arXiv:2301.01412v3 [stat.ME] 9 Feb 2023

一种适用于大规模周期性数据的可扩展高斯过程

Yongxiang Li、Yuting Pu

工业工程与管理系，

中国上海上海交通大学。

Changming Cheng

机械系统与振动国家重点实验室，

中国上海上海交通大学。

Qian Xiao*

统计系，

美国佐治亚州雅典佐治亚大学

摘要

周期高斯过程（PGP）因其高精度，已越来越多地用于对周期性数据建模。然而，计算PGP的似然具有较高的计算复杂度 $\mathcal{O}\left(n^3\right)$ （n 为数据规模），这限制了它的广泛应用。为解决这一问题，我们针对在信号处理应用中常见的网格上采集的大规模周期性数据，提出了一种新型循环PGP（CPGP）模型。该模型将对数似然分解为

*联系方式：Qian Xiao，qian.xiao@uga.edu，赫蒂大道310号，统计系，佐治亚大学，美国佐治亚州雅典，30602。

PGP的两个计算上可扩展的复合对数似然之和，且不涉及任何近似。对于网格观测，计算CPGP的似然仅需$\mathcal{O}\left(p^2\right)$时间（某些特殊情况下为 $\mathcal{O}$（$p$log p）），其中片段长度 p 与 n无关，且远小于 n。我们通过模拟和实际案例研究，展示了CPGP相较于一些最先进方法的优势，尤其是在需要进行周期性估计的应用中。这种新的建模技术能够极大拓展PGP在众多领域的适用性，并使许多以往难以处理的问题得以建模。[1]

关键词： 循环矩阵，复合似然，优化，周期高斯过程，周期信号。

1 引言

周期性数据在科学研究和工业应用中十分常见，例如语音信号（Nielsen 等人，2017年）、心电图（ECG）信号（Chandola 和 Vatsavai，2011年）以及振动信号（Fan 等人，2018年）。这类信号通常通过线性模型进行分析，例如最大似然音高估计（MLPE）方法（Wise 等人，1976年）和抗噪相关（NRC）方法（Li 等人，2021年）；也可通过非线性模型进行分析，例如非线性最小二乘（NLS）方法（Quinn 和 Thomson，1991年；Nielsen 等人，2017年）。这些方法在许多应用中能够提供理想的估计和预测，但在处理信噪比较低的信号时准确性会降低，而且可能需要很长的信号才能抑制强背景噪声的掩蔽效应（Fan 等人，2018年；Li 等人，2021年）。一个关键原因是，这些方法未能恰当地对周期性数据的周期内循环相关性进行建模，即同一周期内任意两点之间的自相关。

周期性的 高斯过程（PGP）近年来得到越来越广泛的应用，尤

[1]本文的补充材料可在线获取。

其适用于对强噪声环境下的信号进行建模。它能够很好地建模循环性周期内相关性，从而显著提升性能（Zhang 等人，2005；Chandola 和 Vatsavai，2011）。例如，Chandola 和 Vatsavai（2011）提出了一种基于 PGP 的变点检测方法，用于周期时间序列的在线监测，其性能优于许多现有方法。Durrande 等人（2016）调整了 PGP 框架，以检测拟南芥基因组中的周期性；Gu´erin 等人（2020）则基于 PGP 滤波开发了一种稳健的目标检测方法。Koulali 和 Clarke（2021）提出使用 PGP 对大地测量时间序列进行建模，以估算所选 GPS 站点的长期速度。PGP 还用于周期过程的长期预测和周期误差控制（HajiGhassemi 和 Deisenroth，2014；Klenske 等人，2015）。

尽管 PGP 在准确性方面具有优势，但仍面临两项挑战，阻碍了它的广泛应用。首先，计算 PGP 的似然需要达到 $\mathcal{O}(n^3)$ 的高复杂度，其中 n 是信号长度。处理中等规模或大规模数据时，这可能会慢得难以接受，例如在普通计算机上处理 10,000 个数据点需要数小时。其次，PGP 通常要求已知周期，或将周期视为调节参数。然而，在实际应用中，真实周期往往未知，而周期性检测通常是人们关注的重点。目前的文献缺乏用于估计 PGP 中周期的高效方法；这是一个涉及大量局部最优点、颇具挑战性的优化问题。

为解决计算方面的挑战，可以采用当前高斯过程文献中的各种近似方法，概括如下。一种策略是简化高斯过程的协方差结构，包括低秩高斯过程（Cressie和Johannesson，2008；Stein，2008）、诱导点高斯过程（Quinonero-Candela和Rasmussen，2005；Titsias，2009）、局部高斯过程（Choudhury等人，2002；Park等人，2011）、协方差截断（Kaufman等人，2008；Bevilacqua等人，2016）以及稀疏高斯过程（Snelson和Ghahramani，2007；Sang和Huang，2012）。另一种策略是通过条件复合似然（Vecchia，1988；Stein等人，2004；Katzfuss和Guinness，

2021年）或分块边际复合似然（Caragea和Smith，2007年；Eidsvik等人，2014年）近似高斯过程的完整似然。这些近似方法均可应用于PGP，但不可避免地会造成一定程度的信息损失，从而为了计算便利而牺牲部分模型准确性。此外，这些方法大多不可扩展，其计算复杂度仍取决于数据规模。详见第2节。

在信号处理应用中，数据通常以给定的采样频率在网格上采集。也就是说，时间序列的间隔相等。处理这类数据时，可以采用基于循环和托普利兹协方差结构的一些代数方法，在不作任何近似的情况下加速高斯过程。嵌入循环矩阵（Wood 和 Chan，1994；Coeurjolly 和 Porcu，2018）是一种常用方法，可用于从高斯过程中快速且精确地模拟数据。然而，PGP 的协方差矩阵可能并非循环矩阵，因为信号长度 n 不一定总是片段长度 p 的倍数。迄今为止，基于托普利兹矩阵的 PGP（TPGP）是文献中计算似然时不涉及近似的最高效方法（Zhang 等人，2005；Chandola 和 Vatsavai，2011）。它可以将 PGP 中 $\mathcal{O}(n^3)$ 的计算复杂度加速至 $\mathcal{O}(n^2)$。不过，在处理大规模数据时，这种方法仍可能慢得难以接受。

为应对经典PGP方法中的两项挑战，我们提出了一种用于网格上采集的大规模周期性数据的新型循环PGP（CPGP）模型。所提出的CPGP将全部数据划分为 k 个片段；若 n 不是 k的倍数，则还会有一部分剩余数据。它将PGP的对数似然分解为两个计算上可扩展的复合对数似然之和，从而将计算似然的计算复杂度从PGP中的$\mathcal{O}(n^3)$加速至CPGP中的 $\mathcal{O}(p^2)$，其中片段长度 p 独立于数据规模 n，且远小于数据规模。当 p 整除 n 时，还可进一步提升至 $\mathcal{O}(p\log\ p)$。请注意，给定周期后，CPGP与PGP具有完全相同的似然，因此与PGP相比，它不涉及任何近似。

不同于需要已知周期的经典PGP方法，所提出的估计

CPGP的方法包含一种定制的优化算法，可在周期事先未知时高效地估计周期。考虑到在CPGP中搜索未知周期的成本，模型拟合的总计算复杂度为 $\mathcal{O}(d^3 p_{max}^3)$，其中 p_{max} 是 p 的最大搜索长度， d 是控制周期估计小数精度的调节参数。此外，我们还表明，CPGP能够提供与PGP中最佳线性无偏预测（BLUP）完全相同的预测，且其计算复杂度为$\mathcal{O}(p^2)$。它明显优于经典复合似然预测器；后者可能会遭受不可忽视的信息损失（Eidsvik等人，2014年）。总之，CPGP中的似然计算、模型拟合和模型预测都具有可扩展性，而且与PGP不同，这些过程无需近似。

当周期性数据是在网格上采集时，只要适用 PGP 模型，就应使用所提出的 CPGP。它使得准确的 PGP 框架能够应用于具有大规模数据或未知周期的实际问题。值得注意的是，我们不主张将 CPGP 用于非网格采集的数据。与一些最先进的非高斯过程方法相比，包括非线性最小二乘（Nielsen 等人，2017）、NRC（Li 等人，2021）和 MLPE（Wise 等人，1976），CPGP 在处理低信噪比数据（如振动信号）时表现更好，同时计算效率也很高。除了能够高效处理长信号外，CPGP 对于快速、准确地检测周期性尤其有用，即利用数量适中的数据，在极短时间内准确识别周期。尽管 CPGP 是为严格周期信号而设计的，也可用于分析伪周期信号，实际案例研究对此进行了说明。此外，我们在模拟研究中讨论了 CPGP 的一个近似版本，该版本速度更快，但总体表现通常不如 CPGP。

本文其余部分安排如下。第2节简要回顾当前关于高斯过程、PGP 以及若干用于建模周期性数据的非高斯过程方法的文献。第3节详细介绍可扩展 CPGP 的似然构造、参数估计和模型预测。第4节通过合成周期信号的模拟研究，

评估 CPGP 的性能。第5节讨论两个实际案例研究，以展示 CPGP 的优越性。第6节总结本研究，并讨论一些后续工作。所有证明和补充技术细节均见补充材料。

2 简要文献综述

高斯过程是一种随机过程，其中由时间或空间索引的任意有限个随机变量都服从多元正态分布。高斯过程模型也称为克里金法，最初由地统计学领域提出（Sacks等人，1989），随后广泛应用于科学与工程的诸多领域（Fang等人，2006；Gramacy，2020）。它能够学习时间和/或空间相关性，并取得理想的性能。从数学上讲，高斯过程模型（即泛克里金法）可定义为

$$y\left(t\right)=\boldsymbol{f}^{T}\left(t\right)\boldsymbol{\beta}+z\left(t\right)+\epsilon\left(t\right), \tag{1}$$

其中，$\boldsymbol{f}(t)$ 是回归函数，$\boldsymbol{\beta}$ 是系数向量，$z(t)$ 是零均值高斯过程，其方差为 σ^2，平稳相关函数为 $\varphi_{\boldsymbol{\phi}}(t,\ t')$（$\boldsymbol{\phi}$ 是该函数中的参数向量），而 $\epsilon(t)\sim\mathcal{N}(0,\ \sigma^2\delta^2)$ 是独立同分布噪声。

高斯过程模型中的参数（$\boldsymbol{\phi}$、$\boldsymbol{\beta}$、σ^2和 δ^2）通常通过最大似然估计（MLE, Santner等人，2013年）获得，即通过最大化以下对数似然函数（忽略常数项）

$$\ell\left(\boldsymbol{\phi},\boldsymbol{\beta},\sigma^{2},\delta\right)=-\frac{\log\left|\sigma^{2}\boldsymbol{K}_{\delta}\right|+\frac{(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta})^{T}\boldsymbol{K}_{\delta}^{-1}(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta})}{\sigma^{2}}}{2}, \tag{2}$$

其中，$\boldsymbol{F}=[\boldsymbol{f}(t_1)\ \ \cdots,\boldsymbol{f}(t_n)]^T$是回归矩阵，而 $\sigma^2\boldsymbol{K}_\delta$是高斯过程的协方差矩阵。此处，协方差 $\boldsymbol{K}_\delta=\boldsymbol{K}+\delta^2\boldsymbol{I}_n$，其中相关矩阵 $\boldsymbol{K}$ 是根据选定的相关函数 $\varphi_{\boldsymbol{\phi}}(t,t)$由 n 个样本点计算得出的。

令式（2）中的对数似然偏导数等于零，可得 $\boldsymbol{\beta}$ 和 σ^2 的参数估计值

$$\begin{cases} \hat{\boldsymbol{\beta}} & = \left(\boldsymbol{F}^T \boldsymbol{K}_\delta^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{F}^T \boldsymbol{K}_\delta^{-1} \boldsymbol{y} \\ \hat{\sigma}^2 & = \frac{1}{n}\left(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}\right)^T \boldsymbol{K}_\delta^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}\right) \end{cases}. \tag{3}$$

然后，我们可以通过一些搜索算法求解最小化问题，得到 $\boldsymbol{\phi}$ 和 δ的最大似然估计的剖面似然：$\hat{\boldsymbol{\phi}}, \hat{\delta} = \text{最小化}_{\boldsymbol{\phi},\ \delta}\{n\log \hat{\sigma}^2 + \log\ |\boldsymbol{K}_\delta|\}$。给定参数估计值后，可以通过最佳线性无偏预测进行模型预测

$$\hat{y}(t) = \boldsymbol{f}^T(t)\hat{\boldsymbol{\beta}} + \boldsymbol{r}^T(t)\boldsymbol{K}_{\hat{\delta}}^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}\right), \tag{4}$$

其中相关向量 $\boldsymbol{r}(t) = \left[\varphi_{\hat{\boldsymbol{\phi}}}(t,\ t_1)\ \cdots,\ \varphi_{\hat{\boldsymbol{\phi}}}(t,\ t_n)\right]^T$。此外，其方差为 $\text{Var}(\hat{y}(t)) = \sigma^2(1 - \boldsymbol{r}^T(t)\ \boldsymbol{K}^{-1}{}_{\hat{\delta}}\ \boldsymbol{r}(t) + \boldsymbol{\omega}^T\left(\boldsymbol{F}^T\boldsymbol{K}_\delta^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{\omega})$，其中 $\boldsymbol{\omega} = \boldsymbol{f}(t) - \boldsymbol{F}^T\boldsymbol{K}^{-1}{}_{\hat{\delta}}\ \boldsymbol{r}(t)$。关于高斯过程推导的更多细节，请参阅 Santner 等人（2013）和 Gramacy（2020）。

许多常用的相关函数，例如高斯核和 Matérn 核，可能并不适合用于建模周期性数据，因为它们没有考虑周期依赖性。在 PGP 文献中，MacKay（1998）提出的周期相关函数得到了广泛应用，其定义如下：

$$\psi_{\boldsymbol{\phi}}(t, t') = \exp\left(-\theta^2 \sin^2\left(\frac{\pi(t - t')}{T}\right)\right), \tag{5}$$

其中，$\boldsymbol{\phi} = \{\theta, T\}$、$\theta$是粗糙度（长度尺度）参数，$T$是周期。式（5）中的相关函数能够很好地建模周期性数据的循环相关结构，并且任意周期内起点和终点之间的相关性趋近于 1。更多讨论和示例参见 Rasmussen 和 Williams（2006）。

模型拟合　　以及对 GP 模型（包括 PGP）的预测，都需要计算

协方差矩阵的逆和行列式，二者的计算复杂度均为 $\mathcal{O}(n^3)$。对于中等规模或大型数据集，这一过程可能非常耗时。在实践中，周期性数据（如语音、心电图和振动信号）通常包含超过 100,000 个数据点，对于如此大的数据规模，拟合经典 GP 或 PGP 几乎是不可能的。

为了加快高斯过程方法，常见策略之一是通过某些近似简化协方差结构。例如，低秩高斯过程（Cressie and Johannesson, 2008; Stein, 2008; Finley et al., 2009; Xiong, 2021）通过构造分解 $\boldsymbol{K} \approx \boldsymbol{Q}_{n\times p}\boldsymbol{K}_{p\times p}^{-1}\boldsymbol{Q}_{n\times p}^{T}$ 来近似相关矩阵 $\boldsymbol{K}$，其中 $\boldsymbol{K}_{p\times p}$ 是由 p 个诱导点构成的相关矩阵（Quinonero-Candela and Rasmussen, 2005; Titsias, 2009）。它可以将高斯过程的计算复杂度降至 $\mathcal{O}(np^2)$。此外，Tebbutt et al.（2016）将 $\boldsymbol{K}_{p\times p}$ 转化为循环矩阵，以进行伪循环高斯过程近似，从而进一步将计算复杂度降至 $\mathcal{O}(np\log p)$。此外，针对大规模高斯过程，还提出了局部近似方法（Choudhury et al., 2002; Park et al., 2011），通过在数据子集上拟合局部高斯过程模型实现近似。协方差截断（Kaufman et al., 2008; Bevilacqua et al., 2016）则通过利用稀疏协方差结构，提升计算效率。稀疏高斯过程（Snelson and Ghahramani, 2007; Sang and Huang, 2012）旨在结合低秩高斯过程与局部高斯过程。另一种策略是采用分而治之方法与复合似然。具体而言，复合似然被用于近似高斯过程的完整似然，以进行参数估计（Vecchia, 1988; Heagerty and Lele, 1998; Stein et al., 2004; Caragea and Smith, 2007）和模型预测（Eidsvik等人, 2014）。Li and Zhou（2016）提出了一种序贯成对建模方法，以实现多变量高斯过程的出色可扩展性。

所有这些 GP 近似方法都可应用于 PGP，但其中大多数尚未在现有文献中实现。一个挑战在于，由于近似处理，这些方法可能造成不可忽略的信息损失，从而影响 PGP 的准确性。在信号处理应用中，周期性数据通常在网格上采集。对于此类数据，嵌入循环矩阵法（Wood 和 Chan，1994；Davies 和 Bryant，2013）

可在不进行任何近似的情况下扩展 GP，其中 GP 的托普利兹协方差矩阵可以嵌入循环协方差矩阵中。计算协方差矩阵的逆矩阵和行列式分别需要 $\mathcal{O}(n\log\ n)$ 和 $\mathcal{O}(n^2)$ 的时间，因此，计算似然的复杂度为 $\mathcal{O}(n^2)$（Zhang 等人，2005；Chandola 和 Vatsavai，2011）。该方法已成功应用于高斯过程模拟（Dietrich 和 Newsam，1997；Coeurjolly 和 Porcu，2018）以及对数高斯 Cox 过程（Diggle 等人，2013；Taylor 和 Diggle，2014）。

具体而言，Chandola 和 Vatsavai（2011）提出的 TPGP 目前是处理在网格上采集的周期性数据最高效的 PGP，即 $t_i = i/f_s$，其中 f_s 是采样频率。其协方差矩阵 $\sigma^2\boldsymbol{K}_\delta$ 被证明对于任意周期 T 都是托普利兹矩阵（对角线元素恒定的矩阵），因为相关矩阵 $\boldsymbol{K}$ 中的任意第 i 个和第 j 个元素

$$\psi_{\boldsymbol{\phi}}\left(t_i, t_j\right) = \exp\left(-\theta^2 \sin^2\left(\frac{\pi\left(i/f_s - j/f_s\right)}{T}\right)\right). \tag{6}$$

显然，$\boldsymbol{K}$ 的所有向下对角线上的元素均为常数，$\boldsymbol{K}_\delta = \boldsymbol{K} + \delta^2\boldsymbol{I}_n$也如此。因此，可以使用托普利兹矩阵求逆算法（Trench，1964），在 $\mathcal{O}(n^2)$ 时间内加速似然的计算。

所提出的 CPGP 与当前的 GP（或 PGP）方法有以下四个方面的不同。首先，与所有 GP（或 PGP）近似方法不同，对于在网格上采集的周期性数据，CPGP 与 PGP 具有完全相同的似然，因此能够保持相同的准确性。其次，CPGP 是一种可扩展的方法，计算似然仅需 $\mathcal{O}(p^2)$ 时间（特殊情况下为 $\mathcal{O}(p\log\ p)$），其中 p 独立于 n，且远小于 n。相比之下，大多数当前的 GP 方法即使经过加速，仍无法扩展，例如，低秩高斯过程和 TPGP 的计算复杂度分别为 $\mathcal{O}(np\log\ p)$ 和 $\mathcal{O}(n^2)$。第三，CPGP 可以处理 n 不是 p 的整数倍的情况，而当前基于循环矩阵的方法无法处理这种情况，尽管它们都利用快速傅里叶变换（FFT）来加速计算。第四，与当前的 PGP

方法通常假设周期已知不同，CPGP 的估计过程包含一种高效优化方法，用于识别未知周期。需要注意的是，除了基于高斯过程的方法之外，我们还将通过数值研究，将所提出的 CPGP 与一些最先进的非高斯过程方法进行比较，包括快速非线性最小二乘（FNLS，Nielsen 等人，2017）、NRC（Li 等人，2021）和 MLPE（Wise 等人，1976）。

3 循环周期高斯过程

在信号处理应用中，大规模周期性数据通常以给定的采样频率在网格上采集。为了构建能够准确建模此类数据的 PGP 框架，我们提出了一种称为 CPGP 的可扩展建模方法，该方法在参数估计和模型预测方面都很快速。按照 Zhang 等人（2005）和 Chandola 和 Vatsavai（2011）中的设定，我们假设周期性数据以采样频率 f_s 在网格上采集，且周期可写为 $T = p/（df_s）$，其中 p 和 d 是互质整数。调节参数 d 用于允许一个周期内包含小数个点（即 p/d 个点），并控制 T 的估计精度。在实践中，用于采集数据的采样频率 f_s 通常是已知的，而信号的真实周期 T 通常未知。将 $T = p/（df_s）$代入式（6），相关函数 $\psi_{\boldsymbol{\phi}}（t_i，t_j）$可写为

$$\psi_{\boldsymbol{\phi}}\left(t_i, t_j\right)=\exp\left(-\theta^2 \sin^2\left(\frac{\pi d\left(i-j\right)}{p}\right)\right), \tag{7}$$

其中 $i，j = 1，……，n$和 $\boldsymbol{\phi} = \{\theta, p, d\}$。

3.1 精确似然分解

在 CPGP 中，我们将网格上采集的周期信号划分为 $k = \lfloor n/p \rfloor$ 段，每段包含 p 个数据点，其中 $\lfloor n/p \rfloor$ 表示不大于 n/p的最大整数。

r段记为 $\boldsymbol{y}_r = \left[y\left(t_{(r-1)p+1}\right), ..., y(t_{rp})\right]^T$，其中 $r = 1$，2，$\cdots$，k；无法组成一段的剩余数据点记为 $\boldsymbol{y}_* = [y(t_{kp+1})$ ……，$y(t_n)]^T$。

CPGP 的一项关键创新在于，它将 PGP 中的对数似然 ℓ 分解为两个计算上可扩展的复合对数似然之和，且不作任何近似。具体而言，通过应用条件概率，我们有

$$\ell(\theta, \delta, p, d) = \ell_1 + \eta\ell_2, \tag{8}$$

其中，$\ell_1 = \log \Pr(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k)$ 是 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k$ 的对数似然，$\ell_2 = \log \Pr(\boldsymbol{y}_* \mid \boldsymbol{y}_1, \cdots, \boldsymbol{y}_k)$ 是在给定 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k$ 的条件下，$\boldsymbol{y}_*$ 的对数似然；若 $n \bmod p = 0$，则 $\eta = 0$，否则为 $\eta = 1$。值得注意的是，此处不应忽略 ℓ_2，因为 CPGP 是 PGP 的精确版本（而非近似版本）。一方面，ℓ_2 是使不同大小的 p 之间的 ℓ_1 具有可比性所必需的，因为 ℓ_1 中的样本量会随 p 变化。另一方面，省略 ℓ_2 可能会导致性能变差，以下示例以及第4节中讨论近似 CPGP 的模拟研究均说明了这一点。

示例 1. 考虑 Fan 等人（2018）中的合成周期信号，其数据规模为$n = 4000$，采样频率为 $f_s = 1$ Hz，信噪比为 -18 dB；详情请参阅第4节。我们使用公式（7）中的相关函数及参数 $\theta = 15$、$\delta = 3$ 和 $d = 1$，对公式（1）中的经典 PGP 进行拟合。图 1 展示了其完整对数似然ℓ 和第一个复合对数似然 ℓ_1 随 p 变化的波形。可以看出，随着 p 变化，ℓ_1会剧烈波动，这是因为ℓ_1中使用的信号长度（即 kp）也随之变化。在图 1 的下图中，我们进一步展示了归一化后的 ℓ_1（即 $\ell_1/(kp)$），它仍然有显著波动。完整似然 ℓ 的最高值位于 $p = 200$（即真实周期）。然而，ℓ_1 或 $\ell_1/(kp)$的最高值却位于大得多的 p 处。显然，这里的第二个复合对数似然 ℓ_2 不应忽略。

将信号划分为 k 个片段 $\boldsymbol{y}_r (r = 1、2、\cdots、k)$ 和剩余的 $\boldsymbol{y}_*$ 后，

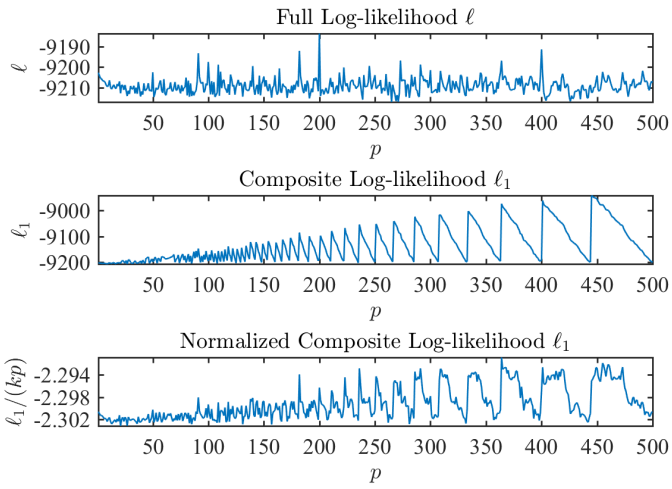


图 1： ℓ 和 ℓ_1 随 p变化的取值。

$\boldsymbol{y}_r$ 与 $\boldsymbol{y}_s$ 之间的协方差矩阵为 $\sigma^2(\boldsymbol{R}+\delta^2\boldsymbol{I}_p)$（当 $r=s$ 时），当 $r\neq s$ 时则为 $\sigma^2\boldsymbol{R}$，其中

$$\boldsymbol{R}=\begin{bmatrix}\psi_{\boldsymbol{\phi}}(t_1,t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_1,t_p)\\ \vdots & \ddots & \vdots\\ \psi_{\boldsymbol{\phi}}(t_p,t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_p,t_p)\end{bmatrix};$$

$\boldsymbol{y}_r$ 和 $\boldsymbol{y}_*$ 之间的协方差矩阵为 $\sigma^2\boldsymbol{R}_\bullet$，其中

$$\boldsymbol{R}_\bullet=\begin{bmatrix}\psi_{\boldsymbol{\phi}}(t_1,t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_1,t_{n-kp})\\ \vdots & \ddots & \vdots\\ \psi_{\boldsymbol{\phi}}(t_p,t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_p,t_{n-kp})\end{bmatrix};$$

$\boldsymbol{y}_*$ 的协方差矩阵为 $\sigma^2(\boldsymbol{R}_*+\delta^2\boldsymbol{I}_{n-kp})$，其中

$$\boldsymbol{R}_*=\begin{bmatrix}\psi_{\boldsymbol{\phi}}(t_1,t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_1,t_{n-kp})\\ \vdots & \ddots & \vdots\\ \psi_{\boldsymbol{\phi}}(t_{n-kp},t_1) & \cdots & \psi_{\boldsymbol{\phi}}(t_{n-kp},t_{n-kp})\end{bmatrix}.$$

接下来，我们得到关于相关矩阵 $\boldsymbol{K}$ 和

矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{R}_*$性质的一个关键结果。

命题 1. 根据式 (7)，相关矩阵 $\boldsymbol{K}$ 可分解为

$$\boldsymbol{K}=\begin{bmatrix} \boldsymbol{R} & \cdots & \boldsymbol{R} & \boldsymbol{R}_{\bullet} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{R} & \cdots & \boldsymbol{R} & \boldsymbol{R}_{\bullet} \\ \boldsymbol{R}_{\bullet}^{T} & \cdots & \boldsymbol{R}_{\bullet}^{T} & \boldsymbol{R}_{*} \end{bmatrix}, \tag{9}$$

其中， $\boldsymbol{R}$ 是对称循环矩阵， $\boldsymbol{R}_*$ 是对称托普利兹矩阵。

片段 $\boldsymbol{\Upsilon}=\left[\boldsymbol{y}_1^T,\cdots,\boldsymbol{y}_k^T\right]^T$的联合分布为

$$\begin{aligned} \boldsymbol{\Upsilon} &\sim \mathcal{N}\left(\boldsymbol{\Gamma}\boldsymbol{\beta},\sigma^2\boldsymbol{\Sigma}\right) \\ &\sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\Gamma}_1\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\Gamma}_k\boldsymbol{\beta} \end{bmatrix},\sigma^2\begin{bmatrix} \boldsymbol{R}+\delta^2\boldsymbol{I}_p & \cdots & \boldsymbol{R} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R} & \cdots & \boldsymbol{R}+\delta^2\boldsymbol{I}_p \end{bmatrix}\right), \end{aligned}$$

其中，对于 $i=1,\ \cdots,\ k,\ \ \boldsymbol{\Gamma}_i=\left[\boldsymbol{f}\left(t_{1+(i-1)p}\right),\cdots,\boldsymbol{f}\ \left(t_{ip}\right)\ \right]^T$。然后，第一个复合对数似然 ℓ_1（忽略常数项）可写为

$$\ell_1=-\frac{1}{2}\left(\log\left|\sigma^2\boldsymbol{\Sigma}\right|+\frac{\left(\boldsymbol{\Upsilon}-\boldsymbol{\Gamma}\boldsymbol{\beta}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Upsilon}-\boldsymbol{\Gamma}\boldsymbol{\beta}\right)}{\sigma^2}\right). \tag{10}$$

给定 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_k,\ \ \boldsymbol{y}_*$的分布为 $\boldsymbol{y}_*\mid\boldsymbol{\Upsilon}\sim\mathcal{N}(\boldsymbol{\mu},\sigma^2\boldsymbol{\Pi})$，其均值为$\boldsymbol{\mu}=\boldsymbol{\Gamma}_*\boldsymbol{\beta}+\boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}($ $\boldsymbol{\Upsilon}-\boldsymbol{\Gamma}\boldsymbol{\beta})$，协方差矩阵为 $\boldsymbol{\Pi}=\boldsymbol{R}_*+\delta^2\boldsymbol{I}_{n-kp}-\boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}$，其中 $\boldsymbol{\Xi}=\left[\boldsymbol{R}_{\bullet}^T\cdots\boldsymbol{R}_{\bullet}^T\right]^T$和 $\boldsymbol{\Gamma}_*=[\boldsymbol{f}(t_{kp+1})\ \ \cdots,\boldsymbol{f}(t_n)]^T$。于是，第二个复合对数似然 ℓ_2 除去某些常数后，可写为

$$\ell_2=-\frac{1}{2}\left(\log\left|\sigma^2\boldsymbol{\Pi}\right|+\frac{\left(\boldsymbol{y}_{\bullet}-\boldsymbol{\Gamma}_{\bullet}\boldsymbol{\beta}\right)^T\boldsymbol{\Pi}^{-1}\left(\boldsymbol{y}_{\bullet}-\boldsymbol{\Gamma}_{\bullet}\boldsymbol{\beta}\right)}{\sigma^2}\right), \tag{11}$$

其中 $\boldsymbol{y}_\bullet = \boldsymbol{y}_* - \boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Upsilon}$和 $\boldsymbol{\Gamma}_\bullet = \boldsymbol{\Gamma}_* - \boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}$。注意，若 $p > n$，则复合对数似然 ℓ_1 化简为零，而复合对数似然 ℓ_2 化简为$\ell_2 = -1/2(\log\ |\sigma^2\boldsymbol{R}_*| + (\boldsymbol{y}_* - \boldsymbol{\Gamma}_*\boldsymbol{\beta})^T\boldsymbol{R}_*^{-1}(\boldsymbol{y}_* - \boldsymbol{\Gamma}_*\boldsymbol{\beta})\ /\sigma^2)$。

3.2 可扩展参数估计

根据公式（10）和（11），ℓ_1 和 ℓ_2 的估计分别需要求出 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Pi}$的行列式与逆矩阵。由于 $\boldsymbol{\Pi} = \boldsymbol{R}_* + \delta^2\boldsymbol{I}_{n-kp} - \boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}$，关键挑战在于计算 $|\boldsymbol{\Sigma}|$ 和 $\boldsymbol{\Sigma}^{-1}$。在命题 2 中，我们证明，利用矩阵求逆引理（Sherman 和 Morrison，1950）和 Sylvester 行列式定理（Akritas 等人，1996），可以大幅简化这两项计算。矩阵求逆引理也称为 Woodbury、Sherman 和 Morrison 公式。

命题 2. $\boldsymbol{\Sigma}$ 的逆矩阵与行列式可以简化为

$$\boldsymbol{\Sigma}^{-1} = \frac{k\boldsymbol{I}_{kp} - \boldsymbol{V}\left(\boldsymbol{I}_p - \boldsymbol{R}_\delta^{-1}\right)\boldsymbol{V}^T}{\delta^2 k}, \tag{12}$$

and

$$|\boldsymbol{\Sigma}| = \delta^{2kp}\left|\boldsymbol{R}_\delta\right|, \tag{13}$$

其中 $\boldsymbol{R}_\delta = \boldsymbol{I}_p + k\boldsymbol{R}/\delta^2$和 $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{I}_p & \cdots & \boldsymbol{I}_p \end{bmatrix}$。

根据命题 1，矩阵 $\boldsymbol{R}$ 是对称循环矩阵。因此，命题 2 中的矩阵 $\boldsymbol{R}_\delta$ 和 $\boldsymbol{R}_\delta^{-1}$ 是对称循环矩阵。随后，可以应用快速傅里叶变换方法（Brigham，1988年），以计算 $\boldsymbol{R}_\delta^{-1}$ 和 $|\boldsymbol{R}_\delta|$，其计算复杂度为 $\mathcal{O}(p\log\ p)$。可使用 MATLAB 工具箱“smt”（Redivo-Zaglia 和 Rodriguez，2012）进行实现。

通过 pl 将公式（12）代入上文定义的 $\boldsymbol{y}_\bullet$、$\boldsymbol{\Gamma}_\bullet$ 和 $\boldsymbol{\Pi}$ 的公式中，我们得到 e

$$\begin{cases} \boldsymbol{y}_\bullet & = \boldsymbol{y}_* - \frac{k}{\delta^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{y}} \\ \boldsymbol{\Gamma}_\bullet & = \boldsymbol{\Gamma}_* - \frac{k}{\delta^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{\Gamma}} \\ \boldsymbol{\Pi} & = \boldsymbol{R}_* + \delta^2\boldsymbol{I}_{n-kp} - \frac{k}{\delta^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_\delta^{-1}\boldsymbol{R}_\bullet \end{cases},$$

其中 $\bar{\boldsymbol{y}} = (\boldsymbol{y}_1 + \cdots + \boldsymbol{y}_k)/k$ 和 $\bar{\boldsymbol{\Gamma}} = (\boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\Gamma}_k)/k$。这里，$\boldsymbol{R}^T\boldsymbol{R}_\delta^{-1}\boldsymbol{R}$ 是一个对称循环矩阵，因为循环矩阵的乘积仍是循环矩阵。矩阵 $\boldsymbol{R}_\bullet^T\boldsymbol{R}_\delta^{-1}\boldsymbol{R}_\bullet$（$\boldsymbol{R}^T\boldsymbol{R}_\delta^{-1}\boldsymbol{R}$ 的一个子矩阵）是一个对称托普利兹矩阵，$\boldsymbol{\Pi}$ 也一样。接着，由于只需通过 FFT 计算 $\boldsymbol{\Pi}$ 的首行/首列，因此可以在 $\mathcal{O}(p\log\ p)$ 时间内计算出 $\boldsymbol{\Pi}$。我们可以使用半正定 Toeplitz 矩阵的 Cholesky 分解算法（Stewart，1997年）计算 $\boldsymbol{\Pi}^{-1}$ 和 $|\boldsymbol{\Pi}|$，所需时间为 $\mathcal{O}\left((n-kp)^2\right)$。很容易证明，这一复杂度始终小于 $\mathcal{O}(p^2)$。请注意，我们还可以通过将 $\boldsymbol{\Pi}$ 嵌入某个循环矩阵中，在 $\mathcal{O}(p\log\ p)$ 时间内进一步加速 $\boldsymbol{\Pi}^{-1}$ 的计算，但 $|\boldsymbol{\Pi}|$ 的计算无法进一步改进。因此，嵌入循环矩阵法无法进一步降低计算 ℓ_2（同时涉及 $\boldsymbol{\Pi}^{-1}$ 和 $|\boldsymbol{\Pi}|$）的复杂度。这里我们仅考虑 Cholesky 分解算法。

通过求解 $\partial\ell/\partial\boldsymbol{\beta} = 0$ 和 $\partial\ell/\partial\sigma^2 = 0$，我们得到 $\boldsymbol{\beta}$ 和 σ^2的估计值

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{S}_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}} + \eta\boldsymbol{\Gamma}_\bullet^T\boldsymbol{\Pi}^{-1}\boldsymbol{\Gamma}_\bullet\right)^{-1}\left(\boldsymbol{S}_{\boldsymbol{\Gamma}\boldsymbol{Y}} + \eta\boldsymbol{\Gamma}_\bullet^T\boldsymbol{\Pi}^{-1}\boldsymbol{y}_\bullet\right), \tag{14}$$

and

$$\hat{\sigma}^2 = \frac{\left(\boldsymbol{S}_{\boldsymbol{Y}\boldsymbol{Y}} + \eta\boldsymbol{y}_\bullet^T\boldsymbol{\Pi}^{-1}\boldsymbol{y}_\bullet\right) - \hat{\boldsymbol{\beta}}^T\left(\boldsymbol{S}_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}} + \eta\boldsymbol{\Gamma}_\bullet^T\boldsymbol{\Pi}^{-1}\boldsymbol{\Gamma}_\bullet\right)\hat{\boldsymbol{\beta}}}{n}, \tag{15}$$

其中

$$\begin{cases} \boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}} & = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma} = \frac{\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} - k\bar{\boldsymbol{\Gamma}}^T\bar{\boldsymbol{\Gamma}} + k\bar{\boldsymbol{\Gamma}}^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{\Gamma}}}{\delta^2} \\ \boldsymbol{S}_{\boldsymbol{\Gamma Y}} & = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \frac{\boldsymbol{\Gamma}^T\boldsymbol{Y} - k\bar{\boldsymbol{\Gamma}}^T\bar{\boldsymbol{y}} + k\bar{\boldsymbol{\Gamma}}^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{y}}}{\delta^2} \\ \boldsymbol{S}_{\boldsymbol{YY}} & = \boldsymbol{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \frac{\boldsymbol{Y}^T\boldsymbol{Y} - k\bar{\boldsymbol{y}}^T\bar{\boldsymbol{y}} + k\bar{\boldsymbol{y}}^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{y}}}{\delta^2} \end{cases}.$$

具体而言，如果 $\boldsymbol{\Gamma}_1 = \cdots = \boldsymbol{\Gamma}_k$, $\boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}} = k\bar{\boldsymbol{\Gamma}}^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{\Gamma}}/\delta^2$和 $\boldsymbol{S}_{\boldsymbol{\Gamma Y}} = k\bar{\boldsymbol{\Gamma}}^T\boldsymbol{R}_\delta^{-1}\bar{\boldsymbol{y}}/\delta^2$。此处，计算 $\hat{}\,\boldsymbol{\beta}$ 和 $\hat{}\sigma^2$ 需要计算 $\boldsymbol{R}_\delta^{-1}$ 和 $\boldsymbol{\Pi}^{-1}$，其维数均不超过 p。这表明，$\hat{}\,\boldsymbol{\beta}$ 和 $\hat{}\sigma^2$ 的估计具有可扩展性。

将式（13）、（14）和（15）代入式（10）和（11），式（8）中的对数似然$\ell(\theta, \delta, p, d)$可写为

$$\hat{\ell}(\theta, \delta, p, d) = -\frac{\log \hat{\sigma}^{2n}\delta^{2kp} + \log|\boldsymbol{R}_\delta|\,|\boldsymbol{\Pi}|^\eta + n}{2}. \tag{16}$$

$\hat{}$此处，计算 $\ell(\theta, \delta, p, d)$需要计算 $\boldsymbol{R}_\delta$和 $\boldsymbol{\Pi}$，分别需要 $\mathcal{O}(p\log\ p)$ 和 $\mathcal{O}(p^2)$ 的时间。因此，一般而言，计算 CPGP 的似然需要 $\mathcal{O}(p^2)$ 的计算复杂度。当 p 整除 n（$\ell = \ell_1$ 和 $\ell_2 = 0$）时，计算 $\hat{}\ell(\theta, \delta, p, d)$只需计算 $\boldsymbol{\Sigma}^{-1}$和 $|\boldsymbol{\Sigma}|$。根据命题 2，这种特殊情形只需 $\mathcal{O}(p\log\ p)$ 的时间。相比之下，PGP 和 TPGP 的复杂度分别为 $\mathcal{O}(n^3)$ 和 $\mathcal{O}(n^2)$。接下来，我们通过一个示例说明 PGP、TPGP 和 CPGP 计算似然的速度。

例 2. 考虑例 1 中的合成周期信号，其长度 n 从 1,000 到 500,000 不等。我们实现了 PGP、TPGP 和 CPGP 模型，用于评估参数为 $\theta = 15$和 $\delta = 10$时的对数似然 $\ell\,\hat{}$。100 次试验的平均计算时间如图 2 所示。与 PGP 和 TPGP 不同，所提出的 CPGP 的计算时间不受信号长度影响。此外，对于较长的信号，CPGP 的速度显著快于 PGP 和 TPGP。另外，我们可以看到，周期 $T = 100$ 预先确定的 CPGP（即 $p = 100$ 和 $d = 1$）比周期为 $T = 11$ 的 CPGP（即 $p = 11$ 和

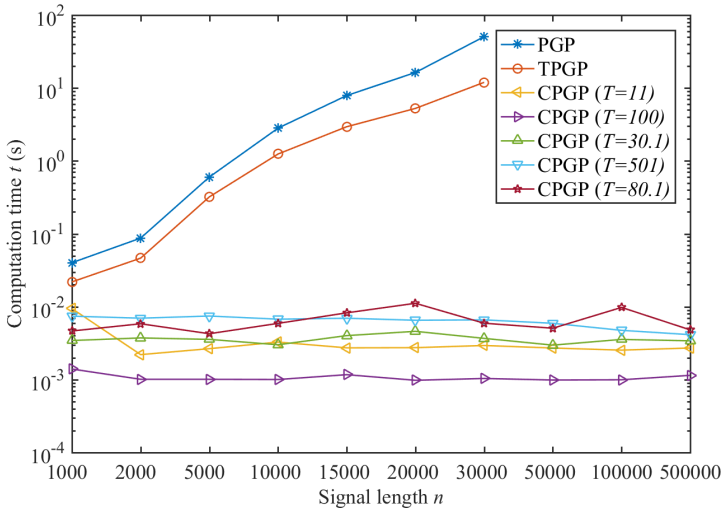


图 2：PGP、TPGP 和 CPGP 的平均计算时间。

$d = 1$）更快，因为当 p 能整除 n时，其复杂度通常会从 $\mathcal{O}(p^2)$ 降至 $\mathcal{O}(p\log\ p)$。CPGP 还可以处理小数周期，如 $T = 30.1$ （$p = 301$ 和$d = 10$）以及 $T = 80.1$ （$p = 801$ 和 $d = 10$）的情况所示。

在实际应用中，用于收集数据的采样频率 f_s 通常预先指定，而周期 $T = p/(df_s)$ 往往未知（片段长度 p也同样未知）。如图 1 所示，似然关于 p通常存在许多局部最优点，因此估计 p 并非易事。本文提出，在应用搜索算法估计 θ 和 δ^2之前，先在 $\mathbf{I} = \{1, 2, \cdots, d_* p_{max}\}$ 中扫描 p，以定位方程 (16) 中 $\ell(\theta, \delta, p, d_*)$ 的最大值 ˆvalue；其中，d_* 是一个调节参数，满足 $1 \leq d_* \leq d$，以节省计算。也就是说，我们将首先在 $\mathbf{I}$ 中扫描 p，以定位最大对数似然 $\hat{\ell}_{\boldsymbol{\mathcal{I}}}(\theta, \delta, d_*) = \max_{p \in \boldsymbol{\mathcal{I}}} \ell(\theta, \delta, p, d_*)$，然后通过某些搜索算法优化 $\ell_{\boldsymbol{\mathcal{I}}}(\theta, \delta, d_*)$，以估计 θ 和 δ^2，即，

$$\hat{\theta}, \hat{\delta} = \text{maximize}_{\theta,\delta} \left\{ \hat{\ell}_{\boldsymbol{\mathcal{I}}}(\theta, \delta, d_*) \right\}. \tag{17}$$

此处可以使用许多启发式优化算法，例如梯度下降算法；我们采用了 MATLAB

工具箱“dace”中实现的 Hooke-Jeeves 搜索算法（Lophaven 等，2002）。

将式（17）中优化后的 θ 和 δ 代入式（16）中的 $\ell(\theta,\ \delta,\ p,\ d)$，可得（给定 p 和 d）

$$\hat{\ell}\left(\hat{\theta},\hat{\delta},p,d\right)=-\frac{\log\hat{\sigma}^{2n}\hat{\delta}^{2kp}+\log\left|\hat{\boldsymbol{R}}_{\hat{\delta}}\right|\left|\hat{\boldsymbol{\Pi}}\right|^{\eta}+n}{2},\tag{18}$$

其中，$\boldsymbol{R}_{\hat{\delta}}$ 和 $\boldsymbol{\Pi}$ 可通过将 θ 和 δ 代入 $\boldsymbol{R}_{\delta}$ 和 $\boldsymbol{\Pi}$获得。式（18）中 $\hat{\ell}\left(\hat{\theta},\ \hat{\delta},\ p,\ d\right)$ 的波形峰值位于真实周期 $\hat{T}=\hat{p}/(df_s)$ 及其倍数处；示例见图 1。由于存在大量局部最优点，为了进行优化，需要在 $\boldsymbol{\mathcal{T}}$中枚举p ，即：

$$\hat{p}=\arg\max_{p\in\boldsymbol{\mathcal{T}}}\left\{\hat{\ell}\left(\hat{\theta},\hat{\delta},p,d\right)\right\},\tag{19}$$

其中，$\boldsymbol{\mathcal{T}}=\{1,2,\cdots,dp_{max}\}$ 是 p的搜索范围。具体而言，如果 $d=d_*$，则$\boldsymbol{\mathcal{I}}=\boldsymbol{\mathcal{T}}$。有关式（17）和式（19）中所提出优化方法的详细说明和伪代码，请参阅补充材料。如果适用，可采用并行或分布式方式计算 p 的枚举步骤，以进一步缩短计算时间。

$\boldsymbol{\mathcal{I}}$ 和 $\boldsymbol{\mathcal{T}}$的设置取决于所需的估计精度与可用计算预算之间的权衡。当采样频率 f_s 相当高时（也就是说，一个周期内的采样点数足够多，例如超过 100），设置 $d=d_*=1$（即 $\boldsymbol{\mathcal{T}}=\boldsymbol{\mathcal{I}}$）便已足够，这将得到整数周期估计。否则，我们可以设置 $d>1$ 和 $d\geq d_*\geq 1$ ，以便估计出带小数位的周期。实证研究表明，$d=d_*$ 并非总是必需的，而在我们所有的数值研究中，$d_*=1$ 便已足够。由于处理小数周期时，p 的搜索范围通常会比处理整数周期时包含更多元素，因此计算小数周期通常需要更多时间；据此，常见的建议取值包括 $d_*=1,d=5$和 $d_*=1,d=10$。$\boldsymbol{\mathcal{I}}$ 和 $\boldsymbol{\mathcal{T}}$ 中的调节参数 p_{max} 通常根据对信号的先验知识来选择

使得 p_{max}/f_s 至少大于真实周期。例如，由于浊音语音的基频大约在 80 Hz 到 300 Hz 之间，我们应设置 p_{max} ，使得 f_s/p_{max} 小于 80 Hz。此外，我们可以根据似然函数的波形来猜测真实周期 T （因此也能猜测 p_{max}）：逐渐增大 p_{max}，直到似然函数呈现出近似周期性的局部最优点。

值得指出的是，通常情况下，计算 CPGP 的似然所需的计算复杂度为 $\mathcal{O}(p^2)$；当 p 能整除 n 时，所需的计算复杂度为 $\mathcal{O}(p\log\ p)$。考虑周期估计时，CPGP 的总计算复杂度为 $\mathcal{O}(d^3p_{max}^3)$。由于实际上 p 和 p_{max} 均独立于 n ，且远小于 n ，因此 CPGP 仍具有可扩展性。在本研究中，我们更强调计算似然的复杂度，因为我们希望公平比较 CPGP 与当前假设周期已知的 PGP 模型。尽管可以将 CPGP 中提出的周期估计嵌入 PGP，但实际上并不可行，因为式（19）中的优化通常需要数千次评估似然函数，而 PGP 和 TPGP 的运行速度都会慢到难以接受。

3.3 可扩展模型预测

正如 Jones 等人（1998）所示， $y(t)$ 的最佳线性无偏预测可通过最大化关于 $y(t)$ 的联合似然 $y(t)$ 和 $\boldsymbol{y}$ 得到，即 $\hat{}y(t) = \arg\max_{y(t)} \Pr(y(t) \quad \boldsymbol{y})$。与第 3.1 节中的似然分解类似，联合分布 $\Pr(y(t) \quad \boldsymbol{y})$ 可分解为两部分

$$\Pr\left(y\left(t\right),\boldsymbol{y}\right)=\Pr\left(y\left(t\right),\boldsymbol{y}_{*}|\boldsymbol{\Upsilon}\right)\Pr\left(\boldsymbol{\Upsilon}\right),$$

其中，$\mathrm{Pr}\left(y(t)\ \ \boldsymbol{y}_{*}|\boldsymbol{\Upsilon}\right)$ 是在给定 $\boldsymbol{\Upsilon}$ 的条件下，$\{y(t)\ \ \boldsymbol{y}_{*}\}$ 的联合分布。因此，我们有

$$\hat{y}\left(t\right)=\arg\max_{y(t)}\mathrm{Pr}\left(y\left(t\right),\boldsymbol{y}_{*}|\boldsymbol{\Upsilon}\right),$$

且 $\mathrm{Pr}\left(y(t)\ \ \boldsymbol{y}_{*}|\boldsymbol{\Upsilon}\right)$ 可进一步写为

$$\begin{bmatrix}y\left(t\right)\\ \boldsymbol{y}_{*}\end{bmatrix}\sim\left(\begin{bmatrix}\mu\left(t\right)\\ \boldsymbol{\mu}\end{bmatrix},\sigma^{2}\begin{bmatrix}\pi\left(t\right) & \boldsymbol{\gamma}_{\bullet}^{T}\left(t\right)\\ \boldsymbol{\gamma}_{\bullet}\left(t\right) & \boldsymbol{\Pi}\end{bmatrix}\right),$$

其中

$$\begin{cases}\mu\left(t\right) & =\boldsymbol{f}^{T}\left(t\right)\boldsymbol{\beta}+\boldsymbol{\gamma}^{T}\left(t\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Upsilon}-\boldsymbol{\Gamma}\right)\\ \pi\left(t\right) & =1+\delta^{2}-\boldsymbol{\gamma}^{T}\left(t\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\left(t\right)\\ \boldsymbol{\gamma}_{\bullet}\left(t\right) & =\boldsymbol{\gamma}_{*}\left(t\right)-\boldsymbol{\Xi}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\left(t\right)\end{cases}. \tag{20}$$

这里，$\boldsymbol{\gamma}(t)$ 和 $\boldsymbol{\gamma}_{*}(t)$ 分别表示 $y(t)$ 与 $\boldsymbol{\Upsilon}$ 之间的相关性，以及 $y(t)$ 与 $\boldsymbol{y}_{*}$ 之间的相关性。因此，给定 $\boldsymbol{\beta}$、σ、θ、δ，和 p，我们得到预测 $\hat{}y(t)=\mu(t)+\boldsymbol{\gamma}_{\bullet}^{T}(t)\ \boldsymbol{\Pi}^{-1}(\boldsymbol{y}_{*}-\boldsymbol{\mu})$ 及其方差 $\mathrm{Var}(\hat{}y(t))=\sigma^{2}\left(\pi(t)\ -\boldsymbol{\gamma}_{\bullet}^{T}(t)\ \boldsymbol{\Pi}^{-1}\boldsymbol{\gamma}_{\bullet}(t)\right)$。

代入公式（12）后，公式（20）可简化为

$$\begin{cases}\mu\left(t\right) & =\boldsymbol{f}^{T}\left(t\right)\boldsymbol{\beta}+\frac{k}{\delta^{2}}\bar{\boldsymbol{\gamma}}^{T}\left(t\right)\boldsymbol{R}_{\delta}^{-1}\left(\bar{\boldsymbol{y}}-\bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}\right)\\ \pi\left(t\right) & =1+\delta^{2}-\frac{k}{\delta^{2}}\bar{\boldsymbol{\gamma}}^{T}\left(t\right)\boldsymbol{R}_{\delta}^{-1}\bar{\boldsymbol{\gamma}}\left(t\right)\\ \boldsymbol{\gamma}_{\bullet}\left(t\right) & =\boldsymbol{\gamma}_{*}\left(t\right)-\frac{k}{\delta^{2}}\boldsymbol{R}_{\bullet}^{T}\boldsymbol{R}_{\delta}^{-1}\bar{\boldsymbol{\gamma}}\left(t\right)\end{cases},$$

$\hat{}$，其中 $\bar{\boldsymbol{\gamma}}(t)$ 表示 $y(t)$ 与 $\boldsymbol{y}_{1}$ 之间的相关性。因此，将 $\boldsymbol{\beta}$ 替换为 $\boldsymbol{\beta}$ 后，我们可以得到一种计算效率较高的最佳线性无偏预测公式

$$\begin{aligned}\hat{y}\left(t\right)=\boldsymbol{f}^{T}\left(t\right)\hat{\boldsymbol{\beta}}+\frac{k}{\delta^{2}}\bar{\boldsymbol{\gamma}}^{T}\left(t\right)\boldsymbol{R}_{\delta}^{-1}\left(\bar{\boldsymbol{y}}-\bar{\boldsymbol{\Gamma}}\hat{\boldsymbol{\beta}}\right)\\ +\boldsymbol{\gamma}_{\bullet}^{T}\left(t\right)\boldsymbol{\Pi}^{-1}\left(\boldsymbol{y}_{\bullet}-\boldsymbol{\Gamma}_{\bullet}\hat{\boldsymbol{\beta}}\right).\end{aligned} \tag{21}$$

公式（14）意味着 Var(ˆ$\boldsymbol{\beta}$) $= \sigma^2 \left(\boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}} + \eta \boldsymbol{\Gamma}_{\bullet}^T \boldsymbol{\Pi}^{-1} \boldsymbol{\Gamma}_{\bullet}\right)^{-1}$，因此，考虑到 $\boldsymbol{\beta}$ 后，ˆ ˆ$y(t)$ 的方差为

$$\begin{aligned} \operatorname{Var}\left(\hat{y}\left(t\right)\right) = \sigma^2 \left(\pi\left(t\right) - \boldsymbol{\gamma}_{\bullet}^T\left(t\right) \boldsymbol{\Pi}^{-1} \boldsymbol{\gamma}_{\bullet}\left(t\right)\right) \\ + \sigma^2 \boldsymbol{\omega}^T \left(\boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}} + \eta \boldsymbol{\Gamma}_{\bullet}^T \boldsymbol{\Pi}^{-1} \boldsymbol{\Gamma}_{\bullet}\right)^{-1} \boldsymbol{\omega}, \end{aligned}$$

其中 $\boldsymbol{\omega} = \boldsymbol{f}(t) - \frac{k}{\delta^2} \bar{\boldsymbol{\Gamma}}^T \boldsymbol{R}_{\delta}^{-1} \bar{\boldsymbol{\gamma}}(t) - \boldsymbol{\Gamma}_{\bullet}^T \boldsymbol{\Pi}^{-1} \boldsymbol{\gamma}_{\bullet}(t)$。

与上述方法类似，可应用用于循环矩阵的 MATLAB 工具箱“smt”（ Redivo-Zaglia 和 Rodriguez，2012）以及用于半正定 Toeplitz 矩阵的 Cholesky 分解算法（Stewart，1997）来计算该最佳线性无偏预测，其计算复杂度为 $\mathcal{O}(p^2)$。

4 模拟研究

PGP 框架主要用于周期性检测应用。这里，我们开展一项常见的周期性检测模拟，以评估所提出的 CPGP 的有效性。我们将其与一些最先进的方法进行比较，包括 TPGP（ Zhang 等人，2005）、FNLS（Nielsen 等人，2017）、NRC（Li 等人，2021）和 MLPE（Wise 等人，1976）。为进行公平比较，式（17）和（19）中提出的优化方法也应用于 TPGP，以实现周期估计。

此外，我们还考虑了 CPGP 的一种变体，称为近似 CPGP（ACPGP），以研究 ℓ_2 在（8）中的重要性。ACPGP 的似然仅包括归一化的 ℓ_1，即:

$$\ell\left(\theta, \delta, p, d\right) = \frac{1}{kp} \ell_1. \tag{22}$$

在 ACPGP 中， $\boldsymbol{\beta}$和 σ^2的最大似然估计分别为 ˆ$\boldsymbol{\beta} = \boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}}^{-1} \boldsymbol{S}_{\boldsymbol{\Gamma Y}}$和 ˆ$\sigma^2 = \frac{1}{n} \left(\boldsymbol{S}_{\boldsymbol{YY}} - \hat{\boldsymbol{\beta}}^T \boldsymbol{S}_{\boldsymbol{\Gamma\Gamma}} \hat{\boldsymbol{\beta}}\right)$，

与公式 (18) 类似，我们有 $\hat{}_\ell\left(\hat{}\theta,\ \hat{}\delta,\ \ p,\ \ d\right) = -\frac{1}{2kp}\left(\log \hat{}\sigma^{2kp}\hat{}\delta^{2kp} + \log\left|\hat{}\boldsymbol{R}_{\hat{\delta}}\right| + kp\right)$，并且可以使用公式 (17) 和 (19) 中相同的优化算法进行参数估计。值得注意的是，计算 ACPGP 的似然需要 $\mathcal{O}(p\log\ p)$ 时间，通常比 CPGP 更快，但代价是牺牲一定的准确性。

考虑 Fan 等人（2018）中广泛讨论的一项关于合成信号的模拟研究，这些周期瞬态信号由以下方程生成：

$$x\left(t\right) = \sum_{i=0}^{\left\lfloor \frac{l}{T_0} \right\rfloor} e^{\frac{-\zeta 2\pi\omega(t-iT_0)^2}{\sqrt{1-\zeta^2}}} \sin 2\pi\omega\left(t - iT_0\right), \tag{23}$$

其中，$\zeta = 0.01$ 为阻尼比，$\omega = 0.055$ Hz 为固有频率，周期为$T_0 = 200$ 秒，l 为信号的时长。信号 $x(t)$ 的采样频率为 $f_s = 1$ Hz，因此每个周期包含 200 个信号点。受噪声污染的信号 $y(t)$（即模拟数据）通过向周期信号 $x(t)$ 添加高斯白噪声生成。这里关注的是周期性检测。

在 CPGP 和 ACPGP 中，由于每个周期包含 200 个点，p的搜索范围设为 $[1, 500],$，即 $\boldsymbol{\mathcal{I}} = \boldsymbol{\mathcal{T}} = \{1, 2, \cdots, 500\}$、$p_{max} = 500$和 $d_* = d = 1$。参数 δ和 θ的搜索范围分别设为 $[2, 20]$ 和 $[1, 30],$。在优化算法中，它们的初始值设为二维搜索空间 $[2, 20] \times [1, 30]$中九个网格候选点里 $\ell_{\boldsymbol{\mathcal{I}}}(\theta, \delta, d_*)$ 的最大ˆ值所对应的值。通过简单的网格搜索选取最佳初始值，可以提高 CPGP 的稳健性，尤其是在缺乏先验信息（领域知识）时，因为似然函数可能存在许多误导性的局部最优点。为简便起见，本文在 CPGP 中使用$\boldsymbol{f}(t) = 1$ 。在本研究中，CPGP 和 TPGP 使用相同的初始参数设置，因此得到相同的参数估计。二者的区别在于计算时间，CPGP 比 TPGP 快得多。在 FNLS 中，最大谐波数设为 $L = 30$。我们重复分析 100 次，每次生成的信号均相互独立且随机。

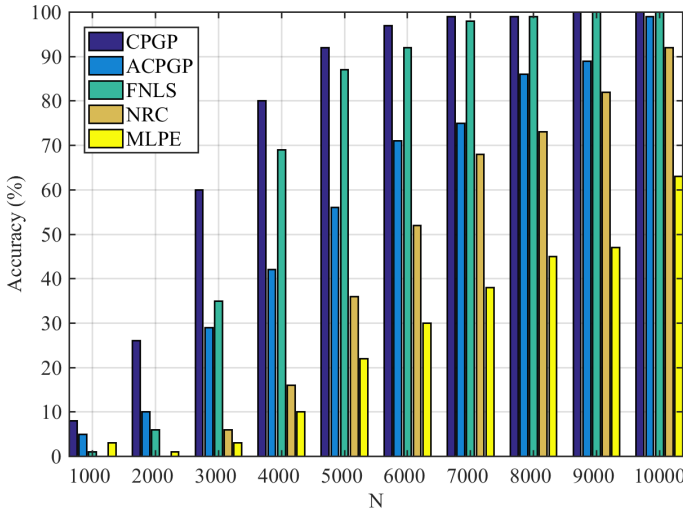


图 3：模拟研究中，CPGP、ACPGP、FNLS、NRC 和 MLPE 的周期估计准确率随信号长度的变化。

ˆ 在图 3 中，我们展示了各方法处理不同信号长度（从 1000 到 10000）时的周期估计准确率（即 100 次试验中 $T = T_0$ 的概率），其中信噪比为 -21 dB。在图 4 中，我们展示了各方法处理不同信噪比（从 -25 dB 到 -11 dB）的信号时的周期估计准确率，其中信号长度为 5000。值得注意的是，由于 TPGP 与 CPGP 的周期估计结果相同，因此图中未展示 TPGP 的结果。二者的区别在于计算时间，见表1。

在图 3 和图 4 中，与 CPGP 相比，ACPGP 在所有情况下的周期估计准确率都较低，这表明第二个复合似然 ℓ_2 在式 (8) 中可能发挥着重要作用。可以看出，ACPGP 仅在信号较长且信噪比较高时表现较好，在其他情况下表现不佳。实际应用中，鉴于 CPGP 的高准确率和稳健性，我们只推荐使用 CPGP 进行周期检测。如果我们预先知道信号长度将会非常长且信噪比足够高，则可以考虑使用速度更快的 ACPGP。

此外，从图 3 可以看出，对于所有信号长度，CPGP 的准确率都显著高于 FNLS、NRC 和 MLPE。随着信号长度增加，所有方法最终都能达到 100% 的准确率，而 CPGP 的收敛速度明显更快。显然，

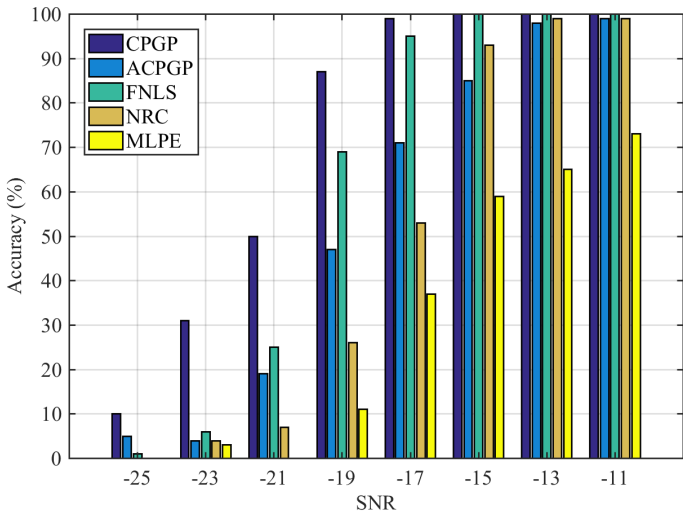


图4：模拟研究中，CPGP、ACPGP、FNLS、NRC 和 MLPE 的周期估计准确率随信噪比变化。

CPGP 的表现优于其他方法，能够在早期准确检测周期性。此外，从图 4 可以看出，当信噪比较低时，CPGP 的准确率显著高于其他方法；随着信噪比增加，所有方法的准确率都会提高。总之，与 FNLS、NRC 和 MLPE 相比，CPGP 使用更少的信号点，就能更有效地处理信噪比较低的信号。

值得注意的是，NRC 和 MLPE 的准确性较低，主要是因为它们没有考虑周期内相关性，而 CPGP 对此进行了精细建模。与使用多个傅里叶基来近似真实信号的 FNLS 不同，CPGP 采用非参数方式表示信号，并优化粗糙度参数 θ，以自适应地对周期内相关性建模。如图 5 所示，PGP（以及 CPGP）的相关函数比 FNLS 的更灵活。FNLS 中的大多数相关性都很小，相关矩阵中的对角元素占主导地位，远大于非对角元素。这些非对角相关元素大多为负值，尤其是在 L较大时。然而，对于两个相距较远的点，假定它们之间的相关性为零（或接近零）应该更合理、更灵活。此外，在给定调节参数 L 的情况下，FNLS 协方差矩阵中这些负元素的位置是固定的（尽管 L 可以

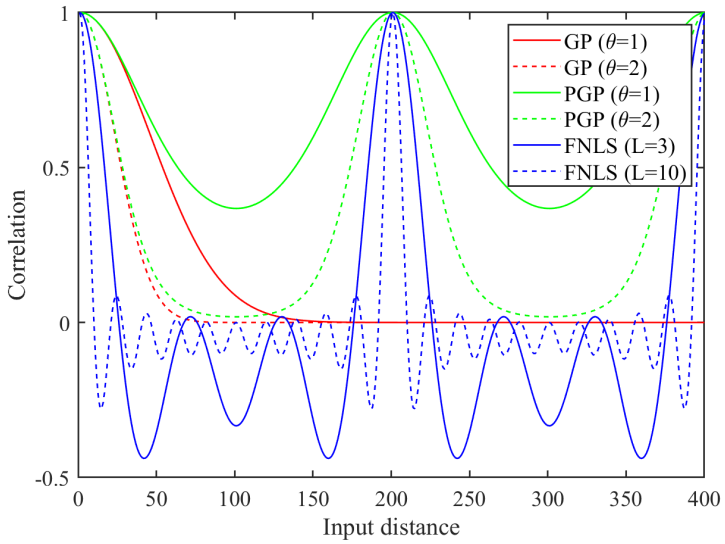


图5：高斯过程、PGP（CPGP）和 FNLS 的相关函数

通过模型选择准则进行优化）。显然，FNLS 中这样的相关结构可能无法有效利用周期性数据的空间或时间相关性。CPGP 的优越性能主要归因于它能够对周期性数据的周期内循环相关性建模。

表1报告了各方法处理不同信号长度时，100 次试验的平均计算时间。其中，TPGP 最慢；即使对长度适中的信号（例如 $n = 5000$）进行建模，也可能需要数小时。值得注意的是，TPGP 比经典周期高斯过程模型快得多，与第2节讨论的采用近似方法的其他基于周期高斯过程的方法相比，速度也相当快。相比之下，本文提出的 CPGP 比 TPGP 快得多，即使处理非常长的信号也只需约 18 秒。如表1所示，CPGP 和 ACPGP 的计算时间不会随信号长度增加而增加，这表明 CPGP 和 ACPGP 的计算复杂度具有可扩展性。与 CPGP 相比，ACPGP 速度更快，但精度较低。

在表1中，尽管 FNLS 处理短信号所需的时间比 CPGP 少，但处理长信号（例如 $n \geq 10,000$）时所需的时间更长。这是因为其计算复杂度为 $\mathcal{O}(nL)$，并且会随信号长度增长。与 CPGP 和 FNLS 相比，NRC 和

表 1：100 次试验的平均计算时间（单位：秒）。

n	TPGP	CPGP	ACPGP	FNLS	NRC	MLPE
1000	189.91	18.744	5.9217	0.6314	0.0117	0.0198
2000	897.01	18.960	5.6760	1.6991	0.0113	0.0216
3000	1771.5	18.809	5.7584	3.6112	0.0139	0.0247
4000	3167.2	19.308	6.0763	5.6430	0.0177	0.0304
5000	4899.6	19.634	5.9567	7.6847	0.0199	0.0328
6000	6129.8	18.307	5.9955	10.418	0.0216	0.0343
7000	8324.0	18.056	5.9258	11.620	0.0239	0.0359
8000	10878	18.172	6.0060	14.971	0.0265	0.0399
9000	13848	17.935	6.0506	15.912	0.0313	0.0453
10000	17076	18.337	6.1132	20.098	0.0341	0.0488

MLPE 的速度更快，但准确度低得多（如图3和图4所示），因为它们忽略了周期内相关性。

5 真实案例研究

5.1 轴承故障检测

作为旋转机械的核心部件之一，滚动轴承直接影响旋转机械的产品质量和可靠性。然而，在复杂工况下，轴承容易发生故障，从而引发周期性振动。轴承故障检测的一种常见方法是应用周期估计方法，从振动信号中提取故障周期。由于复杂的背景噪声可能会掩盖故障轴承的周期信号，因此需要采用准确的方法，尽早识别故障周期。

在本案例研究中，振动数据下载自凯斯西储大学 (CWRU) [2]。所使用的轴承型号为 6205-2RS JEM SKF，测试轴承的参数列于表 2 中。轴承故障设置在外圈，信号由试验台采集，采样频率为 48 kHz。根据

[2]https://engineering.case.edu/bearingdatacenter/

表2：测试轴承 6205-2RS JEM SKF 的参数

Type	6205-2RS JEM SKF
滚动体数量	9
内径（英寸）	0.9843
外径（英寸）	2.0472
厚度（英寸）	0.5906
滚珠直径（英寸）	0.3126
节圆直径（英寸）	1.537
电机转速约值（rpm）	1747
故障尺寸（mm）	0.021
故障频率约值（Hz）	104
故障周期约值（ms）	9.6

表 2，故障特征频率为 104 Hz，故障周期为 $T_0 = 461$。

振动信号包含 100,000 个样本点。由于背景噪声复杂，无法通过常规频谱方法直接定位故障特征频率。因此，我们尝试从时域检测轴承故障。采用所提出的 CPGP方法从振动信号中估计故障周期。由于采样频率相当高， p 的搜索范围设为 $[1, 500]$,，即 $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{I}} = \{1, 2, \cdots, 1, 000\}$, $p_{max} = 1,000$ 和 $d = d_* = 1$。CPGP 中参数 δ和 θ的搜索范围分别设为 $[0.01, 5]$ 和 $[15, 20]$,。δ 和 θ 的初始值以与第4节类似的方式获取。为进行比较，采用 NRC、MLPE 和 FNLS 估计故障周期。FNLS 的最大谐波数设为 $L = 30$。为衡量各方法的准确率，从完整振动信号中随机抽取一段信号，其信号长度 n 介于 2,000 至 20,000 之间。按照 Li等人（2021）的做法，将估计周期落入 $\{0.95T_0, 1.05T_0\}$ 的概率记为检测准确率。对于每个介于 2,000 至 20,000 之间的 n ，重复实验 100 次，并在图6中展示其故障检测准确率。

图6显示，所提出的 CPGP 明显优于其他方法，尤其是在信号长度较短（小于 6,000）时，这表明它

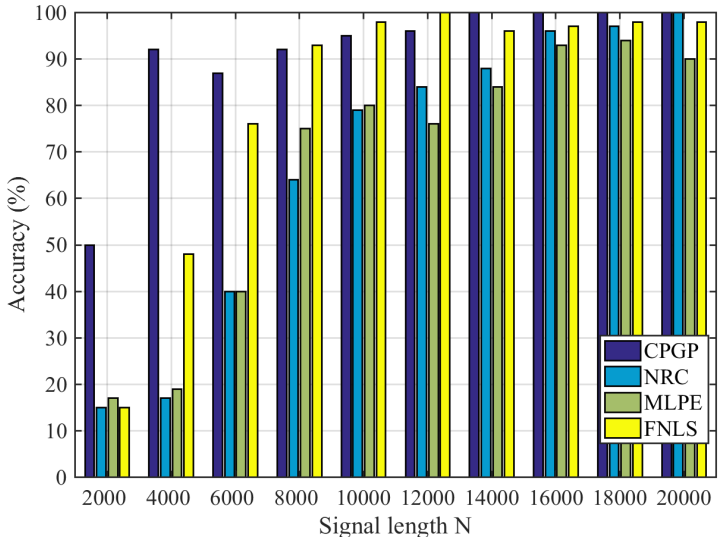


图 6：CPGP、NRC、MLPE 和 FNLS 对轴承振动信号的周期估计准确率。

可用于早期检测轴承故障。与 NRC 和 MLPE 的准确率相比，CPGP 的准确率明显更高，并且随着信号长度增加（超过 14,000），能更快地收敛至 100%。尽管在信号长度适中时，FNLS 的表现略优于 CPGP，但即使信号足够长，其准确率也无法收敛至 100%。这主要是因为真实振动信号并非严格周期性，且受到显著噪声的影响。CPGP 的稳定收敛表明，它能够在真实信号上实现稳健的性能。

5.2 音高估计

音高估计（也称为基频估计）是音频信号处理领域一个引人入胜的研究课题（Nielsen 等人，2017年）。研究人员旨在生成一系列频率值，对应语音或音乐录音的音高。对于单声部音乐或语音录音，人类很容易辨识其音高。然而，对人工智能而言，完成这项任务仍然颇具挑战，因为信号通常包含复杂的信号成分和环境噪声。可以采用基于自相关的方法（例如 NRC）或基于似然的方法来估计音高

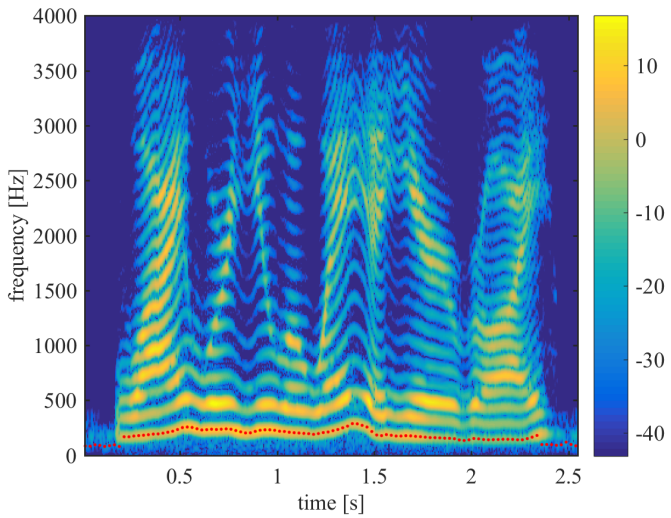


图 7：语音信号的基频估计值和声谱图。

（例如 MLPE）。然而，NRC 和 MLPE 都未考虑周期内相关性，因此需要一种更复杂的方法来将周期内相关性纳入考量。在本案例研究中，CPGP 被用于估计一段人声录音和一段由中提琴演奏的单声部音乐录音的音高。

将信号划分为较短的信号子集（语音信号划分为$n = 200$，音乐信号划分为$n = 240$），并分别分析每个子集，以展示基频的变化过程。由于这些信号的采样频率（8000 Hz）相对较低，因此通过设置 $d_* = 1$、 $d = 5$和 $p_{max} = 120$，为 p设置的搜索网格为$\boldsymbol{\mathcal{I}} = \{2, 3, \cdots, 120\}$和$\boldsymbol{\mathcal{T}} = \{2, 3, \cdots, 600\}$。由于音频信号的波形平滑得多，且噪声强度较弱，因此将 CPGP 中参数δ和 θ的搜索范围分别设为 $[0.3, 0.5]$ 和 $[2, 3],$。我们发现，利用领域知识指定搜索范围有助于提升性能。

如图 7 和图 8 所示，两种信号的基频估计值以红点标记在声谱图上（Mitra 和 Kuo，2006）。红点与声谱图中所示的模式吻合良好，这表明 CPGP 能够准确估计音频和单声部音乐信号的周期。本案例研究

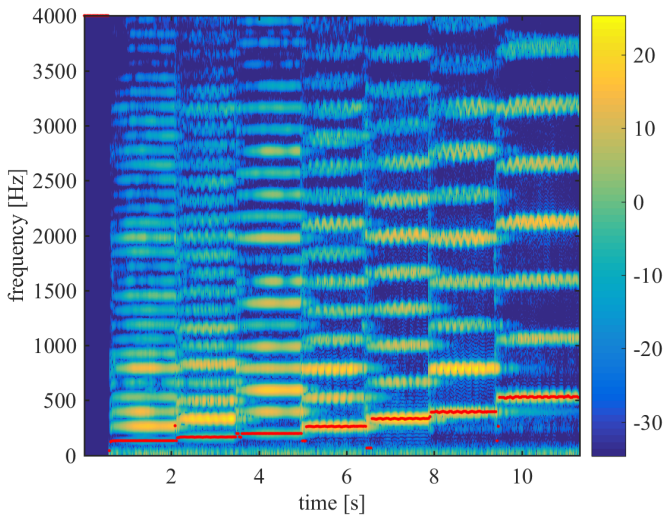


图 8：音乐信号的基频估计值和声谱图。

表明，将 CPGP 应用于音频和音乐信号的音高估计与旋律提取具有巨大潜力。

6 结论

本文提出了一种名为 CPGP 的可扩展建模方法，用于加速 PGP 的参数估计和模型预测，从而大幅提升其处理大规模、低信噪比周期性数据的适用性。当周期性数据采集于网格上时，计算 CPGP 的似然通常只需 $\mathcal{O}(p^2)$ 的计算复杂度；若 p 能被 n 整除，则只需 $\mathcal{O}(p\log\ p)$。考虑周期估计在内，其总计算复杂度为 $\mathcal{O}(d^3p_{max}^3)$；实际应用中，分段长度$p$ 和 p_{max} 独立于数据规模 n，且远小于该规模。与传统近似方法不同，CPGP 的似然与 PGP 完全相同，因此能够保持相同的精度。CPGP 在模拟和案例研究中表现出色，验证了其在现实世界应用中的实用性。

一个值得进一步探讨的主题是，如何在保证精度的前提下进一步加速 CPGP。在模拟研究中可以看到，变体 ACPGP 的运行速度快于 CPGP，但

其性能并不理想。单纯牺牲第二个复合对数似然ℓ_2 来提高速度并不是好办法。一种替代方法是对数据进行零填充，使得 n 可以整除 p，从而将计算似然的复杂度加速至 $\mathcal{O}(p \log p)$。然而，直接应用零填充在实践中效果并不好，尤其是在信号具有非恒定趋势时。未来的工作将探索一种经过精心设计且有坚实理论依据的零填充方法，该方法采用循环嵌入法（Wood 和 Chan，1994；Davies 和 Bryant，2013）以及托普利兹矩阵的循环近似。

所提出的 CPGP 假设数据具有严格的周期性，尽管它已成功应用于实际场景，而这些场景中的信号可能是伪周期性的。对于重复周期存在显著拉伸或扭曲变化的伪周期信号，可能需要一种新的高效建模方法。另一个限制性假设是噪声为白高斯噪声。这些假设可能会限制所提出的 CPGP 的应用范围。未来的研究中，我们将探索 CPGP 的一些扩展，使其能够处理受有色噪声污染的伪周期性数据。

补充材料

补充材料包含理论结果的技术证明、所提出的参数估计优化方法的详细信息（包括伪代码），以及用于复现本文所有图表的代码。

参考文献

Akritas, A. G., Akritas, E. K. 和 Malaschonok, G. I.（1996），《Sylvester（行列式）恒等式的各种证明》，*Mathematics and Computers in Simulation*，42，585–593。

Bevilacqua, M., Fass`o, A., Gaetan, C., Porcu, E. 和 Velandia, D.（2016），《多元高斯随机场估计的协方差截断》，*Statistical Methods & Ap-plications*，25，21–37。

Brigham, E. O.（1988年），快速傅里叶变换及其应用，Prentice-Hall, Inc.

Caragea, P. C.和Smith, R. L.（2007年），《一类多元正态模型的计算高效替代估计量的渐近性质》，《多元分析杂志》，98，1417–1440。

Chandola, V. 和 Vatsavai, R. R.（2011年），《一种用于监测周期时间序列的基于高斯过程的在线变化检测算法》，载于*2011*年工业与应用数学学会国际数据挖掘会议论文集，工业与应用数学学会，第95–106页。

Choudhury, A.、Nair, P. B. 和 Keane, A. J.（2002），《一种用于大规模高斯过程建模的数据并行方法》，载于*2002*工业与应用数学学会国际数据挖掘会议论文集，工业与应用数学学会，第95–111页。

Coeurjolly, J.-F. 和 Porcu, E.（2018年），《通过嵌入循环矩阵快速且精确地模拟复值平稳高斯过程》，《计算与图形统计学杂志》，27，278–290。

Cressie, N. 和 Johannesson, G.（2008），《面向超大型空间数据集的固定秩克里金法》，《皇家统计学会杂志：*B*辑（统计方法学）》，70，209–226。

Davies, T. M. 和 Bryant, D.（2013年），《关于 R语言中高斯随机场的循环嵌入》，《统计软件杂志》，55，1–21。

Dietrich, C. R. 和 Newsam, G. N.（1997年），“通过协方差矩阵的循环嵌入快速且精确地模拟平稳高斯过程”，《工业与应用数学学会科学计算杂志》，18，1088–1107。

Diggle, P. J.、Moraga, P.、Rowlingson, B. 和 Taylor, B. M.（2013年），“空间与时空对数高斯 Cox 过程：扩展地统计学范式”，*StatisticalScience*，28，542–563。

Durrande, N., Hensman, J., Rattray, M., 和 Lawrence, N. D.（2016），《使用高斯过程检测周期性》，*PeerJ Computer Science*，2，e50。

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M., 和 Niemi, J.（2014），《基于块复合似然的空间模型估计与预测》，*Journal of Computationaland Graphical Statistics*，23，295–315。

Fan, W., Li, Y., Tsui, K.L., 和 Zhou, Q.（2018），《一种用于检测噪声信号周期的抗噪相关性方法》，*IEEE Transactions on Signal Processing*，66，2700–2710。

Fang, K. T., Li, R., 和 Sudjianto, A.（2006），计算机实验的设计与建模，Chapman & Hall/CRC 计算机科学与数据分析系列。

Finley, A. O., Banerjee, S., Waldmann, P. 和 Ericsson, T. (2009)，“大型空间试验数据集加性和显性遗传方差的层次空间建模”，*Biometrics*,65, 441–451。

Gramacy, R. B. (2020)，代理模型：应用科学中的高斯过程建模、设计与优化，Chapman and Hall/CRC。

Gu´erin, J., de Paula Canuto, A. M. 和 Goncalves, L. M. G. (2020)，“利用高斯过程滤波稳健检测周期运动中的物体”，载于*2020*年第*19*届*IEEE*机器学习与应用国际会议，IEEE，第685–692页。

HajiGhassemi, N.和Deisenroth, M.（2014年），“利用周期高斯过程进行解析长期预测”，载于人工智能与统计，PMLR，第303–311页。

Heagerty, P.J.和Lele, S. R.（1998），“二元空间数据的复合似然方法”，美国统计学会会刊，93，1099–1111。

Jones, D. R., Schonlau, M., and Welch, W. J.（1998），“昂贵黑箱函数的高效全局优化”，全球优化杂志，13，455–492。

Katzfuss, M. and Guinness（2021），“高斯过程 Vecchia 近似的一般框架”，*Statistical Science*，36，124–141。

Kaufman, C. G., Schervish, M. J., and Nychka, D. W.（2008），“大型空间数据集中基于似然的估计的协方差截断”，美国统计学会会刊，103，1545–1555。

Klenske, E.D.,Zeilinger, M.N., Sch¨olkopf, B., and Hennig, P.（2015），“用于周期性误差校正的基于高斯过程的预测控制”，*IEEE*控制系统技术汇刊，24，110–121。

Koulali, A.and Clarke, P.(2021), “使用高斯过程对大地测量时间序列中的准周期信号进行建模”， 国际地球物理学杂志, 226, 1705–1714.

Li, Y., Zhao, H., Fan, W., and Shen, C. (2021), “用于伪周期信号周期估计的扩展抗噪相关法”， *IEEE*仪器与测量汇刊, 70, 1–11.

Li, Y. and Zhou, Q. (2016), “使用非可分协方差函数对多变量输出计算机模型进行成对元建模”， *Technometrics*, 58, 483–494.

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J. (2002), *MATLAB*工具箱*DACE*的相关方面, Citeseer.

MacKay, D.J.（1998），《高斯过程导论》，北约高级研究院系列*F*：计算机与系统科学，168，133–166。

Mitra, S. K. 和 Kuo, Y.（2006），数字信号处理：一种基于计算机的方法，第2卷，麦格劳-希尔，纽约。

Nielsen, J. K., Jensen, T. L., Jensen, J. R., Christensen, M. G., 和 Jensen, S. H.（2017年），“快速基频估计：使统计效率高的估计器在计算上也高效”，信号处理，135，188–197。

Park, C., Huang, J. Z., 和 Ding, Y.（2011年），“用于大型空间数据集快速高斯过程回归的区域分解方法”，机器学习研究杂志，12，1697–1728。

Quinn, B. 和 Thomson, P.（1991），“周期函数频率的估计”，*Biometrika*，78，65–74。

Quinonero-Candela, J. 和 Rasmussen, C. E.（2005），“稀疏近似高斯过程回归的统一观点”，机器学习研究杂志，6，1939–1959。

Rasmussen, C. E. 和 Williams, C. K. I.（2006），《机器学习中的高斯过程》，麻省理工学院出版社。

Redivo-Zaglia, M. 和 Rodriguez, G.（2012），《smt：用于结构化矩阵的 MATLAB 工具箱》，数值算法，59，639–659。

Sacks, J.、Schiller, S. B. 和 Welch, W. J.（1989），《计算机实验的设计》，*Technometrics*，31，41–47。

Sang, H. 和 Huang, J. Z.（2012），《大规模空间数据集协方差函数的全尺度近似》，《皇家统计学会杂志：*B*辑（统计方法学）》，74，111–132。

Santner, T.J.、Williams, B.J. 和 Notz, W. I.（2013年），计算机实验的设计与分析，施普林格科学与商业媒体。

Sherman, J. 和 Morrison, W. J.（1950年），“对给定矩阵中一个元素发生变化时相应逆矩阵的调整”，《数学统计年鉴》，21，124–127。

Snelson, E. 和 Ghahramani, Z.（2007年），“局部与全局稀疏高斯过程近似”，载于人工智能与统计，PMLR，第 524–531 页。

Stein, M. L.（2008），“一种针对大型空间数据集的建模方法”，《韩国统计学会杂志》，37，3–10。

Stein, M. L.、Chi, Z. 和 Welty, L. J.（2004），“大型空间数据集的似然近似”，《皇家统计学会杂志：B辑》，66，275–296。

Stewart, M.（1997年），“半正定 Toeplitz 矩阵的 Cholesky分解”，线性代数及其应用，254，497–525。

Taylor, B. M. 和 Diggle, P. J.（2014年），“INLA 还是 MCMC？对对数高斯 Cox 过程空间预测的教程与比较评估”，《统计计算与模拟杂志》，84，2266–2284。

Tebbutt、W.、Bui、T.D. 和 Turner、R.E.（2016），“用于扩展高斯过程的循环伪点近似”，近似贝叶斯推断进展，*NIPS 2016* 年研讨会，1–5。

Titsias、M.（2009），“稀疏高斯过程中诱导变量的变分学习”，载于人工智能与统计，PMLR，第567–574页。

Trench、W. F.（1964），“有限Toeplitz矩阵求逆算法”，《工业与应用数学学会杂志》，12，515–522。

Vecchia, A. V.（1988年），“连续空间过程的估计与模型识别”，《皇家统计学会杂志》*B*辑，50，297–312。

Wise, J., Caprio, J. 和 Parks, T.（1976），“最大似然音高估计”，*IEEE*《声学、语音与信号处理汇刊》，24，418–423。

Wood, A. T. and Chan, G. (1994), “在 $[0, 1]$维中模拟平稳高斯过程”，*Journal of Computational and Graphical Statistics*,3, 409–432.

Xiong, S. (2021), “重构方法：从插值到回归”，计量学, 63, 225–235.

Zhang, Y., Leithead, W. E., and Leith, D. J. (2005), “基于托普利兹计算的时间序列高斯过程回归，运算量为O (N 2)、存储量为O (N)级”，载于第*44*届*IEEE*决策与控制会议论文集, IEEE,pp.3711–3716.