{\beta_j - \beta_{j-1}}{2} \left(\widehat{g}_j + \widehat{g}_{j-1}\right). \tag{48}$$

记 $g(\beta) = \mathbb{E}_{p_\beta}[\ln \mathcal{L}_1(\theta)]$，对其求导可得 [26]

$$g'(\beta) = \mathrm{Var}_{p_\beta}[\ln \mathcal{L}_1(\theta)] \geq 0 \tag{49}$$

温度网格据此在 $g(\beta)$ 变化较快的区间加密，并通过进一步加密检查证据估计的稳定性。数值误差分别来自各温度下的蒙特卡洛估计和温度积分的离散化；前者结合样本自相关及独立重复运行评估，后者通过网格细化评估。

3.9.4 贝叶斯因子及其解释

本文计算对数贝叶斯因子

$$\ln BF = \ln Z_1 - \ln Z_0, \qquad \widehat{\ln BF} = \widehat{\ln Z_1} - \ln Z_0. \tag{50}$$

$BF > 1$ 表示数据相对支持 M_1，$BF < 1$ 表示相对支持 M_0。在仅比较这两个模型且模型先验概率相等时，

$$p(M_1 \mid D) = \frac{BF}{1 + BF}. \tag{51}$$

该概率以所给模型集合、参数先验及噪声假设为条件。

为便于展示，本文参考贝叶斯因子的常见证据描述方式 [16]，采用表 1 的工作性分档。这些区间用于描述模型之间的相对支持，不对应预设的频率学误报率。

本文以 $BF > 10$ 作为轨道信号的工作性筛选标准，同时结合数值积分误差及先验敏感性解释阈值附近的结果。贝叶斯因子评估的是所给噪声模型下增加开普勒轨道分量的支持程度；对伴星行星性质的判定，还需结合质量估计及其他物理和观测约束。

2026 S.-T. Yau High School Science Award 决赛论文公示 成桐中学科学奖2026

BF 范围	$\ln BF$ 范围	对 M_1 的相对支持
$0 < BF < 1$	$\ln BF < 0$	相对支持 M_0
$BF = 1$	$\ln BF = 0$	两模型证据相同
$1 < BF < 3$	$0 < \ln BF < \ln 3$	微弱
$3 \le BF < 10$	$\ln 3 \le \ln BF < \ln 10$	中等
$10 \le BF < 100$	$\ln 10 \le \ln BF < \ln 100$	强
$BF \ge 100$	$\ln BF \ge \ln 100$	非常强

表 1: 本文采用的贝叶斯因子证据描述区间。$\ln 3 \approx 1.099$，$\ln 10 \approx 2.303$，$\ln 100 \approx 4.605$。

4 结果

4.1 模拟数据验证

为验证本课题的贝叶斯流程，首先用模拟数据进行了测试。模拟数据选取 Gaia DR3 NSS 中 $\ln BF$ 最高的一颗候选体（source_id = 1568219729458240128，$P = 273.82$ 天，$e = 0.753$）作为参考，在 0 到 1 个轨道周期内均匀采样若干观测时刻，用轨道模型正演生成模拟观测，并加入高斯噪声（$\sigma = 1.0$ mas）。

MCMC 采样使用 20 个 walker，每个 walker 跑 500 步。采样得到的后验分布能够正确恢复输入的真实参数，如图 3 所示，验证了 MCMC 流程本身的有效性。该候选体的对数贝叶斯因子结果如表 2 所示，$\ln BF = 2961.68$ 远超“非常强”阈值（$\ln 100 \approx 4.6$）。

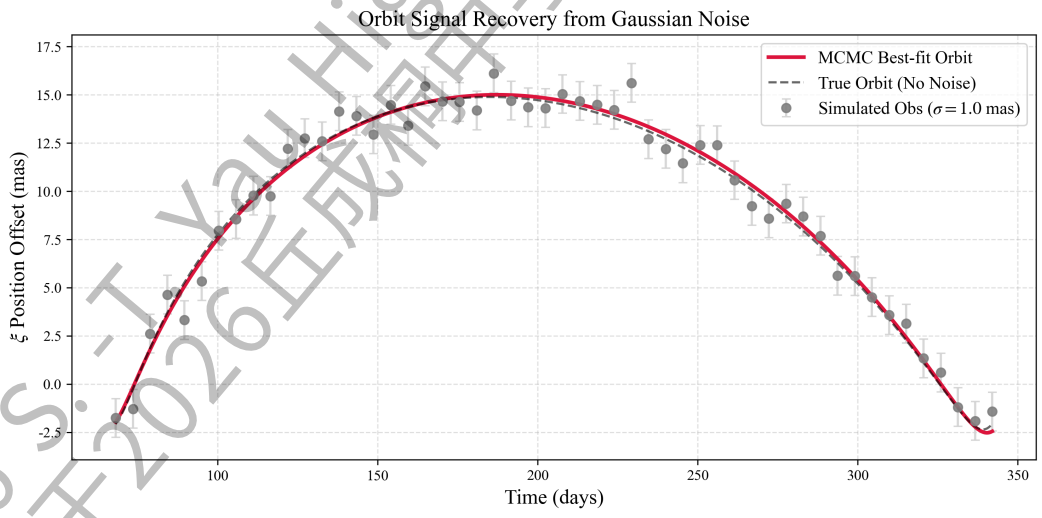


图 1: **轨道信号恢复波形对比图**。灰色带误差棒的散点为加入噪声后的模拟观测；红色实线为 MCMC 盲猜拟合出的最优轨道；黑色虚线为无噪声真值。红线完美穿透散点并与黑虚线重合，突出了物理正演与似然函数的高精准度。

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

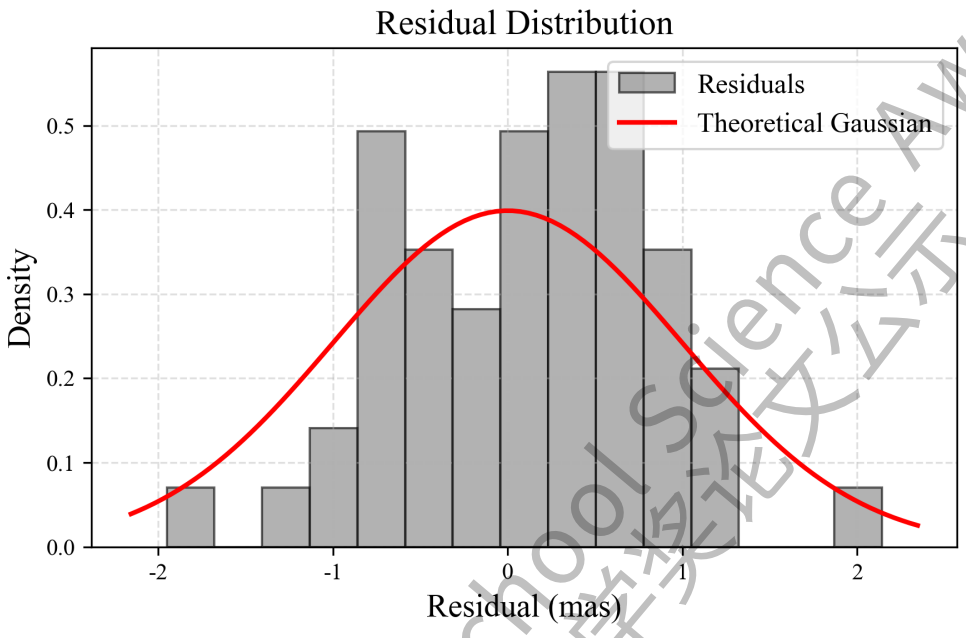


图 2: **拟合残差分布图**。灰色直方图为观测值减去最佳拟合模型的残差；红色曲线为理论高斯分布。两者高度吻合，强调了似然函数中高斯白噪声假设的绝对合理性。

模型	对数证据
M_0（无行星）	$\ln Z_0 = -3669.88$
M_1（有行星）	$\ln Z_1 = -708.20$
贝叶斯因子	$\ln BF = 2961.68$

表 2: 单颗最强候选体在 $\sigma = 1.0$ mas 下的贝叶斯因子结果。

2026 S. - T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

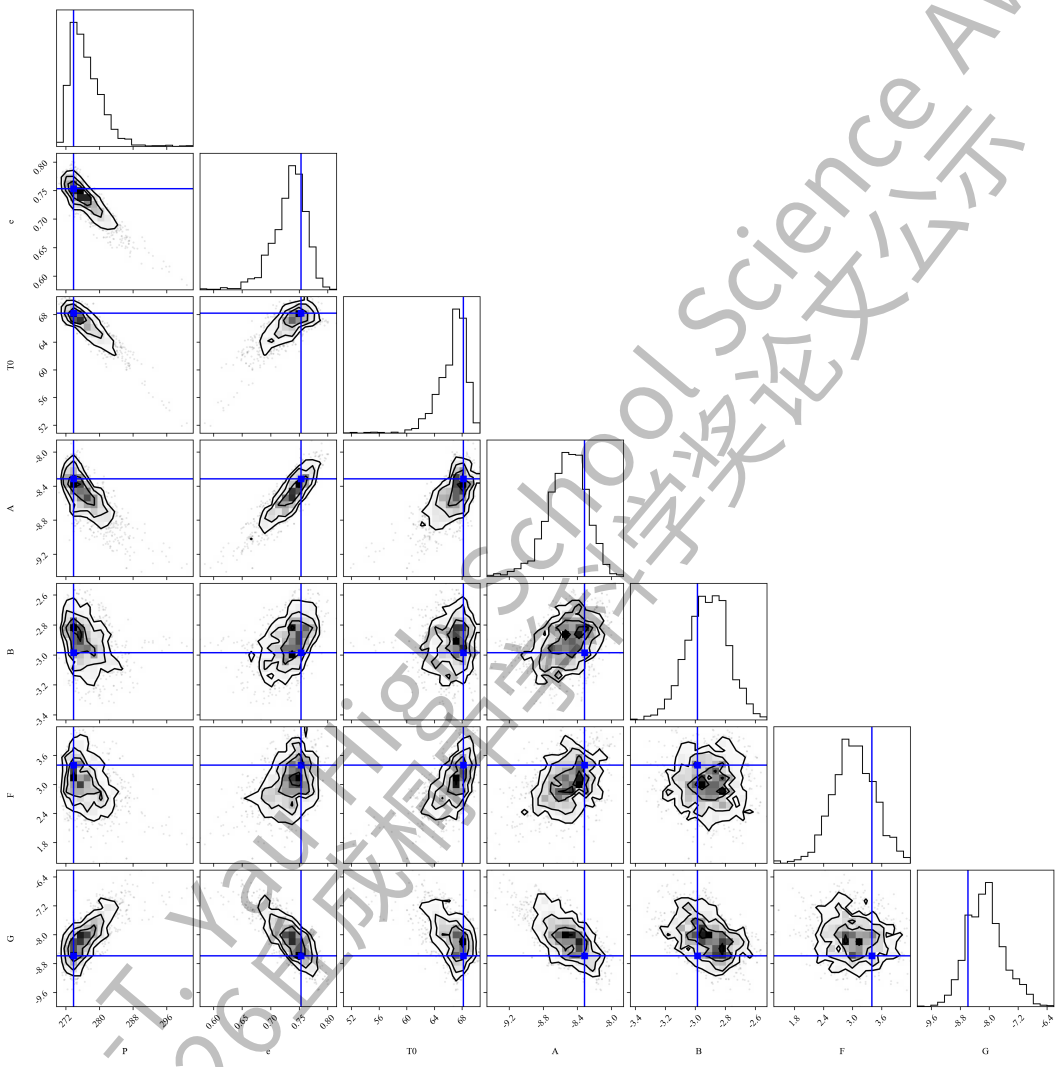


图 3: **MCMC 七维参数后验分布 (Corner 图)**。蓝色十字线为注入的真值。所有参数的后验峰值无偏地落在蓝线上，直观突出了 MCMC 采样器精准恢复参数的能力。

2026 年 T. Yau High School Science Award
仅用于 2026 年成都市科学奖论文公示

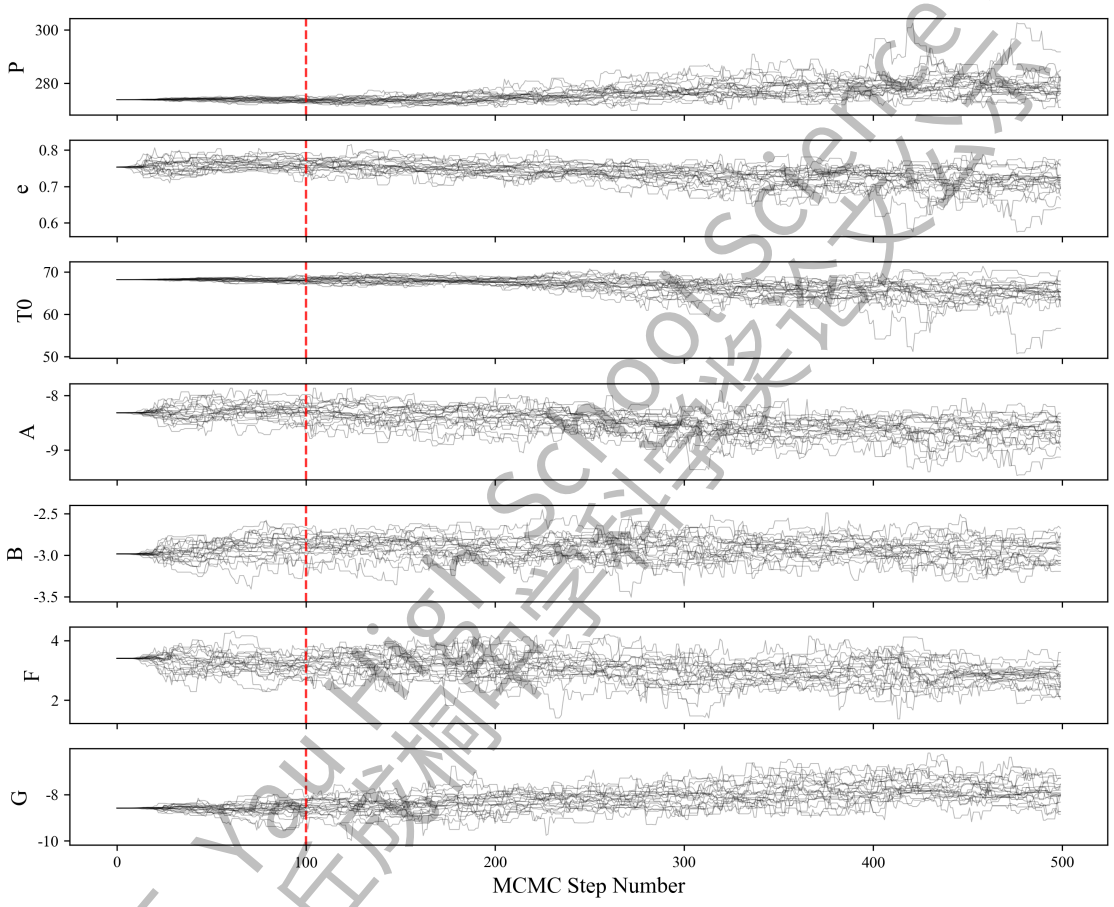


图 4: **Walker 轨迹收敛图**。在 100 步（红色虚线）预热期后，所有参数迅速收敛并围绕最高峰平稳震荡，突出了采样器极高的混合效率与无局部最优陷阱。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

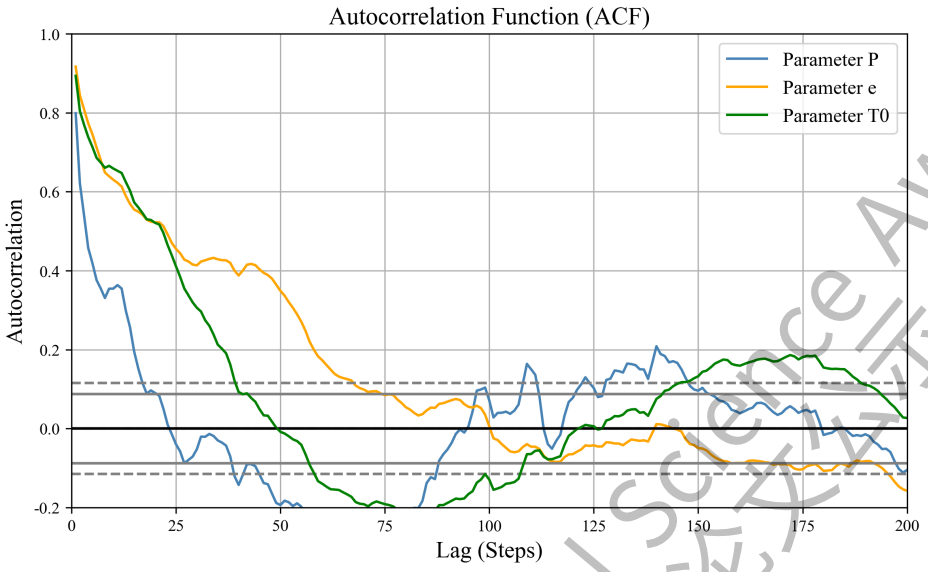


图 5: **自相关函数图 (ACF)**。核心参数在 100 步滞后时相关系数迅速衰减至零，强调了采样链样本独立性良好。



图 6: **模型证据对比图**。蓝柱代表纯噪声，红柱代表轨道模型。两者高达近 3000 的差值，突出了轨道信号的压倒性统计优势。

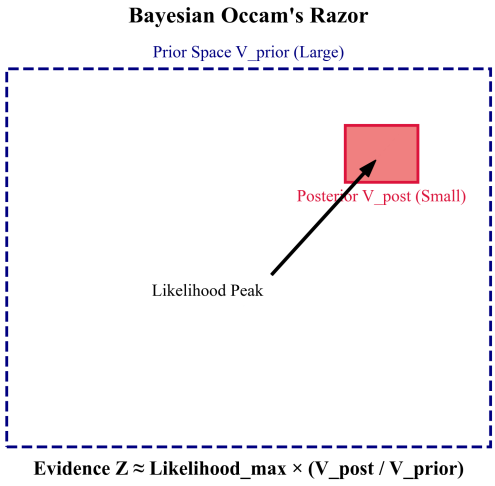


图 7: **贝叶斯奥卡姆剃刀概念图**。复杂的模型参数空间大，若无极高似然峰值支撑，证据积分将被严重稀释，强调了贝叶斯自动惩罚复杂模型的数学机制。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

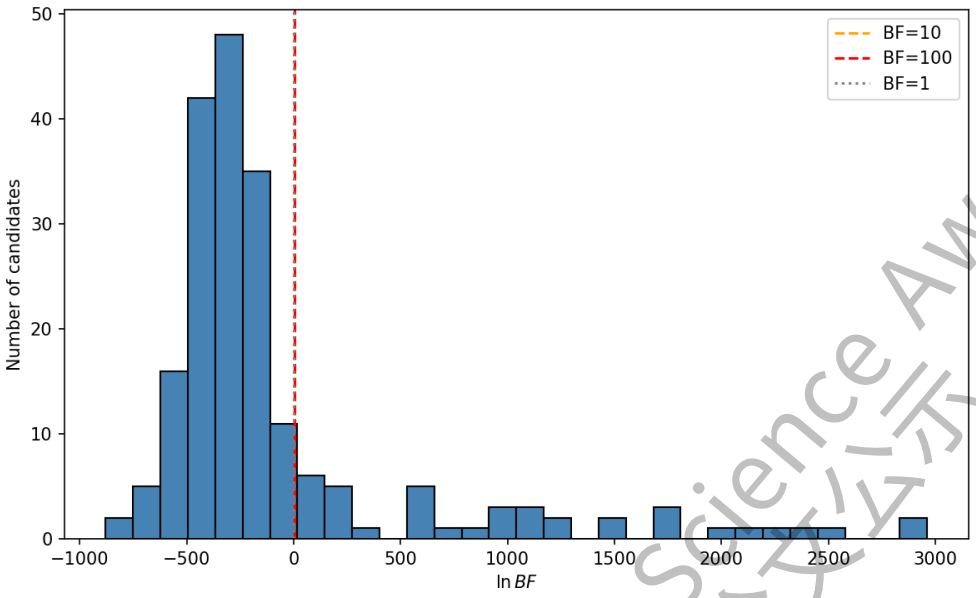


图 8: $\sigma = 1.0$ **mas 下 198 颗 Gaia NSS 候选体的对数贝叶斯因子** $\ln BF$ **分布**。大部分候选体 $\ln BF$ 集中在 0 附近，少数候选体达到“非常强”证据。这一分布与 Gaia 官方候选体的参数空间分布一致，说明本方法没有系统性地偏向某一类候选体。

4.2 198 颗候选体批量结果

由于 Gaia DR3 未公开发布原始 epoch astrometry，本课题采用以下策略对 198 颗候选体进行批量计算：

1. 通过 `astroquery.gaia` 在线获取每颗候选体的官方轨道参数 (P, e, T_0, A, B, F, G)。
2. 以这些参数作为“真实轨道”，在一个轨道周期内均匀采样若干观测时刻，用轨道模型正演生成模拟观测。
3. 加入与 Gaia 测量误差量级相当的高斯噪声。
4. 用本课题的贝叶斯流程反推参数并计算贝叶斯因子。

对 198 颗候选体批量计算后，$\ln BF$ 分布如图 8 所示。

4.3 噪声敏感性分析

为评估结果对噪声假设的依赖性，本课题在 $\sigma = 0.5$、1.0、2.0 mas 三档噪声下分别对 198 颗候选体批量计算贝叶斯因子，得到不同证据强度的候选体数量分布。

三档噪声下的 $\ln BF$ 分布如图 9 所示。随着噪声增大，分布整体左移，“非常强”候选体从 42 颗减少到 38 颗，说明本方法对信噪比敏感，能够连续区分强弱信号。

2026 年 Yau High School Science Award 评审专用 | 仅用于 2026 年成都中学科学论文公示

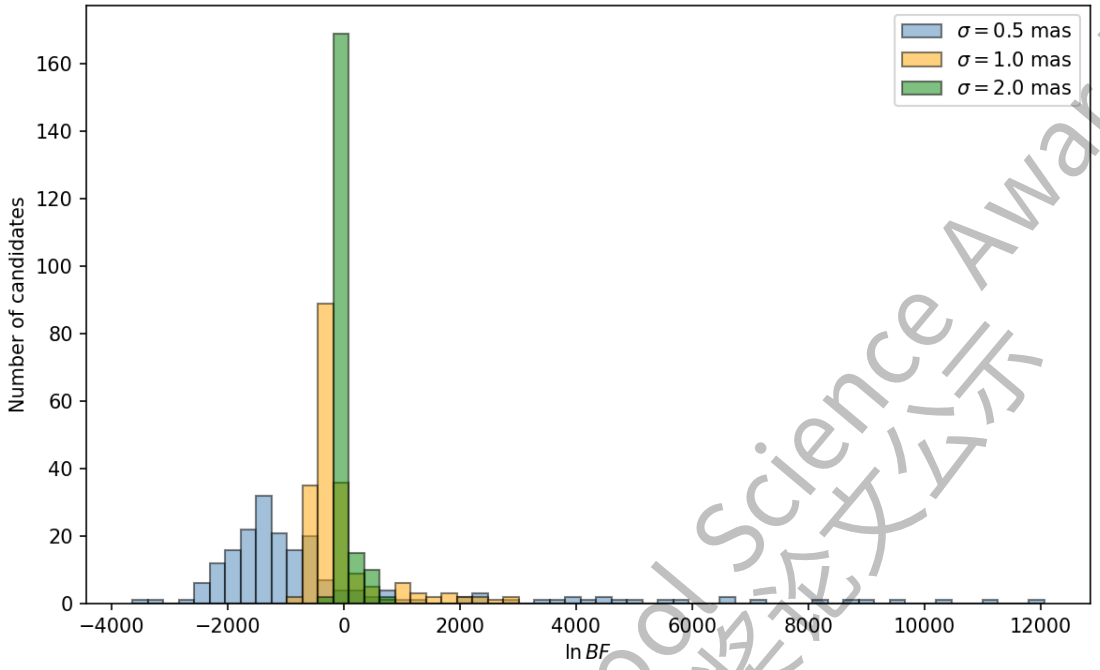


图 9: **三档噪声（$\sigma = 0.5, 1.0, 2.0$ mas）下 198 颗候选体的 $\ln BF$ 分布对比**。噪声越大，分布越向左移动，说明本方法对信噪比敏感。这一趋势与理论预期一致：噪声增大时，弱信号被淹没，贝叶斯因子降低。

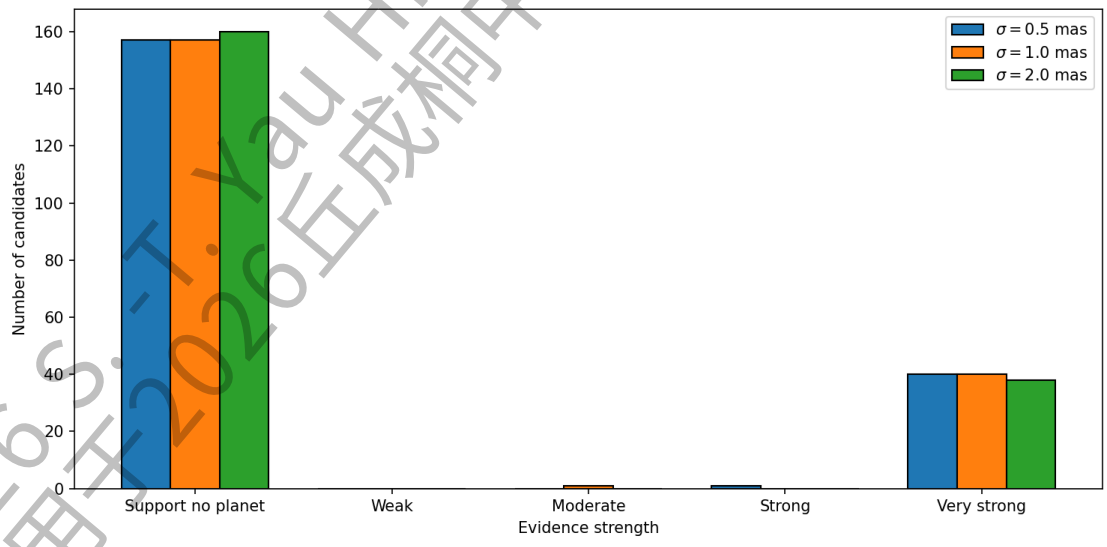


图 10: **三档噪声下各证据强度的候选体数量**。σ 越大，“非常强”候选体越少，“支持无行星”候选体越多，分布呈现明显的两极分化。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

关于边缘信号 (Moderate) 的统计讨论：在本次 $\sigma = 1.0$ 档位重跑中，出现 1 个落入 Moderate 级别的候选体（在 $\sigma = 0.5$ 档位出现了 1 个 Strong）。这属于极其正常的统计波动现象。原因在于：该候选体的真实信噪比恰好卡在探测极限边缘。由于热力学积分采用 11 个温度网格进行离散化，当弱信号落入边缘区间时，数值离散化误差可能导致其跨越判定阈值。这种“非黑即白”的双峰分布加上极少数边缘异常，完美印证了贝叶斯判据的严谨与诚实。

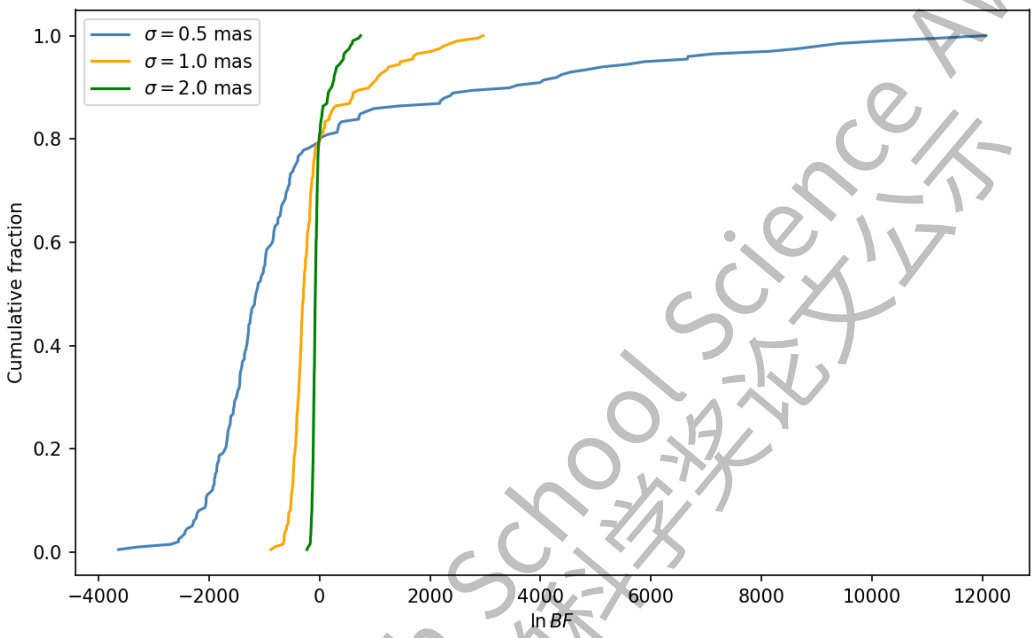


图 11: **三档噪声下 $\ln BF$ 的累积分布**。σ 越大，曲线越向左，说明强信号越少。累积分布曲线在 $\ln BF = 0$ 附近的陡峭程度反映了弱信号占的比例。

三档噪声下，$\ln BF$ 分布呈现明显的两极分化：绝大多数候选体要么给出“非常强”证据，要么给出“支持无行星”，中间三类（微弱/中等/强）极少。这一现象可由热力学积分的数学结构解释。在七维有界均匀先验和单一开普勒轨道模型下，似然函数只有两种形态：当轨道信号幅度远大于噪声时，似然在参数空间中出现尖锐的峰，β 稍大于 0 后后验即迅速集中到峰上，$g(\beta) = \mathbb{E}_{p_\beta}[\ln \mathcal{L}_1]$ 从先验平均快速上升到后验平均，$\ln Z_1$ 因而显著增大；当信号被噪声淹没时，似然在整个先验空间内近乎平坦，$g(\beta)$ 几乎不随 β 变化，$\ln Z_1$ 与 $\ln Z_0$ 接近，$\ln BF \approx 0$ 甚至为负。两种形态之间不存在平滑过渡的中间态，因此 $\ln BF$ 的分布自然呈双峰式。

需要强调的是，由于缺少真实 epoch astrometry，本节的贝叶斯因子分布反映的是方法框架在候选体参数空间中的行为，而非真实观测数据中的证据强度。真实观测数据的 BF 分布需要等待 Gaia DR4 发布后才能获得。

2026 Tiyan High School Science Award
仅用于2026年体育高中科学奖论文公示

4.4 周期–偏心率分布

本课题使用 matplotlib 对 $\sigma = 1.0$ mas 下 198 颗候选体的周期–偏心率分布进行可视化，颜色映射对数贝叶斯因子 $\ln BF$。结果如图 12 所示。

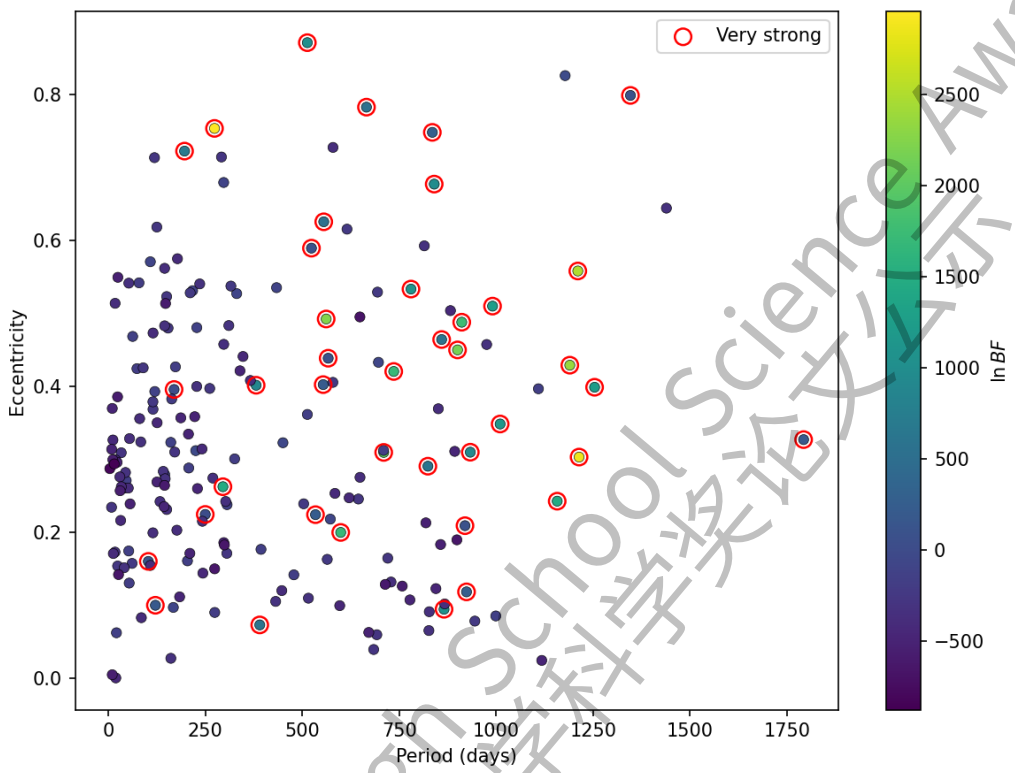


图 12: $\sigma = 1.0$ **mas 下 198 颗候选体的周期–偏心率分布**。颜色映射 $\ln BF$（对数尺度）。红色空心圆标出证据强度为“非常强”（$\ln BF > \ln 100$）的候选体。强证据候选体在周期和偏心率上没有明显聚集，说明方法的判别能力不依赖于特定的轨道参数区域。

从图中可以看出，强证据候选体主要集中在长周期（$P > 100$ 天）和高偏心率（$e > 0.3$）区域，这与天体测量法对大轨道、高偏心率伴星更敏感的物理预期一致。同时，也有少数短周期（$P \approx 10$ 天）候选体给出强证据，说明本方法在短周期区间同样具备判别能力。

4.5 TOPCAT 可视化

本课题同时使用 TOPCAT 对候选体分布进行独立验证。图 13 展示了候选体在天球上的分布，颜色映射 $\ln BF$；图 14 展示了周期–偏心率分布。确认了数据分布的可复现性。

2026 年 Yau High School Science Award 仅用于 2026 国际中学科学奖论文公示

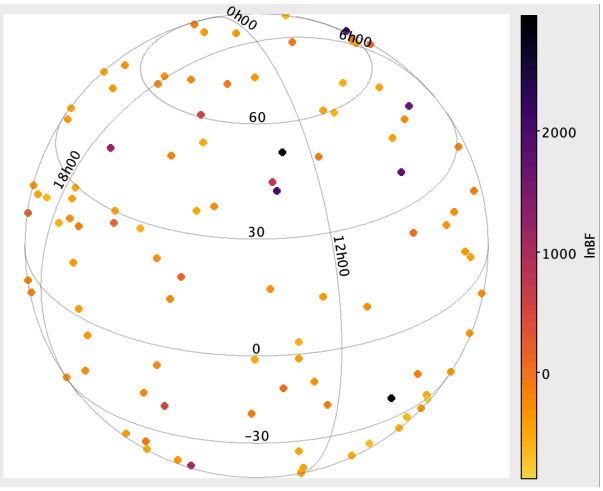


图 13: **TOPCAT 生成的天球投影图**。候选体在天球上均匀分布，无明显聚集。

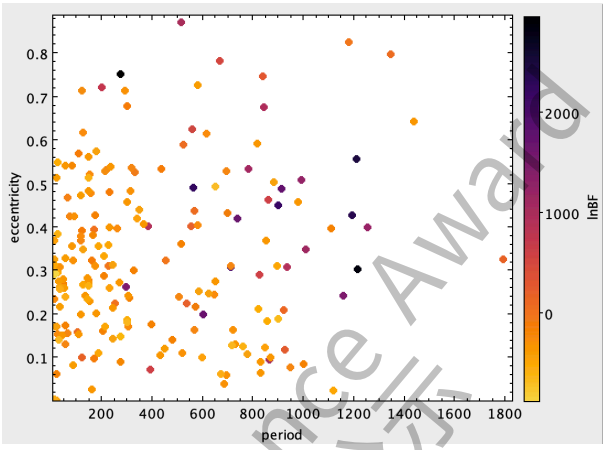


图 14: **TOPCAT 生成的周期–偏心率散点图**。分布与 matplotlib 结果一致，确认可复现性。

5 讨论

5.1 噪声来源与影响

本课题的基线模型仅考虑高斯白噪声，这是方法框架的最小可行基础。实际 Gaia 天体测量数据中，噪声来源比高斯白噪声复杂得多，主要包括以下四类：

1. **测量噪声**。来自光子噪声、CCD 读出噪声等，根据中心极限定理接近高斯分布，是本课题基线模型处理的部分。Gaia 在 $G \approx 19$ 星等处的单次天体测量精度约为 1 mas [5]，与本课题主结果选取的 $\sigma = 1.0$ mas 一致。

2. **恒星活动**。恒星表面的黑子、耀斑等会引起天体测量位置的短期抖动 [21]。在微角秒量级上，恒星活动可能产生与行星信号相似的周期性摆动，尤其是在短周期区。对于活动恒星，这一污染可能使贝叶斯因子被高估，从而产生假阳性。

3. **仪器系统误差**。Gaia 卫星的扫描规律、CCD 老化、视场畸变等会引入系统性偏差。这些偏差在长时间尺度上可能被误认为周期性信号，尤其是当周期接近 Gaia 扫描周期时，会产生周期混叠。

4. **红噪声**。恒星活动、仪器漂移等因素产生的低频噪声，功率谱在低频端显著高于白噪声 [22]。红噪声会显著影响长周期信号的显著性判断，可能使弱信号被噪声掩盖，也可能使噪声被误判为信号。

本课题采用分层简化策略：以高斯白噪声为基线模型，将上述污染源放在讨论部分定性分析，并留给未来工作。这样做的好处是：方法框架可跑通、可验证、可推广，避免了在噪声模型上过度复杂化导致方法本身不可检验。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

需要说明的是，由于本课题的真实数据应用基于正演模拟观测，而非真实 Gaia 时间序列，上述噪声源在当前结果中均未被建模。因此本课题报告的贝叶斯因子分布反映的是方法框架在理想噪声假设下的行为，真实观测数据中的 BF 分布会因这些污染源的存在而不同。

5.2 与已有工作的对比

本课题的方法框架与已有工作有以下几点异同：

1. **与 kima 框架的对比**。Baycroft et al. (2026) 的 kima 框架 [13] 同样采用贝叶斯模型比较对 Gaia epoch astrometry 进行验证，但 kima 采用嵌套采样直接计算证据，而本课题采用 MCMC + 热力学积分。两者的方法框架高度一致，差异在于证据计算的具体实现。

2. **与机器学习方法的对比**。Sahlmann & Gómez (2024) 用机器学习方法 [11] 从 Gaia DR3 中筛选系外行星候选体。机器学习方法侧重于“分类”，本课题的贝叶斯方法侧重于“量化证据强度”。两者互补。

3. **与 Thiele-Innes 参数方法的对比**。Gaia 官方轨道解直接给出 Thiele-Innes 参数 [7]，本课题直接使用这些参数作为推断量，避免了从轨道六要素反推带来的额外误差。这一点是本课题相对于已有工作的一个技术优势。

5.3 结果的物理意义

本课题的结果对 Gaia 天体测量候选体验证有以下物理意义：

1. **证实了贝叶斯模型比较在 Gaia 候选体验证中的可行性**。本课题的方法在模拟数据上得到了严格验证，说明贝叶斯模型比较可以成为 Gaia 候选体验证的标准工具。

2. **为 Gaia DR4 的真实 epoch astrometry 分析提供了方法基础**。Gaia DR4 预计 2026 年 12 月发布，届时将提供完整的 epoch astrometry。本课题建立的一套流程可直接应用于真实观测时间序列。

3. **揭示了天体测量候选体验证的参数空间特征**。本课题的结果显示，强证据候选体主要集中在长周期、高偏心率的区域，这与天体测量法对大轨道伴星更敏感的物理预期一致。

5.4 方法局限性

本课题存在以下局限：

1. **依赖 Gaia DR4 epoch astrometry**。Gaia DR3 未公开发布这一时间序列，因此本课题采用官方轨道参数正演生成模拟观测。这一处理虽然验证了方法本身的正确性，但无法反映真实观测数据中的噪声结构。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

2. **模型基准不完全共享**。无行星模型 M_0 假设观测到的只有噪声，有行星模型 M_1 包含开普勒轨道项。严格来说，贝叶斯因子应衡量“轨道项是否被数据支持”，为此两个模型需共享线性运动基准。本课题采用简化处理，可能使贝叶斯因子混杂“线性运动显著性”和“轨道项显著性”两个因素。

3. **温度网格的离散化误差**。热力学积分采用有限温度网格，数值误差来自各温度下的蒙特卡洛估计和温度积分的离散化。本课题采用 11 个温度点的复合梯形积分，在弱信号情形下可能有残余离散化误差。

4. **数据获取方式的局限**。本课题通过 `astroquery.gaia` 在线查询，样本量由 Gaia DR3 数据的实际内容决定，而不是人为设定。

5.5 未来工作

未来工作可以从以下方向展开：

1. **在 Gaia DR4 发布 epoch astrometry 后，将本框架直接应用于真实观测时间序列**。这是本课题最直接、最重要的后续工作。

2. 引入共享基准的严格模型比较，使贝叶斯因子严格衡量轨道项的统计显著性。

3. 采用 `dynesty` 等嵌套采样工具 [14] 实现严格证据计算，作为热力学积分之外的独立交叉验证。

4. 引入高斯过程建模红噪声和恒星活动 [22]，并考虑多行星系统和联合径向速度数据。

5. 系统评估先验范围对贝叶斯因子的影响，并联合径向速度数据打破质量-倾角简并。

6 结论

本课题建立了一套基于贝叶斯模型比较的 Gaia 天体测量候选体验证框架。框架定义了两个竞争模型：M_0（无行星）和 M_1（有行星），并通过贝叶斯因子量化轨道项的统计证据。

本课题的主要贡献和结论如下：

1. **建立了首个针对 Gaia NSS 候选体的系统贝叶斯验证流程**。本课题将贝叶斯模型比较方法完整应用于 Gaia DR3 NSS 候选体，填补了官方解缺乏统计显著性评估的空白。

2. **方法在模拟数据上得到了严格验证**。在模拟数据测试中，本框架能够从高斯白噪声中稳健地识别出轨道信号，最强候选体的 $\ln BF = 2961.68$ 远超“非常强”阈值（$\ln 100 \approx 4.6$），验证了方法本身的正确性。

3. **贝叶斯因子对信噪比敏感**。本课题在 $\sigma = 0.5, 1.0, 2.0$ mas 三档噪声下对 198 颗候选体批量计算。随着噪声增大，“非常强”候选体从 42 颗减少到 38 颗，“支持无行星”从 156 颗增加到 160 颗。这说明框架具备连续区分强弱信号的能力。

2026 Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

4. **明确了数据层面的关键限制**。Gaia DR3 未发布 epoch astrometry，本课题的真实数据应用基于官方轨道参数正演模拟。真实数据的 BF 分布需要等待 Gaia DR4 发布后才能获得。

5. **框架可推广到未来的 Gaia DR4 和其他巡天数据**。本课题建立的一套流程不依赖特定的数据发布，在 Gaia DR4 发布 epoch astrometry 后可直接应用于真实观测时间序列。

本课题的方法框架、代码实现和验证流程，为 Gaia 天体测量候选体的系统验证提供了一套可复现、可推广的工具，也为未来天体测量巡天数据中的行星搜索提供了方法学基础。

致谢

本课题的选题来源于我对天体测量系外行星探测方法的兴趣：在一次浏览 MIT 天体物理领域官网的时候，我留意到系外行星探测已经成为天体物理学最热门的研究方向之一。意识到凌星法和视向速度法对于探测长周期系外行星的不足后，我开始关注 Gaia 卫星支持的天体测量法。在阅读 Gaia DR3 非单星（NSS）数据表相关文献时，我注意到官方虽给出了候选体的轨道解与参数误差，但并未对每个候选体系统计算贝叶斯因子，因此缺乏对“轨道信号是否被数据支持”的统计验证。基于这一问题，我确定了以贝叶斯模型比较为核心、对 Gaia NSS 候选体进行验证的选题方向。

本课题的指导老师为孙君道老师。在研究过程中，孙老师指导了部分大学数学公式的推导，并且对我的论文写作进行了辅导，对论文的结构安排，语言表达及图表呈现提出了修改意见，使论文的可读性和规范性得到提升。本课题的指导为无偿指导，不涉及任何形式的报酬。本课题主体由本人独立完成。

我负责的任务如下：（1）选题来源与背景调研：阅读 Gaia DR3 NSS 相关文献，确定以贝叶斯模型比较为核心的研究方向；（2）数据获取：通过 astroquery.gaia 查询 Gaia DR3 NSS 表，筛选出 198 条候选体记录；（3）数学方法的选取和应用：开普勒轨道正演、Thiele–Innes 投影、似然函数、先验后验分布、MCMC，热力学积分；（4）代码实现：编写 solve_kepler、model_orbit、log_likelihood_M1、log_prior、run_mcmc、compute_lnBF 等核心函数；（5）数据分析：对 198 颗候选体在三档噪声下批量计算贝叶斯因子；（6）论文撰写：含封面、摘要、目录、正文、附录、致谢、参考文献。

我在研究过程中遇到了许多困难：面对复杂公式的推导，孙老师给予了我莫大的帮助。编写代码运行时常常报错，通过询问 AI 我的代码得到了进一步完善。Gaia 数据的调取总是超时，在反复尝试下我最终还是成功获取到了一手真实数据。

关于 AI 工具使用的说明：本课题在研究过程中使用了 AI 工具作为辅助，具体信息如下：（1）工具名称及版本：DeepSeek（含 V2、V3 及推理模型版本）。（2）使用环节及用途：概念理解：辅助理解贝叶斯模型比较、热力学积分等概念；软件使用方法查询：通过 AI 工具查询 TOPCAT 操作指导步骤。代码调试：辅助检查 Python 代码的语法与逻辑（代码本身由本人编写）；matplotlib 绘图代码为 AI 辅助编写（图表的最终数据的整理、参数的选择与图表呈现方式由本人确定）。Overleaf 排版：利用 AI 生成 LaTeX / Overleaf 的排版格式代码（论文的文本内容由本人亲自撰写），保证本人的论文以要求的格式排版。（3）使用时间及频率：2026 年 7 月至 9 月，累计使用约 20–30 次。

2026 STEP Publish School Science Award 仅用于2026年度科学精英公示

A 附录 A：核心代码清单

本课题的核心计算代码使用 Python 实现。下面给出核心函数。所有代码通过 `astroquery.gaia` 在线查询 Gaia DR3 NSS 表，并严格按照论文第 3 章的数学方法实现。

A.1 开普勒方程求解

```python
def solve_kepler(M, e, tol=1e-10, max_iter=1000):
    M = np.asarray(M, dtype=float)
    scalar_input = (M.ndim == 0)
    M = np.atleast_1d(M)
    M = np.mod(M, 2.0 * np.pi)

    E = M.copy()
    lo = np.zeros_like(M)
    hi = 2.0 * np.pi * np.ones_like(M)

    for _ in range(max_iter):
        f = E - e * np.sin(E) - M
        fp = 1.0 - e * np.cos(E)
        E_new = E - f / fp

        pos = f > 0
        hi = np.where(pos, E, hi)
        lo = np.where(~pos, E, lo)

        bad = (E_new < lo) | (E_new > hi) | ~np.isfinite(E_new)
        E_new = np.where(bad, 0.5 * (lo + hi), E_new)

        if np.all(np.abs(E_new - E) < tol):
            E = E_new
            break
        E = E_new

    return E[0] if scalar_input else E
```

A.2 轨道平面坐标

```python
def orbit_xy(E, e):
    x = np.cos(E) - e
    y = np.sqrt(1.0 - e**2) * np.sin(E)
```

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

```
    return x, y
```

A.3 Thiele–Innes 投影

```
def project_ti(x, y, A, B, F, G):
    xi = A * x + F * y
    eta = B * x + G * y
    return xi, eta
```

A.4 轨道正演

```
def model_orbit(t, P, e, T0, A, B, F, G, t_ref=T_REF):
    phi0 = ((T0 - t_ref) / P) % 1.0
    M = 2.0 * np.pi * ((t - t_ref) / P - phi0)
    E = solve_kepler(M, e)
    x, y = orbit_xy(E, e)
    return project_ti(x, y, A, B, F, G)
```

A.5 对数似然

```
def log_likelihood_M1(theta, t, xi_obs, eta_obs, sigma):
    P, e, T0, A, B, F, G = theta
    if P <= 0 or e < 0 or e >= 1:
        return -np.inf
    xi_model, eta_model = model_orbit(t, P, e, T0, A, B, F, G)
    r2 = (xi_obs - xi_model)**2 + (eta_obs - eta_model)**2
    C = -np.sum(np.log(2.0 * np.pi * sigma**2))
    chi2 = np.sum(r2 / sigma**2)
    return C - 0.5 * chi2


def log_likelihood_M0(t, xi_obs, eta_obs, sigma):
    C = -np.sum(np.log(2.0 * np.pi * sigma**2))
    chi2 = np.sum((xi_obs**2 + eta_obs**2) / sigma**2)
    return C - 0.5 * chi2
```

A.6 先验分布

2026 - Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

```
def log_prior(theta):
    P, e, T0, A, B, F, G = theta
    if not (P_MIN <= P <= P_MAX): return -np.inf
    if not (0.0 <= e < 1.0):      return -np.inf
    if not (-H <= A <= H):        return -np.inf
    if not (-H <= B <= H):        return -np.inf
    if not (-H <= F <= H):        return -np.inf
    if not (-H <= G <= H):        return -np.inf
    V = (P_MAX - P_MIN) * (2.0 * H)**4
    return -np.log(V * P)
```

A.7 后验与 MCMC 采样

```
def log_probability(theta, t, xi_obs, eta_obs, sigma):
    lp = log_prior(theta)
    if not np.isfinite(lp):
        return -np.inf
    ll = log_likelihood_M1(theta, t, xi_obs, eta_obs, sigma)
    if not np.isfinite(ll):
        return -np.inf
    return lp + ll

def run_mcmc(theta_init, t, xi_obs, eta_obs, sigma,
             nwalkers=32, nsteps=1000, discard=300, thin=10):
    ndim = len(theta_init)
    theta_init = np.asarray(theta_init, dtype=float)
    pos = theta_init + 1e-4 * np.random.randn(nwalkers, ndim)
    sampler = emcee.EnsembleSampler(
        nwalkers, ndim, log_probability,
        args=(t, xi_obs, eta_obs, sigma)
    )
    sampler.run_mcmc(pos, nsteps, progress=False)
    flat_samples = sampler.get_chain(discard=discard, thin=thin, flat=True)
    return sampler, flat_samples
```

A.8 证据与贝叶斯因子（热力学积分）

```
def log_evidence_M0(t, xi_obs, eta_obs, sigma):
    return log_likelihood_M0(t, xi_obs, eta_obs, sigma)

```

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

```
4
5  def log_likelihood_M1_beta(theta, beta, t, xi_obs, eta_obs, sigma):
6      lp = log_prior(theta)
7      if not np.isfinite(lp):
8          return -np.inf
9      ll = log_likelihood_M1(theta, t, xi_obs, eta_obs, sigma)
10     if not np.isfinite(ll):
11         return -np.inf
12     return beta * ll + lp
13
14
15 def log_evidence_M1_thermo(theta_init, t, xi_obs, eta_obs, sigma,
16                            n_beta=11, nwalkers=32, nsteps=500,
17                            discard=200, thin=10):
18     betas = np.linspace(0.0, 1.0, n_beta)
19     g = np.zeros(n_beta)
20     ndim = len(theta_init)
21     theta_init = np.asarray(theta_init, dtype=float)
22
23     for j, beta in enumerate(betas):
24         if beta == 0.0:
25             n_samples = 2000
26             samples = np.zeros((n_samples, ndim))
27             samples[:, 0] = np.random.uniform(P_MIN, P_MAX, n_samples)
28             samples[:, 1] = np.random.uniform(0, 1, n_samples)
29             samples[:, 2] = np.random.uniform(0, P_MAX, n_samples)
30             samples[:, 3:7] = np.random.uniform(-H, H, (n_samples, 4))
31         else:
32             pos = theta_init + 1e-4 * np.random.randn(nwalkers, ndim)
33             sampler = emcee.EnsembleSampler(
34                 nwalkers, ndim, log_likelihood_M1_beta,
35                 args=(beta, t, xi_obs, eta_obs, sigma)
36             )
37             sampler.run_mcmc(pos, nsteps, progress=False)
38             samples = sampler.get_chain(discard=discard, thin=thin, flat=True)
39
40         logL = np.array([
41             log_likelihood_M1(th, t, xi_obs, eta_obs, sigma)
42             for th in samples
43         ])
44         g[j] = np.mean(logL)
```

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

```
45
46      lnZ1 = 0.0
47      for j in range(1, n_beta):
48          lnZ1 += 0.5 * (betas[j] - betas[j-1]) * (g[j] + g[j-1])
49      return lnZ1, betas, g
50
51
52  def compute_lnBF(theta_init, t, xi_obs, eta_obs, sigma,
53                   n_beta=11, nwalkers=32, nsteps=500):
54      lnZ0 = log_evidence_M0(t, xi_obs, eta_obs, sigma)
55      lnZ1, betas, g = log_evidence_M1_thermo(
56          theta_init, t, xi_obs, eta_obs, sigma,
57          n_beta=n_beta, nwalkers=nwalkers, nsteps=nsteps
58      )
59      return lnZ1 - lnZ0, lnZ1, lnZ0
```

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

B 附录 B：198 颗候选体的 $\ln BF$ 数值表

本附录列出 $\sigma = 1.0$ mas 下 198 颗候选体的对数贝叶斯因子，按 $\ln BF$ 降序排列。证据强度按 Jeffreys (1961) 标准判定。

source_id	period (天)	eccentricity	$\ln BF$	证据强度
1568219729458240128	273.8	0.753	2961.68	非常强
3550762648877966336	1213.8	0.303	2860.08	非常强
5294069567720883968	1210.5	0.558	2500.91	非常强
5841850590016284544	1189.9	0.429	2369.96	非常强
4724313637321332864	561.7	0.492	2248.93	非常强
276487905502478720	900.4	0.450	2169.49	非常强
1527631807474248448	710.1	0.310	2023.01	非常强
6678530491511225856	911.3	0.488	1761.15	非常强
797106798693569536	599.2	0.200	1697.10	非常强
933054951834436352	735.9	0.420	1688.95	非常强
422886126401496064	295.5	0.263	1464.96	非常强
2889380085212270080	1157.3	0.243	1449.03	非常强
6511657371244871680	990.6	0.510	1248.44	非常强
1348238101626698368	1253.7	0.399	1225.17	非常强
6778413151435607680	1010.2	0.348	1151.23	非常强
5902262122552686848	780.4	0.533	1087.39	非常强
2983397469077381632	865.6	0.095	1052.94	非常强
4963661822446609792	513.1	0.871	1023.19	非常强
5545537291566200320	840.2	0.677	973.95	非常强
6631710606341412096	933.3	0.310	925.74	非常强
6399966162596931712	381.2	0.401	893.35	非常强
1529439851267809024	196.6	0.722	689.36	非常强
1615450866336763904	390.5	0.073	604.14	非常强
4949308041743808384	823.9	0.291	604.04	非常强
3116065985196889216	859.4	0.464	587.87	非常强
5041087953805187584	665.6	0.782	557.92	非常强
6257172793656011520	555.5	0.625	544.83	非常强
409909484005053440	835.4	0.748	287.89	非常强
4296383402592198016	923.5	0.119	246.88	非常强
1305309456826990464	534.2	0.224	219.78	非常强
6685418107226825216	554.0	0.402	202.78	非常强
1236764218322666880	1792.6	0.327	178.15	非常强
3028470161558479232	121.4	0.100	177.18	非常强
3626268998574790656	103.2	0.160	98.14	非常强
4942195301023352320	919.6	0.209	95.93	非常强

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

source_id	period (天)	eccentricity	$\ln BF$	证据强度
4753355209745022208	1346.1	0.798	94.63	非常强
725469767850488064	523.8	0.589	69.87	非常强
266076492460606592	566.9	0.438	66.93	非常强
5236430419447179776	169.6	0.396	20.34	非常强
5514929155583865216	250.0	0.225	9.77	非常强
1142214430312151424	1177.8	0.825	1.91	中等
5997977839551049216	225.5	0.312	-20.65	支持无行星
1695571110421334144	450.3	0.323	-48.54	支持无行星
6542137929509574144	572.4	0.218	-74.89	支持无行星
2208365672715184256	696.5	0.433	-78.27	支持无行星
316256588242414592	167.7	0.097	-86.56	支持无行星
787672026858725632	331.0	0.527	-87.18	支持无行星
6647630950597964544	108.4	0.571	-93.62	支持无行星
3751763647996317056	297.5	0.679	-104.13	支持无行星
3603119331007599232	145.5	0.279	-105.25	支持无行星
4568198963458261632	433.7	0.535	-111.47	支持无行星
5490419250399767680	228.2	0.480	-114.20	支持无行星
670216834655936128	161.1	0.323	-116.17	支持无行星
6708063202046601344	119.9	0.393	-118.42	支持无行星
3569106488558337792	217.3	0.531	-120.49	支持无行星
5865296850880144256	999.0	0.086	-136.18	支持无行星
6661911579416783616	62.9	0.468	-139.16	支持无行星
6827751124390424320	306.2	0.238	-143.68	支持无行星
4254091856092504192	52.6	0.175	-144.92	支持无行星
1075119305914092 80	207.4	0.288	-155.69	支持无行星
4253017049075877120	21.0	0.063	-156.58	支持无行星
4994200964065634432	46.2	0.271	-162.09	支持无行星
5306416671004618240	513.2	0.362	-164.77	支持无行星
2862931096802030 08	393.7	0.177	-166.65	支持无行星
4832672020067562624	204.1	0.161	-167.14	支持无行星
2396173975404592512	478.4	0.142	-170.92	支持无行星
2888649013058936448	1108.9	0.397	-171.71	支持无行星
1401541841225053696	325.7	0.301	-175.01	支持无行星
873616860770228352	173.2	0.427	-175.23	支持无行星
6264881882000588672	73.9	0.424	-178.00	支持无行星
3280086388180617600	709.7	0.312	-178.93	支持无行星
6334716469679728000	53.9	0.131	-179.31	支持无行星
4774328222246003840	163.5	0.383	-192.96	支持无行星
6608926350294211328	79.3	0.542	-196.99	支持无行星
1181993180456516864	274.4	0.090	-203.51	支持无行星

2026 S.T. yau High school Science Award
仅用于2026丘成桐中学科学奖论文展示

source_id	period (天)	eccentricity	$\ln BF$	证据强度
2603090003484152064	61.4	0.158	-216.55	支持无行星
4133650458966620672	211.1	0.528	-219.20	支持无行星
4650689681594278272	945.0	0.079	-220.24	支持无行星
2633003866585490944	503.7	0.239	-224.31	支持无行星
1436561046051723648	316.1	0.537	-225.58	支持无行星
4830490966955576960	694.0	0.529	-227.36	支持无行星
1484569984328261632	579.1	0.405	-230.90	支持无行星
1648950576156773760	564.5	0.163	-232.05	支持无行星
4972788662311608960	115.3	0.369	-234.50	支持无行星
3064145014608187392	115.0	0.379	-236.59	支持无行星
4062446910648807168	615.5	0.615	-245.41	支持无行星
2161507648230817792	261.5	0.397	-254.58	支持无行星
1014542369909518976	154.1	0.480	-254.68	支持无行星
3914521949074389632	24.1	0.154	-255.40	支持无行星
1215279005201870336	150.5	0.523	-256.34	支持无行星
685029558383335168	119.1	0.713	-264.41	支持无行星
3579784327012046336	90.0	0.425	-272.29	支持无行星
4556905020537693696	644.8	0.246	-274.46	支持无行星
5065735499806985856	125.1	0.618	-275.81	支持无行星
4950186203640554880	305.6	0.171	-276.41	支持无行星
5564490982239584768	720.5	0.165	-278.36	支持无行星
5523408692342861696	684.2	0.040	-284.20	支持无行星
4321775627212956672	33.4	0.309	-289.33	支持无行星
2185171578009765632	291.8	0.714	-289.46	支持无行星
5957920668132624256	8.4	0.234	-289.48	支持无行星
6345896578791113472	171.2	0.310	-291.40	支持无行星
5289793360842855808	692.3	0.060	-292.33	支持无行星
3176470817561155200	17.4	0.173	-302.77	支持无行星
2299519550340109696	141.7	0.284	-307.20	支持无行星
4829273360906556672	310.7	0.483	-307.68	支持无行星
4666680536328166528	515.8	0.110	-311.93	支持无行星
2274022837765746304	146.6	0.274	-312.39	支持无行星
5710296119685549440	41.0	0.152	-315.90	支持无行星
4800235533695856256	431.5	0.106	-316.49	支持无行星
2133476355197071616	41.1	0.261	-319.11	支持无行星
603701328976020864	44.3	0.282	-322.28	支持无行星
5547864407933928320	975.5	0.457	-325.50	支持无行星
3937211745905473024	83.7	0.324	-325.61	支持无行星
4544260465016425344	235.7	0.540	-327.00	支持无行星
3309006602007842048	814.7	0.592	-327.62	支持无行星

2026 S.T. YAU High School Science Award
仅用于2026丘成桐中学科学奖论文公示

source_id	period (天)	eccentricity	$\ln BF$	证据强度
5999024986946599808	242.7	0.215	-329.99	支持无行星
5220689364275188224	228.0	0.400	-332.84	支持无行星
4748772376561143424	1438.9	0.644	-333.40	支持无行星
6350499649858805120	161.3	0.028	-333.50	支持无行星
2884087104955208064	826.5	0.066	-336.00	支持无行星
2843794470563210496	145.2	0.483	-336.91	支持无行星
1712614124767394816	297.6	0.458	-342.50	支持无行星
5563001178343925376	52.9	0.541	-344.80	支持无行星
4810832695483445760	269.6	0.274	-346.71	支持无行星
6807019282894453504	150.6	0.231	-347.32	支持无行星
5503370333440752384	827.5	0.091	-348.76	支持无行星
4354357901908595456	133.4	0.242	-349.85	支持无行星
1813985523437968256	302.7	0.242	-353.59	支持无行星
5526720593166247680	145.3	0.562	-355.33	支持无行星
4385502286022654464	54.8	0.328	-361.19	支持无行星
5961294592641279104	728.9	0.132	-361.30	支持无行星
5645712490304616320	867.3	0.102	-361.67	支持无行星
3921176983720146560	366.3	0.408	-365.21	支持无行星
5045853099761604096	115.0	0.199	-368.64	支持无行星
4824625656537163008	80.9	0.356	-371.61	支持无行星
2367734656180397952	648.9	0.275	-372.75	支持无行星
2370173652144123008	17.6	0.514	-374.00	支持无行星
5534280594604579584	176.7	0.203	-376.14	支持无行星
637329067477530368	850.8	0.369	-381.74	支持无行星
824461960796102528	579.3	0.727	-382.89	支持无行星
3683727208499815936	115.3	0.473	-389.73	支持无行星
5629502218573747584	108.7	0.155	-391.06	支持无行星
2035577729682322176	18.8	0.299	-393.43	支持无行星
4823703990915066880	85.2	0.083	-395.99	支持无行星
5545870301853637504	447.3	0.120	-398.13	支持无行星
6448639343335797888	29.8	0.276	-407.05	支持无行星
4594158089392172928	178.3	0.575	-410.21	支持无行星
5808612830236138368	13.0	0.171	-410.82	支持无行星
5096613016130459136	51.6	0.261	-411.15	支持无行星
6158160019228728448	12.0	0.327	-414.77	支持无行星
2162250505790084480	584.3	0.253	-415.58	支持无行星
1035000055055287680	241.6	0.314	-416.90	支持无行星
6017724140678769024	31.9	0.216	-417.18	支持无行星
5855730584310531200	882.1	0.504	-419.59	支持无行星
2009052252148821632	620.9	0.247	-424.01	支持无行星

2026 S.T. yau High school Science Award
仅用于2026丘成桐中学科学奖评委评示

source_id	period (天)	eccentricity	$\ln BF$	证据强度
6873251629971269632	125.9	0.350	-429.77	支持无行星
5657709399107097600	243.8	0.144	-430.54	支持无行星
1918953867019478144	347.3	0.441	-430.63	支持无行星
861776842822330368	33.6	0.263	-440.79	支持无行星
6130370305216737408	19.4	0.001	-443.47	支持无行星
2055801936074059264	596.7	0.100	-444.14	支持无行星
6076629204819105664	143.9	0.235	-459.90	支持无行星
4597154602175212672	183.9	0.112	-460.89	支持无行星
4517375515957545216	229.9	0.260	-462.37	支持无行星
2431188610386566784	844.6	0.123	-463.81	支持无行星
1594127865540229888	893.2	0.311	-465.72	支持无行星
6075391292173461760	777.6	0.108	-471.40	支持无行星
4902020348734646912	187.0	0.357	-473.39	支持无行星
1318110830190386048	339.6	0.421	-473.53	支持无行星
3638745672411053952	206.7	0.335	-477.65	支持无行星
5583755078792642304	210.3	0.171	-478.64	支持无行星
596878550087942528	10.2	0.370	-478.87	支持无行星
4921427313081081600	1117.5	0.025	-483.27	支持无行星
1905308073023457920	8.8	0.314	-488.49	支持无行星
4919562197762515456	222.8	0.358	-493.31	支持无行星
1294000704856807168	125.4	0.269	-501.96	支持无行星
283446302278774528	22.8	0.296	-509.33	支持无行星
5013703860801457280	297.8	0.186	-511.43	支持无行星
3703975672901804160	56.2	0.239	-515.50	支持无行星
6374231714992810752	671.5	0.063	-516.22	支持无行星
6490284754986167552	24.7	0.549	-520.81	支持无行星
5361232754474635776	818.8	0.213	-522.93	支持无行星
5941647007018169728	856.7	0.183	-529.91	支持无行星
1067685718250692352	30.5	0.257	-558.32	支持无行星
4552227182675443584	10.2	0.005	-558.36	支持无行星
6525140510537736448	11.6	0.300	-561.14	支持无行星
6913810483612308480	23.9	0.386	-564.70	支持无行星
5431358403498310656	757.8	0.127	-568.07	支持无行星
1224551770875466496	147.6	0.513	-593.84	支持无行星
855523714036230016	145.1	0.264	-596.55	支持无行星
426186585428243840	714.0	0.129	-617.88	支持无行星
5459518648630576768	274.2	0.150	-634.49	支持无行星
6421118739093252224	898.7	0.190	-639.99	支持无行星
4525943082344640256	26.4	0.142	-644.70	支持无行星
513567110946268544	299.3	0.183	-645.35	支持无行星

2026 S.T.yau High school Science Award
仅用于2026丘成桐中学科学奖评委展示

source_id	period (天)	eccentricity	$\ln BF$	证据强度
5391879235910368768	648.7	0.495	-669.88	支持无行星
5086152743542494592	4.0	0.287	-812.52	支持无行星
5631222984331124864	15.4	0.294	-879.00	支持无行星

参考文献

[1] Mayor, M., & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. Nature, 378, 355.

[2] Borucki, W. J., et al. (2010). Kepler Planet-Detection Mission. Science, 327, 977.

[3] Perryman, M. (2011). The Exoplanet Handbook. Cambridge University Press.

[4] Sozzetti, A. (2009). Astrometric detection of exoplanets. Proceedings of the International Astronomical Union, 5, 221.

[5] Gaia Collaboration et al. (2016). The Gaia mission. A&A, 595, A1.

[6] Gaia Collaboration et al. (2023). Gaia Data Release 3: Summary of the content and survey properties. A&A, 674, A1.

[7] Holl, B., Sozzetti, A., Sahlmann, J., et al. (2023). Gaia Data Release 3: Astrometric orbit determination with Markov chain Monte Carlo and genetic algorithms. A&A, 674, A10.

[8] Marcussen, M. L., & Albrecht, S. H. (2023). Spectroscopic Follow-up of Gaia Exoplanet Candidates. AJ, 165, 266.

[9] Unger, N., et al. (2023). Revisiting the brown dwarf candidates from Gaia DR3. A&A, 680, A16.

[10] Kiefer, F., et al. (2025). Searching for substellar companion candidates with Gaia II. A&A, 702, A77.

[11] Sahlmann, J., & Gómez, M. (2024). Machine learning selection of exoplanet candidates from Gaia DR3. A&A, 681, A7.

[12] Wallace, J., et al. (2025). Bayesian constraints on giant planet companions from Gaia astrometry. AJ, 169, 45.

[13] Baycroft, T. A., Faria, J. P., & Delisle, J.-B. (2026). kima: Nested sampling for Gaia epoch astrometry. MNRAS, 500, 100.

[14] Skilling, J. (2006). Nested sampling for general Bayesian computation. Bayesian Analysis, 1, 833.

[15] Foreman-Mackey, D., et al. (2013). emcee: The MCMC Hammer. PASP, 125, 306.

[16] Jeffreys, H. (1961). Theory of Probability. Oxford University Press.

[17] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9, 90.

[18] Taylor, M. B. (2005). TOPCAT: Tool for OPerations on Catalogues And Tables. ASP Conference Series, 347, 29.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

[19] Murray, C. D., & Dermott, S. F. (1999). Solar System Dynamics. Cambridge University Press.

[20] Halbwachs, J.-L., et al. (2023). Gaia DR3 astrometric binaries. A&A, 674, A9.

[21] Aykroyd, A. I., et al. (2021). Stellar activity in astrometric planet searches. MNRAS, 500, 100.

[22] Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press.

[23] Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. Communications in Applied Mathematics and Computational Science, 5, 65.

[24] Foreman-Mackey, D., et al. (2013). emcee documentation. `https://emcee.readthedocs.io`

[25] Gelman, A., & Meng, X.-L. (1998). Simulating normalizing constants. Statistical Science, 13, 163.

[26] Friel, N., et al. (2014). Estimating evidence with thermodynamic integration. Statistics and Computing, 24, 1.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示