A&A 643, A160 (2020)
https://doi.org/10.1051/0004-6361/201630164

Astronomy & Astrophysics

Periodicity of quasar and galaxy redshift

Arindam Mal[1], Sarbani Palit[2], Ujjwal Bhattacharya[2], and Sisir Roy[3]

[1] Laboratory For Electro-Optic System, 1st Cross, 1st Stage Peenya Industrial Estate, Bangalore 560013, India
e-mail: arindam.mal1982@gmail.com
[2] Computer Vision And Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India
[3] National Institute of Advanced Studies, IISC Campus, Bangalore, India

Received 29 November 2016 / Accepted 12 August 2020

ABSTRACT

Context. An approach for studying the large-scale structure of the Universe lies in the detection and analysis of periodicity of redshift values of extragalactic objects, galaxies, and quasi stellar objects (QSO), in particular. Moreover, the hypothesis of the existence of multiple periodicities in the redshift distributions deserves exploration. The task is compounded by the presence of confounding effects and measurement noise.
Aims. Studies of periodicity detection in redshift values of extragalactic objects obtained from the Sloan Digital Sky Survey (SDSS) have been conducted in the past, largely based on the Fourier transform. The present study aims to revisit the same thing using the singular value decomposition (SVD) as the primary tool.
Methods. Periodicity detection and the determination of the fundamental period have been performed using a standard spectral approach as well as a SVD-based approach for a variety of simulated datasets. The analysis of the quasar redshift data from DR12 and galaxy redshift dataset of DR10 from SDSS data has been carried out.
Results. A wide range of periodicities are observed in different redshift ranges of the quasar datasets. For redshifts greater than 0.03, a period length of 0.2094 was determined while periodicities of 0.1210 and 0.0654 were obtained for redshift ranges (0.03, 1) and (3, 4), respectively. Twin periodicities of 0.1153 and 0.0807 were obtained for the redshift range (1, 3). Determining the ranges to be examined has been done based on histogram computation; the binwidths of which have been obtained by employing a kernel density estimation. The redshift sequence for the galaxy samples exhibits a somewhat different nature, but still contains periodic components. Twin periodicities of 0.0056 and 0.0079 were observed for a redshift range greater than 0.03.
Conclusions. Galaxy and quasar redshift values form sequences, which are not only discrete in amplitude but also contain periodic components. The superiority of the singular value decomposition method over the spectral estimation approach, in redshift periodicity analysis especially from the point of view of robustness, is demonstrated through simulations. The existence of periodicity for quasar and galaxy families is thus firmly established, lending further support to the Hoyle-Narlikar variable mass theory.

Key words. large-scale structure of Universe – X-rays: galaxies – infrared: galaxies – quasars: absorption lines – quasars: emission lines

1. Introduction

It has been long observed that the spectrum of an extragalactic object often exhibits a shift in the spectral lines towards the red end, that is, the lines appear at wavelengths longer than their standard values. A wavelength of the spectrum typically shifts from λ_0 to λ_*, where $\lambda_* = \lambda_0 + \Delta\lambda$. The relative change in wavelength is termed the redshift z and is defined as $z = \Delta\lambda/\lambda_0$. Initially, astronomers ascribed the redshift of the spectrum to the Doppler effect, based on Einstein's general theory of relativity and the theory of an expanding universe. According to this school of thought, the redshift is thought of as a simple continuous Doppler-like effect as a result of expansion of the universe. However, non-uniformities in the redshift distribution of quasars have been discussed by many scientists, that is, they have observed some peaks near the quasar redshift $z = 0.30, 0.60, 0.96, 1.41, 1.96$, and 0.061 and its multiples (Napier & Burbidge 2003). Cocke & Tifft (1991) found a quantised differential redshift distribution for double galaxies. In 1993, Lehto (1990) suggested a mechanism for predicting periodicities. Subsequently, many other scientists proposed various approaches pointing towards the existence of periodicity of redshift for quasars.

Bell & Fort (1973) observed that quasar redshift can be decomposed into the following two components:

$$1 + z = (1 + z_c).(1 + z_x).$$

The component z_c is defined as the cosmological term, while the component z_x is a redshift of an unknown origin in the source. It may be noted that z_x can arise on account of various causes, such as the effects of passing through space over time, but it is not due to the expansion of the universe or the Doppler effect.

According to Depaquit et al. (1985), the periodicity in the quasar redshift can be attributed to three factors:
– the presence of a selection effect during data sampling;
– the non-randomness of quasar distribution in space; and
– the existence of Dopplerian and non-Dopplerian terms in redshift.

The periodicity can be partly attributed to the non-Dopplerian effect (Holba et al. 1994) as well. This line of approach implies that quasar redshifts derive their origin from non cosmological effects. An explanation of periodicity existing in the data samples of the galaxy can be found in Tifft (1997). The theory

of variable mass hypotheses, proposing the ejection of quasars from galactic nuclei, was proposed by Narlikar & Das (1980), Narlikar & Arp (1993). Recently, the mystery of the redshift was explained for quasar galaxy pairs using the Wolf effect in Roy et al. (2000).

Burbidge found that the periodicity in the redshift distribution is in a discretised form and proposed its formulation as $Z_{\rm obs} = 0.061n$ (Burbidge 1968; Burbidge & Burbidge 1969). Karlsson (1971) noticed the existence of redshift periodicity in $\log(1 + z)$ with a period of 0.089; which gives observed maxima at $z = 0.3$, 0.6, 0.96, 1.41, and 1.96. This is in agreement with the conclusions in Barnothy & Barnothy (1976). An earlier investigation by Arp et al. (1990) of various quasar samples revealed periodicity under some constraint. They arrived at the conclusion that the existence of periodicity of a quasar and galaxies is counter evidence of the currently believed cosmological origin of quasar redshift. In particular, it was concluded in Arp et al. (1990) that "If the quasar redshifts are caused by the expansion of space at large distances then periodicity would violate the cosmological principle that the universe must look the same from all points within it". Duari et al. demonstrated the existence of redshift periodicity for quasars using power spectrum-based techniques as well as statistical methods in Duari et al. (1992), Duari (1997). Recently, Fulton et al. (2018) analysed SDSS7 data for quasars and attempted to understand the nonuniformities in redshift distributions.

It was later pointed out in Roeder & Dyer (1972), Basu (2005) that global redshift distribution cannot be used because of the selection effect. Burbidge and O'Dell observed a peak around the redshift of 0.03 instead of a peak at 0.06 and concluded that periodicity depends on data binning (Burbidge 1972). It was further concluded by Scott that the selection effect is the cause of periodicity (Scott 1991). Hawkins et al. (2002) found the nonexistence of periodicity in the redshifts of the 2dF Quasar Redshift Survey (2QZ) (Hawkins et al. 2002). In 2003, however, Napier and Burbidge challenged Hawkins' methodology (Napier & Burbidge 2003). In 2005, Tang & Zhang selected the Sloan Digital Sky Survey (SDSS) sample of quasars and galaxies and reported that there is no statistical association among quasars and galaxies. They argued that the periodicity claimed by others at around $\Delta \log(1+z) = 0.089$ is a coincidence that does not hold for significantly large (fifteen times larger than the ones used earlier) but selected samples of quasar redshifts (Tang & Zhang 2005).

Though the existence of periodicity in quasar and galaxy redshift distribution has been a matter of much contention, it has always been an important research field from both an observational and statistical point of view. However, the detection of periodicity has commonly and largely proceeded using power spectral analysis (PSA), which chiefly employs the periodogram for estimating the spectral density (Bajan & Flin 2007; Duari et al. 1992). The periodogram is a popularly used technique for spectral estimation, which is most effective when the test signal contains a sinusoid along with some harmonics and noise. However, it may be quite confusing when the periodic part of the data is somewhat irregular, containing a large number of overtones and a relatively weak fundamental component. Further, the periodogram may also not be adequate for the quasiperiodic nature of the redshift data for quasars, which does not even have a fixed period length. Quasiperiodic processes are popularly modelled as having two components – one representing the periodic part, but with both amplitude and period length being time-varying, and the other component being white Gaussian noise. Techniques such as a time series analysis and a prediction of the amplitude and period have been attempted separately, for example, using smoothness priors modelling in Kitagawa & Gersch (1996).

This article attempts to provide an entirely different approach to exploring the existence of periodicity in quasar and galaxy redshift data. The chief tool employed for the computations is the Singular Value Decomposition (SVD), which presents certain important advantages over the previous approaches. Firstly, an SVD-based periodicity determination (Kanjilal & Palit 1995) provides an appropriate tool to give a quantitative measure of how close the quasiperiodic signal is to being periodic and a reasonable estimate of the corresponding period length. Secondly, an SVD analysis is able to break down the quasiperiodic signal into base vectors of any shape. Hence, it is more appropriate for characterising quasiperiodic signals, rather than the Fourier Transform-based periodogram, which only uses sinusoids as the basis functions. Fourier-based methods aim to decompose a signal in terms of complex exponentials or sinusoids. When the constituent periodic components are actually sinusoidal in nature, this approach is indeed the most appropriate. However, when the periodic components are of a more general nature, an approach that can directly express the signal in terms of the periodic component is expected to perform better in terms of accuracy and efficiency. We note that these components may be quasiperiodic in nature with a time-varying amplitude and period.

In this article, both spectral analysis and an SVD-based approach are applied to examine the existence of periodicity and determine the fundamental period, if any. The SVD-based approach employs two different measures for periodicity detection and for the measurement of period length. A performance observed for simulated noisy data, for which the ground truth is known, is used to demonstrate the superiority of the SVD-based measures. Consequently, the SVD-based measures are used to explore the presence of periodicity in the other data sets.

It is important to mention that similar to the discrete redshift, there exist discrete structures at a large scale, for example, in the case of a supercluster and void. Recently, Roberts (2002) proposed the quantum perturbative approach to understand the discrete redshift. However, the assumptions made when using this model need more justifications before considering it as a viable theoretical model for discrete redshifts for a galaxy and quasar.

Section 2 lays out the principles of the methodology used. The histogram and bin width computations from the data are explained in Sect. 2.1. The SVD-based approach and the measures defined thereof have been explained in Sect. 2.2, while the principle behind the PSA approach-based periodogram has been outlined in Sect. 2.3. This study employs the following three kinds of data *viz*: simulated data, mock data sets (Rodríguez-Torres et al. 2017), and actual data sets obtained from the SDSS online resources[1]. The results using simulated data have been reported and discussed in Sect. 3. The results for the application on data sets have been presented in Sect. 4. Section 4.1 contains the results using mock data sets, while those using actual cosmological data have been included in Sect. 4.2. A description of the data sets has been given in Sect. 4.2.1, followed by a discussion on some important issues regarding the selection of the data from the data sets in Sect. 4.2.2. Section 4.2.3 elaborates on the results obtained. A physical interpretation of the results have been provided in Sect. 5. Section 6 concludes the article.

[1] http://vizier.u-strasbg.fr/viz-bin/VizieR-2

2. Proposed methodology

In order to test a redshift data set for the presence of periodicity, a histogram is computed. Periodicity of the histogram would imply the existence of a repeating pattern over a redshift interval. Determining the period length is proposed to be accomplished using the popular periodogram as well as the SVD-based approach. Their performances would be compared for simulated data and, finally, the more efficient technique would be used for analysing mockdata and actual redshift data.

2.1. Computation of the histogram

The selection of an appropriate bin width, for a particular data set, is made by minimising a specified cost function, described in Sect. 2.1.1, for the overall data set.

2.1.1. Optimal bin width selection using kernel density estimator

The bin width of the histogram of the redshift values is estimated using a kernel density estimation. A kernel density estimator provides a non-parametric approach for estimating the probability density function of a random variable. Let $(x_1, x_2, \ldots, x_n)$ be a set of samples with an underlying probability density function (f). The problem at hand requires the computation of the shape of this function as accurately as possible. Then its kernel density estimator is given by:

$$f'_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i). \quad (1)$$

Here, K_h is the kernel defined for the bandwidth h, which is also referred to as the smoothing parameter. The selection of the bandwidth (bin width in case of the histogram) is important for a time series data analysis since both over-smoothing or under-smoothing can lead to a misrepresentation of the data. An approach for determining the optimal bandwidth can be constructed by minimising the mean integrated square error (MISE) (Shimazaki & Shinomoto 2010, 2007), defined as:

$$\mathrm{MISE}(h) = \mathcal{E}[\int (f'_h(x) - f(x))^2 \mathrm{d}x], \quad (2)$$

where $\mathcal{E}$ refers to the expectation operator. Hence, this approach is used to determine the optimal bin width, which yields the "best" histogram of the data. The search for periodicity is conducted on the histogram that has thus been obtained. It may be noted that the existence of a periodicity of N in the histogram would imply the existence of a redshift periodicity of $N * W$, where W stands for the optimal bin width.

2.1.2. Algorithm for optimal bin width computation

The quasar redshift data set under consideration (SDSS DR-7, DR-9, or DR-10) is used for an optimal bin width selection. As already mentioned, the bin size is optimised by minimising the MISE, which defines the selected bin width (Shimazaki & Shinomoto 2010, 2007). The steps of the algorithm used for determining the optimal bin width are outlined now. Step I requires dividing the data range into N bins of width δ and counting the number of quasars K_i that fall into the ith bin; the sequence K_i, $i = 1, \ldots, N$ is the histogram sequence for the bin width δ. In Step II, we calculate the mean μ and variance σ^2 for the overall data range. Computation of the cost function, $C(\delta) = 2\mu - \sigma^2/\delta^2$ is carried out in Step III. Finally, Steps I– III are repeated for varying δ. The optimal bin width δ_{opt} is the value of δ for which $C(\delta)$ is minimised.

2.2. Periodicity determination using SVD

The definition of the SVD and, hence, its use in the analysis of periodic and quasiperiodic signals is presented here. Following this groundwork, the measures for determining the periodicity content of a signal are described. Their computation is explained in Sect. 2.2.2.

2.2.1. Matrix formation and application of SVD

The SVD of an $m \times n$ matrix A can be expressed as:

$$\mathbf{A}_{m\times n} = \mathbf{U}_{m\times m}\,\mathbf{\Sigma}_{m\times n}\,\mathbf{V}^T_{n\times n} = \sum_{i=1}^{p} \mathbf{u}_i \sigma_i \mathbf{v}_i^T$$

where $\mathbf{U}_{m\times m} = [\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_m]$ and $\mathbf{u}_i$, $i = 1, \ldots, m$ are the left singular vectors; $\mathbf{\Sigma}_{m\times n} = [\sigma_1\ \sigma_2 \ldots \sigma_n]$ are the singular values of $\mathbf{A}_{m\times n}$ such that $[\sigma_1 > \sigma_2 > \cdots \sigma_p]$; $\mathbf{V}_{n\times n} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n]$ and $\mathbf{v}_i$, $i = 1, \ldots, n$ are the right singular vectors; and $\mathbf{U}^T_{m\times m}\mathbf{U}_{m\times m} = \mathbf{I}_{m\times m}$, $\mathbf{V}^T_{n\times n}\mathbf{V}_{n\times n} = \mathbf{I}_{n\times n}$, and $p = \min(m, n)$. The rank of $\mathbf{A}_{m\times n}$ is equal to the number of non-zero singular values σ_i, that is, p. When each row of the matrix corresponds to a consecutive section of a perfectly non-periodic data sequence, the matrix is of full rank, that is to say the number of non-zero singular values are the minimum of n and m.

When the data set is perfectly periodic, it has the same repeating pattern. If each row of the data matrix contains a period of the sequence, the matrix has rank one and hence only one singular value. However, when the sequence has a strong primary periodic component with a period of n, then $\sigma_1 \gg \sigma_2$, that is, there is only one large singular value and $(\min(m, n) - 1)$ small singular values.

For the present analysis, the histogram sequence formed from the redshift data may be considered as a quasiperiodic signal with some amount of noise, which is commonly found in spectroscopic data. It follows that SVD may be used to detect the presence of periodicity in the histogram sequence. It is further employed to provide a measure of periodicity that would indicate the strength or weakness of the periodic content of the sequence. In order to model the sequence as the sum of a periodic sequence and noise, the data matrix A_{aper} is formed such that each of its rows is constituted by consecutive sections of the histogram sequence. The data matrix thus formed can be decomposed as:

$$A_{\mathrm{aper}} = A_{\mathrm{per}} + E,$$

where E is the residual and A_{per} corresponds to the data matrix of the periodic part with its row length exactly equalling the period length. Since A_{per} is perfectly periodic, it only has one singular value and has an SVD-based decomposition of $A_{\mathrm{per}} = u_1\sigma_1 v_1^t$. Hence, the residual may be expressed as

$$E = A_{\mathrm{aper}} - u_1\sigma_1 v_1^t.$$

It follows that the SVD decomposition of A_{aper} results in the ratio of σ_1/σ_2 being large when the sequence is almost periodic, and small when aperiodic.

2.2.2. Measures for evaluation of periodicity content

Two different SVD-based measures have been considered for the determination of periodicity. Measure 1 (SVR1) is defined as the square of the ratio of the first two singular values of the data matrix.

$$\mathrm{SVR1} = \left(\frac{\sigma_1}{\sigma_2}\right)^2. \tag{3}$$

When the data consists of a perfectly periodic sequence in the presence of some noise and each row of the matrix contains a period of the noisy data, this measure is very high. So, a plot of SVR1 for a varying row length n of the data matrix registers a peak when the row length matches the correct period length exactly. Peaks would also occur at multiples of the fundamental period, lending strength to the proper identification of the period length.

Measure 2 (SVR2), a second measure of periodicity is derived based on a quantification of the aperiodicity of a signal δ_{aper} (Palit & Kanjilal 1994):

$$\delta_{\mathrm{aper}} = \frac{(\sum_{i=2}^{p} \tau_i)^2}{1 + \sum_{i=2}^{p} \tau_i^2} \quad \text{where} \quad \tau_i = \frac{\sigma_i}{\sigma_1}. \tag{4}$$

Then, SVR2 is defined as:

$$\mathrm{SVR2} = \left(\frac{1}{\delta_{\mathrm{aper}}}\right)\frac{(p-1)^2}{p}. \tag{5}$$

We note that the construction of SVR2 is such that it has the same range of values as SVR, that is $[1, \infty)$. The determination of the period length using SVR2 follows the same procedure exactly as described above for SVR1.

A plot of SVR1 or SVR2 for different candidate period lengths has been referred to as the singular value ratio (SVR) spectrum (Kanjilal & Palit 1995). While SVR1 is a simple indicator of inherent periodicity, there are occasions when SVR2 contains more information. A pathological case occurs when $\sigma_1 = \sigma_2$ and $\sigma_3 = \sigma_4 = \cdots = 0$.

The steps of the algorithm for the determination of periodicity, if any, of a quasar redshift data set and the corresponding period length can now be summarised into the following steps. Step I involves the formation of a histogram sequence from the quasar redshift data set under consideration, using δ_{opt} as the bin width, determined as in Sect. 2.1.1. In Step II, for a candidate period length n, with which the data set may be periodic, the corresponding data matrix A is constructed by partitioning the sequence into sections of length n and placing each section (aligned in phase) as a row of A. Observations from m periods are assumed to be available. Computation of the SVD of A and, hence, SVR1 and SVR2 is performed in Step III. Steps II–III are repeated for $n = 3, \ldots, m/3$. From a plot of SVR1 (or SVR2) versus n, determine n_{per}, the value of n for which the plot shows a peak. The existence of peaks, slowly decreasing in magnitude, at multiples of n_{per} confirm the fundamental period P. This forms the chief component of Step IV. In the last step, that is, Step V, the base period of the periodic component may now be calculated using the optimal bin width δ_{opt} determined as explained in Sect. 2.1.2, according to:

$$\text{Base period} = \delta_{\mathrm{opt}} \times P. \tag{6}$$

2.3. The periodogram

A power spectral analysis or estimation of the power density spectrum for cyclicity determination has been a very popular tool for a long time. The power density spectrum of a stochastic signal $x[n], n = 0, \ldots, \infty$ may be defined as the discrete time Fourier transform (DTFT) of its autocorrelation sequence $\phi_{xx}[n]$. For the convenience of the reader, the definitions for a stationary random signal $x[n]$ are given below. The autocorrelation sequence is defined as:

$$\phi_{xx}[n] = \mathcal{E}\{x[m+n]x^*[m]\}, \tag{7}$$

where $\mathcal{E}$ stands for the expectation operator and $*$ refers to the operation of complex conjugation. The power density spectrum, that is, DTFT of the autocorrelation sequence, is defined as:

$$\Phi_{xx}(e^{j\omega}) = \sum_{n=0}^{\infty} \phi_{xx}[n] e^{-j\omega n}. \tag{8}$$

The periodogram is popularly used, owing to the simplicity of its computation, to estimate the power density spectrum, using the discrete Fourier transform (DFT). For a finite length signal $x[n], n = 0, \ldots, N-1$, the following definitions may be stated for the DFT:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, \quad k = 0, \ldots, N-1. \tag{9}$$

The L-equidistant estimates of the power spectral density ($L \geq N$) may be computed by the periodogram as follows:

$$\begin{aligned} P_{xx}(k/L) &= 1/N\,|X[k]|^2 \\ &= 1/N \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/L} \right|^2, \quad k = 0, \ldots, L-1. \end{aligned} \tag{10}$$

However, since the periodogram is known to be an inconsistent estimator of the power density spectrum, the Hamming window is used to smooth the input data, giving the modified periodogram,

$$P_{xx}^{\mathrm{mod}}(k/L) = 1/N \left| \sum_{n=0}^{N-1} x[n] w[n] e^{-j2\pi kn/L} \right|^2, \quad k = 0, \ldots, L-1, \tag{11}$$

where $w[n]$ stands for the Hamming window, defined as:

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/L), & 0 \leq n \leq L \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Equation (11) has been implemented for the computations here.

3. Results of simulations

In order to test the strength of the SVD-based algorithm regarding its ability to detect periodicity, its validity has been examined for different kinds of data sets. Monte Carlo simulations have been run in order to extensively study the effect of noise and the presence of other periodic components, varying their proportions. Results using the periodogram-based approach have been presented for comparison. Finally, the performance of the SVD-based measures for mock data sets as well as various types of actual quasar and galaxy data sets has been studied.

Table 1. 95% confidence intervals computed over 1000 simulations.

S/N	95% confidence interval	
	Modified periodogram	SVD-based approach
46.75	[24.9, 24.9]	[25, 25]
26.87	[24.9, 24.9]	[25, 25]
15.16	[24.9, 24.9]	[25, 25]
12.88	[8.02, 24.9]	[25, 25]
6.77	[6.07, 24.9]	[25, 25]
0.80	[4.88, 24.9]	[25, 25]
−2.75	[4.08, 24.9]	[25, 25]

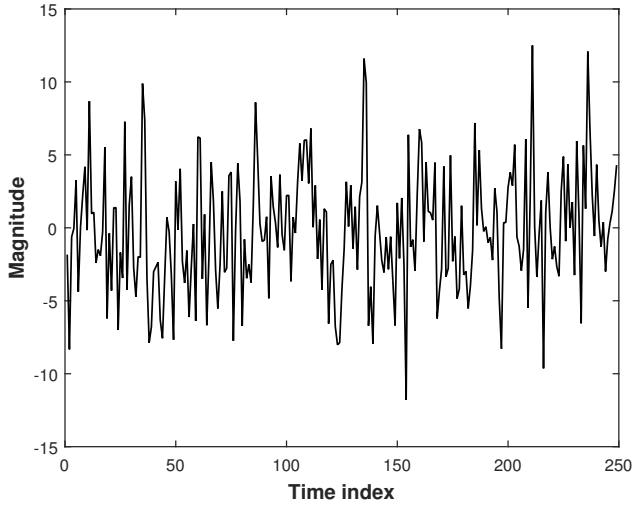

Fig. 1. Simulated data set no. 1.

Two kinds of simulations have been conducted for the present study. For case 1, the data set was generated as a periodic signal with a period length of 25 along with additive white Gaussian noise. The proportion of signal and noise strengths was measured using the quantity signal-to-noise ratio (S/N), defined as the ratio of signal and noise powers, usually expressed in decibels (dB).

$$S/N = 10 \log_{10} \frac{\text{Signal Power}}{\text{Noise Power}}$$

We observe that when the signal and noise strengths are equal, the S/N is zero while it is negative when the noise strength is greater than the signal strength. The determination of the period length using SVD as described in Sect. 2.2.2 was carried out 1000 times and the 95% confidence interval has been reported for each S/N value in Table 1. Figure 1 shows simulated data set no.1 as an example of the data sets constructed for an S/N of 4.4. Figure 2 shows its SVR spectrum, while Fig. 3 shows the corresponding modified periodogram. The red lines indicate the x-axis locations, corresponding to the correct period length.

With regard to case 2, in addition to the two components mentioned above, that is a periodic component with a period length of 25 and noise, the presence of another periodic component with a period length of 17, but with a strength that is 5% of the first (and dominant) one, is considered here. Figure 4 shows simulated data set no. 2 as an example of this type of data set with an $S/N = 4.8$ for the dominant component. The determination of the period length of the dominant component was carried out and the 95% confidence interval reported over

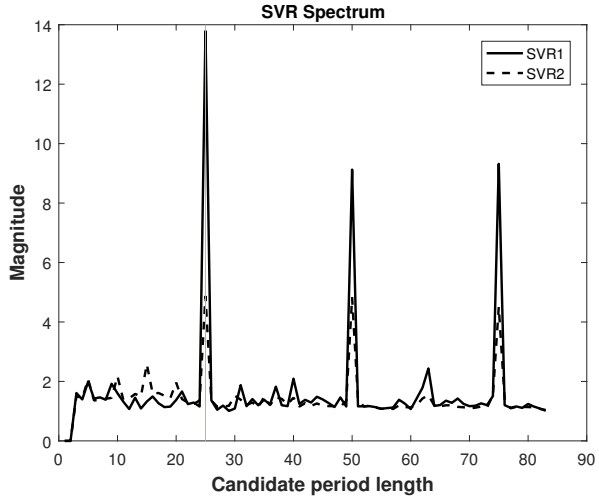

Fig. 2. SVR spectrum of simulated data set no.1: Period length detected = 25.

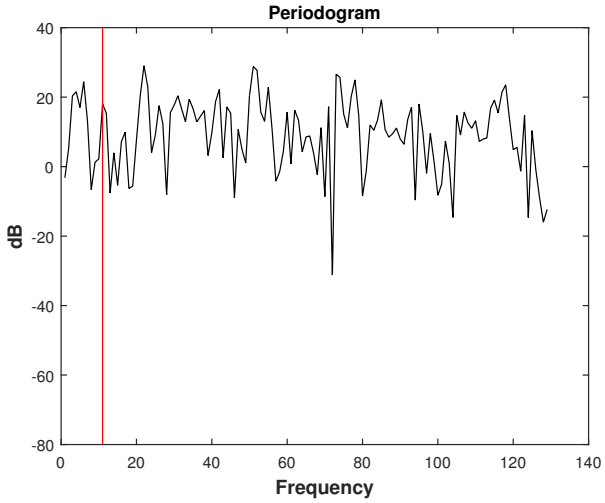

Fig. 3. Modified periodogram of simulated data set no. 1: Correct period length was not detected.

1000 iterations for each S/N value in Table 2. Figure 5 shows the SVR spectrum, while Fig. 6 shows the corresponding modified periodogram. The red lines indicate the x-axis locations corresponding to the correct period length. The SVR spectrum shows peaks at 25 and its multiples. The modified periodogram, on the other hand, exhibits only a local peak at the correct period length and hence fails to identify it.

We note that in Table 1, that is, for the detection of a periodic signal with a period length of 25 and buried in noise, the 95% confidence interval to S/N values as low as −2.75, is [25, 25]. The direct implication of this result is that for S/N values ranging from 46.75 to −2.75, the probability of detecting a spurious peak is not more than 0.05. Furthermore, as seen in Table 2, that is, for the detection of a periodic signal with a period length of 25 from a mixture of noise and another periodic component of strength of 5% of that of the first one, the 95% confidence interval to S/N values as low as −0.83, is [25, 25]. In this case, for S/N values ranging from 25.6 to −0.83, the probability of detecting a spurious peak is not more than 0.05. It may be concluded for the two cases, that for S/N values greater than −0.83, the probability of an erroneous peak detection is not more than 0.05.

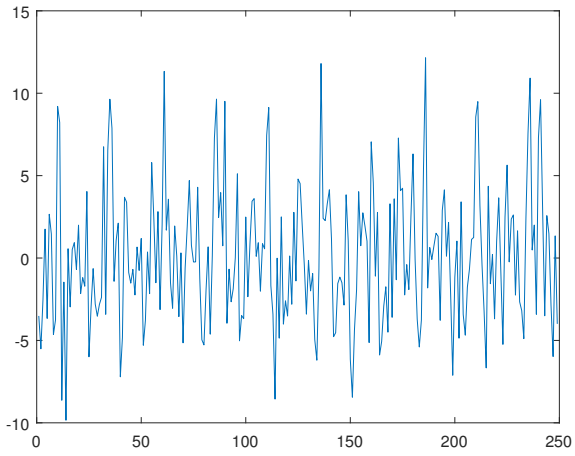

Fig. 4. Simulated data set no. 2.

Table 2. 95% confidence intervals computed over 1000 simulations.

S/N	95% confidence interval	
	Modified periodogram	SVD-based approach
25.6	[24.9, 24.9]	[25, 25]
15.74	[24.9, 24.9]	[25, 25]
13.62	[24.9, 24.9]	[25, 25]
12.22	[24.9, 24.9]	[25, 25]
11.88	[7.88, 24.9]	[25, 25]
6.32	[6.07, 24.9]	[25, 25]
2.96	[6.07, 24.9]	[25, 25]
0.84	[4.86, 24.9]	[25, 25]
−0.83	[4.87, 24.9]	[25, 25]

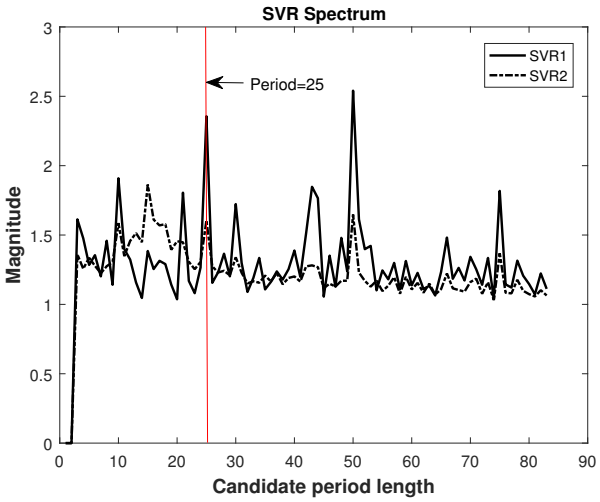

Fig. 5. SVR spectrum of simulated data set no. 2: Period length detected = 25.

4. Application on data sets

Both mock data sets and actual cosmological data sets have been used for the study.

4.1. Performance on mock data sets

The mock data sets used here have been obtained from mock catalogues that were created using the 2.5 Gpc h^{-1} BigMultiDark

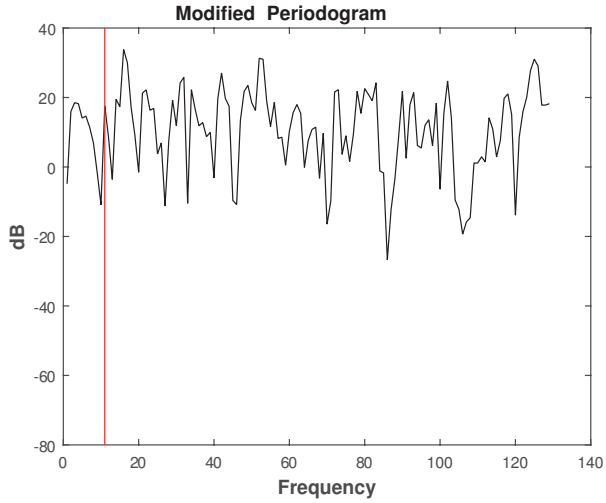

Fig. 6. Modified periodogram of simulated data set no. 2: Correct period length was not detected.

simulation (Rodríguez-Torres et al. 2017). The entire redshift range was simulated by forming the following two different light-cones with the same area (1481.43 $\deg^2$), *viz:* BigMDPL-QSO, which was designed to replicate the redshift range $0.9 < z < 2.2$, using a single set of parameters for the entire range, and BigMDPL-QSOZ, which was designed to include the evolution of the clustering, using different sets of parameters to replicate the clustering in the ranges $0.9 < z < 1.2$, $1.2 < z < 1.5$, $1.5 < z < 1.8$, and $1.8 < z < 2.2$.

The light-cone data files have the ".dat" extension, while the corresponding random catalogues that are also provided have the ".ran" extension. The random catalogues were built by shuffling the redshift from the data sets and then adding them to random angular coordinates. The authors claim that both random catalogues are compatible, having the same dependency of the number density with redshift.

The period lengths obtained by analysing the mock data sets using the SVD based measures SVR1 and SVR2 are presented in Table 3. The results using the periodogram have not been included as the performance of the former has been found to be more efficient for the task at hand.

The data provided by the random catalogues with the ".ran" extension were analysed using the kernel density estimator in order to compute the optimal bin width. However, an optimal value could not be obtained which can be ascribed to the random nature of the data leading to the absence of a definite underlying density. Consequently, the existence of periodicity in these type of data sets can be ruled out.

For each mock data set, the periodicity or fundamental period length, determined as the product of the peak location of the histogram and the optimal bin width (computed using a kernel density estimation) for SVR1 and SVR2 have been mentioned in Col. 4. Column 5 indicates the 95% confidence interval observed for over 1000 experiments. For each of these experiments, re-sampling of the data was performed. A noise component, which was held constant over all of the 1000 experiments, was added in order to ensure that the optimal bin width remained roughly of the same order as the value obtained without re-sampling. A peak location corresponds to the desired period length if multiples of that peak location are peak locations themselves as well.

Table 3. Period length determination in mock data sets.

Data set	Bin size	Method	Period length	95% confidence interval
`BigMDPL-QSO_Y1Q_1481.43deg2.dat.txt`	0.0060	SVR1	0.0660	[0.0541, 0.1138]
		SVR2	1.016	[1.0126,1.017]
`BigMDPL-QSOZ_Y1Q_1481.43deg2.dat.txt`	0.0058	SVR1	0.0756	[0.0545,0.1047]
		SVR2	1.0159	[1.0128,1.0172]

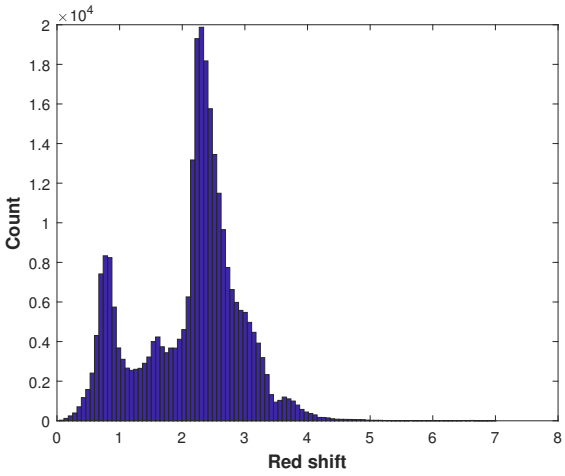

Fig. 7. Histogram of QSO redshifts of DR-12 data set.

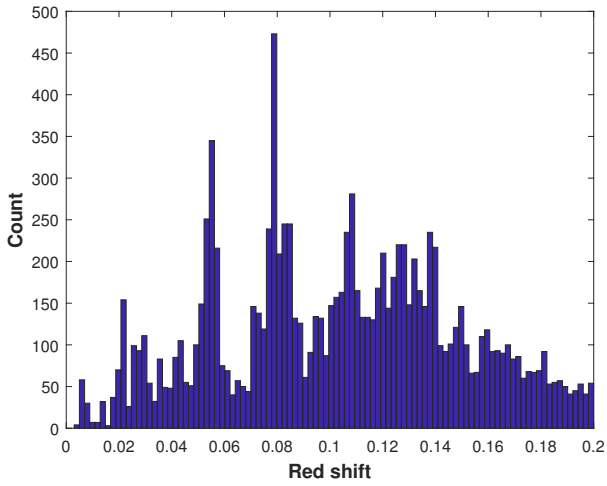

Fig. 8. Histogram of galaxy redshifts of DR-10 data set.

4.2. Performance on cosmological data sets

4.2.1. Description of the data sets

Two real-life data sets, obtained from, which have been tested using the SVD-based approaches, are described now. One of them was the combined data set of quasar DR-12, known as DR12Q catalogue (Pâris et al. 2017) was constructed from the Baryon Oscillation Spectroscopic Survey (BOSS) of the Sloan Digital Sky Survey III. The catalogue also includes known quasars (mostly from SDSS-I and II) that were observed by BOSS. The catalogue contains 166 583 quasars of which 74 454 are new discoveries since SDSS-DR9, which detected over 6373 deg^2. The data set contains redshift values ranging from −0.0049 to 7.0112 and with luminosities whose absolute magnitudes are less than −20.5. The histogram of the data for redshift values greater than 0.03 is shown in Fig. 7.

The other dataset consists of the galaxy redshifts of DR-10. Poudel et al. (2016) cross-matched the SDSS DR10 group catalogue with GAMA Data Release 2 and Wide-field Survey Explorer (WISE) data to construct a catalogue of 1651 groups and 11 436 galaxies, containing photometric information on 15 different wavebands ranging from ultraviolet (0.152 μm) to mid-infrared (22 μm).The redshift values were observed to range from 0.003 to 0.2. The histogram of the redshift values is shown in Fig. 8.

4.2.2. Data selection

For each data set, a bin width selection has been performed using the kernel density-based estimator. Only redshifts greater than 0.03 were used for our computations. For the DR10 and DR12 data sets, only those values for which the flag equals 0 were retained for the analysis as only these data values fall within the 95% confidence interval.

4.2.3. Results

The complete performance results have been tabulated in Table 4. For each range of redshift, a bin-size was fist computed. Here, Col. 5 gives the fundamental period length, while Col. 6 gives the 95 % confidence interval obtained over 1000 experiments. The experiments were performed in a similar fashion to those conducted for mock data. As in Table 3, a strong peak location is identified to correspond to the dominant period length if there are peaks at multiples of this peak location.

For each data set, a plot of the SVR spectrum that was obtained using the SVD-based measure SVR1 corresponding to a single redshift range is presented now for a further illustration and clarification of periodicity. Figure 9 shows the plot of SVR1 for the QSO redshift range $1 < z < 3$ of the DR-12 data. Figure 10 presents the plot for the galaxy redshift data of DR-10. For both the plots, the dominant period length given by the strongest peak location and the peak location at its closest multiple, have been indicated by red lines. The period lengths may be verified from Table 4. Blue lines have been used to indicate the presence of another peak location and its closest multiple. These indicate the presence of additional periodic components, as discussed in Sect. 4.2.3.

Some salient observations are as follows. Firstly, the existence of at least one periodic component is observed for several ranges of redshifts of quasars and galaxies. Secondly, a peak is usually observed at the next multiple of the fundamental period, but not all multiples show peaks. This may be attributed to the effect of noise in the data. A third observation may be made regarding the presence of additional periodic components which are often noticed. This is confirmed by a secondary strong peak and a peak occurring at a location, which is a multiple of the location of the secondary peak. Plots of Figs. 9 and 10 show examples of this phenomenon. Yet another

Table 4. Period lengths of cosmological data sets.

Data set	Redshift range	Bin size	Measure	Period length	95% confidence interval
QSO redshifts of DR12	$z > 0.03$	0.014	SVR1	0.2094	[0.1537, 0.5225]
			SVR2	0.2094	[0.2094, 0.2967]
	$0.03 < z < 1$	0.0019	SVR1	0.1210	[0.1205, 0.1245]
			SVR2	0.1210	[0.1205, 0.1243]
	$1 < z < 3$	0.0058	SVR1	0.1153	[0.0583, 0.1190]
			SVR2	0.0807	[0.0570, 0.1152]
	$3 < z < 4$	0.0044	SVR1	0.0654	[0.0435, 0.1335]
			SVR2	0.0654	[0.0526, 0.0686]
Galaxy redshifts of DR10	$z > 0.03$	0.000397	SVR1	0.0056	[0.0047, 0.0139]
			SVR2	0.0079	[0.0047, 0.0095]

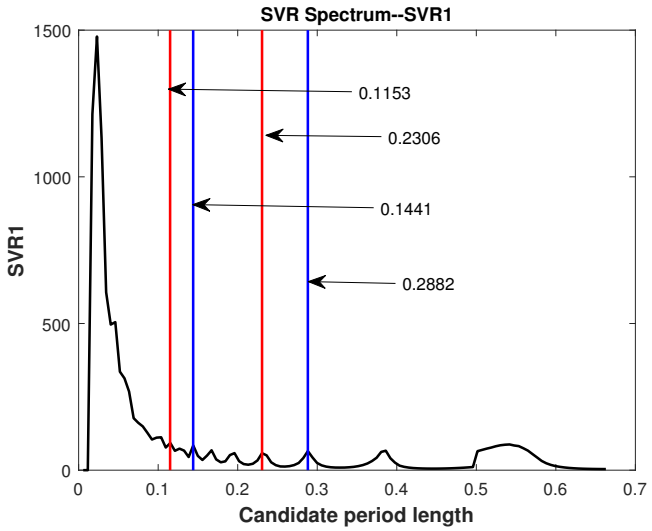

Fig. 9. QSO redshifts of DR-12 data set, $1 < z < 3$, bin width = 0.0058, SVR spectrum using SVR1.

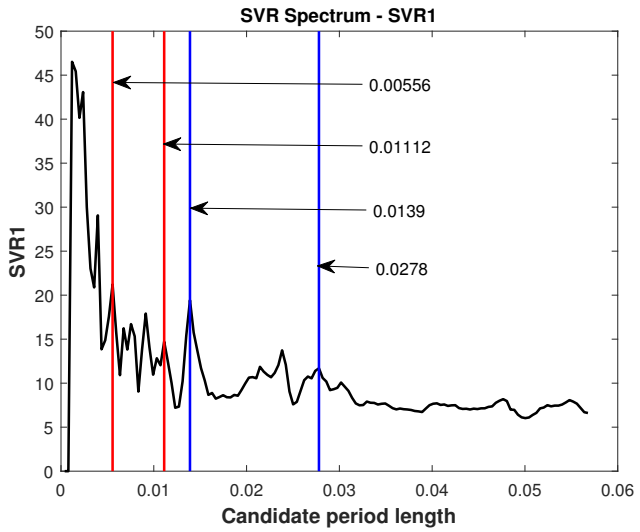

Fig. 10. Galaxy redshifts of DR-10 data set, $z > 0.03$, bin width = 0.000397, SVR spectrum using SVR1.

interesting observation is that the fundamental period of the galaxy redshift data is computed to be less than that of the quasar data by an order of magnitude. We also note that a periodicity of 0.06 is observed for the DR-12 data set for the range $3 < z \leq 4$, while periodicity at integral multiples (about twice, in particular) of this quantity can be observed for the ranges $0.03 < z \leq 1$ and $1 < z \leq 3$. This is in accordance with the findings in Burbidge (1972).

5. Physical mechanism for periodicity of redshift

In a homogeneous and isotropic universe, we expect the redshift distribution of extragalactic objects to approximate a continuous and aperiodic distribution. However, a periodicity with z = 0.031 or 0.062 was found for the QSOs, which cannot be understood in terms of the cosmological hypothesis. It is also shown by Burbridge that the periodicity is not due to the selection effect. Therefore, determining the mechanism behind the periodicity of redshit in quasars is essential. The standard argument behind the redshift for a galaxy and quasar is the existence of the Doppler mechanism, which is due to the expansion of the universe. The pioneering discovery of Wolf on the shift of the frequency of light propagating in a random medium (a medium with a random refractive index) or inhomogeneous medium, also provided the pathway to a solution.

The Wolf effect mimics the Doppler shift even if there is no relative motion between the observer and the object. This mechanism has been tested in various laboratories in the USA, Italy, Japan etc. Wolf applied this mechanism to understand the redshift for galaxies. Here, as this mechanism does not require any relative motion, the concept of expansion of the universe is not needed. It depends on the environments of the astronomical objects. One of the present authors (S. Roy) along with some of his collaborators applied this mechanism to understand the redshift of quasars and quasar-galaxy associations. They analysed Veron Cetty as well as SDSS-3 data (Roy et al. 2007) and showed that this mechanism is needed to understand the high redshift data for quasars. However, no attempt has been made thus far to understand the quantisation of redshift of quasars within this framework.

In 2007, Bajan & Flin (2007) published a comprehensive review on the periodicity of redshift and gave a possible theoretical explanation. He concluded his paper by analysing this issue using a large data base. Here, in our work we found the existence of this kind of periodicity using the large SDSS data base. Now, we should look for a possible theoretical explanation.

Lehto (1990) developed a theoretical model which could predict periods of redshifts. He described basic properties of matter using three-dimensional quantised time. The time unit is Planck time. The redshift quantisation can be obtained based on the

assumption that distances are quantised in Planck's units. However, the concept of three-dimensional quantised time is difficult to understand from the standard interpretation of modern physics.

Roberts (2002) proposed the perturbative approach for understanding the discrete redshift. Here, discrete redshift is supposed to be explained by using the spherical harmonic integer l; this occurs both in the metric or density perturbations and also in the solution of wave equations in Robertson-Walker spacetime. However, justifications of some crucial assumptions are needed before considering it as a viable theoretical explanation.

Hoyle & Narlikar (1964, 1966) proposed the "variable mass hypothesis" within the Machian framework. Broadly speaking, the Machian principle states that inertia of matter arises due the existence of other matter in the universe. The principal idea relies on the fact that the galaxies originate at the zero mass surface and the classical action at the zero mass surface is small and hence physics is dictated by quantum principles. Accordingly, one can argue that the quantum mechanical domain within the material which emerges from a zero mass surface may be with a discrete burst at discrete intervals instead of continuously. This may lead to the quantisation of the redshifts. Here, the quantisation of mass indicates the quantisation of the redshift. However, more comprehensive works are needed to understand this mechanism.

6. Summary and conclusions

A periodicity analysis has been carried out on different types of data sets, including simulations, mock data sets, a quasar data set of DR12, and galaxy data sets of DR10. The analysis indicates the existence of periodicity, irrespective of the selection effect, that is, the range of the redshift that was selected. This periodicity is revealed by the SVD-based approach, which has been demonstrated to be more effective than the periodogram and power spectra-based method, especially in the presence of noise. This work also indicates the presence of multiple periodic components. The difference of fundamental periodicity of QSO and a galaxy can be explained by the Hoyle-Narlikar theory. The physical significance of this periodicity and its interpretation form the scope of future work.

Acknowledgements. The authors express their thanks to Prof. J. V. Narlikar. of Inter University Centre For Astronomy and Astrophysics (IUCCA) for excellent guidance and suggestions provided during the course of this research. The authors are very grateful to H. S. Ravindra and Dr. Prashant Upadhaya of LEOS, ISRO for their encouragement and to Susmita Nandi for her help in carrying out some of the simulations performed to obtain the results reported. The suggestions of Prof. Debasis Sengupta at a crucial stage of the simulation design were especially helpful. One of the authors (SR) is indebted to Homi Bhabha Trust for financial support. Finally, the authors would like to express their gratitude to the anonymous reviewers whose recommendations at various stages have helped tremendously in giving shape to the manuscript in its present form.

References

Arp, H., Bi, H. G., Chu, Y., & Zhu, X. 1990, A&A, 239, 33
Bajan, K., & Flin, P. 2007, Old New Concepts Phys., IV, 159
Barnothy, J. M., & Barnothy, M. F. 1976, PASP, 88, 837
Basu, D. 2005, ApJ, 618, L71
Bell, M. B., & Fort, D. N. 1973, ApJ, 186, 1
Burbidge, G. 1968, ApJ, 154, L41
Burbidge, G. R., & Burbidge, E. M. 1969, Nature, 222, 735
Burbidge, G. R., & O' Dell, S. L. 1972, ApJ, 178, 583
Cocke, W. J., & Tifft, W. G. 1991, ApJ, 368, 383
Depaquit, S., Pecker, J. C., & Vigier, J. P. 1985, Astron. Nachr., 306, 7
Duari, D. 1997, JApA, 18, 441
Duari, D., DasGupta, P., & Narlikar, J. V. 1992, ApJ, 384, 35
Fulton, C. C., Arp, H., & Hartnett, J. G. 2018, Astrophys. Space Sci., 363, 134
Hawkins, E., Maddox, S. J., & Merrifield, M. R. 2002, MNRAS, 336, L13
Holba, Á., Horváth, I., Lukács, B., & Paál, G. 1994, Astrophys. Space Sci., 222, 65
Hoyle, F., & Narlikar, J.V. 1964, Proc. Royal Soc. London, 282, 191
Hoyle, F., & Narlikar, J.V. 1966, Proc. Royal Soc. London, 290, 177
Kanjilal, P. P., & Palit, S. 1995, IEEE Trans. Signal Process., 43, 1536
Karlsson, K. G. 1971, A&A, 13, 333
Kitagawa, G., & Gersch, W. 1996, in Smoothness Priors Analysis of Time Series, (New York, NY: Springer), Lect. Notes Stat., 116
Lehto, A. 1990, Chin. J. Phys., 28, 215
Napier, W., & Burbidge, G. 2003, MNRAS, 342, 601
Narlikar, J. V., & Arp, H. 1993, ApJ, 405, 51
Narlikar, J. V., & Das, P. K. 1980, ApJ, 240, 401
Palit, S., & Kanjilal, P. P. 1994, Signal Process., 40, 269
Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79
Poudel, A., Heinämäki, P., Nurmi, P., et al. 2016, A&A, 590, A29
Roberts, M. D. 2002, Astrophys. Space Sci., 279, 305
Rodríguez-Torres, S. A., Comparat, J., Prada, F., et al. 2017, MNRAS, 468, 728
Roeder, R. C., & Dyer, C. C. 1972, Nat. Phys. Sci., 235, 3
Roy, S., Kafatos, M., & Datta, S. 2000, A&A, 353, 1134
Roy, S., Ghosh, J., Roy, M., & Kafatos, M. 2007, Model. Simul. Sci., 90
Scott, D. 1991, A&A, 242, 1
Shimazaki, H., & Shinomoto, S. 2007, Neural Comput., 19, 1503
Shimazaki, H., & Shinomoto, S. 2010, J. Comput. Neurosci., 29, 171
Tang, S. M., & Zhang, S. N. 2005, ApJ, 633, 41
Tifft, W. G. 1997, ApJ, 485, 465