

Periodograms for Multiband Astronomical Time Series

Jacob T. VanderPlas¹

Željko Ivezić²

ABSTRACT

This paper introduces the *multiband periodogram*, a general extension of the standard Lomb-Scargle approach for detecting periodic signals in time-domain data. In addition to advantages of the Lomb-Scargle method such as treatment of non-uniform sampling and heteroscedastic errors, the multiband periodogram significantly improves period finding for randomly sampled multiband light curves (e.g., Pan-STARRS, DES and LSST). The light curves in each band are modeled as arbitrary truncated Fourier series, with the period and phase shared across all bands. The key aspect is the use of Tikhonov regularization which drives most of the variability into the so-called base model common to all bands, while fits for individual bands describe residuals relative to the base model and typically require lower-order Fourier series. This decrease in the required number of fit parameters is the main reason for improved performance. We use simulated light curves and randomly subsampled SDSS Stripe 82 data to demonstrate the superiority of this method compared to other methods from the literature. The Python implementation of this method, along with code to fully reproduce the results reported here, is available on GitHub.

Subject headings: methods: data analysis — methods: statistical

1. Introduction

Many types of variable stars show periodic flux variability (Eyer & Mowlavi 2008). Periodic variable stars are important both for testing models of stellar evolution and for using such stars as distance indicators (e.g., Cepheids and RR Lyrae stars). One of the first and main goals of the analysis is to detect variability and to estimate the period and its

¹eScience Institute, University of Washington

²Department of Astronomy, University of Washington

uncertainty. A number of parametric and non-parametric methods have been proposed to estimate the period of an astronomical time series (e.g., Graham et al. 2013, and references therein).

The most popular non-parametric method is the phase dispersion minimization (PDM) introduced by Stellingwerf (1978). Dispersion per bin is computed for binned phased light curves evaluated for a grid of trial periods. The best period minimizes the dispersion per bin. A similar and related non-parametric method that has been recently gaining popularity is the Supersmooth routine (Reimann 1994). It uses a running mean or running linear regression on the data to fit the observations as a function of phase to a range of periods. The best period minimizes a figure-of-merit, adopted as weighted sum of absolute residuals around the running mean. Neither the Supersmooth algorithm nor the PDM method require *a priori* knowledge of the light curve shape. We note that neither method produces posterior probability distribution for the period but rather a single point estimate.

The most popular parametric method is the Lomb-Scargle periodogram, which is discussed in detail in Section 2. The Lomb-Scargle periodogram is related to the χ^2 for a least-square fit of a single sinusoid to data and can treat non-uniformly sampled time series with heteroscedastic measurement uncertainties. The underlying model of the Lomb-Scargle periodogram is nonlinear in frequency and basis functions at different frequencies are not orthogonal. As a result, the periodogram has many local maxima and thus in practice the global maximum of the periodogram is found by grid search (for details see, e.g. Ivezić et al. 2014). A more general parameteric method based on the use of continuous-time autoregressive moving average (CARMA) model was recently introduced by Kelly et al. (2014). CARMA models can also treat non-uniformly sampled time series with heteroscedastic measurement uncertainties, and can handle complex variability patterns.

A weakness of the above methods is that they require homogeneous measurements – for astronomy data, this means that successive measurements must be taken through a single photometric bandpass (filter). This has not been a major problem for past surveys because measurements are generally taken through a single photometric filter (e.g. LINEAR, Sesar et al. 2011), or nearly-simultaneously in all bands at each observation (e.g. SDSS, Sesar et al. 2010). For the case of simultaneously taken multiband measurements, Süveges et al. (2012) utilized the principal component method to optimaly extract the best period. Their method is essentially a multiband generalization of the well-known two-band Welch-Stetson variability index (Stetson 1996). Unfortunately, when data in each band are taken at different times, the principal component approach in not applicable. In such cases, past studies have generally relied on *ad hoc* methods such as a majority vote among multiple single-band estimates of the periodogram (e.g., Oluseyi et al. 2012).

For surveys that obtain multiband data one band at a time, such as Pan-STARRS (Kaiser et al. 2010) and DES (Flaugher 2008), and for future multicolor surveys such as LSST (Ivezić et al. 2008), this *ad hoc* approach is not optimal. In order to take advantage of the full information content in available data, it would be desirable to have a single estimate of the periodogram which accounts for all observed data in a manner which does not depend on the underlying spectrum of the object. We propose such a method in this paper.

The proposed method is essentially a generalization of the Lomb-Scargle method to multiband case. The light curves in each band are modeled as arbitrary truncated Fourier series, with the period, and optionally the phase, shared across all bands. The key aspect enabling this approach is the use of Tikhonov regularization (discussed in detail in §4.3) which drives most of the variability into the so-called *base model* common to all bands, while fits for individual bands describe residuals relative to the base model and typically require lower-order Fourier series. This decrease in the required number of fit parameters is the main reason for improved performance.

In Section 2 we provide a brief overview of least-squares periodic fitting, and in Section 3 derive the matrix-based formalism for single-band periodic analysis used through the rest of this work. Section 4 introduces several extensions and generalizations to the single-band model that the matrix formalism makes possible, including floating mean models, truncated Fourier models, and regularized models. In Section 5, we use the ideas behind these extensions to motivate the *multiband periodogram*, and show some examples of its use on simulated data. In Section 6 we apply this method to measurements of 483 RR Lyrae stars first explored by Sesar et al. (2010, hereafter S10), and conclude in Section 7.

2. Brief Overview of Periodograms

The detection and quantification of periodicity in time-varying signals is an important area of data analysis within modern time-domain astronomical surveys. For evenly-spaced data, the *Schuster periodogram*, introduced in 1905, gives a quantitative measure of the periodicity of data as a function of the angular frequency ω . For data $\{y_k\}_{k=1}^N$ measured at equal intervals $t_k = t_0 + k\Delta t$, the Schuster periodogram, which measures the spectral power as a function of the angular frequency, is given by

$$C(\omega) = \frac{1}{N} \left| \sum_{k=1}^N y_k e^{i\omega t_k} \right|^2, \quad (1)$$

and can be computed very efficiently using the Fast Fourier Transform.

Because astronomical observing cadences are rarely so uniform, many have looked at extending the ideas behind the periodogram to work with unevenly-sampled data. Most famously, Lomb (1976) and Scargle (1982) extended earlier work to define the *normalized periodogram*:

$$P_N(\omega) = \frac{1}{2V_y} \left[\frac{[\sum_k (y_k - \bar{y}) \cos \omega(t_k - \tau)]^2}{\sum_k \cos^2 \omega(t_k - \tau)} + \frac{[\sum_k (y_k - \bar{y}) \sin \omega(t_k - \tau)]^2}{\sum_k \sin^2 \omega(t_k - \tau)} \right], \quad (2)$$

where \bar{y} is the mean and V_y is the variance of the data $\{y_k\}$, and τ is the time-offset which makes $P_N(\omega)$ independent of a translation in t (see Press et al. 2007, for an in-depth discussion). Lomb (1976) showed that this time-offset has a deeper effect: namely, it makes P_N identical to the estimate of harmonic content given a least-squares fit to a single-component sinusoidal model,

$$d(t) = A \sin(\omega t + \phi). \quad (3)$$

This long-recognized connection between spectral power and least squares fitting methods was solidified by Jaynes (1987), who demonstrated that the normalized periodogram of Lomb and Scargle is a sufficient statistic for inferences about a stationary-frequency signal in the presence of Gaussian noise. Building on this result, Bretthorst (1988) explored the extension of these methods to more complicated models with multiple frequency terms, nonstationary frequencies, and other more sophisticated models within a Bayesian framework.

While the important features of least squares frequency estimation via Lomb-Scargle periodograms has been discussed elsewhere, we will present a brief introduction to the subject in the following section. The matrix-based formalism we develop here will make clear how the method can be extended to more sophisticated models, including the multiband method.

3. Standard Least Squares Spectral Fitting

In this section we present a quick quantitative introduction to the least squares fitting formulation of the normalized periodogram of Equation (2). We denote N observed data points as

$$D = \{t_k, y_k, \sigma_k\}_{k=1}^N \quad (4)$$

where t_k is the time of observation, y_k is the observed value (typically a magnitude), and σ_k describes the Gaussian errors on each value. Without loss of generality we will assume that the data y_k are centered such that the measurements within each band satisfy

$$\frac{\sum_k w_k y_k}{\sum_k w_k} = 0 \quad (5)$$

where the weights are $w_k = \sigma_k^{-2}$.

3.1. Stationary Sinusoid Model

The normalized periodogram of Equation (2) can be derived from the normalized χ^2 of the best-fit single-term stationary sinusoidal model given in Equation (3). To make the problem linear, we can re-express the model in terms of the parameter vector $\theta = [A \cos \phi, A \sin \phi]$ so that our model is

$$y(t|\omega, \theta) = \theta_1 \sin(\omega t) + \theta_2 \cos(\omega t). \quad (6)$$

For a given ω , the maximum likelihood estimate of the parameters θ can be found by minimizing the χ^2 of the model, which is given by

$$\chi^2(\omega) = \sum_k \frac{[y_k - y(t_k|\omega, \theta)]^2}{\sigma_k^2}. \quad (7)$$

For the single-term Fourier model, it can be shown (See, e.g. Ivezić et al. 2014) that

$$\chi_{min}^2(\omega) = \chi_0^2[1 - P_N(\omega)] \quad (8)$$

where $P_N(\omega)$ is the normalized periodogram given in Equation (2) and χ_0^2 is the reference χ^2 for a constant model, which due to the assumption in Equation (5) is simply $\chi_0^2 = \sum_k (y_k/\sigma_k)^2$.

3.2. Matrix Formalism

The expressions related to the stationary sinusoid model can be expressed more compactly by defining the following matrices:

$$X_\omega = \begin{bmatrix} \sin(\omega t_1) & \cos(\omega t_1) \\ \sin(\omega t_2) & \cos(\omega t_2) \\ \vdots & \vdots \\ \sin(\omega t_N) & \cos(\omega t_N) \end{bmatrix}; \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} \quad (9)$$

With these definitions, the model in Equation (6) can be expressed as a simple linear product, $y(t|\omega, \theta) = X_\omega \theta$, and the model and reference χ^2 can be written

$$\chi^2(\omega) = (y - X_\omega \theta)^T \Sigma^{-1} (y - X_\omega \theta) \quad (10)$$

$$\chi_0^2 = y^T \Sigma^{-1} y \quad (11)$$

The expression for the normalized periodogram can be computed by finding via standard methods the value of θ which minimizes $\chi^2(\omega)$, and plugging the result into Equation (8). This yields

$$P_N(\omega) = \frac{y^T \Sigma^{-1} X_\omega [X_\omega^T \Sigma^{-1} X_\omega]^{-1} X_\omega^T \Sigma^{-1} y}{y^T \Sigma^{-1} y}. \quad (12)$$

We note that this expression is equivalent to Equation (2) in the homoscedastic case with $\Sigma \propto V_y I$.

3.3. Simple Single-band Period Finding

As an example of the standard periodogram in action, we perform a simple single-band harmonic analysis of simulated r -band observations of an RR Lyrae light curve, based on empirical templates derived in S10 (Figure 1). The observations are of a star with a period of 0.622 days, and take place on 60 random nights over a 6-month period, as seen in the left panel.

The upper-right panel shows the normalized periodogram for this source as a function of period. While the power does peak at the true period of 0.622 days, an aliasing effect is readily apparent near $P = 0.38$. This additional peak is due to beat frequency between the true period P and the observing cadence of ~ 1 day. This beat frequency is the first in a large sequence: for nightly observations, we'd expect to find excess power at periods $P_n = P/(1 + nP)$ days, for any integer n . The strong alias in Figure 1 corresponds to the $n = 1$ beat period $P_n = 0.383$. Though it is possible to carefully correct for such aliasing by iteratively removing contributions from the estimated window function (e.g. Roberts et al. 1987), we'll ignore this detail in the current work.

The lower-right panel of Figure 1 shows the maximum likelihood interpretation of this periodogram: it is a measure of the normalized χ^2 for a single-term sinusoidal model. Here we visualize the data from the left panel, but folded as a function of phase, and overplotted with the best-fit single-term model. Here it is apparent that the single-term model is highly biased: RR Lyrae light curves are, in general, much more complicated than a simple sinusoid. Nevertheless, the simplistic sinusoidal model is able to recover the correct frequency to a high degree of accuracy (roughly related to the width of the peak) and significance (roughly related to the height of the peak). For a more complete introduction to and discussion of the single-term normalized periodogram, refer to, e.g. Bretthorst (1988) or Ivezić et al. (2014).



Fig. 1.— An illustration of the basic periodogram and its relationship to the single-term sinusoid model. The left panel shows the input data, while the right panels show the fit derived from the data. The upper-right panel shows the periodogram with a clear peak at the true period of 0.622 days, and the bottom-right panel shows the data as a function of the phase associated with this period. Note in the periodogram the presence of the typical aliasing effect, with power located at beat frequencies between the true period and the 1-day observing cadence (see Section 3.3 for further discussion).

4. Extending the Periodogram

We have shown two forms of the classic normalized periodogram: Equation (2) and Equation (12). Though the two expressions are equivalent, they differ in their utility. Because the expression in Equation (2) avoids the explicit construction of a matrix, it can be computed very efficiently. Furthermore, through clever use of the Fast Fourier Transform, expressions of the form of Equation (2) can be evaluated exactly for N frequencies in $\mathcal{O}[\log N]$ time (Press & Rybicki 1989).

The matrix-based formulation of Equation (12), though slower than the Fourier-derived formulation, is a more general expression and allows several advantages:

1. It is trivially extended to heteroscedastic and/or correlated measurement noise in the data y_k through appropriate modification of the *noise covariance matrix* Σ
2. It is trivially extended to more sophisticated linear models by appropriately modifying the *design matrix* X_ω .
3. It is trivially extended to include Tikhonov/L2-regularization terms (see §4.3 for more details) by adding an appropriate diagonal term to the *normal matrix* $X_\omega^T \Sigma^{-1} X_\omega$.

In the remainder of this section, we will explore a few of these modifications and how they affect the periodogram and resulting model fits.

4.1. Stationary Sinusoid with Floating Mean

As an example of one of these generalizations, we'll consider what Zechmeister & Kürster (2009) call the “generalized Lomb-Scargle method”, and which we'll call the *floating-mean periodogram*. This method adjusts the classic normalized periodogram by fitting the mean of the model alongside the amplitudes:

$$y(t \mid \omega, \theta) = \theta_0 + \theta_1 \sin \omega t + \theta_2 \cos \omega t \quad (13)$$

This model can be more accurate than the standard periodogram for certain observing cadences and selection functions. Zechmeister & Kürster (2009) detail the required modifications to the harmonic formalism of Equation (2) to allow the mean to float in the model. In the matrix formalism, the modification is much more straightforward: all that is required is to add a column of ones to the X_ω matrix before computing the power via Equation (12). This column of ones corresponds to a third entry in the parameter vector θ , and acts as a uniform constant offset for all datapoints.



Fig. 2.— An illustration of the effect of the floating mean model for censored data. The data consist of 80 observations drawn from a sinusoidal model. To mimick a potentially damaging selection effect, all observations with magnitude fainter than 16 are removed (indicated by the light-gray points). The standard and floating-mean periodograms are computed from the remaining data; these fits are shown over the data in the left panel. Because of this biased observing pattern, the mean of the observed data is a biased estimator of the true mean. The standard fixed-mean model in this case fails to recover the true period of 0.622 days, while the floating mean model still finds the correct period.

For well-sampled data, there is usually very little difference between a standard periodogram and a floating-mean periodogram. Where this becomes important is if selection effects or observing cadences cause there to be preferentially more observations at certain phases of the light curve: a toy example demonstrating this situation is shown in Figure 2. The data are drawn from a sinusoid with Gaussian errors, and data with a magnitude fainter than 16 are removed to simulate an observational bias (left panel). Because of this observational bias, the mean of the observed data are a poor predictor of the true mean, causing the standard fixed-mean method to poorly fit the data and miss the input period (upper-right panel). The floating-mean approach is able to automatically adjust for this bias, resulting in a periodogram which readily detects the input period of 0.622 days (lower-right panel).

4.2. Truncated Fourier Models

As mentioned above, the standard periodogram is equivalent to fitting a single-term stationary sinusoidal model to the data. A natural extension is to instead use a multiple-term sinusoidal model, with frequencies at integer multiples of the fundamental frequency. With N Fourier terms, there are $2N + 1$ free parameters, and the model is given by

$$y(t|\omega, \theta) = \theta_0 + \sum_{n=1}^N [\theta_{2n-1} \sin(n\omega t) + \theta_{2n} \cos(n\omega t)]. \quad (14)$$

Because this model remains linear in the parameters θ , it can be easily accommodated into the matrix formalism above. For example, an $N = 2$ -term floating-mean model can be constructed by building a design matrix X_ω with $2N + 1 = 5$ columns:

$$X_\omega^{(2)} = \begin{bmatrix} 1 & \sin(\omega t_1) & \cos(\omega t_1) & \sin(2\omega t_1) & \cos(2\omega t_1) \\ 1 & \sin(\omega t_2) & \cos(\omega t_2) & \sin(2\omega t_2) & \cos(2\omega t_2) \\ 1 & \sin(\omega t_3) & \cos(\omega t_3) & \sin(2\omega t_3) & \cos(2\omega t_3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \sin(\omega t_N) & \cos(\omega t_N) & \sin(2\omega t_N) & \cos(2\omega t_N) \end{bmatrix} \quad (15)$$

Computing the power via Equation (12) using $X_\omega^{(2)}$ will give the two-term periodogram. For larger N , more columns are added, but the periodogram can be computed in the same manner. Figure 3 shows a few examples of this multiterm Fourier approach as applied to the simulated RR Lyrae light curve from Figure 1, and illustrates several important insights into the subtleties of this type of multiterm fit.

First, we see in the right panel that all three models show a clear signal at the true period of $P_0 = 0.622$ days. The higher-order models, however, also show a spike in power



Fig. 3.— The model fits and periodograms for several truncated Fourier models. The data are the same as those in Figure 1. Note that the higher-order models will generally show periodogram peaks at multiples of the true fundamental frequency P_0 : this is because for integer n less than the number of Fourier terms in the model, P_0 is a higher harmonic of the model at $P = nP_0$. Additionally, the increased degrees of freedom in the higher-order models let them fit better at any frequency, which drives up the “background” level in the periodogram.

at $P_1 = 2P_0$: the reason for this is that for and $N > 1$ -term model, the period P_0 is the first harmonic of a model with fundamental frequency $2P_0$, and the higher-order models contain the single-period result.

Second, notice that as the number of terms is increased, the general “background” level of the periodogram increases. This is due to the fact that the periodogram power is inversely related to the χ^2 of the fit at each frequency. A more flexible higher-order model can better fit the data at all periods, not just the true period. Thus in general the observed power of a higher-order Fourier model will be everywhere higher than the power of a lower-order Fourier model.

4.3. Regularized Models

The previous sections raise the question: how complicated a model should we use? We have seen that as we add more terms to the fit, the model will come closer and closer to the observed data. In the extreme case, when the number of model parameters equals the number of data points, the model can fit the data *exactly* regardless of frequency and the periodogram will be everywhere unity (though in most cases, numerical inaccuracies in the fit prevent a truly perfect fit). For very high-order models, the fit becomes very sensitive to the noise in the data, and we say we have *over-fit* the data. This can be addressed by explicitly truncating the series, but we can also use a *regularization* term to mathematically enforce a less complicated model.

A regularization term is an explicit penalty on the magnitude of the model parameters θ , and can take a number of forms. For computational simplicity here we’ll use an L^2 regularization – also known as Tikhonov Regularization (Tikhonov 1963) or Ridge Regression (Hoerl & Kennard 1970) – which is a quadratic penalty term in the model parameters added to the χ^2 . Mathematically, this is equivalent in the Bayesian framework to using a zero-mean Gaussian prior on the model parameters.

We encode our regularization in the matrix $\Lambda = \text{diag}([\lambda_1, \lambda_2 \cdots \lambda_M])$ for a model with M parameters, and construct a “regularized” χ^2 :

$$\chi_{\Lambda}^2(\omega) = (y - X_{\omega}\theta)^T \Sigma^{-1} (y - X_{\omega}\theta) + \theta^T \Lambda \theta \quad (16)$$

Minimizing this regularized χ^2 , solving for θ , and plugging into the expression for P_N gives us the regularized counterpart of Equation (12):

$$P_{N,\Lambda}(\omega) = \frac{y^T \Sigma^{-1} X_{\omega} [X_{\omega}^T \Sigma^{-1} X_{\omega} + \Lambda]^{-1} X_{\omega}^T \Sigma^{-1} y}{y^T \Sigma^{-1} y}. \quad (17)$$

Notice that the effect of this regularization term is to add a diagonal penalty to the normal matrix $X_\omega^T \Sigma^{-1} X_\omega$, which has the additional feature that it can correct ill-posed models where the normal matrix is non-invertible. This feature of the regularization will become important for the multiband models discussed below.

In Figure 4, we compare a regularized and unregularized 20-term truncated Fourier model on our simulated RR Lyrae light curve. We use $\lambda = 0$ on the offset term, and make the penalty λ_j progressively larger for each harmonic component. The result of the regularized model is less over-fitting to the input data, and stronger periodogram peaks.

5. A Multiple-Band Model

In this section we will apply what was learned in the previous sections to construct the *multiband periodogram* which can flexibly account for heterogeneous sources of data for a single object. To compute such a periodogram, we will take advantage of the easy extensibility of the matrix formalism which led to our generalizations above. The model contains the following features:

1. An N_{base} -term truncated Fourier fit which models a latent parameter, which we'll call the “overall variability”.
2. A set of N_{band} -term truncated Fourier fits, each of which models the residual of a single band from this overall variability.

The total number of parameters for K filters is then $M_K = (2N_{base} + 1) + K(2N_{band} + 1)$. As a result, for each band k we have the following model of the observed magnitudes:

$$y_k(t|\omega, \theta) = \theta_0 + \sum_{n=1}^{M_{base}} [\theta_{2n-1} \sin(n\omega t) + \theta_{2n} \cos(n\omega t)] + \theta_0^{(k)} + \sum_{n=1}^{M_{band}} [\theta_{2n-1}^{(k)} \sin(n\omega t) + \theta_{2n}^{(k)} \cos(n\omega t)]. \quad (18)$$

The important feature of this model is that *all bands* share the same base parameters θ , while their offsets $\theta^{(k)}$ are determined individually.

We can construct the normalized periodogram for this model by building a sparse design matrix with M_K columns. Each row corresponds to a single observation through a single band. Columns corresponding to the base model and the matching observation band will have nonzero entries; all other columns will be filled with zeros. For example, the $N_{base} = 1$ and $N_{band} = 0$ model corresponds to one with a simple single-term periodic base frequency,

and an independent constant offset term in each band. The associated design matrix depends on the particular data, but will look similar to this:

$$X_{\omega}^{(1,0)} = \begin{bmatrix} 1 & \sin(\omega t_1) & \cos(\omega t_1) & 1 & 0 & 0 & 0 & 0 \\ 1 & \sin(\omega t_2) & \cos(\omega t_2) & 0 & 0 & 0 & 0 & 1 \\ 1 & \sin(\omega t_3) & \cos(\omega t_3) & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots & \\ 1 & \sin(\omega t_N) & \cos(\omega t_N) & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (19)$$

Here the nonzero entries of the final five columns are binary flags indicating the (u, g, r, i, z) -band of the given observation: for this example, the first row is a u -band measurement, the second is a z -band, the third is a i -band, etc.

On examination of the above matrix, it's clear that the columns are not linearly independent (i.e. X_{ω} is low-rank), and thus the parameters of the best-fit model will be degenerate. Intuitively, this is due to the fact that if we add an overall offset to the base model, this can be perfectly accounted for by subtracting that same offset from each residual. Mathematically, the result of this is that the normal matrix $X_{\omega}^T \Sigma^{-1} X_{\omega}$ will be non-invertible, and thus the periodogram is ill-defined. In order to proceed, then, we'll either have to use a different model, or use a cleverly-constructed regularization term on one of the offending parameters.

We'll choose the latter here, and regularize all the band columns while leaving the base columns un-regularized: for the above X_{ω} matrix, this regularization will look like

$$\Lambda^{(1,0)} = \text{diag}([0, 0, 0, \epsilon, \epsilon, \epsilon, \epsilon, \epsilon]) \quad (20)$$

where ϵ is some small fraction of the trace of the normal matrix $[X_{\omega}^T \Sigma^{-1} X_{\omega}]$. The logic of this choice of regularization is that any component of the model which is common to all bands will be preferentially reflected in the base terms, with independent behavior reflected in the individual band terms. Setting ϵ to some small fraction of the trace ensures that the regularization effect on the remaining model will be small. With this regularization in place, the model is well-posed and Equation (17) can be used to straightforwardly compute the power. The effective number of free parameters for such a regularized $(N_{\text{base}}, N_{\text{band}})$ model with K filters is $M_K^{\text{eff}} = 2N_{\text{base}}^{\text{eff}} + K(2N_{\text{band}} + 1)$ where $N_{\text{base}}^{\text{eff}} = \max(0, N_{\text{base}} - N_{\text{band}})$ is the effective number of base terms.

The final remaining piece to mention is our assumption in Equation (5) that the data are centered. This is required so that the simple form of the reference χ_0^2 remains valid. For the multiband model, this assumption requires that the data satisfy Equation (5) *within each band*: equivalently, we could lift this assumption and compute the reference χ_0^2 of the multiband model with an independent floating mean within each band; the results will be identical.

This multiband approach, then, actually comprises a set of models indexed by their value of N_{base} and N_{band} . The most fundamental models have $(N_{base}, N_{band}) = (1, 0)$ and $(0, 1)$, which we'll call the *shared-phase* and *multi-phase* models respectively. In the shared-phase model, all variability is assumed to be shared between the bands, with only the fixed offset between them allowed to float. In the multi-phase model, each band has independent variability around a shared fixed offset.

5.1. Relationship of Multiband and Single-band approaches

The *multi-phase* ($N_{base} = 0, N_{band} = 1$) model turns out to be a particularly special case. The base model is a simple global offset which is degenerate with the offsets in each band, so that the design matrix X_ω can be straightforwardly rearranged as block-diagonal. A block-diagonal design matrix in a linear model indicates that components of the model are being solved independently: here these independent components amount to the single-band floating-mean model from Section 4.1, fit independently for each of the K bands. This particular multiband model can give us insight into the relationship between single-band and multiband approaches.

For band k , we'll denote the single-band floating-mean periodogram as

$$P_N^{(k)}(\omega) = 1 - \frac{\chi_{min,k}^2(\omega)}{\chi_{0,k}^2} \quad (21)$$

The full multiband periodogram is given by

$$P_N^{(0,1)}(\omega) = 1 - \frac{\sum_{k=1}^K \chi_{min,k}^2(\omega)}{\sum_{k=1}^K \chi_{0,k}^2} \quad (22)$$

and it can easily be shown that the P_N can be constructed as a weighted sum of $P_N^{(k)}$:

$$P_N^{(0,1)}(\omega) = \frac{\sum_{k=1}^K \chi_{0,k}^2 P_N^{(k)}}{\sum_{k=1}^K \chi_{0,k}^2}. \quad (23)$$

We see that this particular multiband periodogram is identical to a weighted sum of standard periodograms in each band, where the weights $\chi_{0,k}^2$ are a reflection of both the number of measurements in each band, and how much those measurements deviate from a simple constant reference model.

5.2. Multiband Periodogram for Simulated Data

Before applying the multiband method to real data, we will here explore its effectiveness on a simulated RR Lyrae lightcurve. The upper panels of Figure 5 show a multiband version of the simulated RR Lyrae light curve from Figure 1. The upper-left panel shows 60 nights of observations spread over a 6-month period, and for each night all five bands (u, g, r, i, z) are recorded. Using the typical approach from the literature, we individually compute the standard normalized periodogram within each band: the results are shown in the upper-right panel. The data are well-enough sampled that a distinct period of 0.622 days can be recognized within each individual band, up to the aliasing effect discussed in Section 3.3. Previous studies have made use of the information in multiple bands to choose between aliases and estimate uncertainties in determined periods (e.g. Oluseyi et al. 2012; Sesar et al. 2010). While this approach is sufficient for well-sampled data, it becomes problematic when the multiband data are sparsely sampled.

The lower panels of Figure 5 show the same 60 nights of data, except with only a *single* band observation recorded each night. The lower-left panel shows the observations as a function of phase, and the lower-right panels show the periodograms derived from the data. With only 12 observations for each individual band, it is clear that there is not enough data to accurately determine the period within each single band. The shared-phase multiband approach (i.e. ($N_{base} = 1, N_{band} = 0$)), shown in the lower-right panel, makes use of all the information in a single model, and clearly recovers the true frequency of 0.622 days.

This shared-phase (1, 0) model is only one of the possible multiband options, however: Figure 6 shows multiband fits to this data for models with various choices of (N_{base}, N_{band}). We see here many of the characteristics noted above for single-band models: as discussed in Section 4.2, increasing the number of Fourier terms leads to power at multiples of the fundamental period. Additionally, increasing the model complexity (roughly indexed by the effective number of free parameters M^{eff}) tends to increase the background level of the periodogram, potentially obscuring significant peaks. For this reason, models with $N_{base} > N_{band}$ are the most promising: they allow a flexible fit with minimal model complexity. Motivated by this, in the next section we'll apply the simplest of this class of models, the (1, 0) shared-phase model, to data from the Stripe 82 of the Sloan Digital Sky Survey.

6. Application to Stripe 82 RR Lyrae

Stripe 82 is a three hundred square degree equatorial region of the sky which was repeatedly imaged through multiple band-passes during phase II of the Sloan Digital Sky

Survey (SDSS II, see Sesar et al. 2007). Here we consider the SDSS II observations of 483 RR Lyrae stars compiled and studied by S10. S10 determined periods for these stars based on empirically-derived light curve templates. Because the template-fitting method is extremely computationally intensive, S10 first determined candidate periods by taking the top 5 results of the Supersmooth (Reimann 1994) algorithm applied to the g -band; template fits were then performed at each candidate period and the period with the best template fit was reported as the true period. In this section, we make use of this dataset to quantitatively evaluate the effectiveness of the multiband periodogram approach.

6.1. Densely-sampled Multiband Data

The full S10 RR Lyrae dataset consists of 483 objects with an average of 55 observations in each of the five SDSS $ugriz$ bands spread over just under ten years. In the upper panels of Figure 7 we show the observed data for one of these objects, along with the periodogram derived with the single-band supersmooth algorithm and the shared-phase $(0, 1)$ -multiband algorithm¹. Here we have a case which is analogous to that shown for simulated data in the top panels of Figure 5: each band has enough data to easily locate candidate peaks, the best of which is selected via the S10 template-fitting procedure.

The lower panels of Figure 7 compare the S10 period with the best periods obtained from the 1-band supersmooth (lower-left) and from the shared-phase multiband model (lower-right). To guide the eye, the figure includes indicators of the locations of beat aliases (dotted lines) and multiplicative aliases (dashed lines) of the S10 period.

The best-fit supersmooth period matches the S10 period in 89% of cases (421/483), while the best-fit multiband period matches the S10 period in 79% of cases (382/483). The modes of failure are instructive: when the supersmooth model misses the S10 period, it tends to land on a multiplicative alias (i.e. the dashed line). This is due to the flexibility of the supersmooth model: a doubled period spreads the points out, leading to fewer constraints in each neighborhood and thus a smaller average residual around model. In other words, the SuperSmoothen tends to over-fit data which is sparsely-sampled. On the other hand, when the multiband model misses the S10 period, it tends to land on a beat alias between the S10 period and the 1-day observing cadence (i.e. the dotted lines). This is due to the fact that the single-frequency periodic model is biased, and significantly under-fits

¹The supersmooth “periodogram” P_{SS} is constructed from the minimum sum of weighted model residuals \bar{r}_{min} in analogy with Equation (8): $P_{SS}(\omega) = 1 - \bar{r}_{min}(\omega)/\bar{r}_0$, where \bar{r}_0 is the residual around a constant model.

the data: it cannot distinguish residuals due to underfitting from residuals due to window function effects.

In both models, the S10 period appears among the top 5 periods 99% of the time: 477/483 for supersmoothen, and 480/483 for multiband.² This suggests that had S10 used the multiband Lomb-Scargle rather than the supersmooth in the first pass for that study, the final results presented there would be for the most part unchanged.

The results of this subsection show that the shared-phase multiband approach is comparable to the supersmooth approach for densely-sampled multiband data, although it has a tendency to get fooled by structure in the survey window. Correction for this based on the estimated window power may alleviate this (see Roberts et al. (1987) for an example of such an approach) though in practice selecting from among the top 5 peaks appears to be sufficient.

6.2. Sparsely-sampled Multiband Data

Above we saw that the multiband model is comparable to methods from the literature for densely-sampled data. Where we expect the multiband approach to gain an advantage is when the data are sparsely sampled, with observations through only a single band at each observation. To simulate this, we reduce the size of the Stripe 82 RR Lyrae dataset by a factor of 5, keeping only a single band of imaging each night: an average of 11 observations of each object per band. This is much closer to the type of data which will be available in future multiband time-domain surveys.

The upper panels of Figure 8 show an example light curve from this reduced dataset, along with the supersmooth and multiband periodograms derived from this data. Analogously to the lower panels of Figure 5, the single-band supersmooth model loses the true period within the noise, while the shared-phase multiband model still shows prominent signal near the S10 period.

The lower panels of Figure 8 show the relationship between the S10 periods (based on the full dataset) and the periods derived with each model from this reduced dataset. It is clear that the supersmooth model is simply over-fitting noise with this few data points:

²We might expect this correspondence to be 100% in the case of the *g*-band supersmooth, which was the model used in the first pass of the S10 computation. This discrepancy here is likely due to the slightly different supersmooth implementations used in S10 and in this work. Objects showing this discrepancy are those with very low signal-to-noise.

the top period matches S10 in only 32% of cases (compared to 87% with the full dataset), and the top 5 periods contain the S10 period only 45% of the time. The failure mode is much less predictable as well: rather than being clustered near aliases, most of the period determinations are scattered seemingly randomly around the space.

The multiband method does much better on the reduced dataset. Even with an 80% reduction in the number of observations, the multiband method matches the S10 period 64% of the time (compared to 79% with the full dataset), and the top 5 peaks contain the S10 period 94% of the time (compared to 99% with the full dataset). This performance is due to the fact that the multiband algorithm has relatively few parameters, but is yet able to flexibly accommodate data from multiple observing bands. In particular, this suggests that with the multiterm periodogram, the S10 analysis could have been done effectively with only a small fraction of the available data. This bodes well for future surveys, where data on variable stars will be much more sparsely sampled.

7. Discussion and Conclusion

We have motivated and derived a multiband version of the classic Lomb-Scargle method for detecting periodicity in astronomical time-series. Experiments on several hundred RR Lyrae stars from the SDSS Stripe 82 dataset indicate that this method outperforms methods used previously in the literature, especially for sparsely-sampled light curves with only single bands observed each night.

The main deficiency of the model is its tendency to be tricked by the survey window function, a result of the very model simplicity which causes its success in the case of sparse multiband data: it cannot disentangle model bias from window bias. This could potentially be accounted for by estimating and filtering the effect of the window function; one potential method for this involves estimating the deconvolution of the window power and the observed power. It may also be possible to propose a multiband extension of CARMA (Kelly et al. 2014), or another forward-modeling approaches to detecting periodicity.

Another potentially fruitful way forward in the multiband case is to make use of physically-motivated priors in the light-curve shape. For example, the paths of RR Lyrae stars through color-color and color-magnitude space are constrained by known astrophysical processes in the structure of the stars (e.g., see Fig. 5 in Szabó et al. 2014), and making use of such a physically-motivated prior may help break degeneracies in period determination.

Looking forward to future variable star catalogs from PanSTARRS, DES, and LSST, there are two important constraints that any analysis method must meet: the methods must

be able to cope with heterogeneous and noisy observations through multiple band-passes, and the methods must be fast enough to be computable on millions or even billions of objects. The multiband method, through its combination of flexibility and model simplicity, meets the first: as shown above, it performs well in the case of sparsely-sampled multiband data. It also meets the second constraint: it consists of a fast linear model at each frequency, compared to a sliding-window model in the case of supersmooth, a nonlinear optimization in the case of template-fitting, and a Markov Chain Monte Carlo in the case of CARMA models. In our own benchmarks, we found the multiband method to be several times faster than the single-band supersmooth approach, and several orders of magnitude faster than the template fitting approach. This suggests that the multiband periodogram could have an important role to play in the analysis of variable stars in future surveys and we make our tools publicly available on GitHub.

7.1. Comparison with Long et al. (2014)

As we were finishing this study, it was pointed out to us that another group had released a preprint independently addressing the multiband periodogram case (Long et al. 2014, hereafter LCB14), and come up with solutions very similar to the ones presented here. They present two methods, the “Multiband Generalized Lomb-Scargle” (MGLS) which is effectively identical to the $(1, 0)$ multi-phase model here, and the “Penalized Generalized Lomb-Scargle” (PGLS), which is similar in spirit to our $(0, 1)$ shared-phase model.

In the PGLS model, they start with a multi-phase model, fitting independent $N = 1$ term fits to each band, and apply a nonlinear regularization term which penalizes differences in the amplitude and phase. In terms of the formalism used in this work, the PGLS model minimizes a regularized χ^2 of the form

$$\chi_{PGLS}^2 = \sum_{k=1}^K \left[\chi_{GLS}^2(D^{(K)}) + J_A(A^{(k)}) + J_\phi(\phi^{(k)}) \right]. \quad (24)$$

for K bands, where $\chi_{GLS}^2(D^{(K)})$ is the χ^2 of the standard floating mean model on the single-band data $D^{(K)}$, and J_A and J_ϕ are regularization/penalty terms which are a function of the amplitude $A^{(k)}$ and phase $\phi^{(k)}$ of each model. In terms of our linear model parameters $\theta^{(k)}$, this amplitude and phase can be expressed:

$$\begin{aligned} A^{(k)} &= \sqrt{(\theta_1^{(k)})^2 + (\theta_2^{(k)})^2} \\ \phi^{(k)} &= \arctan(\theta_2/\theta_1) \end{aligned} \quad (25)$$

The selected form of these regularization terms penalizes deviations of the amplitude and phase from a common mean between the bands; in this sense the PGGLS model can be considered a conceptual mid-point between our shared-phase and multi-phase models. Within the formalism proposed here, such a mid-point may be alternatively attained by suitably increasing the regularization parameter ϵ used in our shared-phase model, though the nature of the resulting regularization will differ.

Computationally, the PGGLS model requires a nonlinear optimization at each frequency ω , and is thus much more expensive than the straightforward linear optimization of our shared-phase model. For this reason, LCB14 propose a clever method by which nested models are used to reduce the number of nonlinear optimizations used: essentially, by showing that the (linear) MGLS χ^2 is a lower-bound of the (non-linear) PGGLS χ^2 , it is possible to iteratively reduce the number of PGGLS computations required to minimize the χ^2 among a grid of frequencies. Such an optimization could also be applied in the case of our shared-phase model, but is not necessary here due to its already high speed. Nevertheless, when applying the method to a very large number of light curves, as in e.g. LSST, such a computational trick may prove very useful.

Given these important distinctions between the models proposed here and in LCB14, in future work we plan to do a detailed comparison of the two means of model regularization.

Acknowledgements: JTV is supported by the University of Washington eScience institute, including grants from the Alfred P. Sloan Foundation, the Gordon and Betty Moore Foundation, and the Washington Research Foundation. The authors thank GitHub for providing free academic accounts which were essential in the development of this work.

REFERENCES

- Bretthorst, G. 1988, Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics (Springer)
- Buitinck, L., Louppe, G., Blondel, M., et al. 2013, arXiv:1309.0238
- Eyer, L., & Mowlavi, N. 2008, Journal of Physics Conference Series, 118, 012010
- Flaugher, B. 2008, in A Decade of Dark Energy: Spring Symposium, Proceedings of the conferences held May 5-8, 2008 in Baltimore, Maryland. (USA). Edited by Norbert Pirzkal and Henry Ferguson. <http://www.stsci.edu/institute/conference/spring2008>
- Graham, M. J., Drake, A. J., Djorgovski, S. G., et al. 2013, MNRAS, 434, 3423

- Hoerl, A. E., & Kennard, R. W. 1970, *Technometrics*, 12, 55
- Ivezić, Ž., Connolly, A., VanderPlas, J., & Gray, A. 2014, *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*, Princeton Series in Modern Observational Astronomy (Princeton University Press)
- Ivezić, Ž., et al. 2008, arXiv:0805.2366
- Jaynes, E. 1987, in *Fundamental Theories of Physics*, Vol. 21, Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems, ed. C. Smith & G. Erickson (Springer Netherlands), 1–37
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7733, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*
- Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, *ApJ*, 788, 33
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- Long, J. P., Chi, E. C., & Baraniuk, R. G. 2014, ArXiv e-prints, arXiv:1412.6520
- Oluseyi, H. M., Becker, A. C., Culliton, C., et al. 2012, *AJ*, 144, 9
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Press, W. H., & Rybicki, G. B. 1989, *ApJ*, 338, 277
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, *Numerical Recipes* 3rd Edition: The Art of Scientific Computing, 3rd edn. (New York, NY, USA: Cambridge University Press)
- Reimann, J. D. 1994, PhD thesis, University of California, Berkeley.
- Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, *AJ*, 93, 968
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Sesar, B., Stuart, J. S., Ivezić, Ž., et al. 2011, *AJ*, 142, 190
- Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, *AJ*, 134, 2236

- Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, ApJ, 708, 717
- Stellingwerf, R. F. 1978, ApJ, 224, 953
- Stetson, P. B. 1996, PASP, 108, 851
- Süveges, M., Sesar, B., Váradi, M., et al. 2012, MNRAS, 424, 2528
- Szabó, R., Ivezić, Ž., Kiss, L. L., et al. 2014, ApJ, 780, 92
- Tikhonov, A. 1963, in Soviet Math. Dokl., Vol. 5, 1035–1038
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engg., 13, 22
- Vanderplas, J. 2015, doi:10.5281/zenodo.14475
- Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47 –54
- Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

A. Python Implementation of Multiband Periodogram

The algorithm outlined in this paper is available in `gatspy`, an open-source Python package for general astronomical time-series analysis³. Along with the periodogram implementation, it also contains code to download all the data used in this work. Code to reproduce this paper, including all figures, is available as well⁴.

`gatspy` is a pure-Python package written to be compatible with both Python 2 and Python 3, and performs fast numerical computation through dependencies on `numpy` (van der Walt et al. 2011)⁵ and `astroML` (Vanderplas et al. 2012)⁶, which offer optimized implementations of numerical methods in Python.

This preprint was prepared with the AAS L^AT_EX macros v5.2.

³<http://github.com/jakevdp/gatspy/>

⁴http://github.com/jakevdp/multiband_LS/

⁵<http://www.numpy.org>

⁶<http://www.astroml.org>

The API for the module is largely influenced by that of the `scikit-learn` package (Pedregosa et al. 2011; Buitinck et al. 2013)⁷, in which models are Python class objects which can be fit to data with the `fit()` method. Here is a basic example of how you can use `multiband_LS` to download the data used in this paper, fit a multiband model to the data, and compute the power at a few periods:

```
from gatspy.periodic import LombScargleMultiband
import numpy as np

# Fetch the Sesar 2010 RR Lyrae data
from gatspy.datasets import fetch_rrlyrae
data = fetch_rrlyrae()
t, mag, dmag, filts = data.get_lightcurve(data.ids[0])

# Construct the multiband model
model = LombScargleMultiband(Nterms_base=0, Nterms_band=1)
model.fit(t, mag, dmag, filts)

# Compute power at the following periods
periods = np.linspace(0.2, 1.4, 1000) # periods in days
power = model.periodogram(periods)
```

Other models are available as well. For example, here is how you can compute the periodogram under the supersmoother model; this implementation of the supersmoother periodogram makes use of the `supersmoother` Python package (Vanderplas 2015).

```
from gatspy.periodic import SuperSmoother

# Construct the supersmoother model
model = SuperSmoother()
gband = (filts == 'g')
model.fit(t[gband], mag[gband], dmag[gband])

# Compute power at the given periods
power = model.periodogram(periods)
```

The models in the `gatspy` package contain many more methods, and much more functionality than what is shown here. For updates, more examples, and more information, visit <http://github.com/jakevdp/gatspy/>.

⁷<http://scikit-learn.org>

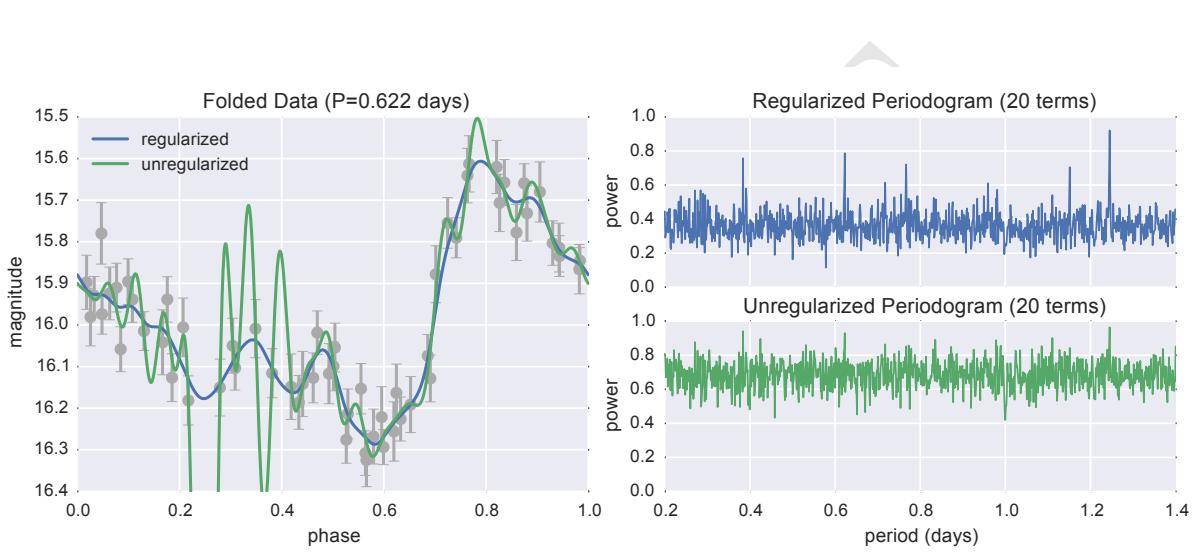


Fig. 4.— The effect of regularization on a high-order model. The data is the same as those in Figure 1. We fit a 20-term truncated Fourier model to the data, with and without a regularization term. Without regularization, the model oscillates widely to fit the noise in the data. The regularization term effectively damps the higher-order Fourier modes and removes this oscillating behavior, leading to a more robust model with stronger periodogram peaks.

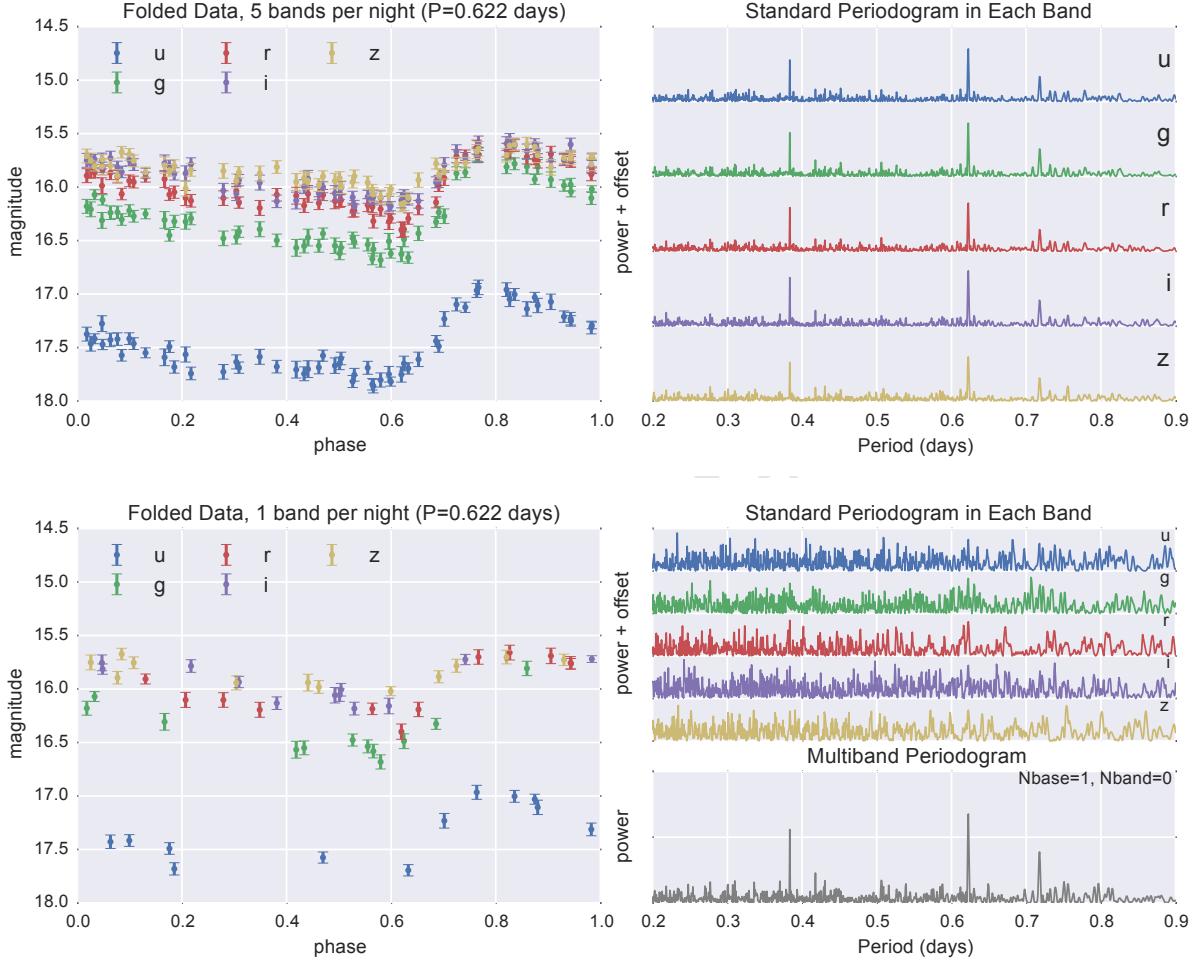


Fig. 5.— An illustration of the performance of the multiband periodogram. The upper panels show simulated *ugriz* observations of an RR Lyrae light curve in which all 5 bands are observed each night. With 60 observations in each band, a periodogram computed from any single band is sufficient to determine the true period of 0.622 days. The lower panels show the same data, except with only a single *ugriz* band observed each night (i.e. 12 observations per band). In this case, no single band has enough information to detect the period. The shared-phase multiband approach of Section 5 (lower-right panel) combines the information from all five bands, and results in a significant detection of the true period.

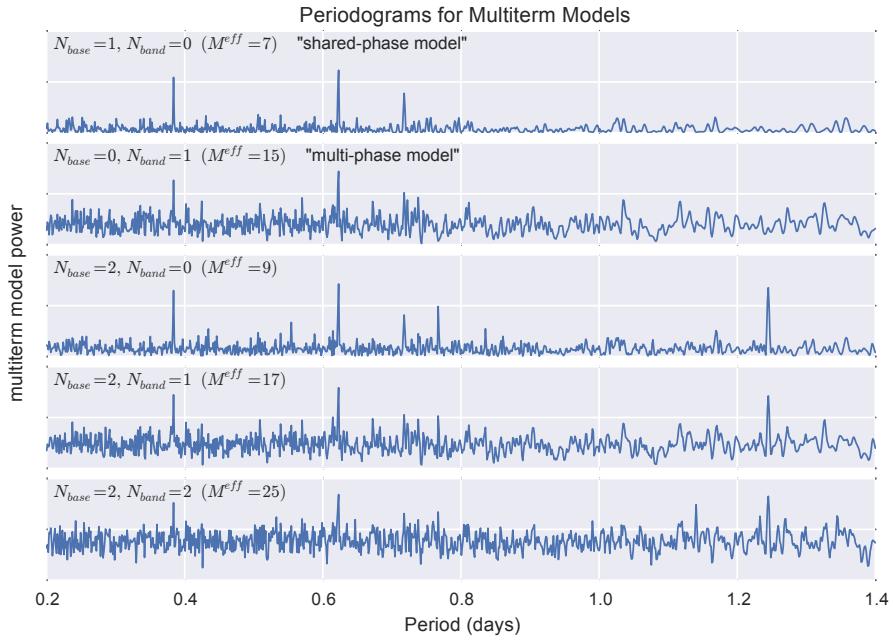


Fig. 6.— Comparison of the periodograms produced by various multiband models. The data is the same as that used in Figure 5. N_{base} gives the number of Fourier terms in the base model, and N_{band} gives the number of Fourier terms used to fit the residuals around this model within each band. The characteristics discussed with previous figures are also seen here: in particular, the level of “background noise” in the periodogram grows with the model complexity M ,

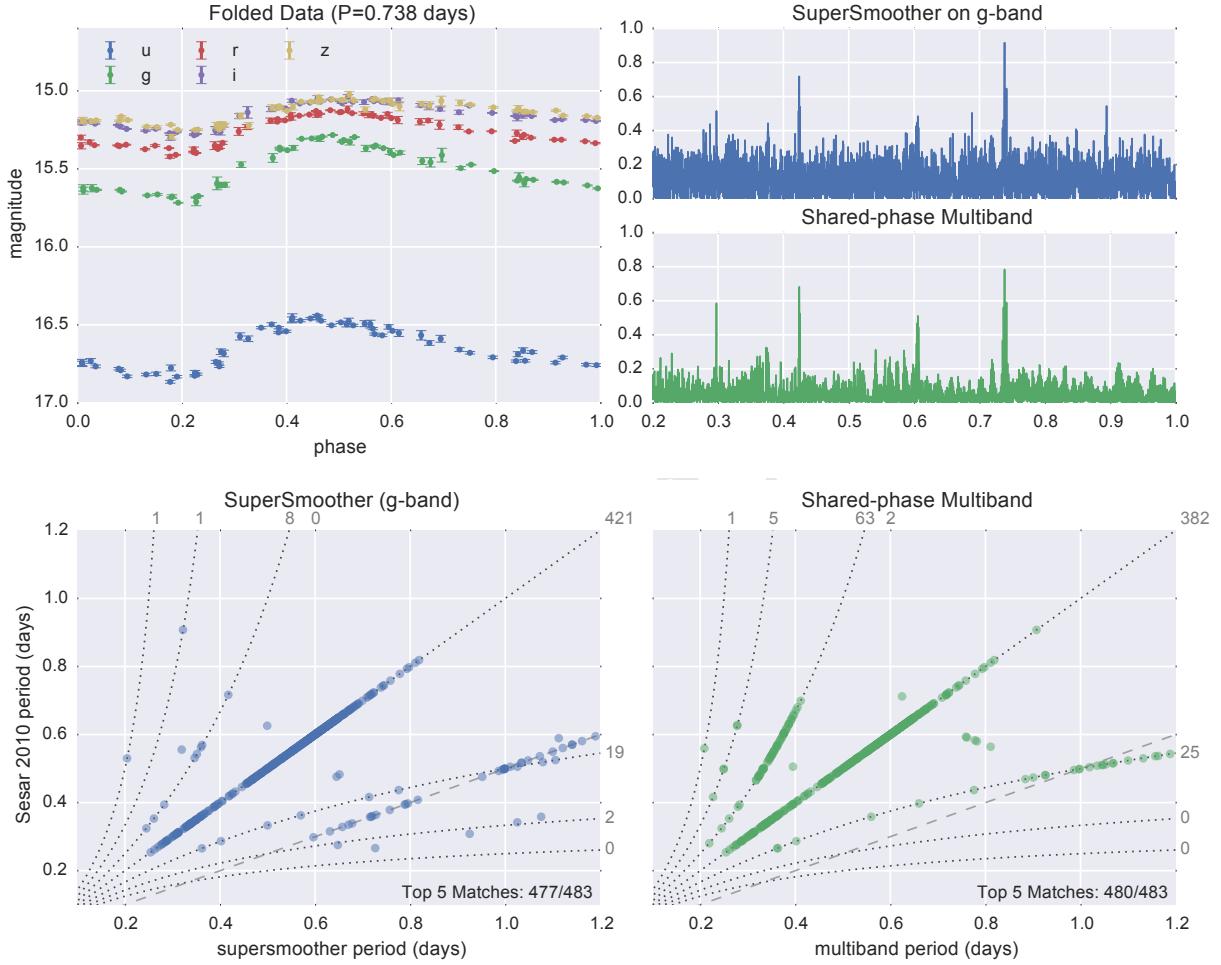


Fig. 7.— Comparison of the Multiband algorithm and single-band supersmooth algorithm on 483 well-sampled RR Lyrae light curves from Stripe 82. The upper panels show a representative lightcurve and periodogram fits, while the bottom panels compare the derived periods to the template-based periods reported in S10. Shown for reference are the beat aliases (dotted lines) and the multiplicative alias (dashed lines): numbers along the top and right edges of the panels indicate the number of points aligned with each trend. The single-band supersmooth algorithm tends to err toward multiplicative aliases, while the multiband algorithm tends to err toward beat frequency aliases. Both methods find the correct period among the top 5 significant peaks around 99% of the time.

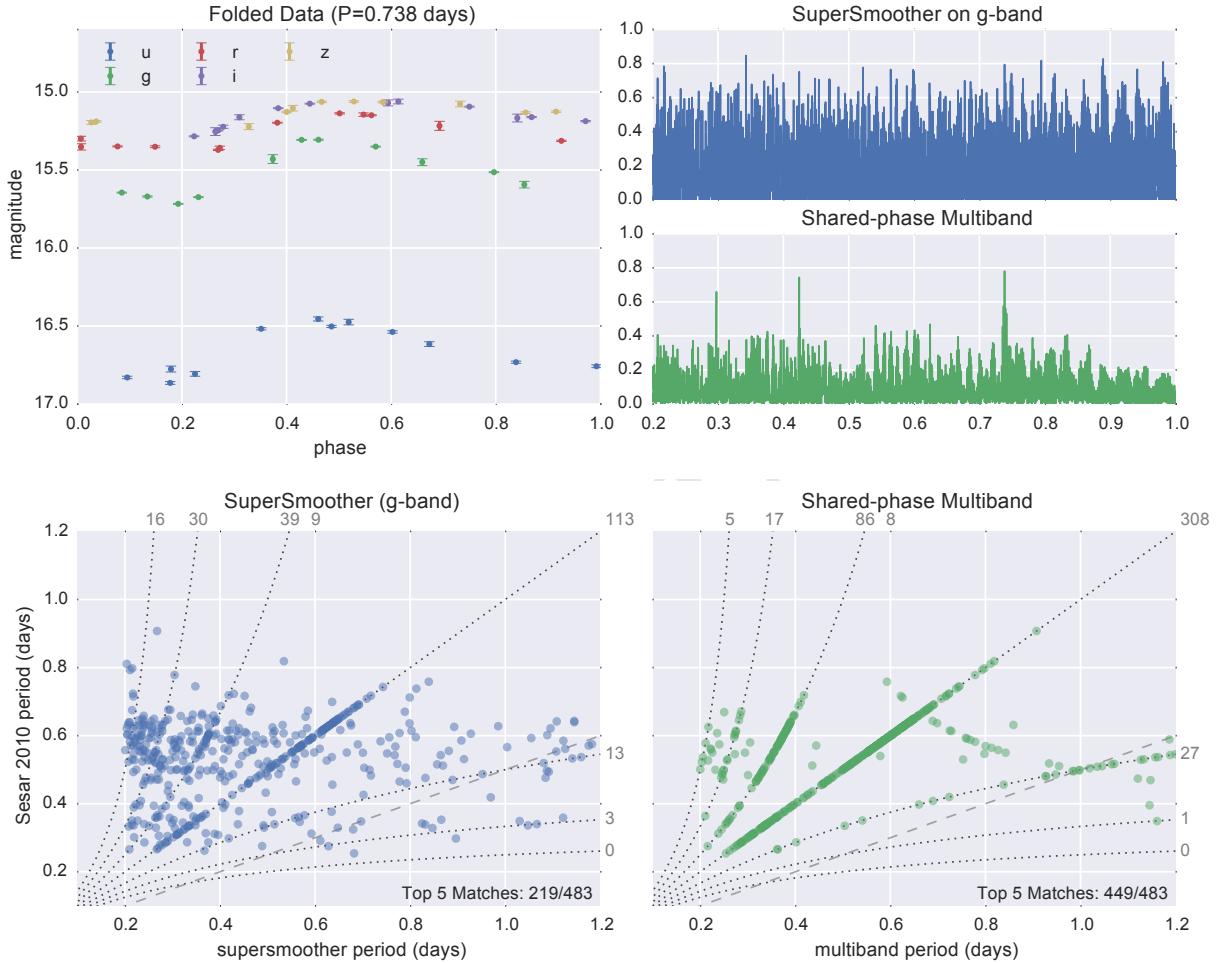


Fig. 8.— This figure repeats the experiment shown in Figure 7 (see caption there for description), but the data is artificially reduced to only a single-band observation on each evening, a situation reflective of the observing cadence of future large-scale surveys. In this case, the single-band SuperSmoothening strategy used as a first pass in S10 fails: there is simply not enough data in each band to recover an accurate period estimate. The correct period is among the top 5 candidates in fewer than 50% of cases. The shared-phase multiband approach utilizes information from all five bands, and returns much more robust results: even with the greatly-reduced data, the true period is among the top 5 candidates in 93% of cases.