FINKER: Frequency Identification through Nonparametric KErnel Regression in astronomical time series

F. Stoppa¹,², C. Johnston¹,³,⁴, E. Cator², G. Nelemans¹,⁴,⁵, and P.J. Groot¹,⁶,⁷

1 Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands
2 Department of Mathematics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands
3 Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85741 Garching bei München, Germany
4 Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium
5 SRON, Netherlands Institute for Space Research, Sorbonnelaan 2, NL-3584 CA Utrecht, The Netherlands
6 Department of Astronomy and Inter-University Institute for Data Intensive Astronomy, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa
7 South African Astronomical Observatory, P.O. Box 9, Observatory, 7935, South Africa

March 28, 2024

ABSTRACT

Context. Optimal frequency identification in astronomical datasets is crucial for variable star studies, exoplanet detection, and asteroseismology. Traditional period-finding methods often rely on specific parametric assumptions, employ binning procedures, or overlook the regression nature of the problem, limiting their applicability and precision.

Aims. We introduce a universal, nonparametric kernel regression method for optimal frequency determination that is generalizable, efficient, and robust across various astronomical data types.

Methods. FINKER uses nonparametric kernel regression on folded datasets at different frequencies, selecting the optimal frequency by minimising squared residuals. This technique inherently incorporates a weighting system that accounts for measurement uncertainties and facilitates multi-band data analysis. We evaluated our method’s performance across a range of frequencies pertinent to diverse data types and compared it with an established period-finding algorithm, conditional entropy.

Results. The method demonstrates superior performance in accuracy and robustness compared to existing algorithms, requiring fewer observations to reliably identify significant frequencies. It exhibits resilience against noise and adapts well to datasets with varying complexity.

1. Introduction

In astronomy, periodic signals buried within time series of flux (photometric light curves) or radial velocity measurements serve as important carriers of scientific information. Their detection is essential for various scientific pursuits, from exoplanet characterisation to the study of variable stars. A plethora of period-finding algorithms exist, from standard techniques such as the classical Lomb-Scargle periodogram (Lomb 1976; Scargle 1982), the generalised Lomb-Scargle (Zeichmeister & Kürster 2009; VanderPlas 2018), and the discrete Fourier transform (Deeming 1975) to popular nonparametric methods such as string length (Dworetsky 1985), analysis of variance (Schwarzemberg-Czerny 1989, 1996), and phase dispersion minimisation (Stellingwerf 1978). In addition to these, smoothing algorithms and statistical hypothesis testing based on regression models and confidence sets have been employed to detect periodicities in radial velocity measurements (McDonald 1986; Toulis & Bean 2021), highlighting the potential of regression techniques for periodicity detection in astronomical data. These methods can be used to search for periodicity via either a grid search or direct optimisation (Reimann 1994).

Recent advancements have also seen the application of Gaussian process regression with specified kernels as a promising approach for analysing periodic data, offering a flexible framework for accounting for the noise characteristics and underlying trends in the time series (Barros et al. 2020; Nicholson & Aigrain 2022; Barragán et al. 2021; Matesic et al. 2024). Furthermore, the introduction of information theory-based methods, such as conditional entropy (Graham et al. 2013) and quadratic mutual information (Huijse et al. 2018), has made significant contributions to this field. While these methodologies align with our work in aiming to identify periodic signals, they differ in terms of application and framework. We discuss these differences in our paper, highlighting the unique advantages and considerations of our approach, which we call Frequency Identification through Nonparametric KErnel Regression (FINKER), compared to these recent advancements.

Comparative analyses have been instrumental in enhancing our understanding of various period-finding algorithms’ capabilities and limitations (Graham et al. 2013). Such studies emphasise the dependence of these methods on the quality of the light curve data. Graham et al. (2013) advocate for a bimodal observation approach, where pairs of observations are taken rapidly each night to retain a high sampling frequency. They also emphasise that algorithms perform notably well for specific types of variables, such as pulsating and eclipsing classes, but find that ensemble methods (e.g. Saha & Vivas 2017; Ranaivomanana et al. 2023), which combine multiple algorithms, do not generally outperform single algorithms. Among the array of algorithms studied, conditional entropy stands out in terms of period recovery and computational time, with analyses of variance and phase dispersion minimisation also showing promise.

Despite these advancements, the search for a universally applicable and assumption-free period-finding method contin-

* Equal contribution.

Article number, page 1 of 16
ues. A noteworthy yet underdeveloped method presented by
Hall et al. (2000) pioneered the exploration of nonparametric
kernel regression techniques for period detection through
grid search least-squares optimisation. Despite its potential, the
method went largely unused in astronomical applications due to
the technological and computational limitations of the time. This
work was later expanded in a series of papers to apply the frame-
work of nonparametric regression estimation to periodograms
using single and multiple sine and cosine components to circum-
vent the alias problem (Hall & Li 2006; Genton & Hall 2007).
While this body of earlier work serves as a conceptual underpin-
ing for our current research, with our approach we have sub-
stantially extended the methodology by developing a more ad-
vanced, nonparametric kernel regression algorithm tailored for
astronomical applications. With it, we can further investigate the
behaviour of the kernel bandwidth when estimating frequencies.

We aim to establish nonparametric kernel regression as a ro-
bust and versatile method for optimal frequency identification,
setting the stage for more precise and efficient analyses in both
astronomical and potentially interdisciplinary fields. The struc-
ture of this paper is as follows: Section 2 contains the mathe-
matical foundations and computational aspects of our proposed
method. Section 3 describes the experimental setup and presents
a comparative analysis with existing methodologies. Section 4
presents the results of FINKER applied to a set of real examples,
such as classical pulsators, short period transiting compact bi-
naries, and radial velocity variations in binaries. Finally, Section
5 offers conclusions and explores avenues for future develop-
ments.

To promote open and reproducible research, FINKER is pub-
licly available on GitHub1 (Stoppa 2024).

2. Method
Nonparametric kernel regression (Nadaraya 1964; Watson 1964;
Hall et al. 2000) serves as the core of our approach. However, be-
fore delving deeper into our method, it is crucial to consider the
work of Schwarzenberg-Czerny (1999) on the comparative capa-
bilities of parametric and non-parametric methods in optimum
period search analyses. Their work highlights that parametric
methods, when aligned with an accurately specified model that
closely mirrors the data’s underlying structure, can yield superior
efficiency and predictive power. However, the effectiveness of
these methods can significantly diminish in scenarios where such
a precise model alignment is lacking, leading to potential misin-
terpretations or overlooked data characteristics. In contrast, non-
parametric methods, not being confined to a predefined model,
stand out for their robust adaptability and resilience and thus
their broader applicability across diverse settings. This inherent
flexibility makes nonparametric methods especially valuable in
situations where the underlying periodic signals in astronomi-
cal time series are complex and not modelled well by standard
parametric forms.

Following the selection of nonparametric kernel regres-
sion for its adaptability and flexibility, this section details the
methodology employed for nonparametric kernel regression-
based frequency identification. We discuss the relevant data pre-
processing, the technical implementation of the nonparametric
regression model, and the determination of the optimal fre-
quency of variability in the time series. Each subsection delves
into the specifics of these steps, elucidating the underlying prin-
ciples and computational strategies.

In general, we considered an arbitrary, sparsely sampled as-
tronomical time series, consisting of \( N \) observations \( y_i \) with
\( i = 1, \ldots, N \). This pertains to both photometric flux and radial
velocity time series.

\[ \phi_i = (t_i - t_0) \mod 1, \]

where \( t_0 \) is the chosen reference epoch2, and \( \nu_j \) is the candi-
date frequency for folding. This technique aligns the repeating
patterns in the data, which can be obscured in the time domain
due to irregular sampling intervals or noise. The periodic signal
becomes more discernible by folding the data at the correct or
assumed frequency, facilitating the analysis of its structure and
properties.

It is important to note that this phase-folding technique intrinsi-
cally assumes the stationarity of the time series.

The phase-folded curve is particularly useful for visualising
and analysing periodic variations in the brightness of vari-
able stars or exoplanets transiting their host stars and for ra-
dial velocity variations of multiple star systems or exoplanet-
ary systems. To illustrate this, Figure 1 showcases a real light
curve alongside its phase-folded counterpart at the literature
frequency (Torrealba et al. 2015). Upon examining this exam-
ple, we notice that by folding the light curve on the under-
lying period of variability, we reduce the scatter between two adja-
cent points. This is the basis for various nonparametric fre-
frequency determination routines, including the analysis of vari-
ance method (Schwarzenberg-Czerny 1989), the Lafler-Kinman
statistic (Clarke 2002), the conditional entropy periodogram
(Graham et al. 2013), and our proposed nonparametric kernel
regression technique.

2.2. Nonparametric kernel regression
Once the astronomical time series is transformed into a phase di-
agram via phase folding, we apply nonparametric kernel regres-
sion to the folded light curve. This application is fundamental to
our methodology, enabling the identification of the optimal fre-
frequency of variability by analysing the periodic signal’s structure
and properties within the phase-flux relationship.

Local constant regression is the simplest variant and most
efficient version of kernel regression. The regression function,

\[ \hat{y}(x) = \sum \frac{\hat{f}(x - x_i)}{\sum \hat{f}(x - x_i)}, \]

2 We note that the reference epoch refers to different moments for dif-
dent types of data. For example, with eclipsing binaries and transit-
ing systems \( t_0 \) is set to coincide with the time of superior conjunction,
whereas in spectroscopic binaries \( t_0 \) refers to the time of periastron
passage.
Fig. 1. MeerLICHT (Bloemen et al. 2016) telescope’s light curve of CRTS J033427.7–271223 on the left and phase-folded version at the literature frequency (Torrealba et al. 2015) on the right. Associated with each observation is an estimated measurement uncertainty.

\[ \hat{f}_K(\phi) = \frac{\sum_{i=1}^{n} K_h(\phi - \phi_i) y_i}{\sum_{i=1}^{n} K_h(\phi - \phi_i)}. \]  

Here, \( y_i \) represents the observed flux corresponding to the phase \( \phi_i \), and the estimated function value \( \hat{f}(\phi) \) is a weighted average of these fluxes. The weights are determined by the kernel function \( K_h \), which measures the closeness of each observed phase \( \phi_i \) to the phase \( \phi \) of interest. This approach is computationally straightforward and only requires one parameter, the bandwidth \( h \) of the kernel \( K_h \), Section 2.4 explores this parameter in detail.

A simple extension to local constant, local linear regression (Fan & Gijbels 1994), enhances the local constant model by incorporating a linear trend within the kernel’s scope at each phase \( \phi \). This method is particularly famous for its ability to better manage boundary problems, an advantage in typical scenarios.

However, in our approach, this boundary issue is already effectively addressed by using a common trick of extending the phase-folded light curve to both sides. This extension effectively mitigates boundary issues, allowing for a more accurate estimation of the regression function near the edges.

Local linear regression, while offering a refined analysis by including linear trends, necessitates the calculation of two additional parameters at each phase: the intercept and slope of the local linear model. Given its considerable computational load and the fact that our extensive frequency folding requirements do not show marked improvements over local constant regression, we opted for the latter. The computational simplicity and efficiency of local constant regression make it the more suitable choice for our frequency optimisation framework.

2.3. Choice of kernel

The array of kernel functions available to researchers is rich, with each kernel bringing its own unique advantages to different data challenges (Epanechnikov 1969; Gasser et al. 1985; Izenman 1991). Despite this variety, our investigation specifically leverages the Gaussian kernel. This decision is informed by the kernel’s prevalent application across various fields and the observation that, for the problems at hand, it tends to provide results that are not significantly different from those obtained using other kernels. The Gaussian kernel is mathematically expressed as

\[ K(\phi, \phi', h) = \exp\left(-\frac{(\phi - \phi')^2}{2h^2}\right). \]  

In this equation, \( \phi \) and \( \phi' \) are points in the feature space, and \( h \) is the bandwidth, or relevant length scale, and can be optimised for specific applications.

One of the primary advantages of the Gaussian kernel is its smoothness. Being infinitely differentiable, it ensures that the estimated function is smooth. This is particularly beneficial for capturing underlying trends in astronomical data, which are often smooth in nature. The Gaussian kernel also has the property of localised influence, assigning significant weight only to points that are close to the target point in the feature space. Overall, the Gaussian kernel is a versatile tool for a wide range of applications.

2.4. Bandwidth selection

In kernel regression, the choice of bandwidth \( h \) is critical, as it directly impacts the estimator’s bias and variance. A well-chosen bandwidth balances the trade-off between the smoothness of the estimated function and the fidelity to the data points. Scott’s rule (Scott 1979) and Silverman’s rule (Silverman 1986) are both prevalent methods for determining the bandwidth of a kernel density estimate. Scott’s rule is typically represented as

\[ h = 3.49 \sigma n^{-1/3}, \]  

where \( n \) is the sample size and \( \sigma \) is the standard deviation of the observations. This bandwidth aims to minimise the mean integrated squared error (MISE) for data that approximates a normal distribution, yielding a smoother density estimate suitable for elucidating the data’s overall structure. Conversely, Silverman’s rule, which is often given by

\[ h = 1.06 \sigma n^{-1/5}, \]  

employs a slightly different formula tailored for Gaussian-like data but allowing for a tighter bandwidth, enhancing the detection of finer structural details within the data distribution.

Both rules presuppose a normal distribution and might not be optimal for datasets that are multi-modal or exhibit heavy-tailed distributions. Given the unique characteristics of our data, we adopted a custom bandwidth formula that better aligns with the non-Gaussian, periodic nature of such datasets:

\[ h = \alpha n^{-1/5}. \]  

This formula is derived from the same principle of minimising the MISE, where the \( n^{-1/5} \) term balances the estimator’s bias and variance as the sample size \( n \) increases. The constant \( \alpha \) replaces the standard deviation component from Silverman’s and Scott’s rules, allowing for adjustment based on the specific characteristics of the periodic data under study. This approach does not assume a Gaussian distribution, making it more adaptable to the heteroskedasticity and periodicity inherent in astronomical time-series data. Section 3.1.1 explains in more depth how we found an empiric optimal \( \alpha \) for folded light curves.

Additionally, we built and tested a more accurate yet computationally expensive adaptive bandwidth strategy that assigns...
In Section 3.1.2, we validate this approach on a set of simulations and show its reliability, comparing it to the known true errors in frequency.

It is crucial to acknowledge that the applicability of bootstrapping for uncertainty estimation in frequency determination rests on specific assumptions regarding the noise characteristics within the data. Particularly, this method assumes that the noise affecting the data is uncorrelated. Violations of this assumption may introduce bias into the bootstrap results, potentially compromising the accuracy of uncertainty estimates.

Furthermore, although a valid and informative method, its results are only meaningful if a frequency value has already been found with a certain degree of accuracy. Suppose the frequency value has been wrongly estimated. In that case, the measure of its uncertainty will remain small due to the search in a localised space around the found minimum, but the resulting value will be meaningless. We do not yet have a solution for this eventuality but will be subject to more in depth statistical research in the future.

2.5. Role of weighted residuals in frequency determination

The sum of squared residuals (SSR) is a commonly used metric for assessing the goodness of fit. In the specific context of this work, it serves as a specialised objective to compare how the fit of the kernel regression behaves for data folded at different periods. The central premise is that a lower SSR indicates a more compact, smoother folded light curve, suggesting a more accurate fit to the inherent periodicity in the data. The SSR is defined as

\[ \text{SSR}_p = \sum_{i=1}^{n} (y_i - \hat{f}(\phi_i; \nu_j))^2, \] (7)

where, \( \hat{f}(\phi_i; \nu_j) \) represents the estimated value at phase \( \phi_i \) when the time-series data are folded at frequency \( \nu_j \), as obtained through a nonparametric regression model.

However, the standard SSR gives equal weight to all residuals, ignoring measurement uncertainties associated with the observations. To address this, we simply used a weighted sum of squared residuals, denoted as \( \text{SSR}_w \), that incorporates these uncertainties:

\[ \text{SSR}_w = \sum_{i=1}^{n} \frac{(y_i - \hat{f}(\phi_i; \nu_j))^2}{\sigma_i^2}. \] (8)

This weighted \( \text{SSR}_w \) accounts for the heterogeneity in observational data quality, making the frequency selection more robust. Moreover, the number of estimated parameters, in our case, the bandwidth, is consistent across each dataset’s folding. This is crucial for mitigating the risk of overfitting, and it allows the use of SSR as a reliable metric for comparison across different foldings.

2.6. Uncertainty estimation in frequency determination via bootstrapping

To enhance the reliability of FINKER, we employed a bootstrap methodology (Efron 1979; Efron & Tibshirani 1986) that provides an estimate of the uncertainty of the predicted frequency. This entails generating multiple datasets from the original by random sampling with replacement, after which FINKER is applied to a finely tuned grid centring on the initial frequency estimate. We use the standard deviation of the bootstrap results as an estimate of the prediction reliability.

The algorithm for optimisation is outlined as follows:

1. Generate a set of candidate frequencies \( \nu_1, \nu_2, \ldots, \nu_m \).
2. For each candidate frequency, \( \nu_j \):
   a. Fold the time-series data on \( \nu_j \) according to Eq. 1.
   b. Apply nonparametric kernel regression to the folded data.
3. Compute \( \text{SSR}_w \) using Eq. 8.
4. Repeat B times:
   a. Sample with replacement from the original dataset.
   b. Apply nonparametric kernel regression to a small range around \( \nu_{opt} \).
5. Identify the frequency \( \nu_{opt} \) that minimises the \( \text{SSR}_w \).

By automating this sequence of operations, FINKER provides a reliable estimation of the intrinsic periodic nature of the data, focusing on the frequency that yields the minimum \( \text{SSR}_w \).

3. Application to synthetic data and benchmarking

This section presents an empirical validation of FINKER and shows its performance on synthetic data mimicking various astronomical phenomena.

3.1. Synthetic data

The generation of synthetic light curves is a critical component of our simulation framework, allowing us to test the robustness and effectiveness of our frequency optimisation algorithm under controlled conditions. These simulations are characterised by several adjustable parameters, each designed to replicate diverse observational scenarios and attributes of astronomical entities.

First, we adjusted the number of data points to emulate both sparsely and densely sampled observations, mirroring the variability encountered in astronomical data acquisition. The total duration of these observations is also configurable and sets the temporal extent for the generated light curves.

Central to our simulations is the emulation of the light curve’s frequency components. We simulated the dominant periodic signal within the data through a primary frequency and...
its corresponding amplitude, which denotes the strength of this signal. Additionally, a secondary frequency and its amplitude are included to model objects exhibiting multiple periodic behaviours.

As discussed in Section 2.4, in our application, the only parameter influencing the performance of the kernel regression model is the kernel bandwidth $h$. As such, we analysed the effect of varying this parameter to find an optimal value that (i) can be fixed for most types of light curves and (ii) would not require any parameters to be fine-tuned when using FINKER.

Our proposed bandwidth, $h = \alpha n^{-1/3}$, is modulated by the multiplicative constant $\alpha$, which scales the base bandwidth, $n^{-1/3}$. The scaling factor based on the sample size $n$ allows us to account for different light curve sizes, and, since the phase space is always in the range [0,1], it also modulates the value of the bandwidth for different levels of $\delta_\phi$, the average distance between observations in phase. This, of course, assumes that in phase space, on average, observations are equidistant. In astronomical time series, this is not always the case, and for these scenarios, we propose a solution later on in this section.

The other bandwidth component, $\alpha$, determines the smoothness of the regression estimate for a light curve folded at a specific frequency, thereby affecting the trade-off between overfitting and underfitting. Figure 3 demonstrates the relationship between the folding frequencies and the residual sum of squares for three $\alpha$ values. An $\alpha$ value of 0.1, around 10% of the phase range ([0,1]), typically achieves a good balance.

Since the sample size is fixed in real scenarios and the phase space is always in the range [0,1], we only need to find an $\alpha$ value that works in most situations. To do so, we built a set of simulations varying different signal-to-noise ratios (S/N) and sample sizes and searched for the alpha value that leads to the smallest estimated frequency error. We found that, on average, an $\alpha$ of 0.06 will allow for an accurate frequency identification and consequentially remove the burden of hand-picking this parameter. This $\alpha$ value will be employed for the subsequent analyses in this study.

Our custom bandwidth approach is an adequate solution for a wide range of scenarios. However, in situations characterised by a limited number of observations, significant sparsity, or data concentrated in specific regions of the phase space, an adaptive bandwidth strategy emerges as a more effective alternative, enhancing accuracy in these particular conditions. This adaptive method, while significantly more computationally intensive, dynamically adjusts the bandwidth in response to local data density, thereby offering a more tailored fit to the underlying structure of the data. In Section 4, we show how, for a select number of examples, an adaptive bandwidth allows the optimal frequency to be recovered with as few as ten observations.

Fig. 2. Simulated light curves at three S/N levels, based on the MeerLICHT telescope data-derived $\sigma$-magnitude function.

Fig. 3. Folding frequencies and squared residuals for kernel regression with different $\alpha$ values. The vertical dashed line indicates the true frequency of 1 $d^{-1}$. All $\alpha$ values lead to frequency estimates that closely converge to the true value.
3.1.2. Bootstrap uncertainty reliability

After having identified the best frequency, FINKER performs a bootstrap resampling of the original light curve and repeats the search process in a small grid around the best frequency. Repeating this process multiple times gives an estimate of the variability of the estimated frequency that can be used as an uncertainty estimate.

In this section, we illustrate the application of FINKER’s bootstrap-based uncertainty estimation method to a synthetic light curve. We demonstrate the method’s potential in providing an uncertainty estimate around the determined frequency and visually assess the nature of these uncertainties in relation to various sample sizes and S/N.

**Figure 4.** Kernel density estimations of bootstrap-estimated frequencies for a single synthetic light curve across varying sample sizes and S/N levels. Each subplot represents the bootstrap frequency distribution of the estimated best frequency juxtaposed with the true frequency of the synthetic light curve. As expected, with larger sample sizes and S/N, FINKER is more accurate, and its bootstrap density better resembles a Gaussian. Consequently, the estimated frequency uncertainty is more reliable.

Figure 4 presents a series of kernel density estimations of the frequencies obtained through the bootstrap method for different combinations of sample size and S/N for a synthetic light curve. The alignment of the best frequency with the bootstrap distribution’s mean suggests an unbiased nature of the estimation process. Notably, the bootstrap distribution deviates from the expected Gaussian shape at a S/N=7.1 with a sample size of 50, indicating that low-S/N signals or a small sample size may pose challenges to the reliability of the bootstrap method.

The standard deviation of the bootstrap frequencies, serving as the uncertainty measure, appears to offer a conservative estimate. This conservatism ensures that the estimated uncertainty is not understated, yet it also hints at the opportunity to refine the method to more accurately capture the true variation. Future iterations of FINKER will focus on improving the uncertainty estimation to ensure a more precise reflection of the underlying variability, particularly in challenging observational scenarios characterised by low S/N and limited sample sizes.

3.1.3. Optimal frequency identification

To validate the accuracy of FINKER, we conducted experiments using synthetic light curves with known true frequencies. Our kernel regression method is benchmarked against the commonly used conditional entropy (Graham et al. 2013), which we employed for a comparative analysis.

In short, conditional entropy gauges the variability of one variable, say the light curve’s flux, given that we have knowledge of another, such as the phase. For a specific folding frequency, a low conditional entropy implies a reduced level of sparseness in the flux for a given phase, consequentially hinting that the frequency value chosen is correct.

In the computation of conditional entropy, a binning process via partition schemes is employed. This involves organizing the data into distinct bins or partitions, which aids in the accurate estimation of the probability distributions involved in the formula. Particularly, a simple rectangular partitioning scheme, with \( i = 1, ..., N \) bins in flux space and \( j = 1, ..., M \) bins in phase space is commonly adopted for computational efficiency during the analysis. The conditional entropy \( H(m|\phi) \) is defined as

\[
H(m|\phi) = \sum_{i,j} p(y_i, \phi_j) \ln \frac{p(\phi_j)}{p(y_i, \phi_j)}
\]

Here, \( p(y_i, \phi_j) \) denotes the estimated probability of a data point occupying the \( i \)th flux bin and the \( j \)th phase bin simultaneously, while \( p(\phi_j) \) signifies the estimated probability of a data point falling within the \( j \)th phase bin, irrespective of its flux. In the context of analysing light curves, conditional entropy is used to infer the frequency value that minimizes this entropy. Algorithmically, frequency identification using conditional entropy follows an approach similar to what is laid out in Section 2.7, but with a different objective function. We fixed the binning size for all subsequent analyses to \( N = 10 \) and \( M = 10 \).

Figures 5 and 6 illustrate the results of the conditional entropy and FINKER, respectively, for the search of the optimal frequency on a sinusoidal light curve. The left figure shows the range of frequencies tested and the objective function values for each method, the SSR for ours and conditional entropy for the benchmark method. The two objective values are not directly comparable, but for visual clarity, we rescaled both metrics in a \([0, 1]\) range for all subsequent plots. The right figure shows the light curve folded at the estimated best frequency and, for FINKER, the associated kernel regression fit.

As shown in Fig. 6, FINKER yields an accurate frequency prediction and has a smooth behaviour in its estimator. The noisy results of the conditional entropy are instead likely attributable to the binning process needed for the calculation of its metric.

Furthermore, our methodology inherently mitigates the common pitfall of mistaking harmonic frequencies for the fundamental frequency – a frequent issue with both other nonparametric approaches and standard parametric approaches. Specifically, when a frequency is an integer multiple of the true frequency, the light curve maintains a semblance of order and, consequently,
Fig. 5. Conditional entropy’s results across a range of frequencies for a sinusoidal light curve with 100 observations, a baseline magnitude of 17, and an amplitude of 0.1 (S/N = 14.3). The plot on the left shows the frequency range searched, and the inset at the minimum entropy shows the noisy behaviour of the estimator. The estimated best frequency and its uncertainty are shown with a black error bar. The plot on the right shows the light curve folded at the found frequency.

Fig. 6. FINKER’s results across a range of frequencies for a sinusoidal light curve with 100 observations, a baseline magnitude of 17, and an amplitude of 0.1 (S/N = 14.3). The plot on the left shows the frequency range searched, and the inset illustrates the smooth behaviour of the residuals around the minimum. The estimated best frequency and its uncertainty are shown with a black error bar. The plot on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid, illustrating the algorithm’s capability to model the periodic signal accurately.

a relatively low entropy. Kernel regression, however, evaluates the goodness-of-fit for a frequency by analysing the residuals post-smoothing. Harmonic frequencies, which introduce regular oscillations, are neutralised by the smoothing process, leading to high residuals similar to those from an incorrect frequency. This unique attribute of kernel regression allows it to identify the true frequency more accurately. This behaviour is demonstrated in Fig. 7.

Despite its robustness against harmonic frequencies, kernel regression seems more sensitive to sub-harmonic frequencies with respect to the conditional entropy method. However, these are usually not misidentified since additional periodicity in the folded light curve will have an overly smoothed estimator and, consequentially, a higher residual. Furthermore, to counter this behaviour, our implementation includes a verification step to examine whether twice the identified optimal frequency yields a significant result. If the doubled frequency also presents a low residual sum of squares, it suggests that the true fundamental frequency may indeed be twice the initially identified one. This additional check enhances the method’s accuracy and is a standard option in our code.

Overall, this additional property of kernel regression not only enhances its reliability in frequency estimation but also reduces the likelihood of the need for manual inspection or secondary validation methods to confirm the fundamental frequency. This robustness is particularly valuable in automated analysis pipelines where large volumes of data preclude detailed individual review.

3.1.4. Sensitivity analysis with small sample sizes

To rigorously assess the robustness and performance of our frequency optimisation algorithm, a sensitivity analysis is conducted. This analysis focuses on two key parameters: the number of data points (\(n_{\text{points}}\)) and the amplitude of the primary frequency.
frequency of variability. These parameters are systematically varied over a predefined range to simulate different observational conditions. For each combination of \( n_{\mathrm{points}} \) and amplitude, the algorithm is executed using two different methods: kernel regression with custom bandwidth and conditional entropy. All simulations are calculated using the same baseline magnitude (17), such that changing the amplitude of the variability will result in different signal-to-noise levels.

The results are visualised for our kernel regression periodogram and the conditional entropy periodogram in Figs. 8 and 7, respectively. For both figures, the x-axis represents the amplitude levels, and the y-axis represents the number of observations simulated. The colour intensity on each heat map indicates the absolute error between the estimated and true frequency (of \( 1 \, \text{d}^{-1} \)).

\[ |f_{\text{true}} - f_{\text{pred}}| \]

4. Application to real data

The efficacy of our frequency optimisation algorithm is further substantiated through its application to a diverse array of real astronomical datasets. These datasets encompass light curves from variable stars, radial velocity measurements, and photometric data of transiting exoplanets. The sources of these datasets are diverse, originating from various telescopes and surveys, including the MeerLICHT (Bloemen et al. 2016) and Zwicky Transient Factory (ZTF; Bellm et al. 2019) photometric missions and the HERMES spectrograph (Raskin et al. 2011).

4.1. Photometric time series

MeerLICHT is a fully robotic telescope located in Sutherland Observatory in South Africa and is equipped with five Sloan \( ugriz \) photometric filters with an additional custom \( g \) filter that is a combination of the \( g \) and \( r \) filters (Bloemen et al. 2016). MeerLICHT nominally integrates for 60 seconds and has several observing strategies aimed at identifying and characterising transients in multiple wavelengths. The combination of these observing strategies has resulted in a large database of heterogeneously sampled, (sometimes) contemporaneous multi-colour photometric time series of millions of objects compiled over the mission five year runtime. As the MeerLICHT light curves are

---

**Fig. 7.** Kernel regression’s fits (red lines) for harmonics and subharmonics of the true frequency. The light curve is folded at four different frequencies: half the true frequency (top left), the true frequency (top right), twice the true frequency (bottom left), and a random frequency (bottom right). Conveniently, for multiples of the true frequency, kernel regression will badly fit the observations resulting in high residuals.

**Fig. 8.** Heat map of the conditional entropy sensitivity analysis, illustrating the absolute error in frequency estimation relative to sample size (y-axis) and amplitude (x-axis). Lighter shades correspond to lower absolute errors, with cells marked by a red cross representing an average absolute error above 0.03, deemed as unreliable estimates. The heat map suggests that approximately ten observations are sufficient for a reliable frequency estimate across various amplitude levels, with the only exception being the lowest amplitude level.

**Fig. 9.** Heat map of the kernel regression sensitivity analysis, depicting the absolute error in frequency estimation as a function of sample size (y-axis) and amplitude (x-axis). Lighter shades correspond to lower absolute errors, with cells marked by a red cross representing an average absolute error above 0.03, deemed as unreliable estimates. The heat map suggests that approximately ten observations are sufficient for a reliable frequency estimate across various amplitude levels, with the only exception being the lowest amplitude level.
highly irregularly sampled, they are ideal candidates for testing

the performance of our algorithm.

Prior to analysis, the MeerLICHT data were subjected to a

standard pre-processing pipeline. This includes the removal of

outliers, correction for atmospheric extinction, and normalisa-
tion to account for instrumental variations (we refer the readers

to de Wet et al. 2021 and Ranaivomanana et al. 2023 for a com-
plete description of the data reduction process).

In addition to MeerLICHT data, we also made use of ZTF
data in the case of compact transiting binaries. ZTF is a robotic
telescope equipped with three photometric filters that surveys the
night sky for transients and periodic variable stars with 30 sec-
ond exposures in each filter (Bellm et al. 2019). While the ma-
jority of the data was obtained in the g and r filters, we only
considered data in the g filter for this work. Due to its observing
strategy, these data are irregularly sampled, but are less sparsely
sampled than the MeerLICHT data. Additionally, ZTF has oc-
casional deep drilling campaigns that cover select fields with a
high cadence over a short period of time. Thus, the ZTF data
have a highly irregular sampling that spans several years.

Table 1 lists a set of known light curves and their literature
frequencies. We also show frequencies and uncertainties deter-
mined by the conditional entropy and nonparametric kernel re-
gression periodograms. To ensure a fair comparison of the two
methods, when a frequency is identified for both methods, half
double of that frequency are further investigated, and the fi-
nal frequency is the one that minimises the respective metrics.

Below is some more information on the type of light curves
shown.

**Classical pulsators:** Classical pulsators are large amplitude,
radially pulsating stars that exist in the classical instability strip,
including δ Scuti variables, RR Lyrae variables, and Cepheid
variables (Aerts et al. 2010; Kurtz 2022). We tested our algo-
rithm on two classical pulsators from the MeerLICHT data and
benchmarked it against the conditional entropy periodogram.
In both cases, FINKER finds a similar optimal frequency as the
conditional entropy periodogram, with a smaller estimated un-
certainty. Crucially, these frequencies are in close agreement
with values reported in the literature, which are often derived
from datasets with higher resolution or more favourable obser-
vation conditions than those available from MeerLICHT. We plot
the kernel regression periodogram for one classical pulsator tar-
get (CRTS J033427.7–271223) in Figure 10.

**Eclipsing binaries:** The majority of stars exist in binaries or
higher order multiples (Moe & Di Stefano 2017; Offner et al.
2023). If the orbital plane is inclined favourably with respect to
us, we can see the stars eclipse one another as they move through
their orbit. This produces periodic decreases in light that can
range from less than 1% to blocking nearly all of the light from
one star. As most stars are in binaries, eclipsing binaries are com-
monly found in photometric time series. Here, we look at two
eclipsing binaries to demonstrate our method’s ability to reliably
identify binaries with more complicated phase behaviour than
classical pulsators and sinusoidal variables. Figure 11 illustrates
the kernel regression periodograms for the eclipsing binary MLT
J033147.60–281307.9.

**Short period transiting compact binaries:** Ultra short pe-
riod binaries that contain at least one compact component (e.g.
a white dwarf) are important contributors to gravitational wave

4.2. Radial velocity time series

In addition to photometric time series, binary stars and exo-
planet systems experience periodic radial velocity shifts due to
orbital motion. Except for dedicated cases, radial velocity time
series are often sparsely sampled due to various scheduling and
weather condition requirements. In addition to variability aris-
ing from orbital motion, several other phenomena can result in
actual or apparent Doppler shifts at the stellar surface, including
stellar pulsations (Aerts et al. 2010), rotation, and winds. As a
result, there are often multiple sources of variability with differ-
et amplitudes in radial velocity measurements. In this work,
we considered radial velocity time series for a series of known spe-
troscopic binaries that were obtained via spectroscopic observa-
tions with the HERMES Echelle spectrograph on the Mercator
Telescope in La Palma, Spain (Raskin et al. 2011). These data
were obtained with various observing strategies and are spaced
out over several months in some cases and several years in oth-
ers.

We find that this method is extremely powerful for search-
ing sparsely sampled radial velocity time series for periodicities.

In the cases listed in Table 1, the regression periodogram unam-
biguously identifies the orbital frequency despite the different
cases having various amounts of data and signal to noise ratios
for different measurements. Furthermore, as our method does not
make assumptions on the morphology of the signal, it can iden-
tify periodic signals originating from both circular and eccentric
orbits. Finally, while we did not attempt it here, this methodol-
ogy is directly applicable to searches for periodic signals arising
from exoplanets as well.

Figure 13 and 14 illustrate the kernel regression peri-
dograms for HD 114520 and V772 Cas. For the latter, FINKER
finds a frequency close to the literature value with only 10 ob-
servations; however, due to the low number of observations, the
uncertainty estimate is unreliable.

To provide a comprehensive evaluation of our method, we
have included in Appendix D a detailed comparison with the
Table 1. Comparison of optimal frequencies.

<table>
<thead>
<tr>
<th>Target</th>
<th>Type</th>
<th>Conditional entropy $[d^{-1}]$</th>
<th>Kernel regression $[d^{-1}]$</th>
<th>Literature frequency $[d^{-1}]$</th>
<th>Sample size</th>
<th>Mag</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRTS J033427.7–271223</td>
<td>Clas. Pulsator</td>
<td>$1.717876 \pm 4.4 \times 10^{-5}$</td>
<td>$1.717831 \pm 6.0 \times 10^{-6}$</td>
<td>$1.717829^{a}$</td>
<td>1414</td>
<td>V=15.5</td>
</tr>
<tr>
<td>CRTS J033500.6–272854</td>
<td>Clas. Pulsator</td>
<td>$1.433518 \pm 1.0 \times 10^{-5}$</td>
<td>$1.433525 \pm 3.4 \times 10^{-6}$</td>
<td>$1.433527^{a}$</td>
<td>1534</td>
<td>V=14.7</td>
</tr>
<tr>
<td>MLT J033147.60–281307.9</td>
<td>Eclips. Binary</td>
<td>$5.277603 \pm 1.1 \times 10^{-5}$</td>
<td>$2.638751 \pm 4.9 \times 10^{-6}$</td>
<td>$2.63875^{b}$</td>
<td>1350</td>
<td>V=14.0</td>
</tr>
<tr>
<td>MLT J162036.64–614110.5</td>
<td>Eclips. Binary</td>
<td>$3.754604 \pm 1.5 \times 10^{-5}$</td>
<td>$3.754548 \pm 8.6 \times 10^{-6}$</td>
<td>$1.87726^{c}$</td>
<td>808</td>
<td>g=15.0</td>
</tr>
<tr>
<td>ZTF J01016.82–033419.5</td>
<td>Comp. Binary</td>
<td>$12.329054 \pm 3.2 \times 10^{-5}$</td>
<td>$12.329054 \pm 1.7 \times 10^{-5}$</td>
<td>$12.329042^{b}$</td>
<td>352</td>
<td>G=17.4</td>
</tr>
<tr>
<td>ZTF J053708.26–245014.6</td>
<td>Comp. Binary</td>
<td>$10.985491 \pm 3.4 \times 10^{-5}$</td>
<td>$3.050674^{f} \pm 1.0 \times 10^{-5}$</td>
<td>$3.050700^{b}$</td>
<td>138</td>
<td>G=16.7</td>
</tr>
<tr>
<td>ZTF J063808.71+091027.4</td>
<td>Comp. Binary</td>
<td>$1.520599 \pm 8.1 \times 10^{-6}$</td>
<td>$1.520579 \pm 5.1 \times 10^{-6}$</td>
<td>$1.520576^{b}$</td>
<td>735</td>
<td>G=19.0</td>
</tr>
<tr>
<td>ZTF J140702.57+211559.7</td>
<td>Comp. Binary</td>
<td>$6.979319 \pm 1.3 \times 10^{-5}$</td>
<td>$6.979318 \pm 5.9 \times 10^{-6}$</td>
<td>$6.979331^{b}$</td>
<td>366</td>
<td>G=18.1</td>
</tr>
<tr>
<td>HD 165246</td>
<td>RV Binary</td>
<td>$0.217709 \pm 8.6 \times 10^{-4}$</td>
<td>$0.217754 \pm 8.7 \times 10^{-5}$</td>
<td>$0.217737^{b}$</td>
<td>95</td>
<td>V=7.6</td>
</tr>
<tr>
<td>V772 Cas</td>
<td>RV Binary</td>
<td>$0.144847 \pm 3.7 \times 10^{-4}$</td>
<td>$0.188593^{a} \pm 1.5 \times 10^{-2}$</td>
<td>$0.199453^{d}$</td>
<td>10</td>
<td>V=6.7</td>
</tr>
<tr>
<td>HD 114520</td>
<td>RV Binary</td>
<td>$0.002284 \pm 7.4 \times 10^{-6}$</td>
<td>$0.002284 \pm 5.5 \times 10^{-6}$</td>
<td>$0.002283^{c}$</td>
<td>71</td>
<td>V=6.8</td>
</tr>
</tbody>
</table>

Notes. Comparison of the optimal frequencies determined for each target using both the conditional entropy and kernel regression periodograms. (a) Torrealba et al. (2015), (b) Brown et al. (2023), (c) Johnston et al. (2021), (d) Kochukhov et al. (2021), (e) Escorza et al. (2019), (*) frequency verified visually, (†) obtained with FINKER’s adaptive bandwidth, (‡) unreliable uncertainty estimate due to small sample sizes.

Fig. 10. FINKER’s frequency search for CRTS J033427.7–271223. The panel on the left shows the frequency range searched, and the inset illustrates the smooth behaviour of the residuals around the minimum. The estimated best frequency and its uncertainty are shown with a black error bar. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid, illustrating the algorithm’s capability to model the periodic signal accurately.

5. Conclusion and future outlooks

We present FINKER, a nonparametric periodogram for determining the frequencies of variability in astronomical time series, specifically sparsely sampled light curves and radial velocity measurements. At its core, this method performs kernel regression on a time series that has been folded over a proposed frequency and calculates the SSR of the data points with respect to the regression fit. A smaller SSR means a more coherent structure and, therefore, corresponds to the likely period of variability. This method’s main strength is its adaptability to localised data structures and inherent robustness to noise. These features enable the method to capture complex, non-sinusoidal periodic trends arising from numerous different physical mechanisms that faster parametric methods may miss.

We evaluated and benchmarked our method by applying it to synthetic datasets, demonstrating a more accurate frequency estimation than the widely used conditional entropy. Moreover, the robustness of kernel regression against the misidentification of harmonic frequencies is highlighted in Section 3.1.3. The empirical analysis of real-world datasets from the MeerLICHT telescope and other sources further validates the algorithm’s efficacy, showcasing its potential as a reliable tool for period determination in the field of astronomy.

Despite its strengths, the method is not without limitations. It is computationally more demanding than some traditional methods, which may restrict its use in processing extremely large datasets or in applications requiring real-time analysis. Additionally, while the method’s susceptibility to half-frequency errors is mitigated by a simple check for significant doubled frequencies, this step requires additional computation and may not be foolproof in all cases. Specifically, this is the case for eclipsing binaries with similar eclipse depths. Future work will investigate whether a hybrid parametric–nonparametric method such as that proposed by Saha & Vivas (2017) can mitigate this. Furthermore, there are clear failure cases. These cases are limited to the situation where we do not achieve a quasi-uniform distribution.
Fig. 11. FINKER’s frequency search for MLT J033147.60–281307.9. The panel on the left shows the frequency range searched, and the inset illustrates the smooth behaviour of the residuals around the minimum. The estimated best frequency and its uncertainty are shown with a black error bar. FINKER’s framework allowed for a search of both the harmonic and sub-harmonic of the best frequency and correctly identified that there is a difference in the eclipses. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid, illustrating the algorithm’s capability to model the eclipsing signal accurately.

Fig. 12. FINKER’s frequency search for ZTF J041016.82–083419.5. The panel on the left shows the frequency range searched, and the inset illustrates the behaviour of the residuals around the minimum. Transit data shows a less smooth behaviour of the residuals around the minimum, but the frequency and its uncertainty, shown as a black error bar, are consistent with the literature value. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid.
Fig. 13. FINKER’s frequency search for HD 114520. The panel on the left shows the frequency range searched, and the inset illustrates the smooth behaviour of the residuals around the minimum. The estimated best frequency and its uncertainty are shown with a black error bar. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid, illustrating the algorithm’s capability to model the periodic signal accurately.

Fig. 14. FINKER’s frequency search for V772 Cas with an adaptive bandwidth. The panel on the left shows the frequency range searched and the wild behaviour of the residuals due to the low number of points. The inset shows that the best frequency agrees with the literature values; however, the estimated uncertainty is unreliable due to the really low number of observations. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid.

Acknowledgements. CJ gratefully acknowledges support from the Netherlands Foundation.

References

Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology (Springer, Dordrecht)
Efron, B. & Tibshirani, R. 1986, Statistical Science, 1, 54
Fan, J. & Gijbe, L. 1994, in Local polynomial modelling and its applications
Groot, P., Bloemen, S., & Jonker, P. 2019, in The La Silla Observatory - From the Inauguration to the Future, 33
Appendix A: Asymptotic properties

The asymptotic properties of the nonparametric kernel regression models are critical for understanding the large-sample behaviour. These have been thoroughly studied in Fan 1993 and Fan & Gijbels 1994. We summarise the key parameters: bias, variance, and mean squared error (MSE) to assess the asymptotic optimality of our methodology.

Bias For our local constant regression model, the bias is of $O(h^2)$, where $h$ is the bandwidth. In practice, the bias can be significantly reduced by choosing an optimal bandwidth through cross-validation methods:

$$\text{Bias}(\hat{f}_h(x)) = O(h^2). \quad (A.1)$$

Variance The variance of the estimator can be described as $O(n^{-1})$ for both local constant and local linear models:

$$\text{Variance}(\hat{f}_h(x)) = O\left(\frac{1}{nh}\right). \quad (A.2)$$

Mean squared error The MSE is a function of both bias and variance and is given by

$$\text{MSE}(\hat{f}_h(x)) = O(h^4 + \frac{1}{nh}). \quad (A.3)$$

The optimal bandwidth minimises this MSE, and cross-validation techniques are commonly employed to find this balance between bias and variance.

Appendix B: Computational efficiency of grid-based kernel regression

Nonparametric kernel regression can be computationally intensive with large datasets. To improve efficiency, we tested a grid-based method. This approach evaluates the regression function at predetermined grid points rather than at every data point. We first created a grid over the data’s range and computed the regression at these points. Then, we interpolated these values to estimate the regression function for the entire dataset.

We evaluated the computation times of kernel regression across various grid sizes using multiple synthetic light curves, both with and without the implementation of the grid method. The results showed that using a grid significantly reduces computation time, especially for large datasets. However, choosing the right grid size is important to ensure accuracy. We analysed how the grid size affects the variability of residuals and found that a grid with about 300 points is sufficient for stable and consistent squared residuals, regardless of sample size, as shown in Figure B.1.

Figure B.1 presents the average computation times without grid and with grids of different sizes. For smaller sample sizes, direct evaluation on the observations is quicker. But for larger samples, using a grid speeds up the process without affecting the regression’s accuracy.

All tests were run on an Alienware Area 51M, Intel Core i9-9900K, 32GB DDR4/2400, Nvidia GeForce RTX 2080.

Appendix C: Controversial cases

In this appendix, we delve into specific instances where FINKER encountered challenges or yielded results that diverged from established literature.

Case study: KIC 5217733 For KIC 5217733, FINKER’s frequency identification process did not align with the frequencies reported in literature (Kirk et al. 2016). As shown in Figure C.1, the algorithm’s search spanned the expected frequency range but failed to accurately recover the known frequency. This discrepancy is possibly attributed to the underestimation of uncertainties and the presence of numerous g-mode pulsations, which introduced additional variations in the line profile.

Case study: MLT J162036.64–614110.5 Another intriguing case is MLT J162036.64–614110.5, where FINKER’s results diverged from established literature. As depicted in Figure C.2, the algorithm executed a comprehensive search, including analyses of both the main frequency it found and its subharmonic. Interestingly, FINKER did not observe any significant difference in the eclipses, in contrast to literature reports. The light curve, when folded at the frequency identified by FINKER, and overlaid with our kernel regression fit, suggests a different interpretation of the data compared to traditional analyses. This case underscores the potential for FINKER’s framework to provide alternative insights into celestial phenomena, although it also highlights the need for cautious interpretation, especially in instances where results significantly deviate from established knowledge.

Appendix D: Comparison between FINKER and Lomb-Scargle periodogram

This appendix presents a comparative analysis of the frequency identification results obtained using the nonparametric kernel
Fig. C.1. FINKER’s frequency search for KIC 5217733 (Kirk et al. 2016). The panel on the left shows the frequency range searched, and the inset illustrates the failed recovery of the literature frequency. The panel on the right shows the light curve folded at the literature frequency (0.006201 d\(^{-1}\)) with the kernel regression fit (green line) overlaid.

Fig. C.2. FINKER’s frequency search for MLT J162036.64–614110.5. The panel on the left shows the frequency range searched and the thorough evaluation around the best frequency and its sub-harmonic. In disagreement with the results in the literature, FINKER did not find a significant difference in the eclipses. The panel on the right shows the light curve folded at the found frequency with the kernel regression fit (green line) overlaid.

regression method implemented in FINKER and those derived from the Lomb-Scargle periodogram.

The datasets used for this comparison are those discussed in Section 4. For the Lomb-Scargle analysis, we used the implementation provided by the Astropy package in Python. Specifically, we conducted the Lomb-Scargle periodogram analysis over the same frequency range used for FINKER. However, taking advantage of the Lomb-Scargle periodogram’s relatively low computational demands, we employed a denser frequency grid of 5 million points to ensure a thorough and detailed frequency search. The frequency identification process for FINKER was conducted as detailed in Section 3.

Table D.1 summarises the outcomes of this comparison, presenting the absolute differences between the frequencies identified by both methods and the established literature frequencies. Overall, FINKER’s nonparametric kernel regression shows lower absolute errors with respect to the literature values than the Lomb-Scargle periodogram. Notably, FINKER shows superior accuracy in identifying the correct frequencies for classical pulsators and eclipsing binaries. Conversely, for compact binaries and certain radial velocity targets, both methods yielded comparable results, with exceptions noted for ZTF J140702.57+211559.7 and HD 114520. In these cases, the Lomb-Scargle periodogram identified an incorrect frequency or double the correct frequency. However, for V772 Cas with only ten observations, the Lomb-Scargle periodogram identified a frequency closer to the literature value than FINKER, illustrating scenarios where parametric methods can outperform nonparametric if the underlying signal closely matches the assumed model.

The comparison between FINKER and the Lomb-Scargle periodogram highlights the strengths and limitations of both methods. FINKER offers advantages in handling complex and noisy data, while the Lomb-Scargle periodogram remains a valuable tool for quick periodicity searches, especially when computational resources are limited. Future work could explore hybrid approaches that combine the robustness of nonparametric methods with the efficiency of parametric periodograms to enhance the accuracy and speed of frequency identification in astronomy.
### Table D.1. Differences in optimal frequencies.

<table>
<thead>
<tr>
<th>Target</th>
<th>$\Delta$ Lomb-Scargle</th>
<th>$\Delta$ Kernel regression</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRTS (1)</td>
<td>$3.9 \times 10^{-4}$</td>
<td>$2.0 \times 10^{-6}$</td>
<td>1414</td>
</tr>
<tr>
<td>CRTS (2)</td>
<td>$8.2 \times 10^{-5}$</td>
<td>$2.0 \times 10^{-6}$</td>
<td>1534</td>
</tr>
<tr>
<td>MLT (1)</td>
<td>$2.638773$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>1350</td>
</tr>
<tr>
<td>MLT (2)</td>
<td>$1.877300$</td>
<td></td>
<td>808</td>
</tr>
<tr>
<td>ZTF (1)</td>
<td>$1.8 \times 10^{-5}$</td>
<td>$1.2 \times 10^{-5}$</td>
<td>352</td>
</tr>
<tr>
<td>ZTF (2)</td>
<td>$2.5 \times 10^{-5}$</td>
<td>$2.6 \times 10^{-5}$</td>
<td>138</td>
</tr>
<tr>
<td>ZTF (3)</td>
<td>$7.8 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-6}$</td>
<td>735</td>
</tr>
<tr>
<td>ZTF (4)</td>
<td>$6.979322$</td>
<td>$1.3 \times 10^{-5}$</td>
<td>366</td>
</tr>
<tr>
<td>HD (1)</td>
<td>$2.6 \times 10^{-5}$</td>
<td>$1.7 \times 10^{-5}$</td>
<td>95</td>
</tr>
<tr>
<td>V772 (1)</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>10</td>
</tr>
<tr>
<td>HD (2)</td>
<td>$0.998183$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>71</td>
</tr>
</tbody>
</table>

**Notes.** Absolute differences between the frequencies identified by the Lomb-Scargle and kernel regression periodograms compared to literature values for the same sources as Table 1.