

地面望远镜数据处理方法: 频域vs.时域

熊 强¹ 李坦达^{1,2†}

(1 北京师范大学物理与天文学院 北京 100875)

(2 北京师范大学天文与天体物理前沿科学研究所 北京 102206)

摘要 地面望远镜的时域观测因昼夜交替和天气影响常导致数据空缺, 占空比较低(典型值约为0.30), 对时域天文研究造成显著影响. 为了对比基于频域分析与时域分析的方法在处理含间隙时域数据时的表现及其在星震学研究中的适用性, 采用了Lomb-Scargle算法和Inpainting插值法作为频域分析方法以及高斯过程(Gaussian Process, GP)方法作为时域分析方法, 对具有类太阳振动特征且占空比介于0.20至0.50之间的模拟时域测光数据进行了分析. 结果显示, 高斯过程方法在还原真实值的准确性和稳定性方面均表现最佳, 优于Lomb-Scargle方法和Inpainting方法. Inpainting方法在处理低占空比数据时可能引入大量假信号, 对信号测量造成干扰. 因此, 高斯过程方法是分析地面望远镜低占空比数据的首选, Lomb-Scargle方法次之, 而Inpainting方法不推荐.

关键词 方法: 数据分析, 恒星: 振动, 恒星: 太阳型

中图分类号: P141; **文献标识码**: A

1 引言

高精度的时域天文研究是当下天文研究的热点和前沿, 其中包括星震学^[1]、系外行星探测^[2]以及恒星转动测量^[3]. 然而, 高精度且连续的测光数据通常来自空间望远镜, 如 *Kepler*^[4]和 *TESS* (Transiting Exoplanet Survey Satellite)^[5]. 尽管这些空间望远镜极大地推动了相关领域的发展, 但其数据也存在局限性. 例如, *Kepler*的观测时长约为4 yr, 对于脉动周期在数天至数年的长周期变星研究而言, 其频率分辨率仍显不足. 在此背景下, 地面望远镜的数据仍然具有不可替代的价值. 例如, OGLE (Optical Gravitational Lensing Experiment)^[6]、MACHO (Massive Compact Halo Objects project)^[7]以及 ASAS-SN (All-Sky Automated Survey for SuperNovae)^[8], 这些项目通常有着更长

的观测时长, 大幅提高了频率分辨率, 在长周期变星研究中具有独特优势, 有助于弥补空间望远镜的不足.

尽管地面望远镜可以提供长时间基线的时域数据, 但由于天气状况和昼夜交替等限制, 数据中常出现大量间隙, 导致单个台站的数据占空比(Duty cycle)很低. 这里的占空比指的是, 天体在一个观测周期内处于活动状态(例如周期现象)的时间与总观测时间的比值. 对于长时间基线的项目如OGLE, 典型的占空比约为0.30. 多台站联合观测有助于将占空比提高至0.60–0.95, 取决于台站的数量以及它们的地理分布. 这种情况对基于频域的分析研究带来巨大影响. 由于数据存在间隙, 对观测数据进行傅里叶变换时会引入窗函数, 该窗函数的频谱与原始信号的真实频谱进行卷积,

2025-01-03收到原稿, 2025-02-13收到修改稿

† litanda@bnu.edu.cn

会产生旁瓣效应导致频谱能量从主峰扩散,从而使信号测量变得困难.因此,如何有效应对数据中的间隙显得尤为重要.处理这一问题的策略主要有3种:一是直接使用适用于非等间隔数据的Lomb-Scargle算法^[9-10],二是在进行频域分析前处理间隙,例如Hekker等^[11]直接去除数据中的间隙或Pires等^[12]通过已有数据点对间隙进行插值;三是使用高斯过程(Gaussian Process, GP)^[13]直接在时域中进行分析.例如Pereira等^[14]使用高斯过程在时域上对米粒组织(Granulation)和振动(Oscillation)进行建模,他们发现此处理方式能够很好地应对数据中存在的间隙.相似的结论还可以从Hey等^[15]的工作中得出,他们在时域中利用高斯过程对稀疏的TESS数据进行了处理,证明了高斯过程在分析存在数据间隙的星震学数据时的应用潜力.

本文聚焦于占空比为0.20–0.50的模拟地面望远镜数据,通过频域和时域两种分析方式对数据进行处理,并对其效能进行系统评估.我们采用了3种具体方法:Lomb-Scargle算法和Inpainting方法用于频域分析,高斯过程方法则完全基于时域进行处理.

2 方法

2.1 数据

本文使用模拟的类太阳振动恒星光变曲线数据.类太阳振动是指恒星外部对流层湍流随机激发的振动机制,与太阳的激发机制相同,因此这些恒星被称为类太阳振动恒星.类太阳振动由于具有随机激发和随机耗散的特点,其频率不稳定,振幅和相位变化具有随机性,因此特征频率具有一定的弥散性,对数据间隙的影响较为敏感,这也对数据处理方法提出了更高要求.模拟数据包含3部分:米粒组织、振动和白噪声(White noise).米粒组织是描述类太阳恒星表面对流元结构的一种观测特征,具体指的是由对流作用引起的细小明暗结构.其在功率谱中的特征可近似由Harvey轮廓^[16]所描述,即,

$$P(f) = \frac{2A_{\text{gran}}}{1 + (f/f_{\text{gran}})^4}, \quad (1)$$

其中, P 表示功率密度(Power Density), f 表示频率, f_{gran} 为特征频率, A_{gran} 为特征频率处所对应的振幅.类太阳振动是压力模或混合压力-重力模的驻波,并且会受到近表层对流的阻尼作用.它们在一定频率范围内被激发,且其振幅大致呈现出钟形分布.它们在功率谱中可近似由洛伦兹轮廓描述,即,

$$P(f) = \frac{A(\gamma f_0)^2}{(f^2 - f_0^2)^2 + (\gamma f)^2}, \quad (2)$$

其中, f_0 表示每个模的中心频率, γ 表示半高全宽, A 表示中心频率处的振幅.我们设置了3个符合方程(2)描述的模式,用以模拟压力模振动(以下简称为信号, Signal).我们设置的参数真实值如表1所示,其中下标“left”、“central”和“right”根据信号的中心频率 f_0 大小,分别对应左信号、中信号和右信号.

表 1 光变曲线中设置的参数及对应真实值
Table 1 Parameters configured in the light curve and their corresponding true values

Parameter	$f/\mu\text{Hz}$	$A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})$	$\gamma/\mu\text{Hz}$
Granulation	0.10	400	-
Signal _{left}	0.40	200	0.06
Signal _{central}	0.60	300	0.06
Signal _{right}	0.80	200	0.07

我们在数据中引入了正态分布的白噪声,其均值为0,方差为 7^2 ppm,即 $\mathcal{N}(0, 7^2)$,以尽可能模拟真实地面数据的信噪比.参考地面望远镜OGLE的数据,我们将光变曲线数据的总时长设置为30 yr,范围为0–11000 d,采样率约为2 d,总共包含5000个数据点.制作的光变曲线及其对应的功率谱示例如图1所示.

为了生成与真实数据相近的间隙,需要构建与实际情况相符的窗(window),其定义为:

$$\text{window}(t) = \begin{cases} 1, & t \text{ 时刻存在数据,} \\ 0, & t \text{ 时刻无数据.} \end{cases} \quad (3)$$

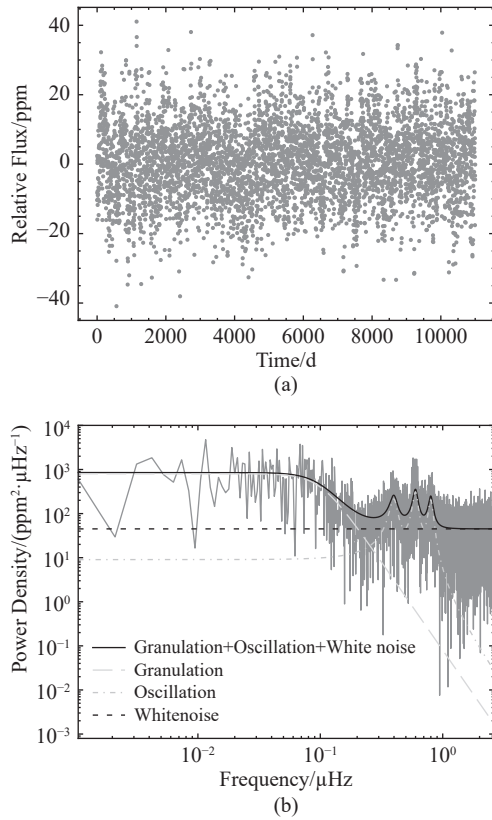


图 1 制作的光变曲线及其对应的功率谱示例, 各个部分标示于图例中。

Fig. 1 An example of the generated light curve and its corresponding power spectrum, with each part labeled in the legend.

为此, 我们同样参考了地面望远镜OGLE的数据. 以下是制作特定占空比窗的流程:

(1) 首先在OGLE-III^①数据库中得到一颗具有较高占空比的真实恒星数据, 如图2 (a)所示, 随后通过光变曲线数据得到观测窗, 如图2 (b)所示;

(2) 将(1)中得到的窗按顺序重复, 直至扩展至与初始光变曲线数据长度一致(11000 d);

(3) 为了模拟更复杂的情况, 我们在窗中分别引入了一个约为1000 d的大间隙以及3个200 d的小间隙. 最终得到的初始窗如图2 (c)所示;

(4) 在(3)中得到的窗中, 随机选取一个位置作为起始索引, 并将该索引及其之后的100个窗值设为0或1. 重复此过程, 直到窗的占空比达到指定值

(实际误差控制在 $\pm 1\%$).

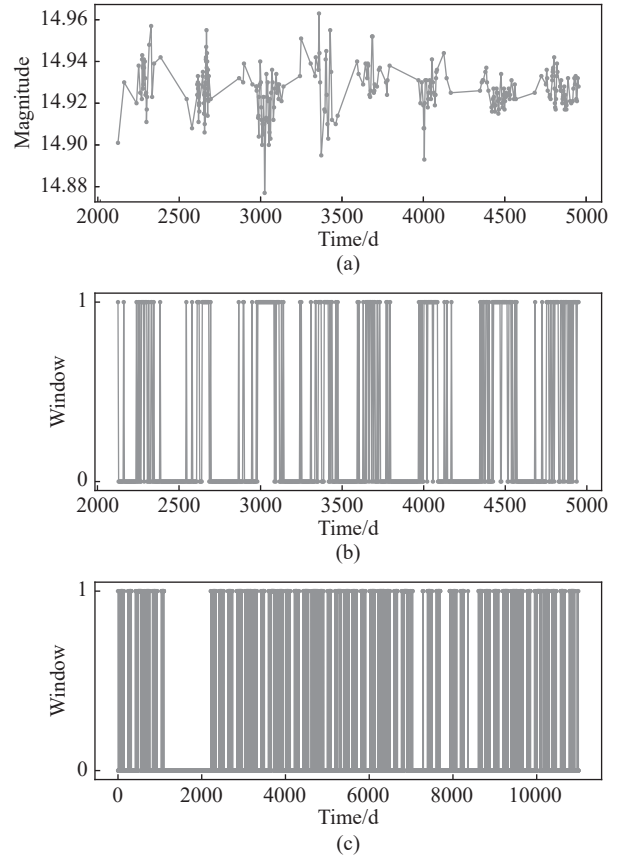


图 2 (a)来自地面望远镜OGLE的一颗特定目标(ID: OGLE-LMC-LPV-35662)的光变曲线数据; (b)与之对应的观测窗; (c)本文最终使用的初始窗。

Fig. 2 (a) Light curve data of a specific target (ID: OGLE-LMC-LPV-35662) from the ground-based OGLE telescope; (b) the corresponding observational window; (c) the initial window ultimately used in this work.

结合初始制作的光变曲线数据和特定占空比的窗, 我们能够得到带有间隙且满足特定占空比的光变曲线数据。

我们分别制作了占空比为0.20–0.50的光变曲线数据, 并根据其在经过Lomb-Scargle算法(详见下文)得到的频谱中是否能够通过目视检查识别出大致振动信号为标准, 在每个占空比下挑选了两条光变曲线, 分别标记(Flag)为“好(good)”以及

^①<https://ogledb.astrouw.edu.pl/~ogle/CVS/>

“差(bad)”. 这样区分的目的是为了比较在数据排列、白噪声以及窗变化引起信号显著变化时, 不

同方法的测量效果. 图3展示了我们在本文使用的所有光变曲线数据.

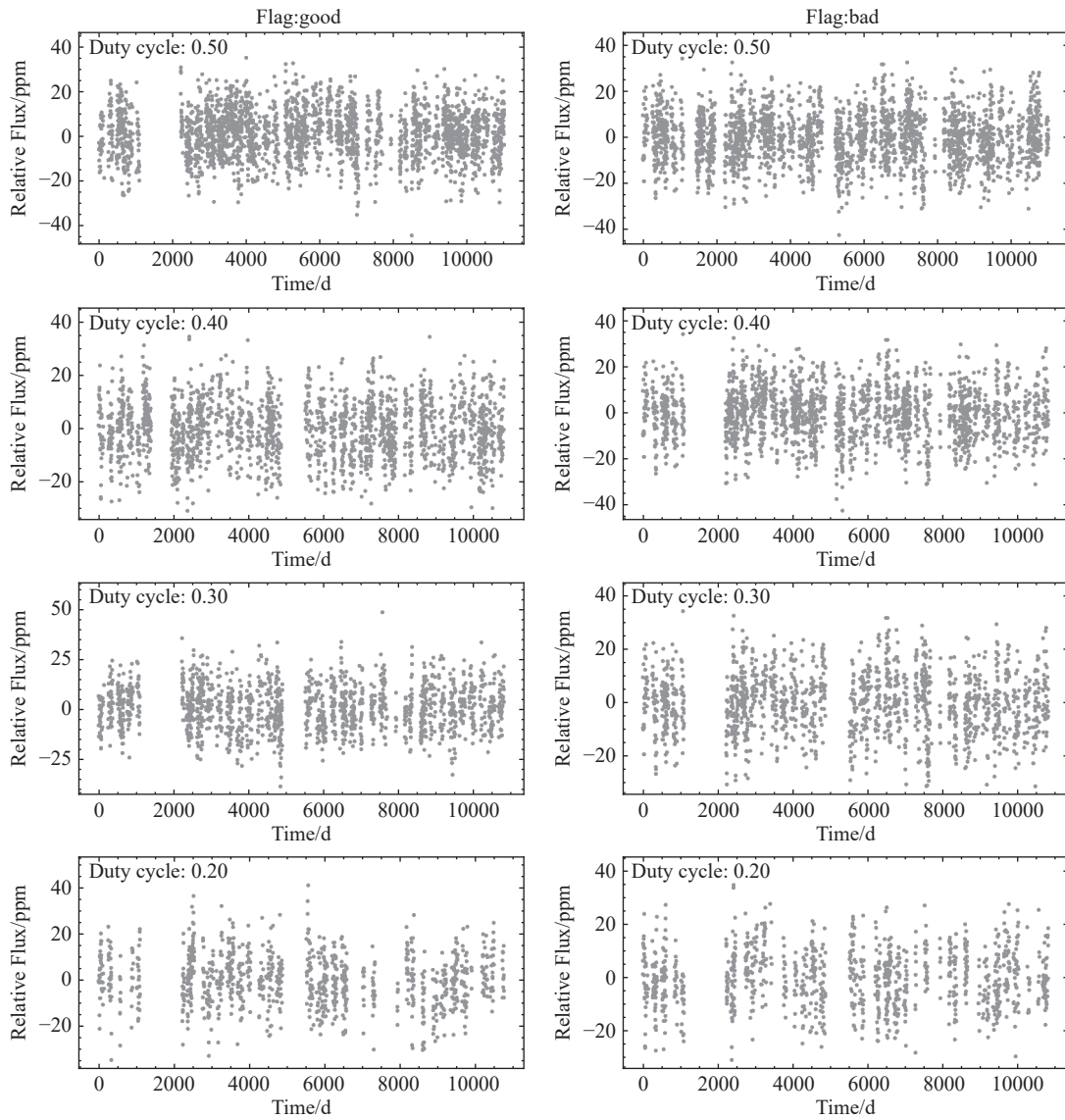


图 3 本文使用的所有光变曲线数据, 其中左侧面板为标记为“好”的数据, 右侧面板为标记为“差”的数据.

Fig. 3 All the light curve data used in this study, with the left panel showing data flagged as “good” and the right panel showing data flagged as “bad”.

2.2 处理方法

2.2.1 Lomb-Scargle

与仅适用于等间隔、均匀采样数据的快速傅里叶变换不同, Lomb-Scargle算法能够处理非等间隔采样的数据. 其基本工作原理是: 对每一个频率

在时域内进行正弦波拟合, 如果该频率的正弦波能够很好地拟合数据(即残差较小), 则在频域中相应的振幅值会较大. 因此, Lomb-Scargle算法能够在一定程度上应对数据中的间隙. 当数据中没有间隙时, Lomb-Scargle与快速傅里叶变换的执行结

果是一样的.

2.2.2 Inpainting

该方法旨在利用数据中的真实值和观测窗, 恢复理想状态下的光变曲线. 通过特定的变换矩阵作用于带有间隙的光变曲线数据, 采用迭代方式寻找变换后系数大部分为零的解, 即最稀疏解. 例如, 对于单一正弦变化的时间序列数据, 其特性可以通过傅里叶空间的系数反映, 因为在这种情况下傅里叶空间中的大部分系数为零, 此时对应的变换矩阵为傅里叶变换. 该方法的具体描述可参见文献[12]. 图4展示了该方法对占空比为0.30, 标记为“好”的光变曲线处理后的结果以及相应的Lomb-Scargle算法作用之后得到的功率谱.

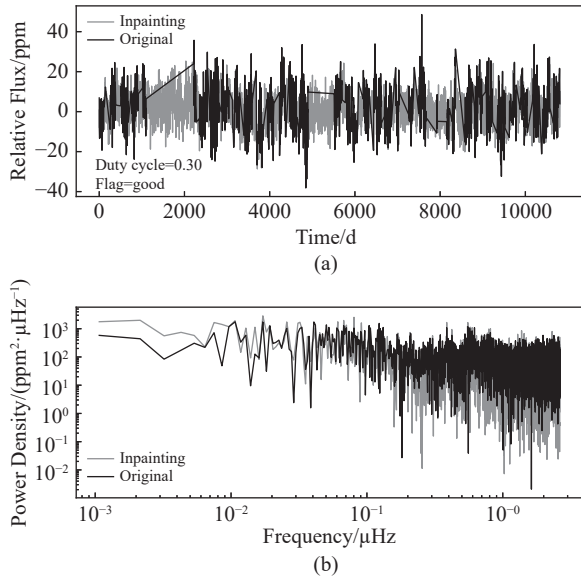


图4 Inpainting方法对占空比为0.30且标记为“好”的光变曲线数据处理结果. 上图为光变曲线, 下图为对应的功率谱.

Fig. 4 The results of the Inpainting method applied to light curve data with a duty cycle of 0.30 and flagged as “good”. The upper panel shows the light curve, while the lower panel presents the corresponding power spectrum.

高斯过程(Gaussian Process, GP)是一种非参数的贝叶斯统计方法, 广泛应用于机器学习领域[13]. 高斯过程是一种随机过程, 其假设每个输入点(如时间或波长)对应的输出值是正态分布的随机变量. 此外, 这些随机变量的任意有限子集服从多元高斯分布. 高斯过程的分布($F(x)$)由均值函数

和协方差函数共同描述, 定义如下:

$$F(x) \sim \mathcal{GP}[m(x), k(x, x')], \quad (4)$$

其中, x 、 x' 指代点, $m(x)$ 为均值函数, 指定了在没有任何观测数据的情况下我们对目标函数的初始猜测, 在本文中, $m(x)$ 被设置为0. $k(x, x')$ 是协方差函数(也称为核函数), 描述的是输入空间中任意两点之间的相似性, 不同的核函数决定了函数不同的先验分布. 如果选择了合适的核函数, 并对其参数进行了有效的优化, 则可以直接在时域内通过高斯过程得到我们感兴趣的参数.

我们使用CELERITE2库^[17-18]中的SHO (Simple Harmonic Oscillator)项作为核函数来构建高斯过程模型, 此核函数描述了阻尼简谐振动过程, 而阻尼简谐振动过程与类太阳振动的物理特征相似, 因此能够满足我们模型构建的需求. 其对应的功率谱为:

$$P(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0 \omega_0^4}{(\omega^2 - \omega_0^2)^2 + \frac{\omega_0^2 \omega^2}{Q^2}}, \quad (5)$$

其中, S_0 为系统的功率谱密度的标度因子, 代表了系统的总功率或能量的强度; ω_0 为特征角频率; Q 为品质因子, 描述了阻尼的大小. 3个参数之间满足 $S(\omega_0) = \sqrt{2/\pi} S_0 Q^2$. 当 $Q = 1/\sqrt{2}$ 时, 此方程刚好对应方程(1), 可用于对米粒组织部分的建模. 注意, 在方程(5)中, 自由参数分别为 S_0 、 ω_0 、 Q (不包括米粒组织), 因此, 有必要与方程(1)和方程(2)建立参数之间的转换关系, 参考Pereira等^[14], 参数之间的转换关系为:

$$\begin{aligned} A_{\text{gran}} &= \sqrt{1/2\pi} S_{\text{gran}}; \\ f_{\text{gran}} &= 2\pi/\omega_{\text{gran}}; \\ f_0 &= 2\pi/\omega_0; \\ \gamma &= \omega_0/2\pi Q; \\ A &= 4S_0 Q^2, \end{aligned} \quad (6)$$

其中, S_{gran} 和 ω_{gran} 为方程(5)中 $Q = 1/\sqrt{2}$ 时的 S_0 和 ω_0 , 是与米粒组织相关的物理量.

2.3 拟合

Lomb-Scargle和Inpainting方法的拟合在频域

中进行. 米粒组织部分由方程(1)描述, 振动部分由方程(2)描述, 同时加入一个常数项(W)以解释白噪声, 拟合方程如下:

$$P(f) = W + \frac{2A_{\text{gran}}}{1 + (f/f_{\text{gran}})^4} + \sum_{i=1}^3 \frac{A_i (\gamma_i f_{0,i})^2}{(f^2 - f_{0,i}^2)^2 + (\gamma_i f)^2}, \quad (7)$$

其中, 变量含义同前文, 对应下标 $i = 1, 2, 3$ 分别对应左、中和右信号. 高斯过程方法的拟合在时域中进行. 我们使用一个 $Q = 1/\sqrt{2}$ 的SHO项来解释米粒组织以及3个常规的SHO项的叠加来解释振动部分, 同时加入一个常数(W)以解释白噪声.

在本文中, 3种方法均采用相同的均匀分布先验. 白噪声的先验设置为 $\ln W \sim [0, 5]$, 其余输入参数及对应先验详见表2.

表 2 Lomb-Scargle方法、Inpainting方法和高斯过程方法的拟合输入参数及相应先验值

Table 2 Fitting input parameters and corresponding priors for the Lomb-Scargle, Inpainting, and Gaussian process methods

Parameter	Priors		
	$f/\mu\text{Hz}$	$\ln[A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})]$	$\gamma/\mu\text{Hz}$
Granulation	[0.01, 0.2]	[0, 7]	–
Signal _{left}	[0.3, 0.5]	[0, 6]	[0.01, 0.2]
Signal _{central}	[0.5, 0.7]	[0, 6]	[0.01, 0.2]
Signal _{right}	[0.7, 0.9]	[0, 6]	[0.01, 0.2]

我们使用PyMC3^[19]通过马尔科夫链蒙特卡洛方法在参数空间中采样以获得参数的后验分布, 其中参数的初始位置由先验分布随机生成. 我们选取了两条链, 首先执行1000步调优(tune), 用于优化采样器的超参数(如步长、接受率等)以提高采样效率和质量. 随后, 我们执行1000次采样(draws), 得到参数的后验分布. 我们通过Gelman-Rubin收敛诊断统计量(\hat{R})^[20]来进行参数收敛判断, 当 $\hat{R} \approx 1$ 时, 意味着两条链中的参数的分布极其相似, 这时参数被认为是收敛的. 我们将参数后验分布的中值作为对该参数的最佳估计. 图5展示了占

空比为0.50, 标记为“好”(上图)以及“差”(下图)的数据的3种方法拟合结果.

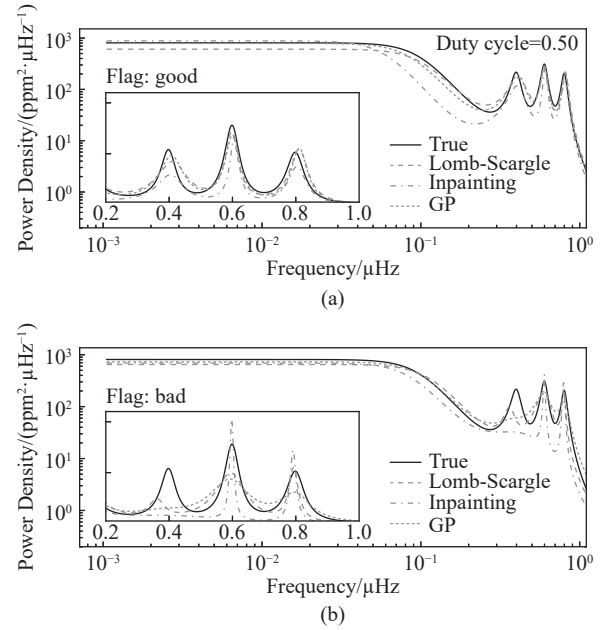


图 5 Lomb-Scargle方法(灰色虚线)、Inpainting方法(灰色点划线)以及高斯过程方法(灰色点线)对占空比为0.50且标记为“好”(上图)及“差”(下图)的数据拟合结果. 黑色线表示真实参数对应的功率谱.

Fig. 5 Fitting results for data with a duty cycle of 0.50 flagged as “good” (upper panel) and “bad” (lower panel), obtained using the Lomb-Scargle method (gray dashed line), the inpainting method (gray dash-dotted line), and the Gaussian process method (gray dotted line). The black line represents the power spectrum corresponding to the true parameters.

值得注意的是, 对于高斯过程方法, 其作用在时域, 图6展示了图5中高斯过程方法作用于光变曲线的示意图, 为了更清晰地展示图中的细节, 仅选取了3000–8000 d的数据及其拟合结果进行展示.

3 结果

本研究采用Lomb-Scargle算法、Inpainting方法和高斯过程方法, 对占空比在0.20至0.50之间、标记(Flag)为“好(good)”和“差(bad)”的数据进行分析, 测定了米粒组织的特征频率和振幅以及振动信号的中心频率、振幅和半高全宽. 所有测量结果列于附录. 参数的最佳估计值由后验分布的

中位数给出, 不确定性则采用绝对中位数偏差表示.

3.1 频率

图7展示了3种方法在占空比0.20–0.50下对频率的测量偏差, 并归一化为相对于真实值的比值. 上图对应标记为“好”的数据, 下图对应标记为“差”的数据. 黑色虚线表示比值等于0的参考位置, 灰色虚线标注了不同的比值参考范围. 对于标记为“好”的数据, 米粒组织、左信号、中信号和右信号的参考值分别为 ± 0.2 、 ± 0.05 、 ± 0.01 和 ± 0.02 ; 对于标记为“差”的数据, 相应值分别为 ± 0.1 、 ± 0.1 、 ± 0.02 和 ± 0.01 .

I. 对于标记为“好”的数据

(1) 米粒组织特征频率的测量: 在准确性方面, 高斯过程方法优于Lomb-Scargle方法和Inpainting方法, 测量偏差在所有占空比下均未超过20%. Lomb-Scargle方法整体优于Inpainting方法, 而后者在占空比0.2时的偏差接近90%. 在稳定性方面, Lomb-Scargle方法略优于高斯过程方法, 且显著优于Inpainting方法. 三种方法对4个占空比条件下

的数据计算得到的标准差(以下简称为标准差)分别为9.8%、11.2%和26.1%.

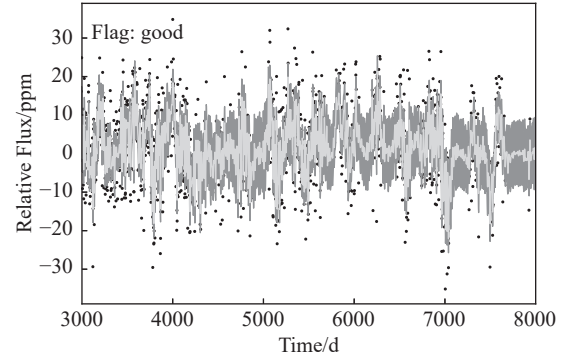


图 6 高斯过程方法对占空比为0.50且标记为“好”的光变曲线数据的直接拟合结果. 黑色点表示原始光变曲线数据, 浅灰线代表每个时刻 t 的预测结果, 深灰色阴影区域表示预测结果的68%置信区间.

Fig. 6 Direct fitting results of the Gaussian process method applied to light curve data with a duty cycle of 0.50 and flagged as “good”. The black dots represent the original light curve data, the lightgray line indicates the predicted values at each time t , and the gray shaded region represents the 68% confidence interval of the predictions.

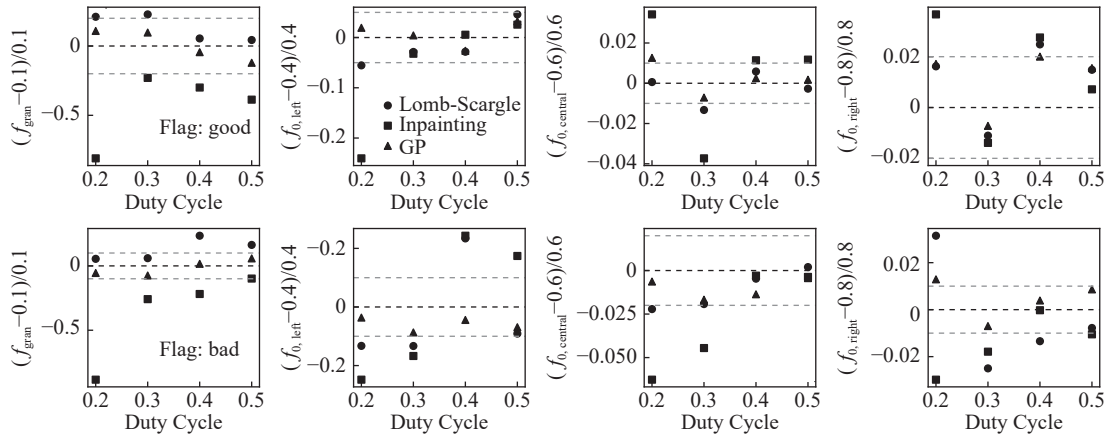


图 7 Lomb-Scargle方法(圆形)、Inpainting方法(方形)和高斯过程方法(三角形)在不同占空比条件下对米粒组织频率(f_{gran})和信号频率(f_0)的测量结果相对于真实值的偏差, 并归一化为相对于真实值的比值.

Fig. 7 The measurement deviations of granulation frequency (f_{gran}) and signal frequency (f_0) relative to the true values, normalized by the true values, using the Lomb-Scargle method (circles), Inpainting method (squares), and Gaussian process method (triangles) under different duty cycles.

(2) 振动信号中心频率的测量: 在准确性方面, 高斯过程方法在所有占空比下均表现最佳, 其测量偏差始终小于5%, 明显优于Lomb-Scargle方法和Inpainting方法. 同时, Lomb-Scargle方法在各占

空比下均优于Inpainting方法, 尤其在占空比0.20时差异较为显著. 在稳定性方面, 高斯过程方法的标准差在左、中、右信号下分别为2.5%、0.8%和1.3%, 整体优于Lomb-Scargle方法(4.4%、0.8%、

1.5%)和Inpainting方法(12.2%、3.0%、2.3%)。其中, Lomb-Scargle方法在中信号的稳定性与高斯过程方法相当,但在其他信号下仍略逊于高斯过程方法,而Inpainting方法的稳定性在所有情况下最差。

II. 对于标记为“差”的数据

(1) 米粒组织特征频率的测量: 在准确性方面, 高斯过程方法表现最佳, 所有占空比下的测量偏差均小于10%。Lomb-Scargle方法优于Inpainting方法, 尤其在占空比0.20时差距明显。稳定性方面, 高斯过程方法优于Lomb-Scargle方法和Inpainting方法, 标准差分别为6.2%、8.7%和35.3%。

(2) 振动信号中心频率的测量: 在准确性方面, 高斯过程方法在所有信号下均表现最佳, 其中左信号的测量偏差始终小于10%, 中信号在2%以内, 右信号在1%或2%以内, 显著优于其他方法。Lomb-Scargle方法整体优于Inpainting方法, 唯有右信号在占空比0.30和0.40时, Inpainting方法略优于Lomb-Scargle方法。在稳定性方面, 高斯过程方法的标准差最小, 在左、中、右信号下分别为

2.3%、0.6%和0.9%, 表现最佳。Lomb-Scargle方法在左信号下的标准差为17.8%, 明显逊于Inpainting方法(24.4%), 但在中信号(1.2%)和右信号(2.5%)下则表现最差, Inpainting方法在这两种情况下的标准差分别为3.0%和1.2%。综合来看, 高斯过程方法在稳定性方面始终优于其他两种方法。

综上, 无论是标记为“好”还是标记为“差”的数据, 高斯过程方法在频率测量的准确性和稳定性方面始终表现最佳。Lomb-Scargle方法整体优于Inpainting方法, 且在占空比0.20和0.30时优势尤为突出。

3.2 振幅

图8展示了3种方法在占空比0.20–0.50下对振幅的测量偏差, 并归一化为相对于真实值的比值。上图对应标记为“好”的数据, 下图对应标记为“差”的数据。黑色虚线表示比值为0的位置, 灰色虚线标注了比值的参考值, 其中, 米粒组织、左信号、中信号和右信号在“好”数据中分别为 ± 0.2 、 ± 0.5 、 ± 0.2 和 ± 0.4 , 在“差”数据中分别为 ± 0.2 、 -0.5 、 ± 0.3 和 ± 0.5 。

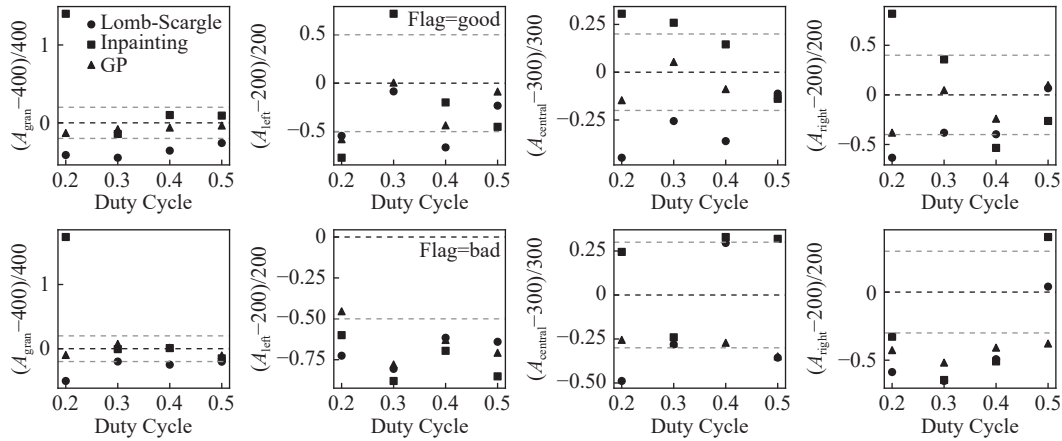


图 8 与图7类似, 但此处展示的是振幅(A)的测量结果相对于真实值的偏差, 并归一化为相对于真实值的比值。

Fig. 8 Similar to Figure 7, but showing the normalized deviations of amplitude (A) measurements relative to the true values.

I. 对于标记为“好”的数据

(1) 米粒组织特征振幅的测量: 在准确性方面, 高斯过程方法的测量偏差均在20%以内, 优于Lomb-Scargle方法和Inpainting方法。尽管Inpainting方法在占空比0.3、0.4和0.5时优于Lomb-Scargle方法, 偏差均在20%以内, 但在占空比0.20时偏

差较大。在稳定性方面, 高斯过程方法表现最佳, 标准差为4.0%, Lomb-Scargle方法和Inpainting方法的标准差分别为8.2%和70.1%。

(2) 振动信号振幅的测量: 在准确性方面, 高斯过程方法在所有信号下均表现最佳, 其中左信号的测量偏差大多在50%以内, 并在占空比0.30和

0.50时小于10%. 中信号的测量偏差始终小于20%, 而右信号的测量偏差则在40%以内. Lomb-Scargle方法整体优于Inpainting方法, 但在左信号的占空比0.40以及中信号的占空比0.40时, Inpainting方法略优. 在稳定性方面, Lomb-Scargle方法在左信号下的标准差最低(26.8%), 略优于高斯过程方法(27.8%), 但在中信号(14.4%对8.9%)和右信号(29.2%对23.0%)下均明显逊色. Inpainting方法的标准差在所有情况下均最大, 分别为63.9%、20.0%和61.0%. 整体来看, 高斯过程方法在多数情况下表现最佳, 具有较高的稳定性.

II. 对于标记为“差”的数据

(1) 米粒组织特征振幅的测量: 在准确性方面, 高斯过程方法优于Lomb-Scargle方法和Inpainting方法, 所有占空比下的偏差均在20%以内. Inpainting方法在占空比0.30、0.40和0.50时优于Lomb-Scargle方法, 但在0.20时偏差接近170%, 偏差显著增大. 在稳定性方面, 高斯过程方法表现最佳, 标准差为8.8%; Lomb-Scargle方法优于Inpainting方法, 标准差分别为14.5%和89.7%.

(2) 振动信号振幅的测量: 在准确性方面, 3种方法在左、右信号下均存在较大偏差, 测量结果与真实值的偏差基本超过30%, 其中左信号的偏差甚至在-50%以上, 仅Lomb-Scargle方法在右信号占空比0.50时的偏差小于10%. 对于中信号, 高斯过程方法的测量偏差整体最小, 除占空比0.50时在-35%以内外, 其余占空比下均在-30%以内. Inpainting方法在中信号的占空比0.20时略优于Lomb-Scargle方法. 在稳定性方面, Lomb-Scargle方法在左信号下的标准差最低(8.7%), 优于Inpainting方法(13.2%)和高斯过程方法(14.0%). 在中信号下, 高斯过程方法的标准差最小(4.2%), 表现最佳, Inpainting方法(27.2%)优于Lomb-Scargle方法(34.2%). 在右信号下, 高斯过程方法的标准差仍最小(6.0%), Lomb-Scargle方法次之(31.2%), 而Inpainting方法的标准差最大(46.8%).

综上, 对于米粒组织的振幅测量, 高斯过程方法在标记为“好”和“差”的数据中均表现出更高的准确性和稳定性, 相比Lomb-Scargle方法和Inpainting方法具有优势. 而对于振动信号的振幅测量, 高斯过程方法在“好”数据下的准确性和稳定性均

优于Lomb-Scargle方法和Inpainting方法; 在“差”数据下, 3种方法的准确性均表现不佳, 而稳定性高斯过程方法最优.

3.3 半高全宽

图9展示了3种方法在占空比0.20-0.50下对信号的半高全宽的测量偏差, 并归一化为相对于真实值的比值. 上图对应标记为“好”的数据, 下图对应标记为“差”的数据. 黑色虚线表示真实值, 灰色虚线标注了比值的参考值, 其中左信号、中信号和右信号在“好”数据中分别为 ± 0.5 、 ± 0.2 和 ± 0.2 , 在“差”数据中分别为0.5、 ± 0.5 和 ± 0.5 .

I. 对于标记为“好”的数据: 在准确性方面, 3种方法在左信号下的测量结果整体较差. 其中, Inpainting方法在占空比0.20时偏差最大, 超过200%. 相比之下, 高斯过程方法的准确性优于其他两种方法, 尤其在占空比0.30时其偏差接近20%. 在中信号下, 3种方法在占空比0.20时均存在较大偏差; 然而, 在占空比0.30、0.40和0.50时, 高斯过程方法的测量偏差均控制在10%以内, 表现最佳, Lomb-Scargle方法次之(偏差在20%以内), 而Inpainting方法的偏差最为显著. 在右信号下, 高斯过程方法在所有占空比下的测量偏差均小于20%, 整体优于其他方法. 此外, 在占空比0.20和0.30时, Lomb-Scargle方法的测量偏差显著小于Inpainting方法. 在稳定性方面, 高斯过程方法在左、右信号下的标准差最小(42.7%和15.0%), 表现最佳, Lomb-Scargle方法次之(88.0%和23.8%), 而Inpainting方法的标准差最大(145.4%和59.3%). 然而, 在中信号下, Inpainting方法的标准差最低(18.3%), 优于Lomb-Scargle方法(40.0%)和高斯过程方法(48.9%).

II. 对于标记为“差”的数据: 在准确性方面, 3种方法在所有信号下的测量结果偏差较大, 整体准确性较低. 在稳定性方面, 高斯过程方法的标准差最小, 在左、中、右信号下分别为31.7%、31.3%和51.3%, 表现最佳. Inpainting方法的标准差较大, 分别为46.1%、53.9%和81.6%, 而Lomb-Scargle方法的标准差最大, 分别达到97.5%、78.8%和92.7%. 总体而言, 高斯过程方法在所有情况下均具有最优的稳定性, 而Lomb-Scargle方法的稳定性最差.

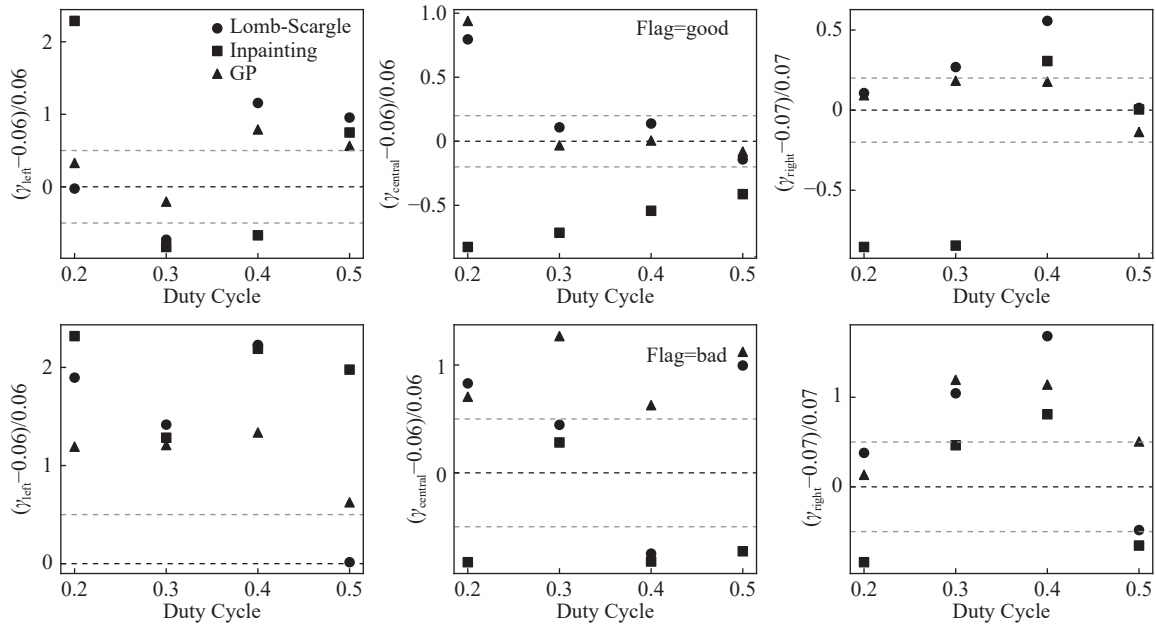


图 9 与图7类似, 但此处展示的是信号的半高全宽(γ)的测量结果相对于真实值的偏差, 并归一化为相对于真实值的比值。

Fig. 9 Similar to Figure 7, but showing the normalized deviations of the signal's full width at half maximum (γ) relative to the true values.

综上, 在振动信号半高全宽的测量中, 对于标记为“差”的数据, 高斯过程方法在准确性和稳定性方面整体表现最佳, 优于Lomb-Scargle方法和Inpainting方法; 而对于标记为“差”的数据, 3种方法的准确性均较差, 但在稳定性方面, 高斯过程方法仍优于Lomb-Scargle方法和Inpainting方法。

3.4 假信号问题

对于Inpainting方法, 我们发现其在处理低占空比(例如0.30)的数据时容易引入大量假信号. 如图10所示, 上图为原始数据经Lomb-Scargle算法得到的频谱(Original), 下图为经过Inpainting方法处理后的频谱, 其中箭头标注了真实信号的频率位置. 在原始频谱中, 由于数据间隙的存在信号被显著削弱, 几乎难以辨识. 而在Inpainting方法处理后的频谱中, 真实信号附近出现了多个假峰. 这些假峰显著干扰了真实信号的测量, 可能导致错误的信号频率判定, 进而影响恒星物理参数的准确测定. 因此, 不建议在低占空比地面望远镜数据处理中采用Inpainting方法。

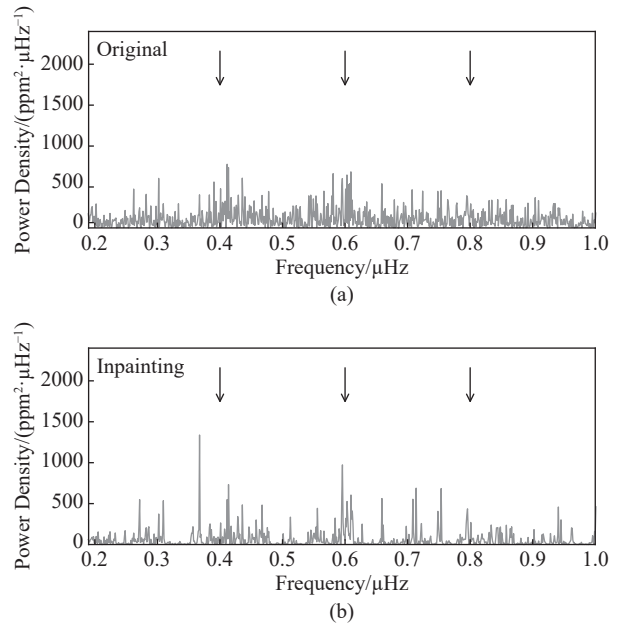


图 10 Inpainting方法对占空比为0.30的光变曲线处理结果, 其中箭头指示真实信号的频率位置。

Fig. 10 Results of the Inpainting method applied to light curve data with a duty cycle of 0.30, where the arrows indicate the frequency positions of the true signals.

4 结论

由于地面望远镜数据中常存在大量间隙, 这对真实信号的测量, 尤其是基于频域的分析方法构成了挑战. 本文针对占空比介于0.20–0.50且具有类太阳振动特征的模拟地面望远镜数据, 比较了3种方法的表现: 基于频域的Lomb-Scargle算法^[9-10]、Inpainting方法^[14]以及基于时域的高斯过程方法^[13]. 通过对数据质量进行分类分析, 我们发现高斯过程方法在频率、振幅和半高全宽的测量中整体表现最佳, 结果更接近真实值且更稳定. 相比之下, Lomb-Scargle方法在部分低占空比数据中准确性尚可, 但稳定性较差; Inpainting方法整体表现最差, 准确性和稳定性均较差, 并且容易引入假信号, 尤其在占空比为0.20和0.30时表现不佳.

因此, 对于低占空比的地面望远镜数据, 基于时域分析的高斯过程方法展现出卓越的鲁棒性和适用性, 成为测量振幅和相位不规律、具有随机激发与耗散特性的周期性信号(如恒星振动、自转调制、对流引发的米粒组织运动等)的首选方法.

参考文献

- [1] Yu J, Huber D, Bedding T R, et al. ApJ, 2018, 236: 42
- [2] Seager S, Deming D. ARA&A, 2010, 48: 631
- [3] Kraft R P. ApJ, 1967, 150: 551
- [4] Koch D G, Borucki W J, Basri G, et al. ApJ, 2010, 713: L79
- [5] Ricker G R, Winn J N, Vanderspek R, et al. JATIS, 2015, 1: 014003
- [6] Udalski A, Szymanski M, Kaluzny J, et al. AcA, 1992, 42: 253
- [7] Alcock C, Allsman R A, Alves D R, et al. ApJ, 2000, 542: 281
- [8] Shappee B J, Prieto J L, Grupe D, et al. ApJ, 2014, 788: 48
- [9] Lomb N R. Ap&SS, 1976, 39: 447
- [10] Scargle J D. ApJ, 1982, 263: 835
- [11] Hekker S, Barban C, Baudin F, et al. A&A, 2010, 520: A60
- [12] Pires S, Mathur S, García R A, et al. A&A, 2015, 574: A18
- [13] Williams C K I, Rasmussen C E. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006: 10-20
- [14] Pereira F, Campante T L, Cunha M S, et al. MNRAS, 2019, 489: 5764
- [15] Hey D, Huber D, Ong J, et al. aXiv, 2024: 2403.02489
- [16] Harvey J. Future Missions in Solar, Heliospheric & Space Plasma Physics. Noordwijk, NL: ESA Scientific & Technical Publications Branch, 1985, 235: 199
- [17] Foreman-Mackey D, Agol E, Ambikasaran S, et al. ApJ, 2017, 154: 220
- [18] Foreman-Mackey D. RNAAS, 2018, 2: 31
- [19] Salvatier J, Wiecki T V, Fonnesbeck C. PeerJ, 2016, 2: e55
- [20] Gelman A, Meng X L, Stern H. Stat Sinica, 1996, 6: 733

Processing Ground-Based Asteroseismic Photometric Data: Frequency Domain vs. Time Domain

XIONG Qiang¹ LI Tan-da^{1,2}

(¹ School of Physics and Astronomy, Beijing Normal University, Beijing 100875)

(² Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206)

ABSTRACT Time-domain observations with ground-based telescopes are often affected by the day-night cycle and weather conditions, leading to data gaps and a relatively low duty cycle (typically around 0.30), which significantly impacts time-domain astronomical studies. To compare the performance of frequency-domain and time-domain analysis methods in handling time-domain data with gaps and their applicability in asteroseismology, the Lomb-Scargle algorithm and the Inpainting interpolation method were employed as frequency-domain approaches, while the Gaussian Process (GP) method was used as a time-domain approach. These methods were applied to simulate light curves

exhibiting solar-like oscillations with duty cycles ranging from 0.20 to 0.50. The results indicate that the Gaussian Process method outperforms both the Lomb-Scargle and Inpainting methods in terms of accuracy and stability in recovering the true values. The Inpainting method, in particular, tends to introduce significant false signals when applied to low-duty-cycle data, leading to potential measurement distortions. Therefore, the Gaussian Process method is the preferred choice for analyzing low-duty-cycle data from ground-based telescopes, followed by the Lomb-Scargle method, while the Inpainting method is not recommended.

Key words methods: data analysis, stars: oscillations, stars: solar-type

附录

拟合结果

表 3 Lomb-Scargle方法、Inpainting方法和高斯过程(Gaussian Process, GP)方法对标记为“好”和“差”的数据的拟合结果

Table 3 The fitting results of the Lomb-Scargle method, Inpainting method, and Gaussian process method for the data flagged as “good” and “bad”

Method	Flag	Feature	Duty Cycle	$f/\mu\text{Hz}$	$A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})$	$\gamma/\mu\text{Hz}$
Lomb-Scargle	good	Granulation	0.2	0.121 (2)	234.9 (3.4)	-
			0.3	0.123 (2)	221.0 (3.4)	-
			0.4	0.105 (2)	257.3 (3.4)	-
			0.5	0.104 (1)	296.4 (3.4)	-
		Signal _{left}	0.2	0.378 (6)	91.2 (15.0)	0.059 (17)
			0.3	0.389 (2)	183.0 (62.4)	0.016 (6)
			0.4	0.389 (7)	67.4 (8.1)	0.129 (24)
			0.5	0.418 (4)	153.6 (8.1)	0.117 (8)
		Signal _{central}	0.2	0.600 (3)	165.8 (7.6)	0.108 (8)
			0.3	0.592 (2)	223.4 (10.1)	0.067 (4)
			0.4	0.604 (2)	191.9 (12.2)	0.068 (8)
			0.5	0.598 (1)	266.4 (12.7)	0.052 (4)
		Signal _{right}	0.2	0.813 (7)	73.3 (9.8)	0.077 (13)
			0.3	0.791 (3)	123.7 (9.4)	0.089 (10)
			0.4	0.820 (4)	120.6 (7.9)	0.109 (10)
			0.5	0.812 (2)	213.4 (11.1)	0.071 (6)

表3 续
Table 3 Continued

Method	Flag	Feature	Duty Cycle	$f/\mu\text{Hz}$	$A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})$	$\gamma/\mu\text{Hz}$		
Lomb-Scargle	bad	Granulation	0.2	0.105 (2)	198.9 (3.5)	-		
			0.3	0.106 (1)	320.1 (3.7)	-		
			0.4	0.123 (1)	300.5 (3.2)	-		
			0.5	0.116 (1)	319.1 (3.5)	-		
		Signal _{left}	0.2	0.347 (11)	54.9 (5.6)	0.174 (16)		
			0.3	0.347 (15)	38.8 (6.8)	0.145 (27)		
			0.4	0.494 (5)	76.9 (5.6)	0.194 (4)		
			0.5	0.364 (6)	72.0 (14.6)	0.061 (17)		
		Signal _{central}	0.2	0.587 (3)	153.4 (7.9)	0.110 (8)		
			0.3	0.588 (2)	215.8 (10.7)	0.087 (7)		
			0.4	0.597 (1)	388.7 (9.6)	0.015 (1)		
			0.5	0.601 (3)	193.2 (9.7)	0.120 (10)		
		Signal _{right}	0.2	0.825 (6)	82.6 (7.9)	0.097 (12)		
			0.3	0.780 (8)	69.5 (8.0)	0.143 (27)		
			0.4	0.789 (6)	101.5 (5.5)	0.188 (8)		
			0.5	0.794 (2)	208.0 (20.8)	0.036 (6)		
		Inpainting	good	Granulation	0.2	0.019 (1)	960.2 (9.3)	-
					0.3	0.077 (2)	342.8 (4.9)	-
					0.4	0.070 (1)	440.2 (5.5)	-
					0.5	0.061 (1)	437.2 (5.8)	-
Signal _{left}	0.2			0.304 (3)	46.1 (4.8)	0.197 (2)		
	0.3			0.387 (1)	343.4 (22.3)	0.010 (1)		
	0.4			0.402 (1)	160.3 (34.5)	0.020 (6)		
	0.5			0.410 (5)	109.9 (9.7)	0.105 (12)		
Signal _{central}	0.2			0.621 (1)	391.8 (7.4)	0.010 (1)		
	0.3			0.578 (1)	377.6 (19.1)	0.017 (1)		
	0.4			0.607 (1)	343.7 (23.5)	0.027 (6)		
	0.5			0.607 (1)	258.3 (15.7)	0.035 (3)		
Signal _{right}	0.2			0.829 (1)	363.6 (19.7)	0.010 (1)		
	0.3			0.789 (1)	271.5 (27.9)	0.011 (1)		
	0.4			0.822 (6)	93.1 (12.2)	0.091 (15)		
	0.5			0.806 (3)	147.4 (11.7)	0.070 (8)		

表3 续
Table 3 Continued

Method	Flag	Feature	Duty Cycle	$f/\mu\text{Hz}$	$A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})$	$\gamma/\mu\text{Hz}$		
Inpainting	bad	Granulation	0.2	0.012 (1)	1095.9 (0.4)	-		
			0.3	0.074 (1)	397.3 (5.4)	-		
			0.4	0.078 (1)	403.8 (5.4)	-		
			0.5	0.090 (1)	339.8 (4.2)	-		
		Signal _{left}	0.2	0.301 (1)	80.0 (7.6)	0.199 (1)		
			0.3	0.333 (16)	24.0 (8.6)	0.137 (38)		
			0.4	0.497 (2)	60.9 (6.2)	0.191 (6)		
			0.5	0.470 (26)	29.7 (6.0)	0.179 (15)		
		Signal _{central}	0.2	0.562 (1)	373.5 (8.7)	0.010 (1)		
			0.3	0.573 (3)	227.9 (11.0)	0.077 (5)		
			0.4	0.598 (1)	398.9 (3.0)	0.011 (1)		
			0.5	0.598 (1)	395.8 (5.4)	0.016 (1)		
		Signal _{right}	0.2	0.776 (2)	134.5 (16.0)	0.011 (1)		
			0.3	0.786 (8)	70.9 (13.9)	0.103 (36)		
			0.4	0.800 (6)	98.2 (11.0)	0.127 (24)		
			0.5	0.792 (1)	280.9 (20.2)	0.024 (3)		
		GP	good	Granulation	0.2	0.111 (10)	348.1 (52.2)	-
					0.3	0.110 (9)	367.8 (49.3)	-
					0.4	0.096 (8)	375.8 (47.4)	-
					0.5	0.088 (6)	386.2 (43.9)	-
Signal _{left}	0.2			0.408 (23)	84.3 (52.7)	0.080 (39)		
	0.3			0.402 (9)	201.1 (65.0)	0.048 (16)		
	0.4			0.389 (10)	113.0 (27.6)	0.107 (32)		
	0.5			0.413 (7)	182.7 (34.7)	0.094 (18)		
Signal _{central}	0.2			0.608 (12)	256.0 (46.4)	0.116 (32)		
	0.3			0.596 (6)	316.1 (45.9)	0.058 (10)		
	0.4			0.602 (5)	273.5 (52.9)	0.060 (13)		
	0.5			0.601 (4)	263.9 (50.3)	0.055 (11)		
Signal _{right}	0.2			0.814 (17)	123.9 (49.7)	0.077 (35)		
	0.3			0.794 (8)	209.6 (51.1)	0.083 (21)		
	0.4			0.816 (8)	152.0 (33.8)	0.082 (18)		
	0.5			0.812 (5)	219.9 (45.2)	0.060 (12)		

表3 续
Table 3 Continued

Method	Flag	Feature	Duty Cycle	$f/\mu\text{Hz}$	$A/(\text{ppm}^2 \cdot \mu\text{Hz}^{-1})$	$\gamma/\mu\text{Hz}$
GP	bad	Granulation	0.2	0.094 (12)	361.1 (59.4)	-
			0.3	0.093 (8)	430.0 (59.6)	-
			0.4	0.102 (7)	406.4 (51.1)	-
			0.5	0.106 (7)	358.2 (39.1)	-
		Signal _{left}	0.2	0.386 (22)	109.2 (32.7)	0.132 (36)
			0.3	0.365 (33)	44.2 (25.4)	0.133 (38)
			0.4	0.382 (22)	74.2 (21.7)	0.140 (36)
			0.5	0.372 (17)	58.5 (24.5)	0.098 (46)
		Signal _{central}	0.2	0.596 (11)	223.6 (54.8)	0.102 (36)
			0.3	0.590 (12)	217.2 (36.8)	0.136 (39)
			0.4	0.592 (7)	218.6 (45.0)	0.098 (36)
			0.5	0.597 (8)	195.3 (32.2)	0.127 (31)
		Signal _{right}	0.2	0.810 (13)	114.8 (43.7)	0.079 (32)
			0.3	0.794 (20)	96.4 (22.3)	0.154 (28)
			0.4	0.803 (13)	118.4 (20.7)	0.150 (27)
			0.5	0.807 (8)	124.5 (27.4)	0.105 (34)