Bayesian Model Selection and Analysis for Cepheid Star Oscillations (Initial Draft)

James O. Berger
William H. Jefferys
Peter Muller
Thomas G. Barnes

ABSTRACT Cepheid variables are a class of pulsating variable stars with the useful property that their periods of variability are strongly correlated with their absolute luminosity. Once this relationship has been calibrated, knowledge of the period gives knowledge of the luminosity. This makes these stars useful as "standard candles" for estimating distances in the universe. Available data consists of photometric and velocity information for a number of Cepheid variables, at unequally spaced points in their periods. Note that photometry and velocity are connected by nonlinear relations involving the physical parameters of interest. Bayesian analysis is used to provide inferences for useful physical features, such as the absolute luminosity of the star, its distance, its radius, and other parameters.
In the absence of reliable physical models of the pulsation of Cepheid variables, we model the photometry and velocity curves as (i) a trigonometric polynomial with an unknown number of terms; or (ii) via a wavelet basis with an unknown number of terms. Bayesian analysis allows computation of the posterior probabilities of these varying dimensional models, and results in inferences on the physical parameters that are based on 'averaging' over the possible models. Computations are done using reversible-jump Markov chain Monte Carlo methodology.

1 Introduction

1.1 Bayesian Model Selection and Model Averaging

The Bayesian approach to hypothesis testing and model selection is conceptually straightforward. One assigns *prior* probabilities to all unknown hypotheses or models, as well as to unknown parameters or quantities within models, and uses probability theory to compute the *posterior* probabilities of the hypotheses or models, given observed data. One attractive feature of this approach is simplicity of interpretation: stating, at the end of the analysis, that the only tenable models are Models 5, 6, and 7, and that they have probabilities 0.34, 0.56 and 0.10, respectively, has appealing clarity.

A second attractive feature of this approach is that one need not choose a fixed model. One could select Model 6 above (it is the model most likely to be true), but the data also gives considerable support to Model 5, and even Model 7 should not be ignored. One deals with this uncertainty by 'Bayesian model averaging,' in which predictions or desired estimates from models are averaged according to the model posterior probabilities. Thus if Models 5, 6, and 7 provided distance estimates (posterior means) to a star of 750, 790, and 800 parsecs, respectively, the 'model-averaged' distance estimate would be $0.34 \times 750 + 0.56 \times 790 + 0.10 \times 800 = 777.4$ parsecs.

The accuracy associated with a model-averaged estimate will also incorporate the model uncertainty. For instance, suppose Model j yields the distance estimate $\hat{d}_j$, with associated posterior variance V_j, and that p_j is the posterior probability of Model j. Then the overall variance of the model-averaged distance estimate $\hat{d}^* = \sum p_j \hat{d}_j$ is given by

$$V^* = \sum p_j [V_j + (\hat{d}_j - \hat{d}^*)^2].$$

For the case in the previous paragraph, if the individual model posterior variances were $V_5 = V_6 = V_7 = 400$ (corresponding to standard errors of 20 parsecs), then the overall variance of $\hat{d}^* = 777.4$ would be 795.24, almost twice the variance that would be associated with any specific model. (Indeed, it is a general advantage of the Bayesian approach that inaccuracies in all unknown parameters are incorporated automatically.)

Note, also, that the Bayesian approach to model selection acts as a natural "Ockham's razor," in the sense of favoring a simpler model over a more complex model if the data provides roughly comparable fits for the models. And this is without having to introduce any explicit penalty for the more complex models. (For an interesting historical example of Ockham's razor, and general discussion and references, see Jefferys and Berger, 1992.)

1.2 Cepheid Star Oscillations

Cepheid variable stars pulsate, varying their luminosity (light output) and size with a very regular periodic behavior. It is possible to measure both the velocity of the surface of the star as it pulsates and the variable luminosity and color of the star. For instance, Figure 1 presents the data concerning the radial velocity of the surface of the star T Moncerotis, at various phases of the star's period. (The actual data are indicated by the x's.)

There is a mathematical relationship between surface velocity, luminosity and color that enables one to determine the distance to the star. The considerable uncertainty in these measurements and the limited data that is available for each star suggest that analysis which fully incorporates these uncertainties is desirable.

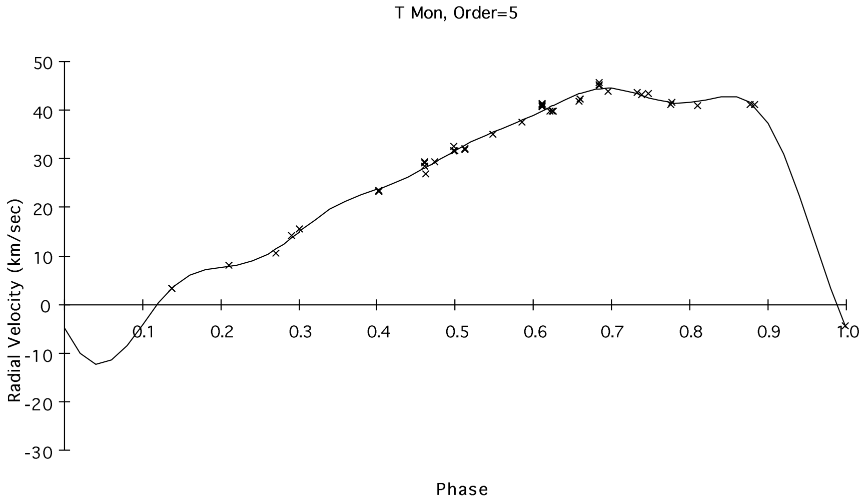

Figure 1. The radial velocity data (the x's) for T Mon, and their fit to a fifth-order trigonometric polynomial.

1.3 Challenges in the Bayesian Approach

There are three significant challenges in implementing the Bayesian approach for complex problems. The first challenge is common to all statistical analyses, namely the need to find appropriate statistical models for the data. For a Cepheid star, the most challenging features to model are the radial velocity and the photometric information. For instance, Figure 1 clearly indicates that the radial velocity of the star is a quite complex function of its phase, but existing physical theories for Cepheid stars do not provide guidance as to the form of this function. Hence one must resort to generic statistical modeling, such as Fourier analysis. Figure 1 shows that a fifth-order trigonometric polynomial fits this particular data quite well, but the needed order of the polynomial changes from star to star and, indeed, there are typically several different orders that fit a particular star well (without overfitting). The different models that will be under consideration in our analysis are simply the different possible orders of the trigonometric polynomials (which will also be used to model the photometric data). Later we will also consider wavelet models of the radial velocity.

The second challenge in Bayesian analysis is to choose prior distributions for unknown quantities in the analysis (for instance, for the unknown Fourier coefficients of the trigonometric polynomial). The most common choices are noninformative or objective priors; these will be primarily utilized in the analysis here and are discussed in the next subsection.

The third challenge is computational. Bayesian analysis can require the computation of high-dimensional integrals, and is especially costly when model selection is involved. (For instance, when using trigonometric polynomials in the Cepheid modeling, it is necessary to compute up to 50-dimensional integrals for up to 40 models; in the wavelet version of the analysis discussed in section 4, these numbers increase by orders of magnitudes.) The modern approach to such computation is Markov chain Monte Carlo (MCMC) analysis. This is a computational paradigm that can be easily described in simple cases, but which requires study and experience for successful application in complex cases (such as that considered here). Thus we will content ourselves, in this paper, with only a higher-level description of the particular steps needed in the Cepheid problem. Recent general books on the subject are Robert and Casella (1999) and Chen, Shao, and Ibrahim (2000).

2 Objective Bayesian Model Selection

2.1 Statistical Notation

The data, $\boldsymbol{Y}$, is assumed to have arisen from one of several possible models $M_1, \ldots, M_k$. Under M_i, the density of $\boldsymbol{Y}$ is $f_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is an unknown vector of parameters.

The Bayesian approach to model selection begins by assigning prior probabilities, $P(M_i)$, to each model; often, equal prior probabilities are used, i.e. $P(M_i) = 1/k$, and this will be done here. It is also necessary to choose prior distributions $p_i(\boldsymbol{\theta}_i)$ for the unknown parameters of each model; sometimes these can also be chosen in a "default" manner, as will be illustrated later. The analysis then proceeds by computing the posterior probabilities of each model, which elementary probability theory (Bayes theorem) shows to be equal to

$$P(M_i|\boldsymbol{y}) = \frac{P(M_i)m_i(\boldsymbol{y})}{\sum_{j=1}^{k} P(M_j)m_j(\boldsymbol{y})}, \tag{1.1}$$

where $m_j(\boldsymbol{y}) = \int f_j(\boldsymbol{y} \mid \boldsymbol{\theta}_j)p_j(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j$ is the *marginal density* of $\boldsymbol{y}$. See Kass and Raftery (1995) for a general discussion of Bayesian model selection.

2.2 Choice of Prior Distributions

It may well be the case that subjective knowledge about the $\boldsymbol{\theta}_i$ is available, and can be incorporated into subjective proper priors for the $\boldsymbol{\theta}_i$. This is clearly desirable if it can be done. Indeed, for Cepheid stars we will see that subjective prior information concerning their distance can be utilized.

For most of the unknown parameters in models it will typically be infeasible to utilize subjective prior distributions. Frequently this is because subjective information is simply unavailable. (Thus, for Cepheid stars, turning the physical principles that underlie oscillatory behavior into models for the velocity and photometric curves is so difficult to accomplish that, in actuality, there is little subjective information about the Fourier coefficients of the curves.) Even if subjective prior information is available, it can be very difficult to utilize in high-dimensional problems.

For these and other reasons, the most popular Bayesian methods are default or 'objective Bayesian' methods. For estimation and prediction problems, objective Bayesian theories are well developed. The most famous of these are the *Jeffreys prior* (cf. Jeffreys, 1961), *maximum entropy* priors (cf. Jaynes, 1999), and *reference priors* (which prove remarkably successful in higher dimensional problems; cf., Berger and Bernardo, 1992).

Testing and model selection have proved to be much more resistant to the development of default Bayesian methods. This is because the objective priors discussed above are typically improper distributions (i.e., their integrals are infinite). This does not typically pose a problem in estimation and prediction, but it does for testing and model selection. See Berger and Pericchi (2001) for discussion of these difficulties and possible solutions. Here are some guidelines for choosing default priors in model selection.

1. *Common Parameters:* If all models have certain common parameters (see Berger and Pericchi, 2001, for discussion of what it means to be 'common') these parameters can typically be assigned the same improper objective prior. For instance, all the models for Cepheid radial velocity will have a common unknown mean radial velocity u_0, and this can be assigned the usual objective (improper) prior $p_i^*(u_0) = 1$.

2. *Conventional proper priors* are sometimes available in the literature. For instance, in the Cepheid problem, we will model the observed radial velocities as arising from a trigonometric polynomial subject to error; in statistical language, the ensuing model can formally be written as a general linear model of the form

$$\boldsymbol{Y} = \theta_0 \mathbf{1} + \boldsymbol{X}\boldsymbol{\theta} + \varepsilon,$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ is the vector of observations (radial velocities in the Cepheid problem), $\boldsymbol{X}$ is the corresponding design matrix of covariates (sines and cosines evaluated at multiples of the phases of the observations, corresponding to the trigonometric polynomial in the Cepheid problem), $\boldsymbol{\theta}$ is an unknown vector of parameters (the Fourier coefficients of the trigonometric polynomial in the Cepheid problem), $\mathbf{1}$ is the column vector of ones, θ_0 is the unknown mean level of the observations, and ε is a multivariate normal vector of errors with mean zero and covariance matrix $\sigma^2\boldsymbol{G}$, $\boldsymbol{G}$ a known matrix (i.e., ε is $\boldsymbol{N}(\mathbf{0}, \sigma^2\boldsymbol{G})$).

The recommended prior (from Zellner and Siow, 1980) for the unknown θ_0 is $p(\theta_0) = 1$, while that for $\boldsymbol{\theta}$, given σ^2, can be written in two stages (for later convenience) as:

$$p(\boldsymbol{\theta} \mid \sigma^2, \tau) \text{ is } \boldsymbol{N}(\mathbf{0}, \tau n \sigma^2 (\boldsymbol{X}' \boldsymbol{G}^{-1} \boldsymbol{X})^{-1}); \;\; p(\tau) = \frac{1}{\sqrt{2\pi}\, \tau^{3/2}} \exp(-\frac{1}{2\tau}). \tag{1.2}$$

3. A *general default prior* for model selection is the "empirical expected posterior prior" (Perez and Berger, 2000). For a given model M_i with unknown parameters $\boldsymbol{\theta}_i$, the most convenient form for this prior, when computation is to be done via the Markov chain Monte Carlo method, arises from introducing 'latent' random variables $\mathbf{y}^*$, which can be thought of as random subsamples of the data with sample size (typically) equal to the dimension of $\boldsymbol{\theta}_i$. Then the desired prior distribution is

$$p_i(\boldsymbol{\theta}_i, \mathbf{y}^*) = f_i(\mathbf{y}^* \mid \boldsymbol{\theta}_i) p_i^*(\boldsymbol{\theta}_i) m^E(\mathbf{y}^*)/m^*(\mathbf{y}^*), \tag{1.3}$$

where $p_i^*(\boldsymbol{\theta}_i)$ is a standard (improper) objective prior, $m^E(\mathbf{y}^*)$ refers to the empirical distribution of subsamples (i.e., choose each subsample of the given size with equal probability), and $m^*(\mathbf{y}^*)$ is the marginal density of $\mathbf{y}^*$ under the prior p_i^*. (The actual prior for $\boldsymbol{\theta}_i$ is the marginal density found by summing over $\mathbf{y}^*$ in (1.3), but it is actually more convenient computationally to work with the 'latent' joint distribution.)

3 Cepheid Stars

3.1 The Model and Likelihood

For a given star, the data consists of m observed radial velocities $U_i, i = 1, \ldots, m$, at unequally spaced phases of the star's period (cf. Figure 1), together with n vectors of photometry data consisting of magnitude $V_i, i = 1, \ldots, n$, and color index $C_i, i = 1, \ldots, n$. (It is to be understood that, attached to each observation, is the phase at which it was observed; note that the radial velocity and photometry data were typically observed at different phases of the star's period.) Each observation has a standard deviation specified by the observer; denote these by σ_{U_i}, σ_{V_i}, and σ_{C_i}, respectively. It is generally wise to be somewhat skeptical of such specified standard errors, and so we take the variances of the data to, instead, be given by $\sigma^2_{U_i}/\tau_u$, $\sigma^2_{V_i}/\tau_v$, $\sigma^2_{C_i}/\tau_c$, where the *parameters* τ_u, τ_v, and τ_c are unknown.

To complete the modeling of the data, let u_i, v_i, and c_i denote the true unknown mean velocity, magnitude, and color index, respectively, corresponding to each data point. We assume normality and independence of the measurement errors, so that

$$U_i \sim N(u_i, \sigma^2_{U_i}/\tau_u), \;\; V_i \sim N(v_i, \sigma^2_{V_i}/\tau_v), \;\text{ and }\; C_i \sim N(c_i, \sigma^2_{C_i}/\tau_c). \tag{1.4}$$

Since the velocities u and photometry (v, c) are periodic functions of time, an obvious strategy is to model them as trigonometric polynomials.

For the velocity u at phase ϕ, this would lead to the representation

$$u = u_0 + \sum_{j=1}^{M} [\theta_{1j} \cos(j\phi) + \theta_{2j} \sin(j\phi)], \tag{1.5}$$

where u_0 is the mean radial velocity of the star and M is the (unknown) order of the trigonometric polynomial. A similar equation holds for the luminosity data v. (We need to do this only for u and v, since the colors c are mathematically related to u and v through (1.7) below.) Let N denote the (unknown) order of the trigonometric polynomial for v. These polynomials contain $2M+1$ and $2N+1$ terms, respectively, including the leading constant terms.

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ denote column vectors of the velocity and luminosity data, respectively; define $\boldsymbol{X}_u$ and $\boldsymbol{X}_v$ to be the $(m \times 2M)$ and $(n \times 2N)$ design matrices consisting of the sines and cosines of multiple angles, evaluated at the phases of the corresponding data; and let $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$ be the corresponding vectors of unknown Fourier coefficients. With the normality assumption above, we can summarize the model (corresponding to M and N) for the velocity and luminosity data by the statistical linear models

$$\begin{aligned} \boldsymbol{u} &= u_0 + \boldsymbol{X}_u \boldsymbol{\theta}_u + \boldsymbol{\varepsilon}_u \\ \boldsymbol{v} &= v_0 + \boldsymbol{X}_v \boldsymbol{\theta}_v + \boldsymbol{\varepsilon}_v, \end{aligned} \tag{1.6}$$

where u_0 and v_0 are the (unknown) mean radial velocity and luminosity, respectively, and $\boldsymbol{\varepsilon}_u$ and $\boldsymbol{\varepsilon}_v$ are independently $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G}_u/\tau_u)$ and $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G}_v/\tau_v)$ multivariate errors, with $\boldsymbol{G}_u$ and $\boldsymbol{G}_v$ being the known diagonal matrices of the variances $\sigma^2_{U_i}$ and $\sigma^2_{V_i}$, respectively. (Note that τ_u and τ_v would have had to be introduced at this stage of the modeling, in any case, to account for the fact that u and v cannot be expected to exactly follow a trigonometric polynomial of finite order.)

The phases in the above likelihoods (entering through the design matrices) were assumed to be known exactly. In practice, however, the velocity data and photometry data are taken independently, and 'translated' to the same phase scale. The period of the star is not known perfectly, however, so that there is an unknown phase error $\Delta\phi$ (the difference between the zero-point of the phase for the velocity data and that for the photometric data) that is introduced. Thus we include that additional unknown in (say) the phase for the photometric data.

The (nonlinear) relationship between the radius of the star and the photometry is given by

$$c_i = a[0.1 v_i - b + 0.5 \log{(\phi_0 + \Delta r_i / s)}], \tag{1.7}$$

where a and b are known constants, ϕ_0 and s are the angular size and distance of the star (the latter being of primary interest to us), and Δr, the change in radius corresponding to phase ϕ, is given by

$$\Delta r = -g \sum_{j=1}^{M} \frac{1}{j} [\theta_{1j} \sin(j(\phi - \Delta\phi)) - \theta_{2j} \cos(j(\phi - \Delta\phi))], \tag{1.8}$$

found by integrating the nonconstant part of (1.5) term by term with respect to the phase; here g is another known constant. These expressions are to be inserted in the likelihood terms arising from the C_i in (1.4).

3.2 Choice of Prior Distributions

The unknown parameters in the above model are:

(1) The orders of the trigonometric polynomials, M and N.

(2) The parameters τ_u, τ_v, and τ_c, adjusting the measurement standard errors.

(3) The angular diameter ϕ_0 and the unknown phase error $\Delta\phi$.

(4) The distance s.

(5) The mean velocity and luminosity, u_0 and v_0, and the Fourier coefficients, $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$.

Some additional 'hyperparameters' will be introduced through the prior distributions for these unknowns, and the hyperparameters will also require prior distributions.

The orders of the models are expected to be modest (given the limited amount of data and the strong Ockham's razor effect of Bayesian analysis); we thus chose a uniform prior on the model orders (M, N) up to some cutoff (e.g., $(10, 10)$), with zero probability assigned to higher orders.

The parameters τ_u, τ_v, τ_c are given the standard objective priors for 'scale parameters,' namely the Jeffreys-rule prior $p(\tau) = 1/\tau$. Similarly, the priors on the 'location parameters' u_0 and v_0 are chosen to be the standard objective priors $p(u_0) = 1$ and $p(v_0) = 1$. Note that we are employing Rule 1 of subsection 2.2; since these parameters are common scale and location parameters for all models, they have an essentially fixed interpretation across models and can be assigned standard objective priors (even though improper).

For the parameters $\Delta\phi$ and ϕ_0, we also chose the objective priors $p(\Delta\phi) = 1$ and $p(\phi_0) = 1$. While it is unclear if these are 'optimal' objective priors for these parameters, preliminary investigations showed that the choice of priors for these parameters is almost irrelevant for the Cepheid data sets, so that additional effort was not expended in their development.

Failure to take the spatial distribution of the stars into account would result in the so-called *Lutz-Kelker bias*, which is a bias in the estimated distance. Bayesian analysis takes care of such biases through the straightforward process of incorporating the cause of the bias in the prior distribution. If Cepheid stars were distributed uniformly over a region, the prior distribution of distances s from the observer would be proportional to s^2. However, the spatial distribution of Cepheid variables is known to be flattened with respect to the galactic plane. We thus modify the s^2 prior by using a spatial distribution of stars that is exponentially stratified as one moves away from the galactic plane. In particular, the prior distribution on the distance s, given a hyperparameter z_0, is

$$p(s) \propto s^2 \exp\left(-|z|/z_0\right),$$

where $z = s \sin\beta$, with β being the galactic latitude of the star (its angle above the galactic plane, another known covariate), and z_0 being the imperfectly known 'scale height.' This 'hyperparameter' z_0 is known to be $z_0 = 97 \pm 7$ parsecs, so we simply assigned it a uniform prior over this range.

The priors on the Fourier coefficients $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$ must be chosen carefully, to avoid overfitting or underfitting. Luckily, the models in (1.6) are exactly of the form (1.2), so that the conventional priors described there can be utilized directly. (We are slightly cheating here, in that $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$ also occur in the likelihood terms arising from the C_i, when (1.7) is used in 1.4), and one could argue that the appropriate default priors should reflect this. We ignore this complication, in part because we think it would make little difference and, in part, because it is unclear how to take this into account in defining a default prior. Also, in the computations reported here, we utilized the simpler hyperprior $p(\tau) = 1/\tau^{3/2}$.)

A possible alternative prior for the Fourier coefficients would be the empirical expected posterior prior, also defined in section 2.2. Note that, for the normal linear model, $m^*(\mathbf{y}^*)$ can be found in closed form. Space precludes our presenting these results here.

3.3 Computation

Space limitations preclude a full description of the MCMC computation that is used to analyze the Bayesian model. We thus limit the discussion to presentation of the major steps in the analysis, especially those that are non-standard. Familiarity with MCMC computation is assumed.

A reversible-jump MCMC algorithm of the type reviewed in Dellaportas *et. al.* (2000) is used to generate posterior distributions and estimates. The full conditional distributions for the variance and precision parameters and hyperparameters are standard χ^2 distributions and so the sampling of these parameters can be accomplished with straightforward Gibbs sampling.

For $\Delta\phi$, ϕ_0 and s, we employ a random-walk Metropolis algorithm using, as the proposal distribution, a multivariate normal distribution centered on the currently imputed parameter values and with a covariance matrix that is proportional to the covariance matrix for the linearized least-squares problem for just these three parameters. (This means linearizing the logarithm in the expression for c_i in (1.7)). This proposal distribution leads to a fast mixing Markov chain, which implies fast convergence of the computational algorithm.

The Fourier coefficients $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$, as well as u_0 and v_0, are sampled via an independence-chain Metropolis step. The natural proposal distributions are found by combining the normal likelihoods in (1.6) with the normal priors (given τ) in (1.2), leading to conjugate normal posterior distributions. Note that these are not the actual full conditionals from the posterior, because of the nonlinear way in which $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$ appear in the full likelihood. However, the acceptance probabilities for these proposals are well over 90%, and the sampling of the Fourier parameter spaces is very effective.

The Metropolis steps for $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_v$ are included within a step that proposes a jump between models. Thus, if the current model has a certain number of parameters, we propose a jump to a model with a (possibly different) number of parameters, and simultaneously propose new values for all the Fourier coefficients. To make the sampling efficient, during the burn-in phase we also estimate the posterior probabilities of the individual models, and use them as the basis for the proposal probabilities of new models during the computation phase of the calculation. Thus models of higher posterior probability are proposed with greater frequency. A total of 10,000 iterations of the MCMC computation were performed.

3.4 Results

Figures 2 and 3 give the posterior probabilities of the orders of the trigonometric models for the radial velocity and the photometry, respectively. The fifth-order model is clearly overwhelmingly preferred for velocity. For the photometry model, on the other hand, the third and fourth-order models are nearly equally supported. The MCMC computational strategy discussed above will automatically perform 'model-averaging' over these models, when computing posterior quantities of interest.

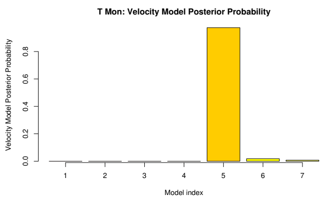

Figure 2. Posterior marginal distribution of velocity models for T Mon.

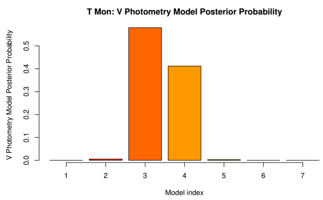

Figure 3. Posterior marginal distribution of photometry models for T Mon.

Estimates, standard errors, etc., for any of the unknowns or parameters in the model are also available from the MCMC computation. Here we simply show, in Figure 4, the posterior distribution of the distribution of the parallax (the inverse of the distance) for T Mon. Figure 5 shows the simulation history of the parallax, i.e., the values of the parallax that were generated at each trial of the MCMC computation. The very random appearance of this history strongly indicates that the MCMC computation was accurate.

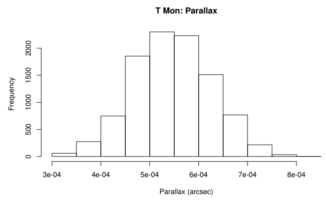

Figure 4. Posterior marginal distribution of the parallax of T Mon.

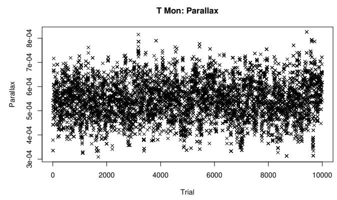

Figure 5. Simulation history of the parallax of T Mon.

4 A Wavelet Approach

Examination of Figure 1 suggests a potential concern. It turns out that the analysis is quite sensitive to the extent of the 'dip' in the velocity curve that occurs between phases 0 and 0.1. Notice also that there is no data between 0.9 and 0.1 (phases 1.0 and 0.0 being, of course, considered to be equal). Because Fourier analysis is non-local (each term in the trigonometric polynomial influencing the curve over the entire domain), there is concern that Fourier analysis may over-accentuate or under-accentuate the dip, in order to find a slightly better fit at points distant from the dip.

An approach that avoids this difficulty is the wavelet approach, since wavelet bases are local. To date, we have only applied this approach to the problem of fitting the velocity curve. Space precludes a detailed description here (see Müller, Berger, and Jefferys, 2001, for details and results), but we can, at least, outline the needed steps.

Step 1. A *function space* prior is needed, i.e., a prior on the space of possible velocity curves. The idea is to develop the prior in terms of intuitively accessible features of the function, and then transform this prior into the wavelet domain (a domain in which it is not as natural to construct priors). Adapting a suggestion of Vannucci and Corradi (1999), we chose the function space prior to be a Gaussian process (since this allows easy transformation into the wavelet domain). We actually construct the prior on differences of the function, since this makes it easier to (i) build periodicity into the Gaussian process and (ii) build smoothness into the function.

Step 2. One *transforms* this Gaussian process prior on the function space into the wavelet domain, using a bivariate wavelet decomposition, as suggested in Vannucci and Corradi (1999). The resulting prior on the wavelet coefficients is multivariate normal with a non-diagonal covariance matrix (i.e., the wavelet coefficients are apriori dependent).
Step 3. A *model* in the wavelet domain is defined by some subset of all the wavelets in the basis. We specify the prior probability of a model through the device of allowing each wavelet coefficient to be zero, with specified probability $p(k)$, where k is the 'level' of the wavelet coefficient. (In practice, we used $p(k) = 1 - \alpha^{k+1}$, and tried various values of α.) Then, with probability $1 - p(k)$, the coefficient would be in the model. The prior distribution of the coefficients in the model is obtained from the multivariate normal prior found in Step 2, by conditioning on the other coefficients being zero.
Step 4. A *Metropolis-Hastings MCMC* analysis is implemented, in which moves are made to adjacent models (i.e., either a nonzero wavelet coefficient is set equal to zero, or a zero coefficient is made nonzero). The key is that, utilizing properties of the multivariate normal distribution, the computations involved in these 'small' steps are of relatively low cost to implement. (Wavelet models are large enough that it would be prohibitively expensive to compute, from scratch, the posterior model probabilities.)

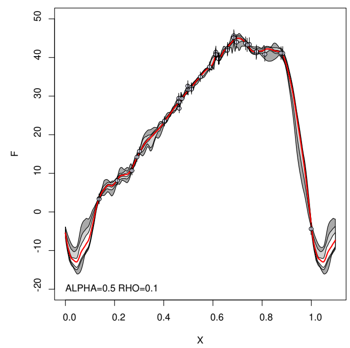

Figure 6. Wavelet fit of velocity data for T Mon. Shown are contour lines for the posterior distribution of the velocity as a function of phase; the thick smooth line in the center is the posterior mean curve. The grey shaded margins show central 50% (light grey) and central 90% (dark grey) intervals. The points are the observed data points, with little error bars showing 2 standard deviations for the measurement error.

5 Acknowledgments

This research was supported by the National Science Foundation, under Grants DMS-9802261 and DMS-0103265.

References

Berger, J. and Bernardo, J. (1992). On the development of the reference prior method. In J. Bernardo, J. Berger, A. Dawid and A. F. M. Smith (editors), *Bayesian Statistics 4*, Oxford University Press, London.

Berger, J. and Pericchi, L. (2001). Objective Bayesian methods for model selection: introduction and comparison (with Discussion). To appear in the IMS Lecture Notes volume on *Model Selection*, P. Lahiri, editor.

Chen, M.H., Shao, Q.M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

Dellaportas, P., Forster, J. and Ntzoufras, I. (2000). On Bayesian model and variable selection using MCMC. To appear in *Statistics and Computing*.

Jaynes, E.T. (1999). *Probability Theory: The Logic of Science.* Accessible at the website http://bayes.wustl.edu/etj/prob.html.

Jefferys, W. H. and Barnes, T. G. (1999). Bayesian analysis of Cepheid variable data. *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 777–783.

Jefferys, W. and Berger, J. (1992). Ockham's razor and Bayesian analysis. *American Scientist* **80**, 64–72.

Jeffreys, H. (1961). *Theory of Probability* (3rd edition), Oxford University Press, London.

Kass, R. and Raftery, A. (1995). Bayes factors and model uncertainty. *J. Amer. Statist. Assoc.*, 90, 773–795.

Müller, P., Berger, J., and Jefferys, W. (2001). Nonparametric regression with wavelet based priors: efficient posterior simulation for unequally spaced data and dependent priors. ISDS Discussion Paper, Duke University.

Pérez, J.M. and Berger, J. (2000). Expected posterior prior distributions for model selection. ISDS Discussion Paper 00-08, Duke University

Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods.* New York: Springer-Verlag.

Vannucci, M. and Corradi, F. (1999). Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *J. Roy. Statist. Soc. B* , 971–986.

Zellner, A. and Siow (1980). Posterior odds for selected regression hypotheses. In *Bayesian Statistics 1*, eds. J.M. Bernardo, et. al., Valencia: Valencia University Press, pp. 585–603.