Statistical Science
2012, Vol. 27, No. 1, 3–10
DOI: 10.1214/11-STS384

Bayesian Nonparametric Shrinkage Applied to Cepheid Star Oscillations

James Berger, William H. Jefferys and Peter Müller

Abstract. Bayesian nonparametric regression with dependent wavelets has dual shrinkage properties: there is shrinkage through a dependent prior put on functional differences, and shrinkage through the setting of most of the wavelet coefficients to zero through Bayesian variable selection methods. The methodology can deal with unequally spaced data and is efficient because of the existence of fast moves in model space for the MCMC computation.

The methodology is illustrated on the problem of modeling the oscillations of Cepheid variable stars; these 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.

Key words and phrases: Nonparametric regression, wavelets, shrinkage prior, sparsity, variable selection methods.

1. INTRODUCTION

1.1 Nonparametric Bayesian Shrinkage

Bayesian analysis has long been a major methodological vehicle for implementation of shrinkage ideas in complex scenarios. There are two primary ways in which such shrinkage is implemented. The first is through use of prior distributions which shrink the unknowns in some fashion—to prespecified locations or prespecified subspaces, depending on the problem and type of prior. Thus an unknown normal mean could be shrunk toward a specified prior mean; a collection of unknown normal means could be shrunk toward the hyperplane in which the means are equal; and an unknown real function could be shrunk toward the subspace of monotonic functions. This is the Bayesian version of the type of shrinkage originating with Stein (1956) and James and Stein (1961).

The second major Bayesian vehicle for shrinkage is Bayesian variable selection, which sets some of the unknown parameters to zero. This is often an overly drastic shrinkage, but is certainly not so in the context of model selection, or in the context of nonparametric function estimation. In the latter setting, the unknown parameters that are set to zero are typically coefficients of basis elements from a basis representation of the function, and sparsity considerations strongly encourage such shrinkage.

Both of these shrinkage concepts are herein utilized in nonparametric function estimation with dependent wavelets. The motivating application is to Cepheid variable stars and is described in the next subsection; the functions to be estimated can have arbitrary shapes, but are quite smooth. It is to induce sufficient smoothness that will utilize both types of shrinkage discussed above.

1.2 The Astronomical Problem

There is a class of stars, called Cepheid variables, that pulsate with a regular and distinctive periodic sig-

James Berger is the Arts and Sciences Professor of Statistics, Department of Statistical Science, Duke University, Durham, North Carolina 27708-0251, USA (e-mail: berger@stat.duke.edu). William H. Jefferys is Harlan J. Smith Centennial Professor of Astronomy (Emeritus), The University of Texas at Austin and Adjunct Professor of Statistics, The University of Vermont, 253 Strong Road, Moretown, Vermont 05660, USA (e-mail: bill@astro.as.utexas.edu). Peter Müller is Professor, Department of Mathematics, The University of Texas at Austin, 1 University Station, C1200, Austin, Texas 78712, USA (e-mail: pmueller@math.utexas.edu).

nature. The stars actually grow larger and then smaller, and as a result their luminosities vary periodically along with their colors. Since there is a physical relationship between the star's linear diameter, its luminosity, and its color, there are actually two independent periodically varying quantities.

A very interesting and useful property of these stars is that their mean luminosities are highly correlated with their pulsation period, in that the shorter-period stars are less luminous than the longer-period ones. This is very well approximated as a linear relation between the log of the period and the log of the luminosity. As a consequence, if one knows the slope and intercept of this relationship, and measures the period of a Cepheid (which is trivial), one can infer the luminosity with quite high precision. This makes these stars very useful as "standard candles," because knowledge of a star's luminosity as well as its observed brightness allows us to compute the distance from the inverse square law. Knowing the distance to the individual Cepheid also gives us the distance to the galaxy or cluster of stars in which it is embedded. Thus, these stars are fundamental in setting the distance scale of the universe.

The most challenging feature of the problem statistically is that the key photometry and radial velocity curves for a star are unknown, and have no simple structure. In Barnes et al. (2003), Fourier polynomials of finite (but unknown) degree were used to represent these two curves. 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 $\times$'s) together with a fifth-order trigonometric polynomial fit to the data. Because of the possibility of quite arbitrary shapes for the photometry and velocity curves for Cepheid variable stars, we instead desired to model the curves via much more flexible wavelet decompositions.

1.3 Computational Implementation

Posterior inference in this setup is formally equivalent to variable selection in a normal linear regression problem with massively many candidate covariates. Posterior simulation requires averaging and/or selection across alternative models defined by the set of basis functions (wavelets) which are included in the model. In the context of normal-linear regression, common approaches are guided search in the model space using the Occam's Window principle (Madigan and Raftery, 1994; Raftery, Madigan and Hoeting, 1997); Markov chain Monte Carlo simulation across the model space (George and McCulloch, 1997; Smith and Kohn, 1996); and importance sampling or Gibbs sampling based on analytic approximations to the marginal posterior distribution on the model indicator (Clyde, DeSimone and Parmigiani, 1996; Clyde, Parmigiani and Vidakovic, 1998). See, for example, Clyde (1999), Hoeting et al. (1999) and Clyde and George (2004) for reviews. In this paper we introduce a scheme for fast posterior simulation across the model space, marginalizing over the wavelet coefficients. We use a computational strategy similar to that used by George and McCulloch (1997) and Smith and Kohn (1996) to allow fast computation of marginal model probabilities when considering models differing by only one wavelet basis function.

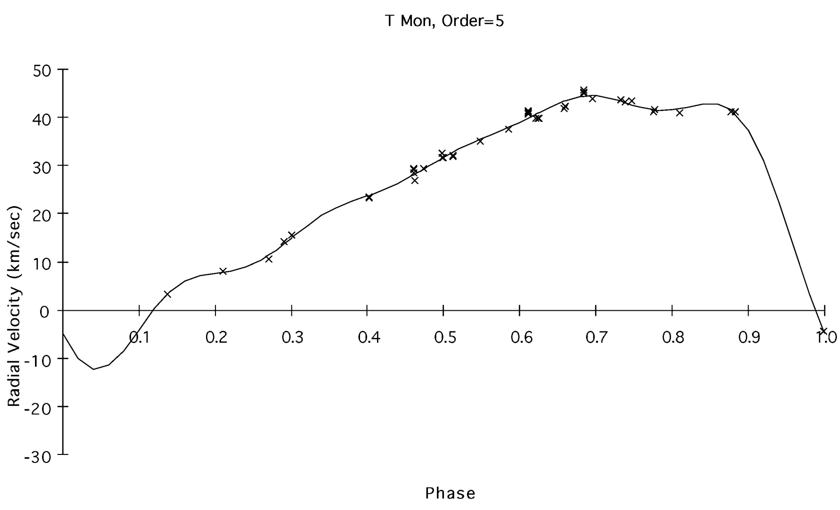


FIG. 1. *The radial velocity data (the $\times$'s) for T Mon, and their fit to a fifth-order trigonometric polynomial.*

2. WAVELET REPRESENTATION

Wavelet decomposition allows representation of any square integrable function $f(x)$ as

$$f(x) = \sum_{k \in Z} c_{J_0 k} \phi_{J_0 k}(x) + \sum_{j \geq J_0} \sum_{k \in Z} d_{jk} \psi_{jk}(x). \tag{1}$$

Here $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ and $\phi_{jk}(x) = 2^{j/2} \cdot \phi(2^j x - k)$ are wavelets and scaling functions at level of detail j and shift k. In the context of statistical modeling, (1) allows for inference about random functions by defining a probability model for the coefficients $\theta = (c_{J_0 k}, d_{jk},\ j \geq J_0;\ \ k \in Z)$, that is, (1) provides a parameterization of a random function f in terms of the wavelet coefficients θ. See, for example, Vidakovic and Müller (1999) or Ferreira and Lee (2007), Chapter 5, for a review of wavelet representations relevant for statistical modeling.

Perhaps the most common application of (1) in statistical modeling is to nonlinear regression where $f(x)$ represents the unknown mean response $E(y|x)$ for an observation y with covariate x. Chipman, Kolaczyk and McCulloch (1997), Clyde, Parmigiani and Vidakovic (1998), Vidakovic (1998), Semadeni, Davison and Hinkley (2004), Tadesse et al. (2005), Wang and Wood (2006), ter Braak (2006) and Abramovich, Angelini and De Canditiis (2007), among many others, discuss Bayesian inference in such models assuming equally spaced data, that is, covariate values x_i are on a regular grid. For equally spaced data the discrete wavelet transformation is orthogonal. Together with assuming independent measurement errors and a priori independent wavelet coefficients this leads to posterior independence of the d_{jk}. Thus the problem essentially reduces to a sequence of univariate problems, one for each wavelet coefficient. See, for example, Yau and Kohn (1999) for a review. Generalizations of wavelet techniques to non-equidistant (NES) design impose additional conceptual and computational burdens. A reasonable approximation is to bin observations in equally spaced bins and proceed as in the equally spaced case. If only few observations are missing to complete an equally spaced grid, treating these few as missing data leads to efficient implementations (Antoniadis, Grégoire and McKeague, 1994; Cai and Brown, 1998). We propose instead an approach which does not depend on posterior independence. Our approach includes informative dependent priors with positive prior probabilities for vanishing wavelet coefficients.

3. SHRINKAGE OF *f*(*x*)

3.1 Shrinkage Toward a Smooth Subspace

Because of the wavelet representation that will be used, a function space prior can be defined by considering the function at the discrete points $\{i/n, i = 1, \ldots, n\}$, where $n = 2^J$. Letting $f_i = f(i/n)$, consider the difference process $d_i = f_i - f_{i-1}$.

A function space prior that "shrinks toward smoothness" can be defined by imposing positive correlations on the d_i. Specifically, let $d = (d_1, \ldots, d_n)$, and define the prior to be $p(d) = N(0, \Delta)$ with $\Delta_{ij} = \lambda \exp(-\beta|i - j|)$; that is, we assume a multivariate normal prior with scale parameter λ and log correlations proportional to distance.

Let $\Delta_{(11)}$ denote the left upper $(n-1) \times (n-1)$ submatrix of Δ and partition Δ into

$$\Delta = \begin{bmatrix} \Delta_{(11)} & \Delta_{(12)} \\ \Delta_{(21)} & \Delta_{(22)} \end{bmatrix}.$$

Let $v = \mathrm{Var}(\sum_{i=1}^n d_i) = \lambda \sum_{i=1}^n \sum_{j=1}^n \exp(-\beta|i - j|)$. Assuming $f_0 \sim N(0, \lambda\sigma_0^2)$ we find

$$p(f_0, \ldots, f_{n-1} | f_0 = f_n) = N(0, \lambda V),$$

with $V = A H_0 A'$,

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & & & \\ 1 & 1 & \cdots & 1 \end{bmatrix},$$

$$H_0 = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & H \end{bmatrix} \quad \text{and}$$

$$H = \Delta_{(11)} - \Delta_{(12)} \Delta'_{(12)} / v.$$

In view of the normalization property, $\|\phi_{jk}\| = 1$, scaling coefficients at the highest level of detail J are approximately proportional to the represented function, $c_{Jk} \approx 2^{-J/2} f_k$. Therefore the multivariate normal prior on $(f_0, \ldots, f_{n-1})$ implies $p(c_J) = N(0, r_J \cdot \lambda V)$ where $r_J = 2^{-J}$. Following common practice in the use of wavelet decomposition, we will ignore the proportionality constant r_J and assume

$$p(c_J) = N(0, \lambda V).$$

As long as we also drop r_J in the reconstruction of $f(x)$, ignoring the proportionality constant will leave the final inference unchanged.

The prior $p(c_J) = N(0, \lambda V)$ implies a dependent multivariate normal prior for the vector of all wavelet

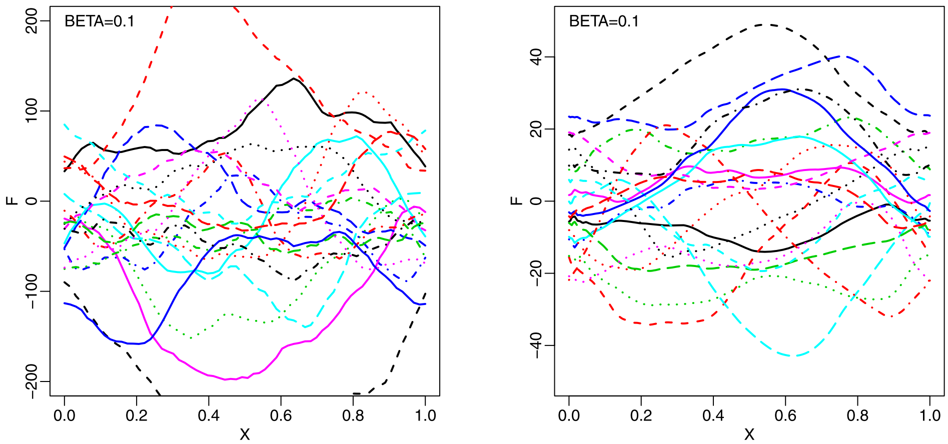


FIG. 2. *For $\beta = 0.1$, the left panel plots simulations from the prior process on the unknown function conditioning on all wavelet coefficients included; the right panel shows for comparison prior simulations conditional on setting those coefficients equal to zero which are excluded by the universal wavelet thresholding rule with $\sqrt{2n}\hat{\sigma}$ of Donoho and Johnstone (1994).*

coefficients $d = (c_{J_0k}, d_{jk}, j = J_0, \ldots, J, k = 0, \ldots, 2^j - 1)$

$$p(d|\gamma = 1) = N(0, \lambda\Lambda). \tag{2}$$

In principle Λ can be found by explicitly computing the linear operator of the wavelet decomposition. But from a computational point of view this is unnecessary and undesirable. Instead Vannucci and Corradi (1999) show how Λ can be derived from V as a bivariate wavelet decomposition of V.

3.2 Shrinkage Through Wavelet Sparsity

One of the important advantages of wavelet bases over alternative bases for L^2 functions is the parsimony property of wavelet representations. Reasonably regular functions are well approximated with only few nonzero wavelet coefficients. Therefore "shrinkage toward smoothness" can also be induced by setting many of the wavelet coefficients to be zero. We thus assume positive prior probability for vanishing wavelet coefficients.

Let $\gamma = (\gamma_1, \ldots, \gamma_l)$ denote the vector of indices of nonzero wavelet coefficients, that is, $d_{jk} = 0$ iff $(jk) \notin \gamma$. We define a prior distribution on γ with geometrically decreasing probability for nonzero wavelet coefficients in higher levels of detail j:

$$\Pr(d_{jk} = 0) = 1 - \alpha^{j+1}.$$

See, for example, Abramovich, Sapatinas and Silverman (1998) for a discussion of the choice of α.

We write θ_γ for the subvector of nonzero wavelet coefficients d_{jk}, and we use $\gamma = 1$ for the full model which includes all coefficients $\gamma = ((jk), j = J_0, \ldots, J$ and $k = 0, \ldots, 2^j - 1)$. The prior $p(\theta_\gamma|\gamma)$ for the wavelet coefficients under model γ is implied from (2) by conditioning the multivariate normal on $\theta_h = 0$, $h \notin \gamma$. Let $\Omega = V^{-1}$ and write $\Omega_{(\gamma)}$ for the submatrix with rows and columns $(\gamma_1, \ldots, \gamma_l)$. Then

$$p(\theta_\gamma|\gamma) = N\big(0, \lambda\Omega_{(\gamma)}^{-1}\big) = N(0, \lambda\Lambda). \tag{3}$$

We use Λ to generically denote $\Omega_{(\gamma)}^{-1}$, suppressing the dependence on γ to simplify notation.

3.3 Illustration of the Shrinkage Effects

Figures 2 and 3 demonstrate the "shrinkage toward smoothness" behavior of the priors in Sections 3.1 and 3.2. The figures give realizations from the priors specified in the two subsections. Figure 2 utilizes $\beta = 0.1$ from the prior in Section 3.1 and Figure 3 utilizes $\beta = 0.9$. The smaller β induces much more dependence, clearly resulting in smoother functions.

The left panel of each figure is generated from use of only the prior in Section 3.1, that is, all the wavelet coefficients are kept. In contrast, the right panels of each figure show what happens when many of the wavelet coefficients are set to zero. (For simplicity, these were produced using a standard wavelet thresholding rule.) Clearly, setting many wavelet coefficients to zero does seem to result in considerable additional shrinkage toward smoothness.

4. POSTERIOR SIMULATION

We implement posterior inference using Markov chain Monte Carlo simulation. Marginalizing over θ_γ, we use the posterior probabilities $p(\gamma|y)$ to define a

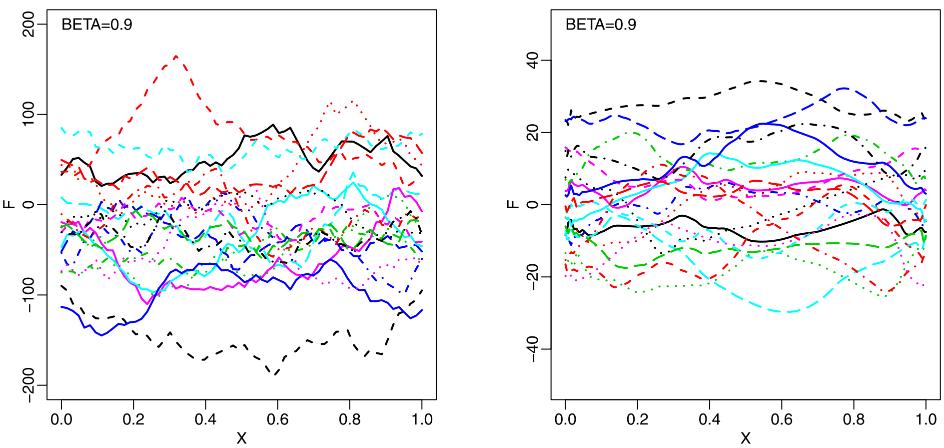


FIG. 3. *Prior simulations as in Figure 2, but using $\beta = 0.9$ (very little dependence).*

Metropolis–Hastings scheme which proposes moves in the model space by adding or deleting one wavelet basis function at a time. The computational effort of the proposed scheme is comparable to that of George and McCulloch (1997) and Smith and Kohn (1996), who suggest schemes based on algorithms by Chambers (1971) and (1979) which allow fast updating of a Choleski decomposition of the cross-product matrix $X'X$. The algorithms proposed by George and McCulloch (1997) and Smith and Kohn (1996) allow computation of marginal posterior probabilities with $O(q^2)$ basic operations, where q is the number of covariates (basis functions) included in the model. We describe a similar efficient updating algorithm in a form suitable for the wavelet regression problem.

Notation. Let A_{ij} be the element in the ith row and jth column of a matrix A, with A_i being its ith column vector. For a vector $\gamma = (\gamma_1, \ldots, \gamma_l)$ we denote with A_γ the submatrix consisting of *columns* $(\gamma_1, \ldots, \gamma_l)$, with $A_{(\gamma)}$ the submatrix consisting of *columns and rows* $(\gamma_1, \ldots, \gamma_l)$, and with $A_{(-\gamma)}$ the submatrix with rows and columns $\gamma = (\gamma_1, \ldots, \gamma_l)$ removed.

Let x_i, y_i, $i = 1, \ldots, N$, denote the observed data. Let $h = 1, \ldots, 2^J$ index the wavelet coefficients $d = (c_{J_0 k}, d_{jk})$ and let X denote the design matrix

$$X_{ih} = \begin{cases} \psi_{jk}(x_i) & \text{for } h = 2^{J_0} + 1, \ldots, n, \\ \phi_{J_0 k}(x_i) & \text{for } h = 1, \ldots, 2^{J_0}, \end{cases}$$

where (jk) are the wavelet indices corresponding to the hth element in the vector d of wavelet coefficients.

Likelihood. For a given model γ the wavelet decomposition of the unknown velocity curve f implies a likelihood

$$(4) \qquad y_i|\theta, \gamma \overset{\text{i.i.d.}}{\sim} N(X_\gamma \theta_\gamma, S), \quad i = 1, \ldots, N,$$

where $S = \operatorname{diag}(\sigma_i^2)$ with known variances σ_i^2, $i = 1, \ldots, N$.

Posterior. Together with prior (3) the likelihood implies a multivariate normal posterior $p(\theta_\gamma | y, \gamma) = N(\mu, \Sigma)$ with

$$\Sigma^{-1} = \underbrace{(X_\gamma)' S^{-1} X_\gamma}_{Q^\gamma} + 1/\lambda \Omega_{(\gamma)} \quad \text{and}$$

$$\mu = \Sigma \cdot \underbrace{(X^\gamma)' S^{-1} y}_{v^\gamma}.$$

Again, to simplify notation we suppress the dependence on γ in μ and Σ.

4.1 Down Move

Assume $\gamma = (\gamma_1, \ldots, \gamma_l)$ and consider a move "down" to the submodel $\gamma^* = (\gamma_1, \ldots, \gamma_{l-1})$. Partition Σ into

$$\Sigma = \begin{bmatrix} \Sigma_{(-l)} & \tilde{\Sigma}_l \\ \tilde{\Sigma}_l' & \Sigma_{ll} \end{bmatrix}$$

and similarly $\mu = (\mu_{(-l)}, \mu_l)$. Then

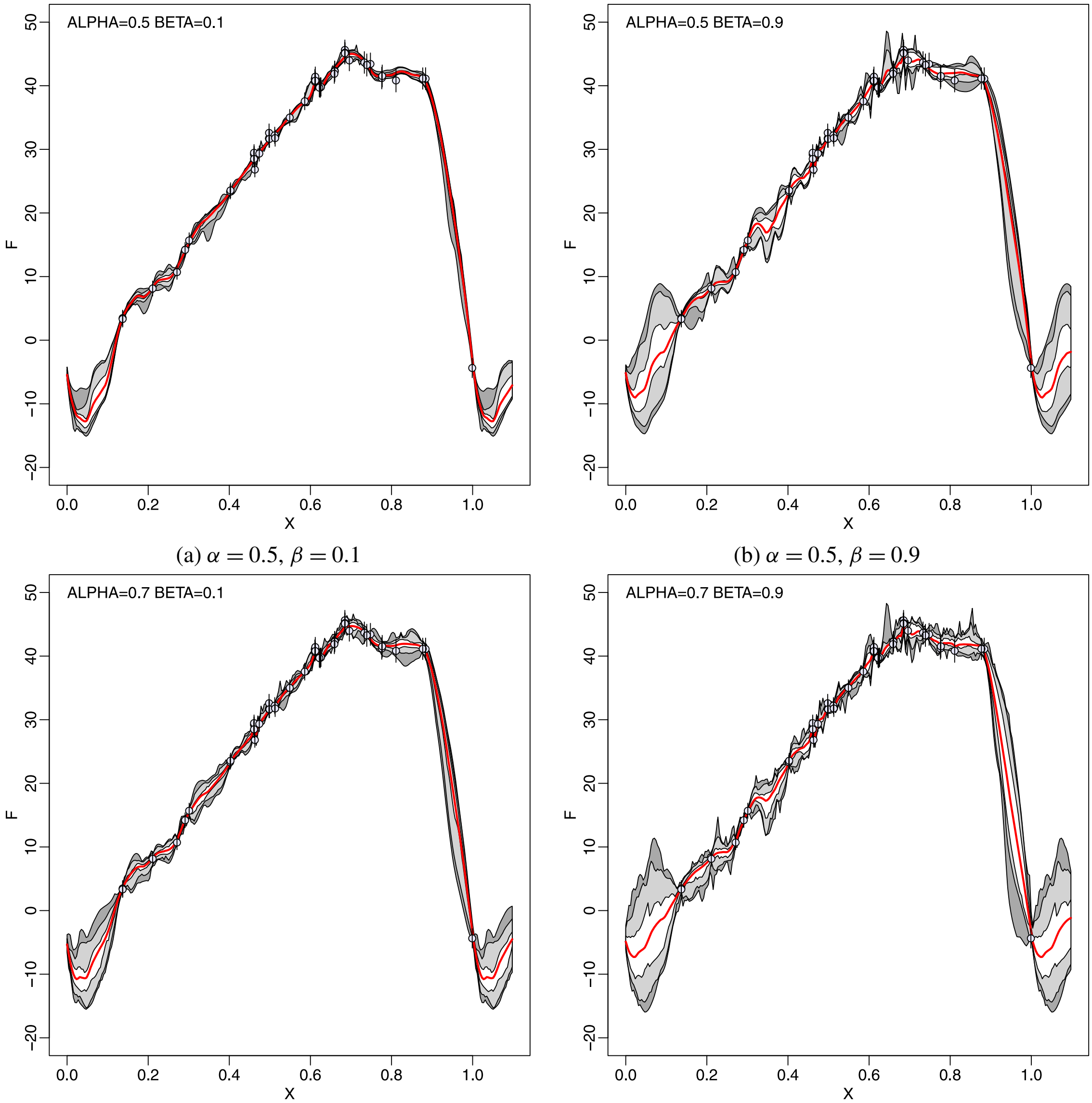
$$p(\theta_{\gamma^*} | y, \gamma^*) = N(\mu^*, \Sigma^*),$$

with $\Sigma^* = \Sigma_{(-l)} - \tilde{\Sigma}_l \Sigma_{ll}^{-1} \tilde{\Sigma}_l'$ and $\mu^* = \mu_{(-l)} + \tilde{\Sigma} \Sigma_{ll}^{-1}(-\mu_l)$. Similarly, $\Lambda^* = \Lambda_{(-l)} - \tilde{\Lambda} \Lambda_{ll}^{-1} \tilde{\Lambda}_l'$.

The corresponding ratio of marginal probabilities is

$$\frac{p(y|\gamma^*)}{p(y|\gamma)} = \left(\frac{\lambda \Lambda_{ll}}{\Sigma_{ll}} \right)^{1/2} e^{-(1/2)\mu_l^2 / \Sigma_{ll}}.$$

This expression is easily verified using the candidate formula $p(y|\gamma) = p(\theta_\gamma|\gamma) p(y|\theta_\gamma, \gamma) / p(\theta_\gamma | y, \gamma)$ and substituting $\theta_\gamma = 0$.



(a) $\alpha = 0.5$, $\beta = 0.1$ (b) $\alpha = 0.5$, $\beta = 0.9$

(c) $\alpha = 0.7$, $\beta = 0.1$ (d) $\alpha = 0.7$, $\beta = 0.9$

FIG. 4. *Posterior inference for T Moncerotis. In all four panels, the thick smooth line shows the posterior mean curve. The gray shaded margins show central 50% (light gray) and central 90% (dark gray) intervals. The points are the observed data points, with little error bars showing 2 standard deviations for the measurement error. Panel* (a) *shows inference under $\beta = 0.1$ and $\alpha = 0.5$. Panels* (b) *through* (d) *show posterior inference using $\beta = 0.9$* (b *and* d) *and $\alpha = 0.7$* (c *and* d). *Fixing $\beta = 0.9$ essentially assumes independence of the d_i and implies less smoothing; setting $\alpha = 0.7$ greatly decreases the number of wavelet coefficients set to zero.*

4.2 Up Move

Consider a move from γ to $\gamma^* = (\gamma_1^*, \gamma)$. Denote with (μ, Σ) and Λ the posterior and prior moments under the (current) model γ:

$$p(\theta_\gamma|\gamma, y) = N(\mu, \Sigma) \quad \text{and} \quad p(\theta_\gamma|\gamma) = N(0, \lambda\Lambda).$$

Similarly, let (μ^*, Σ^*) and Λ^* denote the posterior and prior moments under the (proposed) model γ^*:

$$p(\theta_{\gamma^*}|\gamma^*, y) = N(\mu^*, \Sigma^*) \quad \text{and}$$

$$p(\theta_{\gamma^*}|\gamma^*) = N(0, \lambda\Lambda^*).$$

For posterior simulation we use a lower triangular Choleski decomposition of the posterior variance/covariance matrix, $TT' = \Sigma$ and $T^{*\prime}T^* = \Sigma^*$. The new moments μ^*, Σ^* and Λ^* and the Choleski decomposition T^* are computed using the following expressions.

Let $Q^* = (X^{\gamma^*})' S^{-1} X^{\gamma^*}$, $\Omega^* = \Omega_{(\gamma^*)}$, $Q = (X^{\gamma})' \cdot S^{-1} X^{\gamma}$ and $\Omega = \Omega_{(\gamma)}$ and partition

$$Q^* = \begin{bmatrix} Q^*_{11} & \tilde{Q}^{*\prime}_1 \\ \tilde{Q}^*_1 & Q \end{bmatrix} \quad \text{and} \quad \Omega^* = \begin{bmatrix} \Omega^*_{11} & \tilde{\Omega}^{*\prime}_1 \\ \tilde{\Omega}^*_1 & \Omega \end{bmatrix}.$$

Let $b = \tilde{Q}^*_1 + 1/\lambda \tilde{\Omega}^*_1$, $h = \Sigma b$, $c = \tilde{Q}^*_{11} + 1/\lambda \Omega^*_{11}$, $b_0 = \tilde{\Omega}^*_1$, $h_0 = \Lambda \tilde{\Omega}^*_1$ and $c_0 = \Omega^*_{11}$. Then

$$\Sigma^* = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma \end{bmatrix} + \frac{1}{c - b'h} \begin{bmatrix} 1 & -h' \\ -h & hh' \end{bmatrix} \quad \text{and}$$

$$\Lambda^* = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda \end{bmatrix} + \frac{1}{c_0 - b_0' h_0} \begin{bmatrix} 1 & -h_0' \\ -h_0 & h_0 h_0' \end{bmatrix},$$

$$\mu^* = \begin{pmatrix} 0 \\ \mu \end{pmatrix} + (c - b'h) \Sigma^*_1 \Sigma^{*\prime}_1 v^{(\gamma^*)},$$

and T^* is obtained by augmenting T with a new first column $w = \Sigma^*_1 / \sqrt{\Sigma^*_{11}}$ to

$$T^* = \begin{bmatrix} & 0 \\ w & T \end{bmatrix}.$$

The corresponding ratio of marginal probabilities is, by symmetry to the down move,

$$\frac{p(y|\gamma)}{p(y|\gamma^*)} = \left(\frac{\lambda \Lambda^*_{11}}{\Sigma^*_{11}} \right)^{1/2} e^{-(1/2) \mu_1^{*2} / \Sigma^*_{11}}.$$

5. EXAMPLE

We apply the above methodology to the data for the star T Moncerotis, as shown in Figure 1, for the choices $\beta = 0.1$ (strong dependence of the d_i) and $\alpha = 0.5$ (inducing a moderate level of sparsity). The resulting nonparametric posterior is difficult to summarize; some features of this posterior are presented in Figure 4(a).

It is, of course, one of the strengths of the Bayesian approach to shrinkage that uncertainty in the shrinkage estimate [the posterior mean of $f(x)$, given by the thick center line in Figure 4(a)] can also be given. This is crucial in characterizing the (considerable) uncertainty in the eventual estimate of distance to the star (see Barnes et al., 2003).

Figure 4 also indicates the effect on the T Moncerotis data of each of the shrinkage priors in Sections 3.1 and 3.2. Panel (b) shows the effect of the prior in Section 3.1; setting $\beta = 0.9$ effectively makes the d_i independent. Panel (c) shows the effect of the prior in Section 3.2; setting $\alpha = 0.7$ greatly decreases the number of wavelet coefficients set to zero. In both cases, the posterior functions appear to be unreasonably rough and the uncertainty in the shrinkage estimate appears to be unreasonably large. Panel (d), which effectively uses neither of the shrinkage techniques, is especially unsatisfactory.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grants DMS-01-03265, DMS-06-35449 and DMS-07-57549-001. We are grateful to Thomas Barnes for providing us with the data analyzed herein.

REFERENCES

ABRAMOVICH, F., ANGELINI, C. and DE CANDITIIS, D. (2007). Pointwise optimality of Bayesian wavelet estimators. *Ann. Inst. Statist. Math.* **59** 425–434. MR2388800

ABRAMOVICH, F., SAPATINAS, T. and SILVERMAN, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 725–749. MR1649547

ANTONIADIS, A., GRÉGOIRE, G. and MCKEAGUE, I. W. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.* **89** 1340–1353. MR1310225

BARNES, T. G., JEFFERYS, W. H., BERGER, J., MUELLER, P., ORR, K. and RODRIGUEZ, R. (2003). Bayesian analysis of the cepheid distance scale. *Astrophysical Journal* **592** 539–554.

CAI, T. T. and BROWN, L. D. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.* **26** 1783–1799. MR1673278

CHAMBERS, J. (1971). Regression updating. *J. Amer. Statist. Assoc.* **66** 744–748.

CHIPMAN, H., KOLACZYK, E. and MCCULLOCH, R. (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc.* **92** 440.

CLYDE, M. A. (1999). Bayesian model averaging and model search strategies. In *Bayesian Statistics* 6 (J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith, eds.) 157–185. Oxford Univ. Press, New York. MR1723497

CLYDE, M. A., DESIMONE, H. and PARMIGIANI, G. (1996). Prediction via orthogonalized model mixing. *J. Amer. Statist. Assoc.* **91** 1197–1208.

CLYDE, M. and GEORGE, E. I. (2004). Model uncertainty. *Statist. Sci.* **19** 81–94. MR2082148

CLYDE, M., PARMIGIANI, G. and VIDAKOVIC, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85** 391–401. MR1649120

DONGARRA, J., MOLER, C., BUNCH, J. and STEWART, G. (1979). *Linpack's Users' Guide*. SIAM, Philadelphia.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

FERREIRA, M. A. R. and LEE, H. K. H. (2007). *Multiscale Modeling: A Bayesian Perspective*. Springer, New York. MR2348732

GEORGE, E. and MCCULLOCH, R. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7** 339–374.

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–401. MR1765176

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

MADIGAN, D. and RAFTERY, A. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *J. Amer. Statist. Assoc.* **89** 1535–1546.

RAFTERY, A. E., MADIGAN, D. and HOETING, J. A. (1997). Bayesian model averaging for linear regression models. *J. Amer. Statist. Assoc.* **92** 179–191. MR1436107

SEMADENI, C., DAVISON, A. C. and HINKLEY, D. V. (2004). Posterior probability intervals in Bayesian wavelet estimation. *Biometrika* **91** 497–505. MR2081316

SMITH, M. and KOHN, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75** 317–344.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley. MR0084922

TADESSE, M. G., IBRAHIM, J. G., VANNUCCI, M. and GENTLEMAN, R. (2005). Wavelet thresholding with Bayesian false discovery rate control. *Biometrics* **61** 25–35. MR2129198

TER BRAAK, C. J. F. (2006). Bayesian sigmoid shrinkage with improper variance priors and an application to wavelet denoising. *Comput. Statist. Data Anal.* **51** 1232–1242. MR2297519

VANNUCCI, M. and CORRADI, F. (1999). Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 971–986. MR1722251

VIDAKOVIC, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.* **93** 173–179. MR1614620

VIDAKOVIC, B. and MÜLLER, P. (1999). An introduction to wavelets. In *Bayesian Inference in Wavelet-Based Models*. *Lecture Notes in Statistics* **141** 1–18. Springer, New York.

WANG, X. and WOOD, A. T. A. (2006). Empirical Bayes block shrinkage of wavelet coefficients via the noncentral χ^2 distribution. *Biometrika* **93** 705–722. MR2261452

YAU, P. and KOHN, R. (1999). Wavelet nonparametric regression using basis averaging. In *Bayesian Inference in Wavelet-Based Models*. *Lecture Notes in Statistics*. **141** 95–108. Springer, New York. MR1699888