

CorrSim: a multiwavelength timing observation simulator

J. A. Paice¹, R. Misra⁴ and P. Gandhi³

¹Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

²Department of Physics and Astronomy, University of Manchester, Oxford Rd, Manchester M13 9PL, UK

³Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

⁴Inter-University Centre for Astronomy and Astrophysics, Pune, Maharashtra 411007, India

Accepted 2024 August 5. Received 2024 July 11; in original form 2024 March 12

ABSTRACT

Studying the rapid variability of many astronomical objects is key to understanding the underlying processes at play. However, obtaining data well-suited to this task can be tricky, especially for simultaneous multiwavelength observations. Parameters often have to be fine-tuned while ‘on-site,’ or observations may only be found to not achieve their goals later. Here, we present *CorrSim*; a program tailored to X-ray Binary analysis, and expandable to many kinds of multiwavelength coordinated observations. *CorrSim* takes power spectra, coherence, and lags, and returns a simulated multiwavelength observation. The goals of this are: (i) To simulate a potential observation; (ii) To investigate how different Fourier models affect a system’s variability, including data products like cross-correlation functions; and (iii) To simulate existing data and investigate its trustworthiness. We outline *CorrSim*’s methodology, show how a variety of parameters (e.g. noise sources, observation length, telescope choice) can affect data, and present examples of the software in action. Through *CorrSim*, we also detail the effects of the length of the data train on Fourier and correlation function uncertainties. We also highlight previous *CorrSim* modelling, noting that the ‘pre-cognition dip’ seen in X-ray binaries can be constructed by periodic processes, and discuss this result in the wider context.

Key words: software – data methods – rapid timing astrophysics – black holes.

1 INTRODUCTION

When observed at different wavelengths, many astrophysical sources can show startlingly different signals. While these signals are often rapid and complex, they are often all interrelated; they are the result of some base process (or processes) that created them. By studying these different signals, and uncovering the relationships between them, we can thus begin to decode what the underlying processes are.

This general description can be applied to many fields of astrophysics. Simultaneous (or at least quasi-simultaneous) observations on multiple telescopes have been key to the studies of supernovae (Perley et al. 2019), cataclysmic variables (Wheatley, Mauche & Mattei 2003), millisecond pulsars (Draghis et al. 2019; Papitto et al. 2019), fast radio bursts (Scholz et al. 2016), ultra-luminous X-ray sources (Middleton et al. 2017), gamma ray bursts (Costa et al. 1997), tidal disruption events (van Velzen et al. 2019), Blazars (Acciari et al. 2011), active galactic nuclei (AGNs; McHardy et al. 2014), and even kilonovae (Kasliwal et al. 2017).

A valuable test case is one particular field: the study of rapid variability in X-ray Binaries. Not only are these complex systems, with a compact object being surrounded by a swirling accretion disc being fed from a companion star, but they are also very dense systems; they are only a few hundred thousand to a few million Schwarzschild radii at the largest length scales. To put this in the terms of time, the dynamical time-scale – the time-scale at which matter can flow

between different parts of the system – is on the order of minutes for typical values (Frank, King & Raine 2002). Meanwhile, the systems are only a few tens of lightseconds across, with length scales close to the compact object of around, and below, one lightsecond. This means that two regions – emitting at drastically different wavelengths – can affect one other on time-scales of minutes to milliseconds.

Since these sources are so complex and compact, rapid (often subsecond) multiwavelength observations are *key* to understanding how their different regions interact, and what underlying structures are present. Decades of effort here have resulted in a growing collection of correlated observations, revealing complex lags and relations (for just a few examples, see Kanbach et al. 2001; Durant et al. 2008; Gandhi et al. 2008, 2010, 2017; Veledina et al. 2015; Pahari et al. 2017; Paice et al. 2019; Vincentelli et al. 2021).

However, such studies have had several key problems to overcome. For one, a simultaneous observation between two telescopes is difficult to plan, and prone to many issues such as orbital visibility, day/night cycles, packed schedules, and the initial difficulty of obtaining proposals on all required telescopes. And on the night of an observation, issues may arise that were not accounted for; count rates that are lower than expected, technical difficulties that may lead to increased read noise or less observation time, poor weather conditions that could increase the scintillation in the atmosphere, or a change in the expected Fourier properties of the source, to name just a few.

And, even if data are obtained at a high-enough quality, it is only natural to ask: How representative are these data of the source as a whole? A finite observation typically cannot capture the full model

* E-mail: johnapaice@gmail.com

of a complex system; this is particularly true for X-ray Binaries, where the behaviour of a source can vary suddenly and dramatically over the course over several weeks (see e.g. Veledina et al. 2017; Paice et al. 2021; Thomas et al. 2022), or even minutes (Gandhi et al. 2016, 2017). Middleton et al. (2017) covers many of these problems in greater detail, and notes many fields of study where solving them could lead to significant advances in our understanding (including a case study where simultaneous observations lead to a revolution in the understanding of one source, V404 Cyg).

To summarize, the main problems are thus; *it is difficult to plan an observation that will deliver data of a high enough quality for us to investigate our questions*, and even if an observation is carried out, *it is not trivial to know if those data correspond well to a model of the source*.

With these issues in mind, we have developed a program, `CorrSim`. Given both source and observational properties, `CorrSim` simulates an observation and return various data products, such as light curves, cross-correlation functions (CCFs), and cross-spectral analyses. This will allow for the testing not only of observational setups, but also various models and how they affect the data produced.

In specific, the motivations of the program are:

(i) To quantify the minimum data required to perform an analysis; i.e. the minimum length of an observation that would be required, and what magnitude of measurement errors would still produce a reliable result.

(ii) To find the conditions under which the shape and significance of obtained CCFs can be trusted.

(iii) To understand the effect of different kinds of measurement errors and incoherent broad-band noise on the data analysis.

(iv) For smaller data sets and/or those with higher measurement errors, to ascertain which type of analysis would provide the most reliable results.

(v) To investigate the ambiguity of time lags in the analysis, and in the case of periodic signals where the true time lag is greater than one time period away, how to spot this in real data.

We will first detail the workings and assumptions of `CorrSim` (Section 2). We will then show the result of putting real, observed source properties into `CorrSim`, and quantify how factors such as observation length, noise sources, and choice of telescope alter the resultant data (Section 3). We will finally detail two examples of using the program to (i) inform choices about a hypothetical upcoming observation (Section 4) and (ii) test how modified Fourier components affect the CCFs (Section 5). A glossary of used symbols (Section 6) and description of cross-correlation and Fourier analysis (Appendix A) are also appended at the end of this paper.

1.1 On the use of `CorrSim`

In Sections 3–4, X-ray binaries (XRBs) are used for the examples of the program; this primary use of a single source is partly to make comparison between different sections easier. However, `CorrSim` aims to be usable to many other areas of astrophysics.

In general, if an analysis consists of two simultaneous light curves with consistent Fourier components and time resolution, and the waveband can be adequately described using a count rate, fractional RMS, and a power spectrum, then `CorrSim` should be useful in simulating it – this might be the case for X-ray binaries, binary systems with orbital periods on the scale of hours, neutron stars, cataclysmic variables, or even studying interband lags, to name a few; Figs 4 and 5 show the example of an observation featuring a

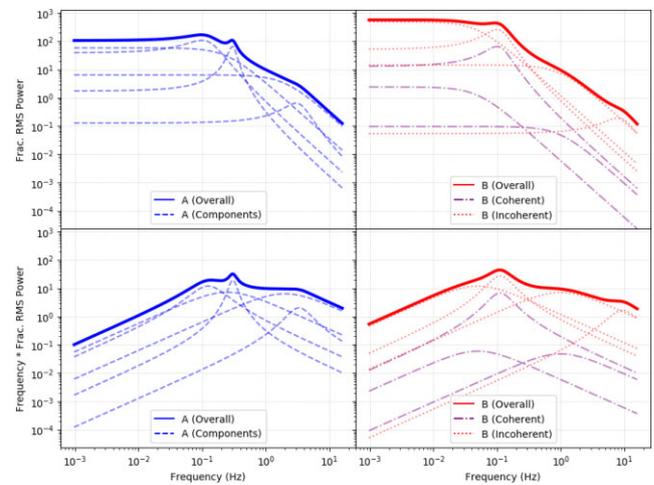


Figure 1. Examples of model input power spectra, made up of several Lorentzians. Top row: Fractional RMS power. Bottom row: Fractional RMS power multiplied by frequency. Each component Lorentzian is plotted, as well as the overall summation. Note the coherent and incoherent Lorentzians in Series B; some have the same shape but have different normalizations.

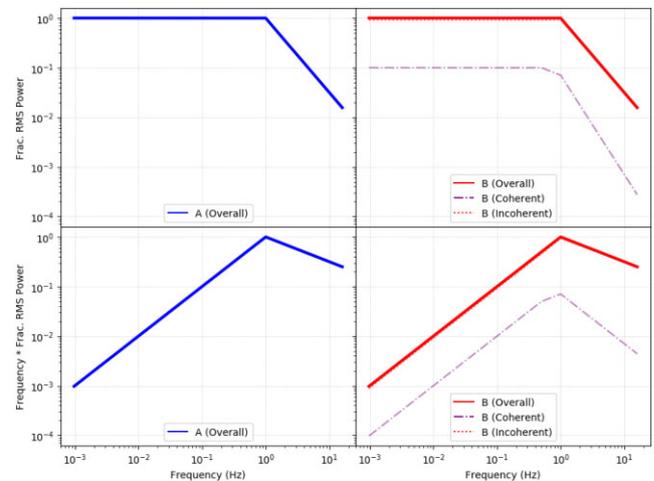


Figure 2. Examples of model input power spectra, made up of a broken powerlaw. Top row: Fractional RMS power. Bottom row: Fractional RMS power multiplied by frequency. Note the coherent and incoherent Lorentzians in Series B, and how the coherent power law has a different shape and break frequency.

system with a strong spin period that’s coherent between both bands, as can be the case for some cataclysmic variables or neutron star systems. Also, we note how `CorrSim` can be used to tune parameters ‘on-site’, during an observation; when run on an average laptop, run-times can be a matter of seconds to minutes for observations with $< 10^{6-7}$ data points, and thus such a use should be possible for all but the shortest observations.

`CorrSim` is less useful in areas where the time resolution is not consistent, such as AGNs studies, or in any studies where the power spectra or phase/time lags change over time, as `CorrSim` assumes these to be unchanging over the course of the observation. However, in the case of the former, we note that removing sections of `CorrSim`’s output light curves (or perhaps taking a random sample of the points) and then analysing them with external codes or tools is a way to extend `CorrSim`’s functionality. In either case, the Supplementary Material features a run similar to the X-ray binary

case, but shifted into a domain more akin to AGNs, to demonstrate the scales that CorrSim is useful over.

We finally note that if one is using CorrSim to prepare for an observation, the program’s modelling can deviate from an eventual observation if the source’s parameters depart from the initial inputs. This can particularly occur with transient sources, which are often dynamic and do not have well-defined nor static Fourier properties, which are often serendipitously observed during an observation (see e.g. Gandhi et al. 2016; Vincentelli et al. 2023 for only two examples). The magnitude of this change cannot be easily stated; for this, we thus advise running CorrSim with probable variations of the source parameters, and aim for a setup which minimizes signal-to-noise and uncertainties.

2 CorrSim

We created the program CorrSim as part of this project.¹ The aim of CorrSim is to present a simulated result of an observation, and then run analysis on that result and compare it to the inputted parameters.

2.0.1 On the definition of coherence

Throughout this paper, we use a definition of coherence to refer to the phase coherence between two time series. This is slightly different to the meaning of ‘coherent’ in some areas of astronomy, such as coherent pulsations in pulsars. See Appendix A2.2 for more details.

2.1 Inputs and outputs

The inputs here are split into two. Source parameters are intrinsic to the object, and essentially unchangeable for a given observation (aside from, say, mean count rate, which is a function of the telescope used). Observation parameters are intrinsic to the observation itself, and often within the control of an observer, e.g. time resolution or length of observation.

(i) *Source parameters:*

- (a) **Mean count rates** (\bar{A} , \bar{B}) – Units of counts per second.
- (b) **Fractional RMS values** ($F_{rms,A}$, $F_{rms,B}$) – units of percent. Defines the variability of the light curves.
- (c) **Model power spectra** ($p_{A,Model}$, $p_{B,Model}$). Two power spectral model types are defined in CorrSim: the summation of several Lorentzians (Fig. 1), and a broken power law (Fig. 2). The shape of the power spectra itself only affects the light curve, but the dependence of the coherence on time strongly affects the shape and strength of the correlation function.
- (d) **Model coherence** (γ_{Model}). This is defined in CorrSim by the relative coherence of different components in the power spectra.
- (e) **Model phase/time lags** (δ_{Model}), as a function of Fourier frequency. Several different ways to define the lags have been provided, relating to the dependence of phase or time with Fourier frequency in either log–log or semi-log space.
- (f) **Red Noise** – Applies ‘red’ noise, i.e. noise dependent upon frequency. Uses two subparameters: ‘Fractional RMS,’ which defines the amount of the noise; and ‘Slope,’ which defines its dependence on frequency.
- (g) **Poisson Noise** – Boolean.

Inputs

Source Parameters:	Models:	Observational Parameters:
• Mean Count Rate	• Power Spectra	• Observation Length
• Fractional RMS	• Lorentzian or BPL	• Time Resolution:
• Noises:	• Coherence	• Noises:
– Red	• Lags	– Scintillation
– Poisson	– Phase or Time	– Readout

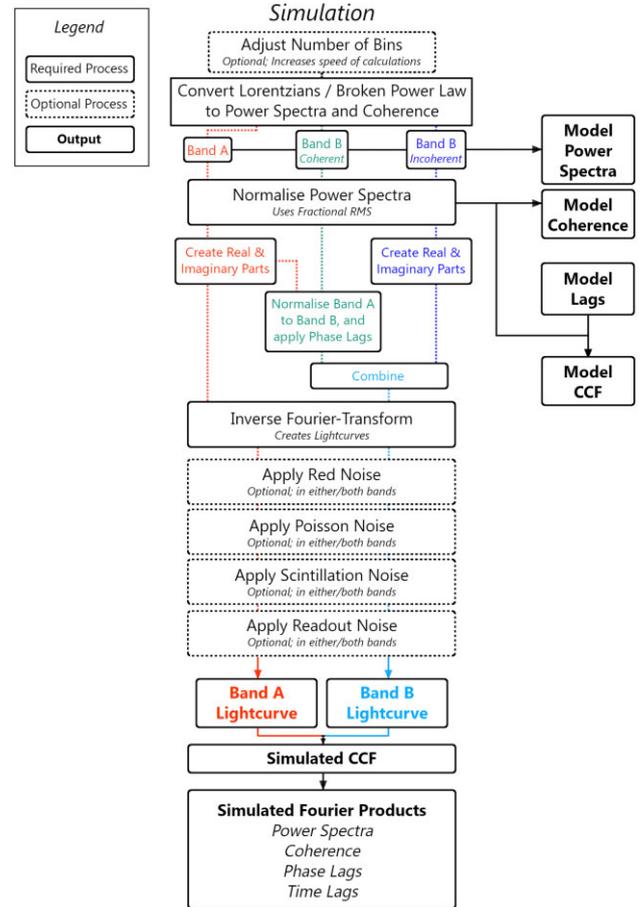


Figure 3. Flowchart demonstrating the process that CorrSim uses to create its outputs.

(ii) *Observation parameters:*

- (a) **Observation length** (T) – units of seconds.
- (b) **Time resolution** (dT) – units of seconds.
- (c) **Scintillation noise** – Simulates noise from atmospheric scintillation; only necessary for simulating ground-based optical and infrared observations. Uses several sub-parameters: ‘Telescope Diameter’ (m); ‘Telescope Altitude’ (m); ‘Exposure Time’ (s); ‘Target Altitude’ (°); ‘Atmospheric Turbulence Height’ (m); and an empirical value (Osborn et al. 2015).
- (d) **Readout noise** – Units of counts per bin. Necessary for any telescope that uses a CCD with a non-negligible readout noise.

The outputs are:

- (i) Simulated light curves for each band
- (ii) Correlation function analysis
- (iii) Fourier analysis (i.e. power spectra, coherence, and phase and time lags as a function of Fourier frequency)

¹<https://github.com/JohnAPaice/CorrSim>

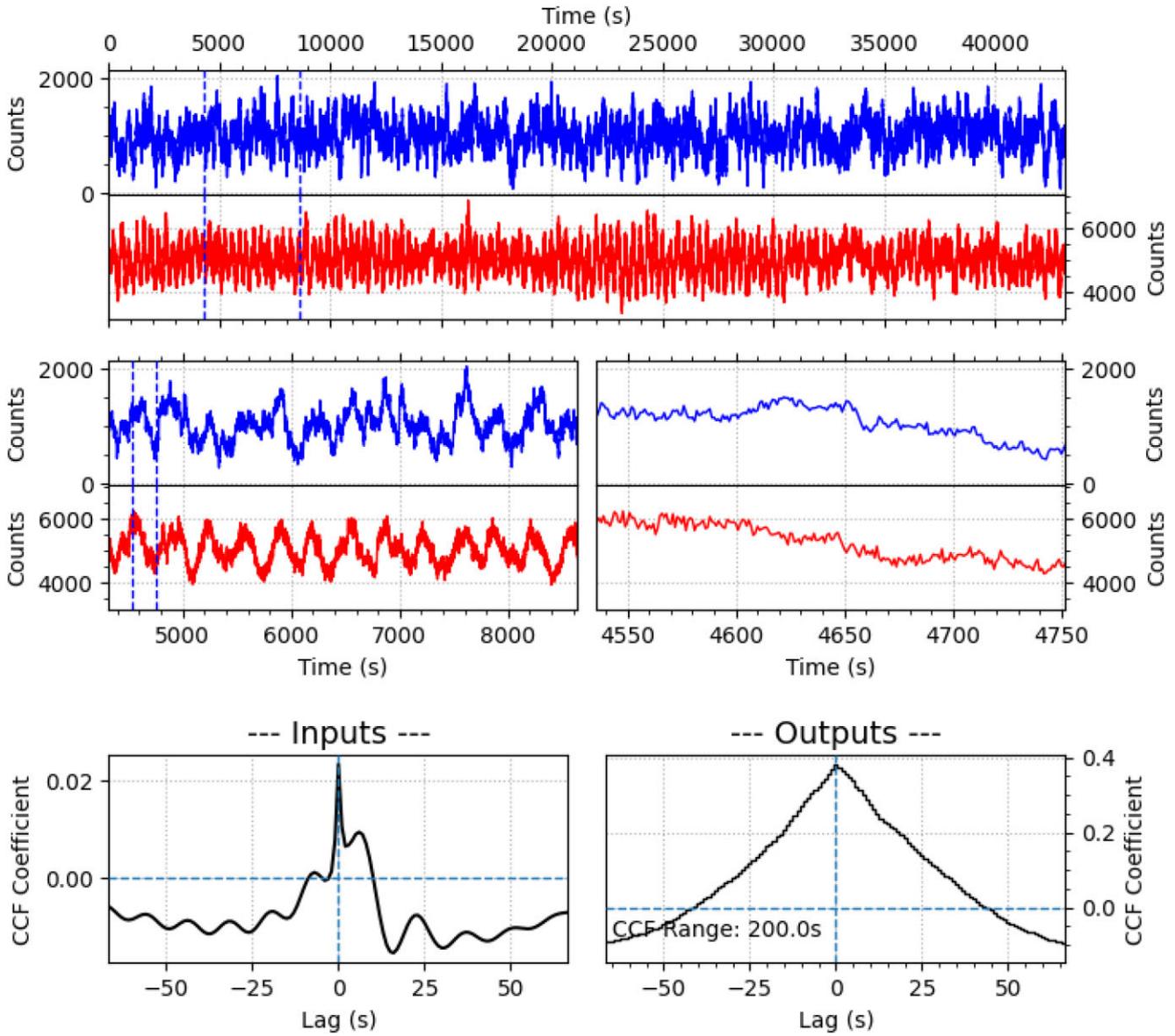


Figure 4. The first half of a compilation plot provided by *CorrSim*. Top, Middle: 100 s, 10 s, and 2 s segments of the produced light curves (Band A, blue, top panels; band B, red, bottom panels). The dashed lines represent the ranges of the insets. Bottom: Model (left) and simulated (right) correlation functions (Band B versus band A, i.e. a peak at positive lags shows band B lagging band A), with the latter also showing the range over which it was averaged. The model correlation function was calculated from the input coherence, power spectra, and lags (see Fig. 5, while the reproduced correlation function was produced from the light curves.

Some of these outputs have subparameters for controlling their behaviour, such as segment size and binning for both correlation function and Fourier analysis.

2.2 Methodology

A flowchart showing the methodology of *CorrSim* is shown in Fig. 3. Below, we go into detail about each of these steps.

Initially, *CorrSim* (optionally) adjusts the input observation length in order to maximize the speed of the calculations, which go faster with smaller factors. This is done either by adjusting the number of bins to the closest ‘7-smooth’ number (a number with no prime factor larger than 7; Berndt 1994).

CorrSim then creates three power spectra from the details provided – one for band A (p_A), and one for each of the coherent

and incoherent parts of Band B ($p_{B,coh}$ and $p_{B,inc}$ respectively). It then normalizes all three to match the fractional RMS values given.

In the next few steps, *CorrSim* will create the two simulated light curves from the power spectra (S_A and S_B), using methodology set out in Timmer & Koenig (1995). *CorrSim* first draws two random numbers each for band A and the incoherent part of band B, with a standard deviation equal to the respective power spectra. These numbers become the real and imaginary parts of a complex series (S_A and $S_{B,inc}$). To make $S_{B,coh}$, the band A complex series is copied and normalized to band B, and the phase lags are applied:

$$S_{B,coh} = C \times S_A \times e^{-i\delta_{Model,Phase}} \tag{1}$$

where $C = \sqrt{p_{B,coh}/p_A}$, i is the imaginary number, and $\delta_{Model,Phase}$ are the model phase lags (not the time lags). The first bin of S_A

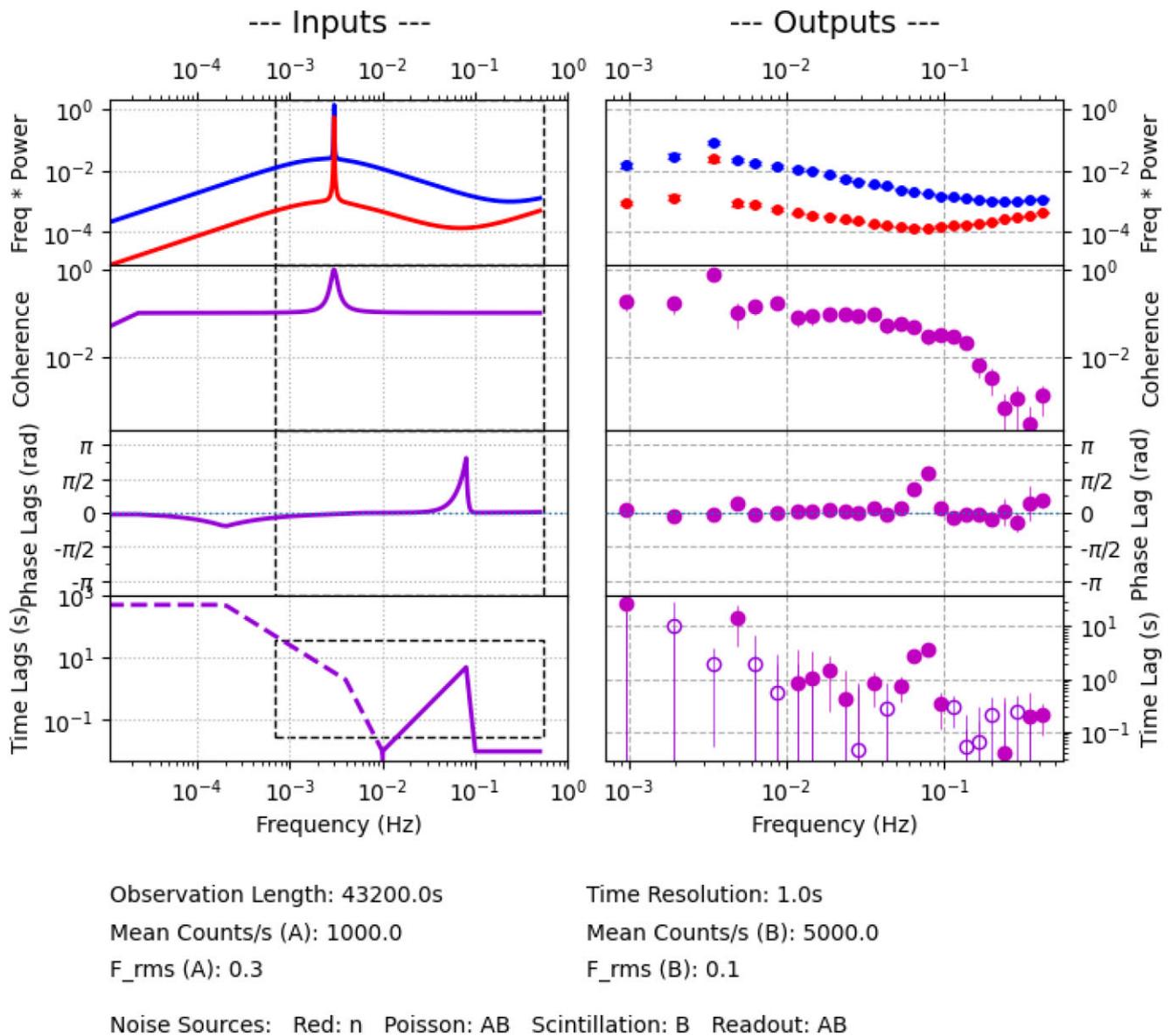


Figure 5. The second half of a compilation plot provided by `CorrSim` – input Fourier properties (left) and results from Fourier analysis of the light curves (right). Since the outputs are calculated by splitting up the light curve and averaging over several segments, they only cover a smaller range; black dashed boxes on the left represent this range. First Row: Power Spectra of band A (blue, upper data) and B (red, lower data). Second Row: The coherence between the bands. Note how the presence of uncorrelated noise reduces the input coherence. Third Row: The phase lags. For the output, since phase lags outside of $\pm\pi$ are meaningless, the plot wraps around at these ranges. Fourth Row: The time lags. Negative time lags are represented as a dashed line for the input, and open circles for the output. In `CorrSim`, a reference point can be given for calculating the time lags from the phase lags, i.e. stating at which frequency the phase lags can be trusted to be the true value, and not multiples of 2π prior or hence. Bottom Information: Key input values. For the noise sources, the letters (A, B, AB, or n) represent which bands that noise is present in, if any.

is adjusted to be equal to the mean count rate multiplied by the number of bins, $\bar{A}N$ (and $\bar{B}N$ for $S_{B,\text{coh}}$, while the first bin of $S_{B,\text{inc}}$ is set to zero). $S_{B,\text{coh}}$ and $S_{B,\text{inc}}$ are then combined into $S_{B,\text{total}}$. Both S_A and $S_{B,\text{total}}$ are inverse Fourier transformed to turn them into light curves, and the fractional RMS is checked against the input values.

Finally, noise is (optionally) added: First, red noise is generated using a similar method from Timmer & Koenig (1995), and added on to S_A and $S_{B,\text{total}}$ (in this case, the light curves are remade as above). Then, `CorrSim` adds Poisson noise, calculates and adds scintillation noise, and finally adds readout noise. The result of this is the two final simulated light curves for band A and B.

This is the crux of `CorrSim`; these light curves simulate the data taken of an observed source, i.e. an imperfect representation of the true relationship between signals, and are analogous to a real observation given the various source and observational parameters input. These light curves are not only plotted graphically, but also outputted as comma-separated text files, in case users wish to use them with their own code.

If desired, at this point, `CorrSim` essentially works in reverse; it takes those light curves and runs both cross-correlation and Fourier analysis on them, as one would for real, observed data. Much of the Fourier analysis uses functions from `Stingray` for the calculations. Examples of the plots created are seen in Section 2.4.

2.2.1 Modelling the correlation function

The correlation function (CF) is the inverse Fourier transform of the coherence and lags, and thus a model of the function can be reproduced purely from these properties. This is based on the mathematical definitions of coherence presented in Vaughan & Nowak (1997), and is defined in Section A1. In `CorrSim`, a model of the correlation function is produced from the initial coherence and lag inputs, and the methodology is described here.

A Fourier transform creates a complex series, which can be wholly described by its amplitude and its arguments. The amplitude of this series is defined as:

$$U = \gamma_{\text{Model}} * \sqrt{(p_A)(F_{\text{rms}A}) * (p_B)(F_{\text{rms}B})}, \quad (2)$$

where U is the amplitude, γ_{Model} is the model coherence, and p and F_{rms} are the power spectra and fractional RMS values (for each band A and B). The arguments are then taken to be the phase lags. This complex series is then inverse Fourier transformed, and the real part is taken. The result is normalized with a constant (V) thusly:

$$V = \frac{1}{2 * dT * \sqrt{F_{\text{rms}A} * F_{\text{rms}B}}}, \quad (3)$$

where dT is the time resolution of the data. After the normalization is applied, the final product is the model correlation function.

2.2.2 Converting phase to time

Phase lags, when combined with the frequency they are found at, can be converted to time lags using equation (A11).

However, there is a limitation to phase lags; `CorrSim` adopts the convention that phase lags are given between $\pm\pi$. If the actual phase lags are outside of this range, say between π and 2π radians, then `CorrSim` will show that they are between $-\pi$ and 0 radians due to the cyclical nature of sine waves. This limitation passes on to time lags; a ‘true’ time lag that is between $+\pi$ and $+2\pi$ radians in phase will be represented as a negative time lag.

There is a solution – or, at least, a mitigation – to this. By using the correlation function, we can see where sources of correlation (or anticorrelation) are occurring and find which frequency bins are most likely to be correct. Then, assuming that phase lags follow a reasonably continuous distribution, we shift them by $\pm 2\pi$ radians to minimize discontinuities.

This method is intrinsic to `CorrSim`. A ‘Reference Frequency’ parameter is set (default 1 Hz) which is assumed to be correct. Then, `CorrSim` looks at the next lowest frequency. That point is compared to the *mean of the previous three points*. If the difference is greater than π , it and *all points at lower frequencies than this one* are shifted by $\pm 2\pi$ to minimize the discontinuity. This is repeated for every frequency bin. The same procedure is then conducted, but instead going to higher frequencies. These corrected phase lags² are then converted to time lags using the above equation.

This is a method that has its drawbacks – any processes constrained to just a single frequency bin, such as quasi-periodic oscillations (QPOs), may be more than π away from phase lags in adjacent bins, and this will thus be misrepresented in the time lags. Close inspection of the time lag plots, and comparison with the CF, may be required to investigate this possibility.

²The corrected phase lags themselves are not plotted; the phase lags shown are the ones gained from the Fourier analysis, without any shifting.

2.3 Assumptions, models, and noise

Several assumptions have been made in producing `CorrSim` – these have been for simplicity and ease of use, and none should strongly affect the primary results. They are detailed here.

(i) Power spectra

The broken power law, and its handling of coherence, is a simplification of the behaviour of a source. Lorentzians, while also being limited in their own ways, offer much finer control.

(ii) Phase and time lags

`CorrSim` approximates the phase and time lags using a series of constant, linear, power law, or polynomial distributions. This does not trivially handle the true lag behaviour – which may be better represented by other distributions – but the versatility of `CorrSim` in being able to handle any number of segments mitigates this simplification. In calculating the time lags, `CorrSim` also assumes that phase lags are roughly continuous and do not deviate by more than π radians between frequency bins.

(iii) Simultaneous, continuous light curves

`CorrSim` creates light curves that are simultaneous, sampled at the same rate, and have no gaps, and assumes this is the case during its analysis. While this is not wholly realistic for most multiwavelength observations, it serves as a good approximation, and the simultaneous sampling vastly speeds up the calculation of the correlation function. If a user wishes to analyse a light curve that is affected by gaps, then they may take the light-curve data that `CorrSim` generates and create the gaps manually, and then use their own code to do the analysis.

(iv) Noise

`CorrSim` currently allows for four sources of noise. The way this noise is calculated is a simplification of processes within the source and the instrument. Their methods are detailed here.

(a) *Red noise*. This is noise that is more significant at lower frequencies, a result of random ‘Brownian’ motion, or ‘Random Walk’-type signals in the source. Using methods set out in Timmer & Koenig (1995), this noise is modelled by drawing two random Gaussian numbers and then multiplying them by a value proportional to the model power spectra for each frequency bin. These two resultant numbers then become the real and imaginary parts of the complex array which will later create the light curve.

(b) *Poisson noise*. This noise accounts for ‘shot noise’, dictated by the count rate and the signal-to-noise ratio. This is modelled by taking the light curve, and for each bin, drawing a random number from a Poisson distribution ($\lambda = \text{counts in that bin}$).

(c) *Scintillation noise*. This noise arises from scintillation seen by ground-based telescopes, caused by variations in the atmosphere. `CorrSim` uses a modified version of Young’s approximation, presented in equation (7) of Osborn et al. (2015), which is observation- and telescope-specific; the required parameters include the telescope’s diameter and altitude, the distance of the source from zenith, the height of atmospheric turbulence, exposure time, and an empirical, telescope-specific constant. The equation gives the variance of the scintillation noise. Therefore, this noise is modelled by taking the light curve, and for each bin, drawing a random number from a Gaussian distribution, with the mean equal to the counts in that bin, and the standard deviation equal to the square root of Young’s approximation.

(d) *Readout noise*. This noise arises from the imperfect amplification of a signal in the CCD, where the true charge on a number of electrons is misread. This noise is modelled by drawing from a Gaussian distribution centred on zero, with the standard deviation equal to the readout noise in counts, for each bin, and then adding that number on to the value of the bin. Note, in practice, readout noise is given in electrons; here, counts are used for simplicity.

2.4 Outputs

Fig. 4 shows example produced light curves and both the input and output correlation functions. Fig. 5 shows examples of both inputs and outputs of CorrSim, and some observation parameters.

The Fourier properties of this example are loosely based on a cataclysmic variable with a coherent 300 s spin period (i.e. seen in both bands, with corresponding Lorentzian having 100 per cent coherence). Phase lags are arbitrary, and the source has been modelled assuming an observation with X-ray (band A) and optical (band B) telescopes.

3 DISCUSSION I: COMPARISON OF PARAMETERS

These simulations allow us to compare different aspects of observations, and see their effects on the final products with respect to the initial models. In this section, we will detail how several different parameters can affect the final products. There are many parameters that can be varied with CorrSim, and not all can be shown here; more figures can be seen in the Supplementary Material, found online. As an example, Figs 6–8 demonstrate the effects of varying the observation length, the count rate, and the target altitude.

Each figure shows three different setups and details their effect on the simulated correlation function and Fourier outputs. The models are given by black lines, and the outputs by coloured lines.

The first row is the CCF, a particular kind of correlation function that functions with equally spaced bins in a gapless light curve, with the shaded region indicating the standard error; note that the model CF does not take effects of noise into account, which reduces coherence and thus the correlation coefficient; the model CF therefore may not approximate a CF created from a theoretically infinite observation, and would instead have a greater magnitude.

The second row shows the power spectra. The black dashed lines are the model band A power spectra, with lighter coloured uncertainties representing their outputs, and the black dotted lines are the model band B power spectra with the darker coloured uncertainties representing their outputs. Like the CF, the model power spectra do not take the effects of noise into account.

The third row shows the coherence. These models, like the CF and power spectra, do not take the effects of noise into account.

The fourth row shows the phase lags. This is constrained between $\pm\pi$, and additional models at normalizations of $\pm 2\pi$ are also shown; the ‘middle’ model, dominant at 1 Hz, is the ‘correct’ one. Model phase lags are not affected by noise sources in a predictable way.

The fifth row shows the time lags. Solid black lines show the model time lags, while dashed black lines show the inverse model time lags, for representing negative lags. Similarly, the solid coloured circles show the outputs, and the open coloured circles show their inverse. Time lags are fixed at a defined reference frequency, and then shifted depending on what causes the smallest discontinuities, as described in Section 2.2.2. As they are calculated from the phase

lags, the model time lags are also not affected by noise sources in a predictable way.

Tables 1–3 show the default inputs for these tests (i.e. the values that will be used unless specified otherwise); these approximate a reasonably bright source observed with a reasonably sensitive telescope. The Fourier properties used are based on data from the X-ray Binary MAXIJ1820+070 during its 2018 April hard state (a mode during which higher energy X-ray emission dominates, and the source shows significant optical and X-ray variability). The input parameters, such as the F_{rms} , power spectra, and phase lags, are all approximations to those found in the analysis carried out by Paice et al. (2021), during epoch 4 (25 d after peak X-ray brightness, 37 d after the outburst was first detected). Meanwhile, the noise parameters have been selected to correspond with an X-ray (band A) and optical (band B) observation (taken by the NICER instrument on the ISS and the HiPERCAM instrument at the Gran Telescopio Canarias, Roque de Los Muchachos, La Palma, respectively).

3.1 Error dependence

For the following tests, all parameters were kept as those in Tables 1–3, unless otherwise stated. Note that these data are empirical, obtained from a single run of CorrSim each time, rather than model values.

3.1.1 Error in the correlation function

Another way to illustrate how different parameters affect the observation is by looking at the standard error in the CCF. The assumption here is that if the error is lower, the simulated CCF will be closer to the true model shape.

We ran CorrSim again with eight different lengths of observation (64, 128, 256, 512, 1024, 2048, 4096, and 8192 s). A CCF was constructed from 30 s segments, and standard error was calculated for each bin from those segments, as usual. For this demonstration, the middle third of the CCF was selected (from -10 s to +10 s lags), and then the mean and the standard deviation of the standard error were calculated. The results are plotted in Fig. 9.

3.1.2 Percentage error in the Fourier components

We can also look at the magnitude of the errors. For this, we assume that smaller errors mean that the observation is closer to the model, which we consider a reasonable assumption for illustrative purposes. By varying a parameter, we can thus see how much it affects the accuracy of the observation by quantifying the change in magnitude of the error (relative to the value of the bin).

Similar to the previous section, we ran CorrSim again with eight different lengths of observation (64, 128, 256, 512, 1024, 2048, 4096, and 8192 s), while the Fourier segment length was set to 2^{10} bins. Fourier plots were made for each run. Then, for each of the power spectra, coherence, and time lags, the median absolute percentage error was calculated over all bins. The results are plotted in Fig. 10.

3.1.3 Detecting a subsecond lag

We can also vary how correlated the light curves are, and find how that affects the resultant CCF. For this, we altered the band B Lorentzians: all coherence fractions were set to 0, and then the second Lorentzian was set to a normalization of 10 and a midpoint of 1. Its coherence fraction was then varied, and the resultant CCF was plotted in each instance.

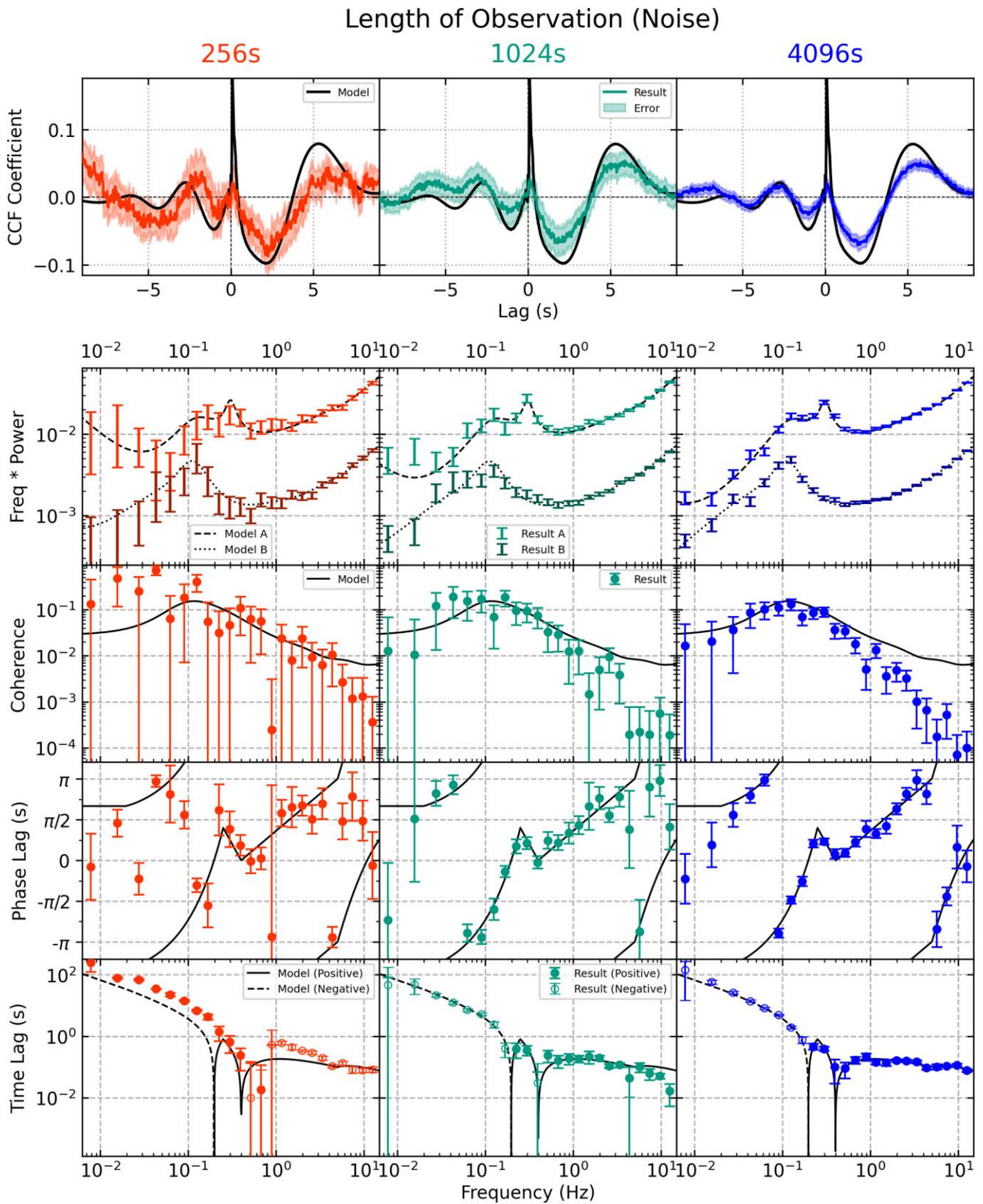


Figure 6. Three different runs of CorrSim with different lengths of observations; 256, 1024, and 4096 s (with a time resolution of 2^{-5} (0.004) s, this is equivalent to 2^{13} , 2^{15} , and 2^{17} bins, respectively) with default noise added. Note the significant affect at lower frequencies.

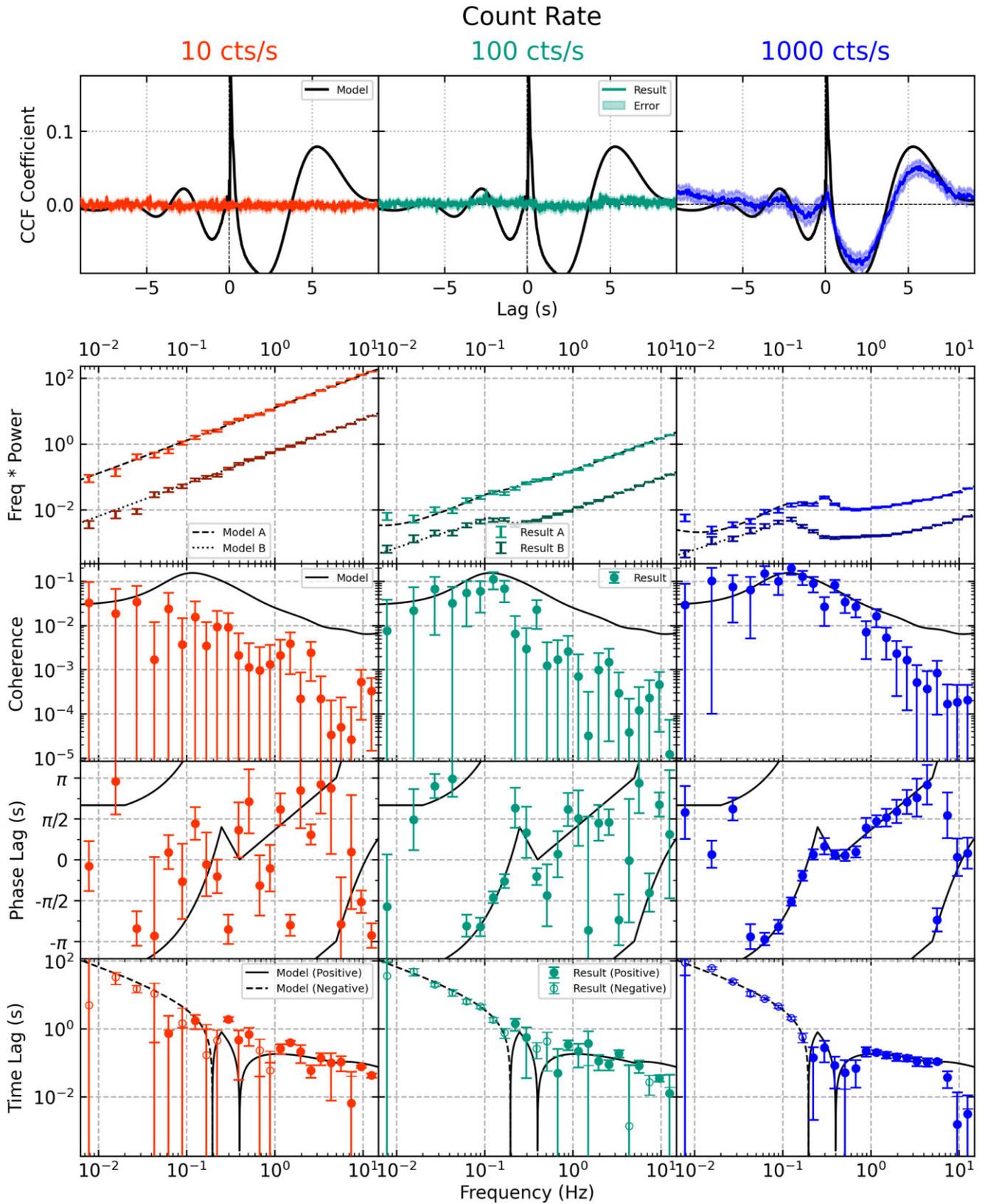


Figure 7. Three different runs of the CorrSim code with different count rates; 10, 100, and 1000 cts s⁻¹ for band A (and 50, 500, and 5000 cts s⁻¹ for band B, respectively). Note how the lower count rates have significantly lower coherence and magnitudes of the CF.

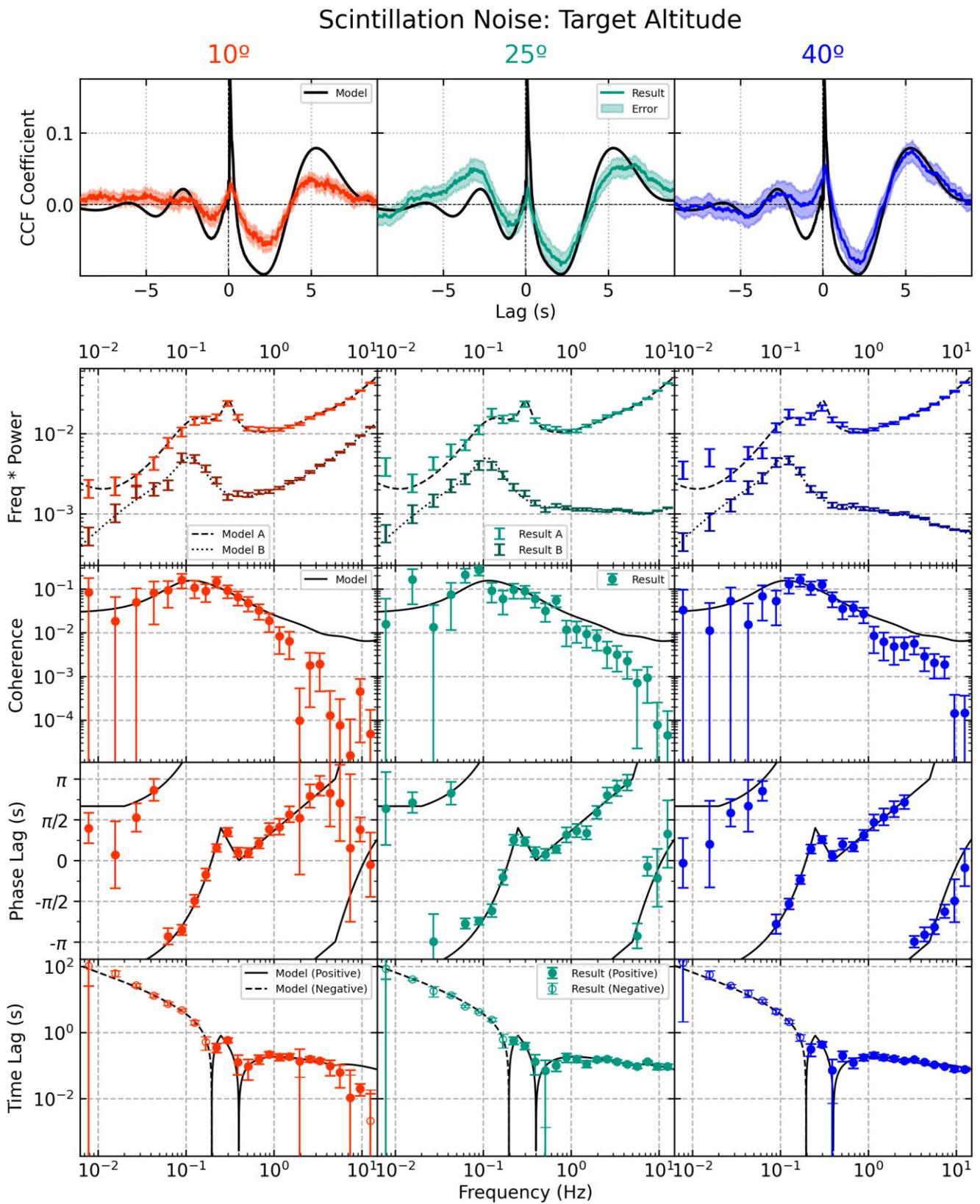


Figure 8. Three runs of CorrSim with different target altitudes and its effect on scintillation noise; 10°, 25°, and 40°. Band A used default noise, while only scintillation noise was applied to band B (to simulate a Space+Ground based observation). Note how scintillation noise strongly affects a target at 10°, yet has minimal effect on a target at 40° (and thus higher altitudes).

Table 1. Default values for the comparison data.

Group	Parameter	Default value	Notes
Observation	Length of observation (s)	2048	–
	Time resolution (s)	0.03125	$= 2^{-5}$
Band A	Mean count rate (cts s ⁻¹)	1000	–
	Fractional RMS (per cent)	0.31	–
	Red noise? (Y/N)	Y	See footnote 1
	Poisson noise? (Y/N)	Y	–
	Readout noise? (Y/N)	Y	4.5 counts
Band B	Scintillation noise? (Y/N)	N	–
	Mean count rate (cts s ⁻¹)	5000	–
	Fractional RMS (per cent)	0.11	–
	Red noise? (Y/N)	Y	See footnote 1
	Poisson noise? (Y/N)	Y	–
Fourier	Readout noise? (Y/N)	Y	4.5 counts
	Scintillation noise? (Y/N)	Y	See footnote 2
	Power spectra model	Lorentzians	See Table 2
	Lag model	Phase	See Table 3
Plotting	CF Range (s)	30	–
	CF Binning	0	–
	Fourier segment length (Bins)	2 ¹²	–
	Fourier rebinning factor	1.3	–
	Reference frequency (Hz)	1	See footnote 3

Notes. 1 For Band A and B, the fractional RMS of the red noise is 0.2 and 0.03, respectively, analogous to red noise from X-rays and optical. The slope of the red noise is -2 for both.

2 Using reasonable values, analogous to the new technology telescope (NTT) at La Silla, Chile (Tarenghi & Wilson 1989):

Telescope diameter = 3.58 m; Telescope altitude = 2400 m; Exposure time = Time resolution - 1.5 ms (i.e. ‘Deadtime’ of 1.5 ms); Target altitude = 40°; Turbulence height = 8000; Empirical coefficient $C_Y = 1.5$.

3 Frequency at which the phase lag is be assumed to be correct (i.e. not shifted by $\pm 2\pi$)

Table 2. Lorentzian parameters.

Norm	Band A		Norm	Band B		Coherence fraction
	Width	Midpoint (Hz)		Width	Midpoint (Hz)	
45	0.5	0	75	0.1	0	1/200
40	4	0	50	0.1	0.1	1/5
25	0.15	0.1	45	2	0	1/150
10	0.1	0.3	6	4	3	0
3	3	3	3	10	8	0
–	–	–	1	20	25	0

Table 3. Phase lag parameters¹.

Distribution	Freq. 1 ²	Lag 1 ³	Freq. 2 ²	Lag 2 ³	Freq. 3 ²	Lag 3 ³
Constant (Phase)	0.001	$-4\pi/3$	0.02	–	–	–
Power	0.02	$-4\pi/3$	0.25	$2\pi/5$	–	–
Linear	0.25	$2\pi/5$	0.4	0	–	–
Linear	0.4	0	5	π	–	–
Polynomial	5	π	200	π	28	$5\pi/2$

Notes. 1 Outside of the specified frequencies, the lag is set to $-4\pi/3$.

2 Units of Hz.

3 Units of Radians.

Fig. 11 shows a selection of the results. As expected, while the subsecond lag is clearly identifiable at a coherence fraction of 1, it is almost completely gone at a fraction of 0.001. Under these source and observational conditions, this graph can thus tell us that a coherence fraction of 0.01 is the minimum to detect a subsecond lag, while values of at least 0.1 are much preferred.

4 DISCUSSION II: EXAMPLE USAGE

We will now demonstrate how an astronomer may use CorrSim to plan and maximize an observation. Let us say they want to observe a source similar to that given in Tables 1 and 2, but with a lower count rate (100 and 1000 counts s⁻¹ in bands A and B, respectively), and they have 1024 s of observation time with a time resolution of 0.03 s

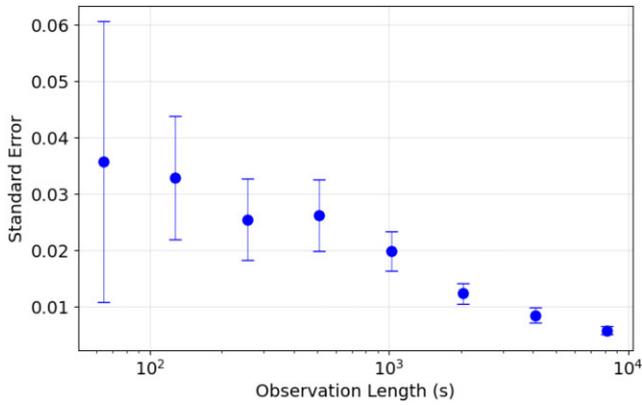


Figure 9. Empirical dependence of the standard error in the CCF – and its standard deviation – on observation length.

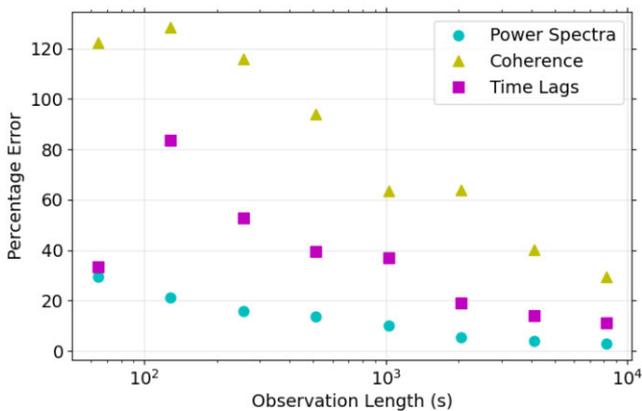


Figure 10. Empirical dependence of the median absolute percentage error on observation length.

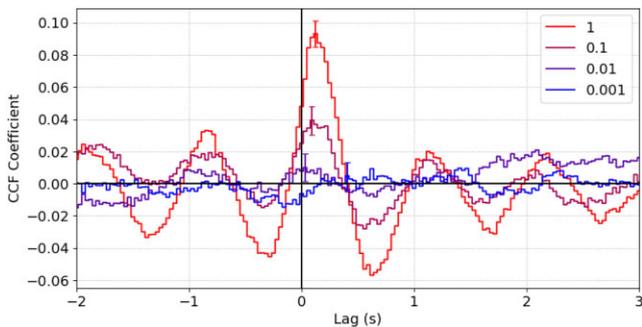


Figure 11. Four CCFs created by varying the coherence fraction of a single Lorentzian at 1 Hz, with all other Lorentzians at 0 coherence. Note how the magnitude of the subsecond lag changes with each coherence fraction. Representative errors are plotted.

(2^{-5}). The source has been known to show a broad precognition dip and a subsecond lag, as well as a small feature in the phase lags (described in Table 4). It is the latter two in particular – the subsecond lag and the feature in the phase lags – that the observer wants to study.

By running `CorrSim` with these parameters, our astronomer gets what is shown in Fig. 12. Initially, they are not pleased with the results – the signal-to-noise ratio (S/N) is too low for their purposes, with the feature in the phase lags indistinguishable from random variations, and the subsecond lag not as clear nor as smooth as hoped. Now they

have seen that their plan would not give them what they want, they make a decision to change the setup of the observation.

What can our astronomer do? They have a selection of possibilities in front of them, but ordinarily, it might mainly involve guesswork (and luck) to decide between them. From their options, they want to investigate three possibilities:

(i) **1: Double the observation length.** By increasing the length of the observation, one would increase the S/N. The Fourier segment size was also increased by a factor of 2, to preserve the number of segments averaged over.

(ii) **2: Decrease time resolution by a factor of four.** Smaller time bins can cause problems like readout noise to become more significant, or missing some photons due to deadtime. Increasing the bin size, and thus lowering the time resolution, would reduce such effects. The Fourier segment size was also reduced by a factor of 4, to preserve the number of segments averaged over.

(iii) **3: Using a better telescope.** Fortunately, our intrepid astronomer knows of another telescope they can use for band A. This telescope has five times the sensitivity (i.e. the mean count rate in band A would become 500) but they can only use it for half the time (i.e. observation length would be 512 s). The Fourier segment size was also reduced by a factor of 2, to preserve the number of segments averaged over. Is there a net benefit to using this telescope?

By altering the parameters, our astronomer runs the program with all these alternative possibilities. Fig. 13 shows the results.

The resulting simulations show that the longer observation gives a decent approximation for the phase lags. Meanwhile, the lower resolution observation gives the highest normalization in the CCF and the closest shape. Finally, the better telescope has a significantly lower white noise floor (i.e. the noise does not dominate the power spectra until higher frequencies than usual; note the difference in power spectra).

These are all important in different ways. For the investigation of the 0.1 s lag, one would need low uncertainties around the 0.1–1 Hz range, the longer observation offers lower uncertainties overall, and the more sensitive telescope has a lower white-noise floor at higher frequencies. Meanwhile, the lower time resolution replicates the frequency range of the unique feature in the phase lags, but does not probe far into the required frequencies. What `CorrSim` thus provides here is additional information with which one can make about their decision.

This comparison shown is somewhat surface-level, and the results shown in Fig. 13 are meant to be indicative of an observation rather than a strong prediction. A more rigorous investigation into predicting an observation would be more complex; for instance, considering the possibility of a fainter source, or different fractional RMS, or a different slope in the power law – any of these could significantly alter the outcome of an observation, and potentially be a problem. In these cases, `CorrSim`'s ability to run such simulations in a matter of seconds to minutes is of value, both for ease of planning and for switching observation strategies during an observation.

5 DISCUSSION III: MODEL TESTING

So far, we have shown the effect of altering parameters – both observational and intrinsic – on resultant light curves and their Fourier products, and in particular, the strength of `CorrSim` in simulating these effects.

Another strength of `CorrSim` is recreating a previously – carried out observation, and then modifying various Fourier parameters to

Table 4. Example-phase lag parameters¹.

Distribution	Freq. 1 ²	Lag 1 ³	Freq. 2 ²	Lag 2 ³	Freq 3 ²	Lag 3 ³
Linear	0.1	$3\pi/4$	0.25	$\pi/5$	–	–
Linear	0.25	$\pi/5$	0.3	$2\pi/3$	–	–
Linear	0.3	$2\pi/3$	0.4	0	–	–
Linear	0.4	0	5	π	–	–
Polynomial	5	π	200	π	28	$5\pi/2$

Notes. 1 Outside of the specified frequencies, the lag is set to $3\pi/4$.

2 Units of Hz.

3 Units of radians.

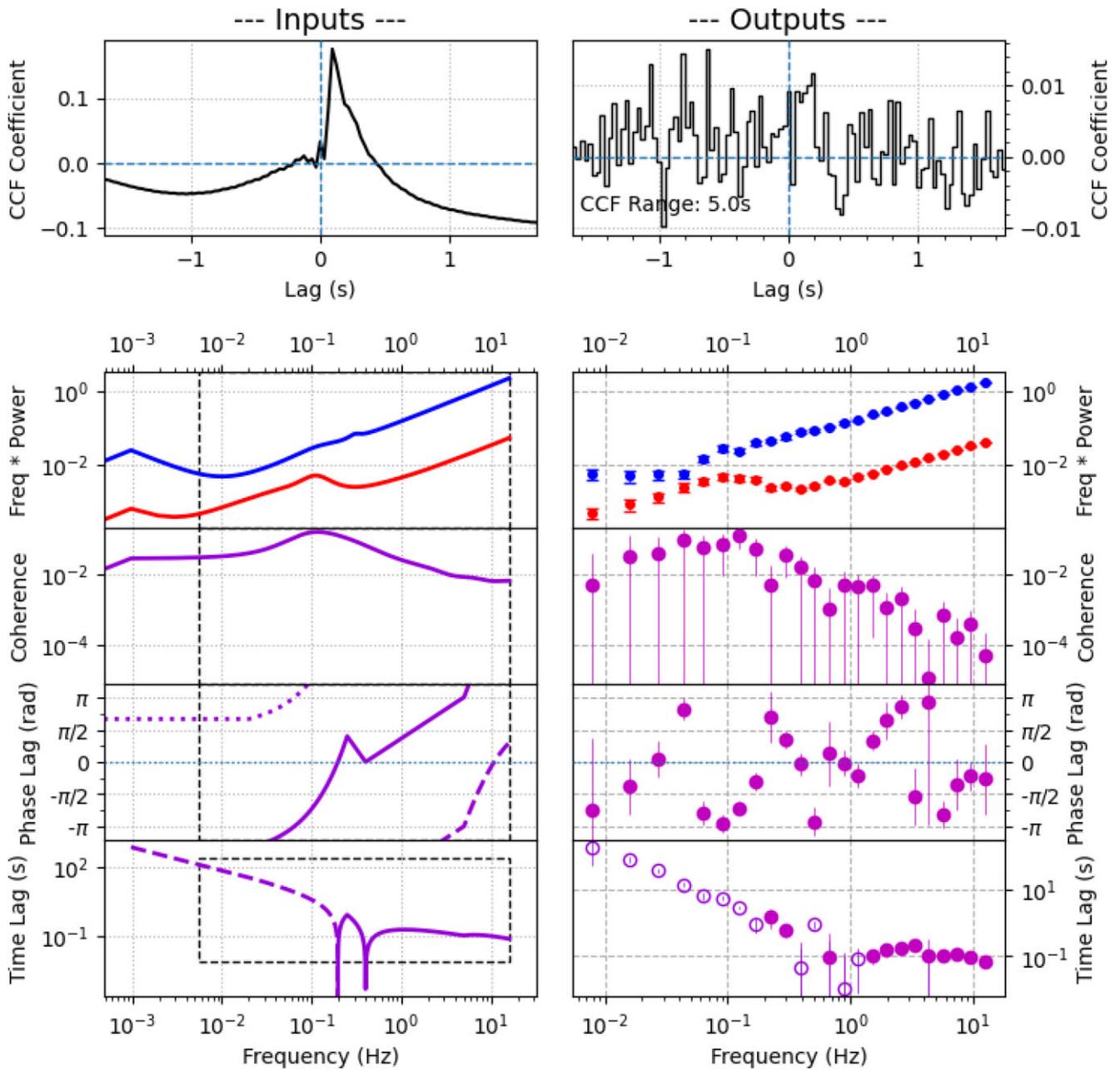


Figure 12. Fourier inputs and outputs for the example observation. Note the lack of clarity in the outputs, especially in the phase lags, compared to the inputs.

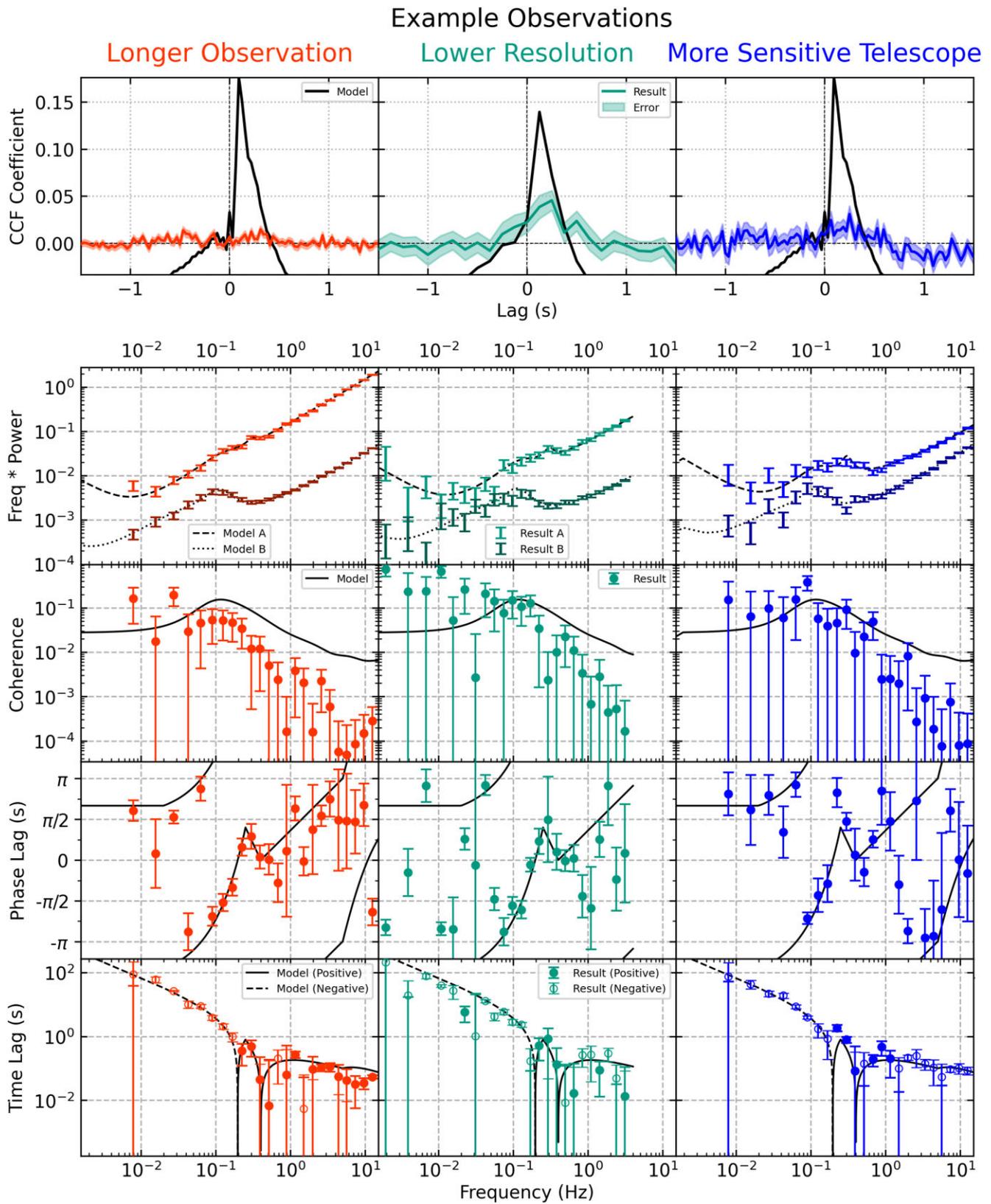
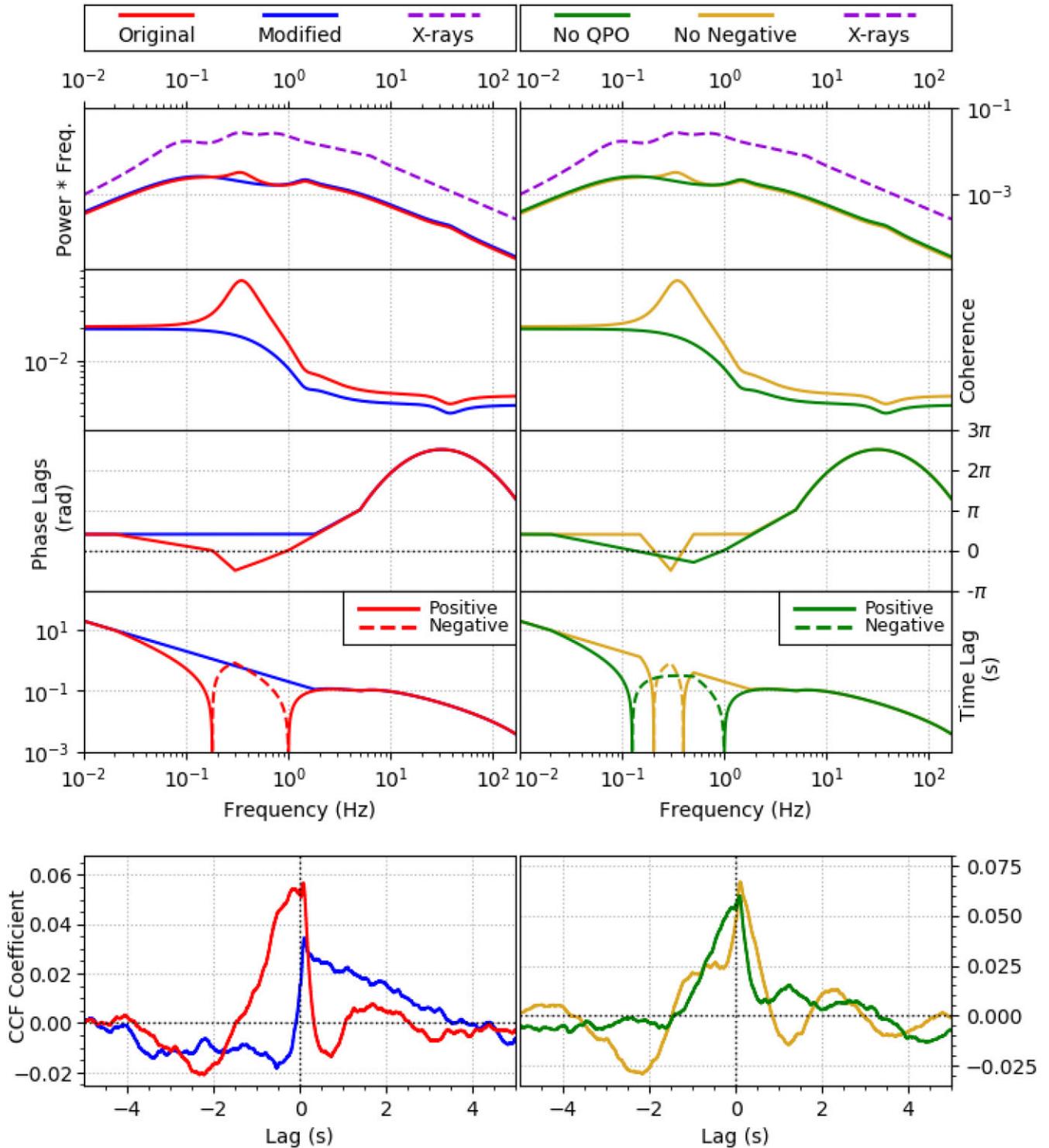


Figure 13. Three modified runs of Fig. 12; double the observation length, four times lower time resolution, and a five times more sensitive telescope with half the observing time. Each simulation has been split into eight segments and averaged. Note how each option has benefits and drawbacks.



Downloaded from https://academic.oup.com/rastai/article/3/1/453/7729280 by guest on 30 August 2024

Figure 14. Various simulations of MAXIJ1820+070, as carried out in Paice et al. (2021). Each column shows two simulations of the i_s band (optical) versus X-rays. Top: Input Fourier components. All y-axes are shared. Bottom: CCFs made by converting the Fourier components into light curves and then cross-correlating. CCFs were averaged over multiple 10s segments. Note that the y-axes are not shared. Left: The red lines are a close representation of the data. The blue lines are a modification that removes the QPO and the negative lags from the i_s band’s Fourier components between 0.02 and 2 Hz. Right: The green lines are a modification that just removes the QPO, and the gold lines are a modification that just removes the negative lags. Note how the behaviour in the CCF changes between -2 and +3 s, showing the significance of each component over this range.

investigate how they affect other outputs, such as the light curves and CCFs. With this, one can test various different models of a source.

This kind of analysis was carried out on MAXIJ1820+070, which can be seen in section 4.5 and appendix A2 of Paice et al. (2021).

The figures showing the simulations have been replicated in Fig. 14. This particular analysis was motivated by a wish to investigate a negative-lag correlation that had appeared in the CCF of one of the epochs; this correlation was not present in any other epoch (most of

Table 5. List of symbols.

$A(t)$	Signal A
$B(t)$	Signal B
$C(f)$	Measured cross-spectrum
$c(\tau)$	Correlation coefficient
D	Normalization for L
dT	Time resolution
F_{rms}	Fractional RMS
f	Frequency; dependent variable of C, L, γ
f_0	Midpoint for L
f_{seg}	Number of frequencies per segment
i	Imaginary number
$L(f)$	Lorentzian
m	Number of segments
N	Number of bins
n	Noise (in power spectra)
p	Power spectra
q	Formula shorthand in error on γ_l^2
P_{rms^2}	Fractional RMS normalization
S	Complex series
s	Signal (noiseless power spectra)
T	Observation length
t	Time; dependent variable of A, B, X
U	Amplitude (in power spectral calculations)
V	Normalization (in power spectral calculations)
$X(t)$	Inverse Fourier transform of $y(f)$
$y(f)$	Fourier transform of $X(t)$
Γ	Full-width at half-maximum for L
$\gamma_l^2(f)$	Intrinsic coherence
γ_{Model}	Model coherence
ΔT_{samp}	Sampling interval
δ	Lags
$\delta\tau$	Time lags
$\delta\phi$	Phase lag
τ	Lag; dependent variable of c

which had even had a negative-lag anticorrelation instead), and was theorized to be connected to a QPO also present in the source. It was hoped that the Fourier components that were creating this feature could be identified.

This analysis was carried out by first finding the Fourier properties of a real observation of MAXI J1820+070, and then modelling them in `CorrSim`. These properties were then altered to create four variations: the original, unedited model of the source; an edited model with the QPO removed; an edited model with a stretch of negative lags removed (but keeping the QPO); and an edited model with both the QPO and the negative lags removed.

By creating simulated CCFs from each of the models, it was thus determined that the negative-lag correlated component was caused by a mix of both the QPO and the negative lags working constructively. Removing each independently created CCFs which still show a negative-lag correlation, but at a lower significance. Removing both completely removes the component, instead giving a CCF profile with a slight anticorrelation at negative lags and an extended correlation at positive lags, similar to those seen earlier in the source, or even in other sources such as Swift J1753.5-0127 (Durant et al. 2008).

From this, the importance of both of these features was concluded: First, that the QPO had a significant effect on the CCF, and such features should be considered in future observations as a possible obfuscating factor; secondly, that a relatively small extent of negative lags can have a significant effect on the resultant CCF, even when coherence is low across that extent.

This was an interesting result for X-ray binaries. Features at negative lags are seen in many X-ray binary systems, such as the ‘pre-cognition dip,’ where optical emission can drop a few seconds prior to an X-ray flare – as if the optical ‘knows’ about future X-ray activity. Such features have been examples of much discussion in the past (see e.g. Kanbach et al. 2001). Here, it’s shown that such features *can* be due to a combination of other periodic features, echoing a result also found by Omama et al. (2021). Does this mean all negative-lag features are the result of periodic processes? Almost certainly not; other theories, such as a two-component flow proposed in Veledina, Poutanen & Vurm (2013), have strong merit for particular features. However, this result highlights that the possibilities behind a particular feature are broader than are sometimes considered.

Studies like this can be valuable for future observations. They can allow for testing of alternative Fourier inputs to see how variability changes – and thus which components are most important. They can also allow for investigating which Fourier inputs and combinations create similar CCFs, thereby testing alternative explanations for already-known features. In either case, these studies have the potential to give valuable insight into the inner processes of XRBs.

6 CONCLUSIONS

First and foremost, we have introduced `CorrSim`, a tool for simulating two correlated light curves based on a variety of parameters, both intrinsic to the source (e.g. mean count rate, power spectra, time lags), and the observation (e.g. observation length and telescope diameter).

Through this program, we have shown how a myriad of different parameters can affect the results of an observation; both parameters intrinsic to the source, and those caused by observation methods. We have also shown how these effects are not always clear; an increase in readout or scintillation noise, for example, can affect the variability and show up clearly in the power spectra. However, it does not typically lead one to imagine a reduction in the correlation function, as we see in Fig. 8 and in the supplementary material.

Some of these parameters can affect observations in significant, but sometimes unclear, ways; the effects of a small telescope or a low altitude target are known to be detrimental, but sometimes such parameters cannot be helped and an astronomer may wonder about the quality of the data they could obtain despite this. This program presents a way to answer these questions.

And these answers can have a very practical side; more than just adjusting expectations, one can adjust the parameters of the observation to make sure that any data gained will be of high-enough quality to answer any questions being investigated.

This is the first aim of the program: To provide astronomers with both qualitative and quantitative metrics with which they can assess the quality of the resultant data, and then change the observational setup to optimize them. In this way, the program can be a tool to actively increase the quality of future observations.

We also demonstrated how this program can be used to simulate different properties of a source, showing a practical example that was carried out in Paice et al. (2021). There, the phase lags were manipulated to remove a range of negative lags, a QPO, and then both, in order to see the effect on the correlation function. From these simulations, it was concluded that the presence of a correlation at negative optical lags is probably not an indication of an optical process occurring before an X-ray process, but instead is more likely a periodic X-ray process having an optical response at a lag greater than π radians in phase.

This demonstrates the second aim of this program: to allow astronomers the ability to take models of already-obtained data and

change them in order to better understand their makeup, assess their validity, and thus inform them of the components of the system.

Through these abilities, CorrSim can be of significant assistance to astronomers across the timeline of an observation and, it is hoped, will help in increasing the frequency, usefulness, and quality of rapid correlated observations in the future.

6.1 Glossary

Table 5 gives a list and definition of all the symbols used in this paper.

ACKNOWLEDGEMENTS

We acknowledge support from STFC and a UGC-UKIERI Thematic Partnership. Over the course of this research, JAP was supported by several sources; a University of Southampton Central VC Scholarship, the ERC under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 715051; Spiders), and STFC consolidated grant ST/X001075/1. JAP also thanks D. Ashton for spectral timing help, as well as A. Stevens and D. Huppenkothen for help with the Stingray software. Some simulations were based on observations by GTC/HIPERCAM and NICER; we thank the observers for the HIPERCAM observations, Vik Dhillon and Stuart Littlefair, and we also thank Keith Gendreau, Zaven Arzoumanian, and the rest of the NICER team for their assistance. SMARTNet helped us to coordinate those observations. This research made use of software and web tools from the High Energy Astrophysics Science Archive Research Center (HEASARC).

DATA AVAILABILITY

All data used within this paper (i.e. all simulated data on which the figures were generated) will be shared on reasonable request to the corresponding author. The program used to generate these data is available at <https://github.com/JohnAPaice/CorrSim>.

REFERENCES

- Acciari V. A. et al., 2011, *ApJ*, 738, 25
 Bendat J. S., Piersol A. G., 2000, *Meas. Sci. Technol.*, 11, 1825
 Berndt B., 1994, *Ramanujan’s Notebooks. Part IV*. Springer, Germany
 Costa E. et al., 1997, *Nature*, 387, 783
 Draghis P., Romani R. W., Filippenko A. V., Brink T. G., Zheng W., Halpern J. P., Camilo F., 2019, *ApJ*, 883, 108
 Durant M. et al., 2008, *ApJ*, 682, L45
 Frank J., King A., Raine D. J., 2002, *Accretion Power in Astrophysics: 3rd edn*. Cambridge Univ. Press, Cambridge
 Gandhi P. et al., 2008, *MNRAS*, 390, L29
 Gandhi P. et al., 2010, *MNRAS*, 407, 2166
 Gandhi P. et al., 2016, *MNRAS*, 459, 554
 Gandhi P. et al., 2017, *Nat. Astron.*, 1, 859
 Huppenkothen D. et al., 2019, *ApJ*, 881, 39
 Kanbach G. et al., 2001, *Nature*, 414, 180
 Kasliwal M. M. et al., 2017, *Science*, 358, 1559
 Lewin W. H. G., van der Klis M., 2006, *Compact Stellar X-ray Sources*, Vol. 39. Cambridge Univ. Press, Cambridge
 McHardy I. M. et al., 2014, *MNRAS*, 444, 1469
 Middleton M. J. et al., 2017, *New Astron. Rev.*, 79, 26
 Misra R., Bora A., Dewangan G., 2018, *Astron. Comput.*, 23, 83
 Miyamoto S., Kimura K., Kitamoto S., Dotani T., Ebisawa K., 1991, *ApJ*, 383, 784
 Omama T., Uemura M., Ikeda S., Morii M., 2021, *PASJ*, 73, 716

- Osborn J., Föhring D., Dhillon V. S., Wilson R. W., 2015, *MNRAS*, 452, 1707
 Pahari M. et al., 2017, *MNRAS*, 469, 193
 Paice J. A. et al., 2019, *MNRAS*, 490, L62
 Paice J. A. et al., 2021, *MNRAS*, 505, 3452
 Papitto A. et al., 2019, *ApJ*, 882, 104
 Perley D. A. et al., 2019, *MNRAS*, 484, 1031
 Scholz P. et al., 2016, *ApJ*, 833, 177
 Tarengi M., Wilson R. N., 1989, in Roddier F. J., ed., *Proc. SPIE Conf. Ser. Vol. 1114, Active Telescope Systems*. SPIE, Bellingham, p. 302
 Thomas J. K., Charles P. A., Buckley D. A. H., Kotze M. M., Lasota J.-P., Potter S. B., Steiner J. F., Paice J. A., 2022, *MNRAS*, 509, 1062
 Timmer J., Koenig M., 1995, *A&A*, 300, 707
 Uttley P., Casella P., 2014, *Space Sci. Rev.*, 183, 453
 van Velzen S. et al., 2019, *ApJ*, 872, 198
 van der Klis M., 1997, in Babu G. J., Feigelson E. D., eds, *Statistical Challenges in Modern Astronomy II*. Springer, New York, p. 321
 Vaughan B. A., Nowak M. A., 1997, *ApJ*, 474, L43
 Vaughan S., Edelson R., Warwick R. S., Uttley P., 2003, *MNRAS*, 345, 1271
 Veledina A., Poutanen J., Vurm I., 2013, *MNRAS*, 430, 3196
 Veledina A., Revnitsev M. G., Durant M., Gandhi P., Poutanen J., 2015, *MNRAS*, 454, 2855
 Veledina A., Gandhi P., Hynes R., Kajava J. J. E., Tsygankov S. S., Revnitsev M. G., Durant M., Poutanen J., 2017, *MNRAS*, 470, 48
 Venables W. N., Ripley B. D., 2002, *Modern Applied Statistics with S*, 4th edn. Springer, New York
 Vincentelli F. M. et al., 2021, *MNRAS*, 503, 614
 Vincentelli F. M. et al., 2023, *MNRAS*, 525, 2509
 Wheatley P. J., Mauche C. W., Mattei J. A., 2003, *MNRAS*, 345, 49
 Zu Y., Kochanek C. S., Peterson B. M., 2011, *ApJ*, 735, 80

SUPPORTING INFORMATION

Supplementary data are available at *RASTAI* online.

CorrSim_Accepted_And_Supplementary_Source.zip CorrSim_Accepted_Supplementary_Material.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: METHODS

Analysis of the relationship between two signals can be split into two main themes; those that take place in the time domain (e.g. cross-correlation analysis), and those that take place in the frequency domain (e.g. Fourier analysis). Each have their own benefits and drawbacks.

Cross-correlation analysis gives a coefficient for how two signals are ‘correlated’ based on some lag (measured in time, or bins). Cross-correlation analysis can be useful even with relatively low numbers of bins (compared with Fourier analysis; Fig. 6 demonstrates how input Correlation models can be reasonably reproduced at small observation lengths, while Fourier models cannot). However, the result of cross-correlation analysis does not trivially translate into quantitative constraints on source properties. For example, multiple interfering processes may ‘average-out’ the resultant coefficient at certain lag ranges, and it can be difficult to separate source features from noise, as well as estimate confidence intervals. Cross-correlation analysis also does not work well when the lags between the two bands are caused by some, potentially variable, periodic process – the resultant values may not show this unique cause.

Cross-frequency analysis, however, ideally solves these two issues; it allows one to determine the time lag as a function of

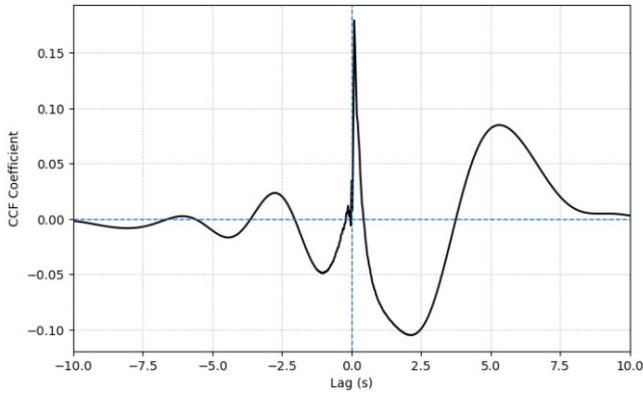


Figure A1. An example CF from cross-correlation analysis, based on data from the black hole X-ray binary MAXIJ1820+070 (Paice et al. 2019). Note how an anticorrelation can be seen at a lag of +2 s, while a correlation can be seen at +5 s.

frequency, and the errors on the coherence and the time lags are better quantified (e.g. Vaughan & Nowak 1997). However, this analysis typically requires a greater number of bins in the light curves, and a higher signal-to-noise ratio (S/N). Additionally, if the time-lag is larger than the time period (i.e. the response in the second light curve occurs more than one oscillation away from the cause in the first light curve), the reported lag will be much shorter, and the true lag obscured; in this case, it is not clear what the full implications are for the measured value and its error. Investigating these questions and determining which analysis is most ideal is one of the motivations behind this work.

Here we will detail the theoretical and mathematical underpinning of both methods.

A1 Cross-correlation analysis

Correlation functions (CFs) find the relationships between two signals as a function of lag. They produce a ‘correlation coefficient’, between -1 and +1, at each of a range of lag bins; -1 indicating pure anticorrelation, and +1 indicating pure correlation. For example, if there is a sharp rise in Signal A, and then a similar sharp rise in Signal B delayed by 10 s, then a CF of signal B versus Signal A will show a positive correlation coefficient at +10 s. Inversely, if Signal B shows a dip instead of a rise, then the correlation coefficient at +10 s will be negative instead – an anticorrelation. An example CF, with its component signals, is seen in Fig. A1.

CorrSim uses the CCF. This particular CF is described in Venables & Ripley (2002), and is best suited to simultaneously sampled light curves (which we will be using):

$$c(\tau) = \frac{1}{N} \sum_{s=\max(1, -\tau)}^{\min(N-\tau, N)} [A(t + \tau) - \bar{A}][B(t) - \bar{B}], \quad (\text{A1})$$

where $A(t)$ and $B(t)$ refer to the two signals dependent upon time t , τ is the dependent variable (the ‘lag’ in units of bins) where $\tau = 1, \dots, N$, and N is the total number of bins in the signals.

The CCF can be modelled mathematically from Fourier components, and is done so by CorrSim. We describe this process in Section 2.2.1.

³Venables & Ripley (2002) use X_i and X_j for these, and define the coefficient as $c_{ij}(t)$; These have been changed so that symbols are consistent across this paper.

There is no straightforward method to evaluate uncertainties or errors in CFs. CorrSim uses a method of splitting up the light curves into segments, creating a CF from each segment, and then finding the standard error between these individual CFs; this is the method also used by `crosscorr` in Xspec. However, there are other methods available, though not programmed into CorrSim; Bootstrapping, for instance, is a method of drawing a random selection of points (sometimes with replacement) and then calculating properties from that selection; here, that can take the form of finding the spread of values from a bootstrapped selection of individual CFs. More advanced methods include error analysis carried out by the `Javelin` code (see Zu, Kochanek & Peterson 2011), or from Misra, Bora & Dewangan (2018) who calculate an analytical formulation for CF errors.

A2 Fourier analysis

Fourier analysis, also known as cross-spectral analysis, is founded on the idea that any signal can be constructed by summing together a series of sinusoids of varying phases and amplitudes. Taking two simultaneous light curves, we can pair up those component sinusoids which are at the same frequency. One can then determine the coherence and lags between two signals as a function of frequency, rather than having a value at one time lag for all frequencies.

These results can then be used to help disentangle the processes between bands which might be difficult to uncover otherwise. For example, in cross-correlation analysis, strongly correlated processes could hide weakly correlated ones; however, if the stronger processes only occur at lower frequencies and the weaker ones at higher frequencies, then Fourier analysis would separate them out. This is especially useful for sources which have processes across the Fourier spectrum—again, such as X-ray binaries (see Lewin & van der Klis 2006, Ch. 2). Additionally, the errors on the coherence and the time lags are better quantified than cross-correlation analysis (Vaughan & Nowak 1997).

The Fourier transform that is used in our work is the fast Fourier transform from the `SCIPY` package.⁴ There, the equation for a Fourier transformed series y dependent upon frequency f is:

$$y(f) = \sum_{t=0}^{N-1} e^{-2\pi i \frac{ft}{N}} X(t) \quad (\text{A2})$$

along with its inverse,

$$X(t) = \frac{1}{N} \sum_{f=0}^{N-1} e^{2\pi i \frac{ft}{N}} y(f), \quad (\text{A3})$$

where $X(t)$ is the series to be Fourier transformed, of length N , and i is the imaginary number.⁵ For Fourier analysis, N should be a power of two; this can be problematic when trying to analyse at the lowest frequencies (which may involve up to half of the data not being used), but for the higher frequencies, the light curves can be easily split into smaller segments and averaged.

Much of this research has made use of the `STINGRAY`⁶ python package (Huppenkothen et al. 2019), which is used in the calculation of Fourier components from the simulated light curves.

⁴<https://docs.scipy.org/doc/>

⁵In the SciPy documentation, these are referred to as $x(n)$, N , and j respectively.

⁶<https://github.com/StingraySoftware/stingray>

A2.1 Power spectra

A power spectrum shows the amount of variability (‘power’) at each frequency, which is dependent upon the amplitude of the sin waves at that frequency.

‘Power’ is an umbrella term, and there are several ways to normalize it. In our work, the normalization that we have used is the fractional RMS. This is defined by van der Klis (1997) (see also Miyamoto et al. 1991), and is given by the formula:

$$R_{\text{rms}}^2 = \frac{2\Delta T_{\text{samp}}}{\bar{X}^2 N}, \quad (\text{A4})$$

where ΔT_{samp} is the sampling interval of the data (i.e. the ‘exposure time’ of the telescope), \bar{X} is the mean rate of the series, and N is the total number of bins in the series. The units of this normalization are $(\text{rms}/\text{mean})^2 \text{ Hz}^{-1}$. This normalization is useful, particular for our research, since integrating the power spectrum yields the fractional variance of the data (Vaughan et al. 2003).

Power spectra can often be described by a series of Lorentzians, which are distributions given by the form:

$$L(f) = \frac{D}{\pi} \frac{\frac{1}{2}\Gamma}{(f - f_0)^2 + (\frac{1}{2}\Gamma)^2}, \quad (\text{A5})$$

where D is the normalization (controls its magnitude), Γ is the full width at half maximum (controls its width), and f_0 is the midpoint ($f_0 = 0$ indicates a zero-centred Lorentzian). For an example of a power spectrum being modelled by a series of Lorentzians, see Fig. 1.

A2.2 Coherence

Coherence is the magnitude of the complex-valued cross-spectrum; i.e. how much the amplitudes of the sin waves relate to each other. A higher coherence means the two signals are more correlated at that frequency. This value is constrained between 0 (completely incoherent) and 1 (completely coherent).

The coherence is given by Vaughan & Nowak (1997), and mathematically can be given thusly. Assume a power spectra $p = |s|^2 + |n|^2$, where $|s|^2$ is the power of the signal, and $|n|^2$ is the noise. The coherence function is then:

$$\gamma_I^2(f) = \frac{|\langle C(f) \rangle|^2}{\langle |s_A(f)|^2 \rangle \langle |s_B(f)|^2 \rangle} \quad (\text{A6})$$

$$|\langle C(f) \rangle|^2 = |\langle s_A^* s_B \rangle + \langle s_A^* n_B \rangle + \langle n_A^* s_B \rangle + \langle n_A^* n_B \rangle|^2, \quad (\text{A7})$$

where the subscript I denotes intrinsic coherence between noiseless signals, subscripts A and B refer to the two signals, an asterisk denotes the complex conjugate of the power spectra, and f is frequency.

For XRBs, the variability usually gives high powers with relatively high coherence. Using the error formulation of Vaughan & Nowak (1997), the errors in this case can be computed like so:

$$\gamma_I^2 = \frac{|\langle C \rangle|^2 - q^2}{|s_A|^2 |s_B|^2} \times \left(1 \pm m^{-1/2} \left[\frac{2q^4 m}{(|\langle C \rangle|^2 - q^2)^2} + \frac{|n_A|^4}{|s_A|^4} + \frac{|n_B|^4}{|s_B|^4} + \frac{m\delta\gamma_I^2}{\gamma_I^4} \right]^{1/2} \right) \quad (\text{A8})$$

where m is the number of segments averaged over, and

$$q^2 = \frac{|s_A|^2 |n_B|^2 + |n_A|^2 |s_B|^2 + |n_A|^2 |n_B|^2}{m} \quad (\text{A9})$$

Vaughan & Nowak (1997) also give an equation for high powers and relatively low coherence; however, this option is not included in CorrSim both for simplicity’s sake and because this case does not typically arise in XRBs. This is also true for the case of low powers, combined with the fact that no complete solution exists for confidence values under these conditions.

A2.3 Phase and time lags

The phase lags are the phase angle of the complex-valued cross-spectrum; i.e. the offset between the sin waves at each frequency, as a function of their phase. In units of radians, they lie between $-\pi$ and π , both of which relate to perfect antiphase, while a value of 0 relates to perfect phase.

The errors on the phase lags are calculated from the coherence:

$$\delta_\phi = \delta_\phi \pm \sqrt{\frac{1 - \gamma_I^2}{2\gamma_I^2 f_{\text{seg}} m}}, \quad (\text{A10})$$

where γ_I is the coherence, f_{seg} is the number of frequencies per segment, and m is the number of segments averaged over (Bendat & Piersol 2000; Uttley & Casella 2014).

Time lags are similar to phase lags, but are, unsurprisingly, a function of time. They are calculated thusly:

$$\delta_\tau = \frac{\delta_\phi}{2\pi f}, \quad (\text{A11})$$

where δ_ϕ is the phase lag, and f is the frequency of the bin. Errors on the time lags are calculated similarly. Intrinsic drawbacks of phase lags, and possible solutions to such, are discussed in Section 2.2.2.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.