Prior Elicitation in the Wavelet Domain

Hugh A. Chipman
Lara J. Wolfson

ABSTRACT Bayesian methods provide an effective tool for shrinkage in wavelet models. An important issue in any Bayesian analysis is the elicitation of a prior distribution. Elicitation in the wavelet domain is considered by first describing the structure of a wavelet model, and examining several prior distributions that are used in a variety of recent articles. Although elicitation has not been directly considered in many of these papers, most do attach some practical interpretation to the hyperparameters which enables empirical Bayes estimation. By considering the interpretations, we indicate how elicitation might proceed in Bayesian wavelet problems.

1 Introduction

Bayesian statistical analysis using any statistical method, including wavelets, typically involves specifying priors on components of the problem. Bayesian analysis using wavelets primarily focuses on using priors to determine shrinkage rules – in other words, placing prior distributions in some fashion to indicate which wavelet coefficients, for a given basis, are going to tend towards zero. There are a variety of ways that these priors might be formulated, incorporating increasing levels of complexity and information about the uncertainty inherent in wavelet coefficients at various resolution levels tending towards zero. After using a graphical model to outline some commonly used structures, we then attempt to summarize the different ways that researchers have developed to formulate and apply these priors, focusing particularly on how the parameters for each prior are chosen. In this sense, this is a chapter about the "elicitation" of priors on wavelet coefficients.

When using a Bayesian approach and formulating priors, whether for wavelets or any other problem, one concern to many practitioners of statistics is how the priors are chosen. How should the parametric form, and the actual parameters themselves (usually referred to as the hyperparameters) be determined? The answers, unfortunately, are very difficult to obtain. Considerable attention is given in a paper by Kass and Wasserman (1995) to criteria for choosing a prior; several recent papers (Kadane and Wolfson, 1998; O'Hagan, 1998; and Craig et. al., 1998) attempt to lay out criteria for how the hyperparameters for the prior distribution should be elicited. Each problem tends to be somewhat unique, in that the choice of the prior distribution and its parameters tends to be a function not only of common statistical practice, but of the specific objectives of the statistical analysis being performed.

In wavelets, where subjectivity usually comes in through priors placed on the wavelet coefficients, the process of "eliciting" the values of the hyperparameters is dependent on the form chosen for the prior. One criteria that should therefore be considered in choosing the form of the prior, then, is whether or not the hyperparameters will have meaningful interpretations that will ease the process of elicitation, or if a connection be drawn between the hyperparameters and some quantity that is easier to specify. The existing literature contains examples of both, as shown in Section 3 where we outline the approaches various authors have taken to prior specifications. An alternative, one that seems to be quite popular among those who do research in Bayesian wavelets, is to employ empirical Bayes methods that attempt to determine the hyperparameters of the prior distribution from the data being analyzed. This has some obvious drawbacks, but it can be argued that it eliminates some of the subjectivity from the use of the Bayesian approach.

The wavelet coefficients, however, are not the only subjective area in wavelet analysis. Overlooked in many papers is the actual choice of an appropriate wavelet basis. Most of the papers we examine in this chapter assume that the appropriate basis (Daubechies, Meyer, Haar, etc.) is agreed on in advance, and that the uncertainty lies only in the coefficients estimated for a given basis. Since different bases may more parsimoniously represent signals with different characteristics, priors on the choice of basis seem a natural extension of much of the work on Bayesian model selection. We discuss this further in Section 4.

2 The Structure of a wavelet model and priors

In this section, the parameters of a wavelet model are given, and basic choices of prior distributions discussed. Brief references are given to specific papers; for full details, refer to Section 3. A general framework is introduced for placing prior distributions on these parameters, via a graphical model. This structure is then used to describe and compare particular choices of priors used in a number of recent papers. Where possible, the notation of Chapter 1 is used.

The vector of observed signal values $Y = (Y_1, Y_2, \ldots, Y_N)$ is assumed to have been generated by the model

$$Y = X\theta + \epsilon \tag{1}$$

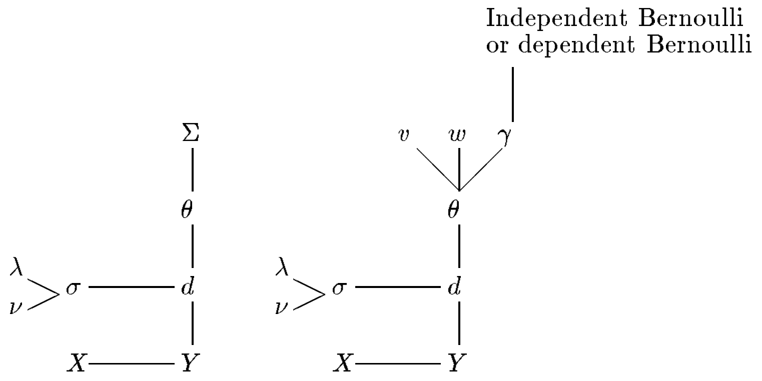

FIGURE 1. Parameters of a wavelet model and corresponding priors. Figure 1(a) (left) involves one level on unobserved wavelet coefficients θ, while 1(b) (right) adds a second hierarchy.

where X is the $(N \times N)$ matrix of wavelet basis functions evaluated at equally spaced $x[i] = i/N$, θ is the $(N \times 1)$ vector of unobserved wavelet coefficients, and ϵ is a N-vector of iid $\mathrm{N}(0, \sigma^2)$ errors. Typically $N = 2^n$ with n integer. Because the matrix X is orthogonal, the N-vector of observed wavelet coefficients

$$d = (X'X)^{-1}X'Y = X'Y \tag{2}$$

is also assumed to be an iid vector with normal errors having variance σ^2. That is,

$$d \overset{iid}{\sim} N(\theta, \sigma^2). \tag{3}$$

Orthogonality of X also implies that the observed coefficients d and the observed signal Y are 1-1 functions of one another, and the observed coefficients may be treated as the response, conditional on the choice of a basis X. Orthogonality also gives the simplification of (2). The power of wavelet models is their ability to decompose a signal in terms of both location and scale ("frequency"). Typically the wavelet coefficients d are indexed as d_{jk}, with j indicating the scale and k indicating location.

The parameters (X, θ, σ^2) define the wavelet model. It is on these parameters that priors are to be placed. In what follows, both variables (such as Y), model parameters ((X, θ, σ^2), and prior parameters) will be generically referred to as "variables".

These variables and relationships (1) and (3) are represented graphically in Figures 1(a) and 1(b). Figure 1(a) represents the simplest case, and is considered initially. In this graphical model, solid lines are used to indicate dependence between variables (represented by the vertices of the graph). For example, the variable Y depends on (d, X) since $Y = Xd$. Conditional independence is represented as follows: if a variable is conditioned upon, the corresponding vertex of the graph and all lines originating at that vertex are removed. Any variables not connected by a line (directly or through other variables) are conditionally independent. For example, conditional on the observed response, X and d are independent. That is, d and X depend on each other only through the observed response Y.

Note that in some models the (unobserved) wavelet coefficients θ are assumed to have a variance proportional to σ^2, necessitating a dependence between θ and σ. Since some models do not assume this dependence, this relationship is not indicated in Figure 1(a) or 1(b).

Prior distributions may be placed on all variables that are at the periphery of Figure 1(a), e.g. X, σ, θ. Priors on σ are simple enough that they are represented in the main part of the figure. Specifically σ is assumed to follow an inverse gamma distribution with parameters ν, λ, or equivalently $\nu\lambda/\sigma^2 \sim \chi^2_\nu$. Special cases of interest include fixing σ to some value s by taking $\nu = \infty, \lambda = s^2$, and an uninformative prior, taking $\nu = 0$.

The wavelet basis, given by the matrix X is usually assumed to be fixed, rather than placing a prior on it. This is further discussed in Section 4.

Several different priors have been considered for the (unobserved) wavelet coefficients θ. Two classes of priors are represented in Figures 1(a) and 1(b), and described below. Since shrinkage of the wavelet coefficients is achieved via a prior on θ, it is to be expected that a number of different prior structures have been considered. It is in this respect that most of the papers surveyed here differ; here we outline most of the general approaches, and refer the reader to the next section for details.

One straightforward possibility is to assign to θ a multivariate normal prior with mean vector 0 and covariance matrix Σ, i.e. $\theta \sim \text{MVN}(0, \Sigma)$, as in Figure 1(a) (Vannucci and Corradi (this volume), Vidakovic and Muller 1995). Another possibility is to have the prior covariance of θ given as $\sigma^2\Sigma$, indicated by the dashed line between θ and σ in Figure 1(a). An important special case is to assume that Σ is the identity matrix, yielding prior independence of the coefficients. Vidakovic (1998) considers a similar case, but with independent t distributions rather than normals. These approaches achieve shrinkage through the choice of Σ. Dependence between coefficients is also introduced via Σ. Coefficients with strong prior correlations would be those which are "close" to each other - i.e. coefficients at similar locations, or similar resolutions.

Shrinkage of the wavelet coefficients can be achieved by other priors. A popular choice for the prior the coefficients θ is the use of a scale mixture of two distributions. One mixture component would then correspond to "negligible" coefficients, the other to "significant" coefficients. This may be achieved in a number of ways, although most articles to date have used either a scale mixture of two normal distributions or a mixture of one normal and a point mass at 0. The latter case is a limiting case of the first. This parameterization is depicted in Figure 1(b), and is given as

$$\theta_{jk} = \gamma_{jk} N(0, v_{jk}^2) + (1 - \gamma_{jk}) N(0, w_{jk}^2), \qquad \theta_{jk}|\gamma_{jk} \text{ independent} \tag{4}$$

Paper	Prior θ	σ	shrinkage
Ruggeri & Vidakovic (1995)	various	fixed	threshold
Vannucci & Corradi (1997)	MVN, correlated	IG	shrink
Vidakovic (1998)	T, indep	exp	shrink

TABLE 1. Comparison of techniques using a single prior on wavelet coefficients. Abbreviations are MVN = multivariate normal, T=Student's t, IG = inverse gamma, exp = exponential.

where $v_{jk}^2 \gg w_{jk}^2$ and γ_{jk} is binary (0/1). The greater magnitude of v_{jk}^2 represents the large coefficients, while the mixture component with smaller variance w_{jk}^2 represents the negligible cases. In such cases, v_{jk}^2 and w_{jk}^2 are chosen, and a prior placed on γ_{jk}. The limiting case, in which $w_{jk} = 0$ reduces one mixture component to a point mass at 0.

An important distinction between the use of two normals versus a normal and a point mass is the type of shrinkage obtained. With $w_{jk} > 0$, no coefficient estimate based on the posterior will be exactly equal to zero. If $w_{jk} = 0$, then it is possible that some wavelet coefficient estimates will be exactly zero (i.e. the coefficients are "thresholded"). Abramovich, Sapatinas, and Silverman (1998) achieve this latter result by using the posterior median rather than the posterior mean. Although $w_{jk} > 0$ yields a smoother shrinkage function, it does not compress the data, since all estimated wavelet coefficients are nonzero. If compression is paramount, the point mass prior may be preferred.

Whether a mixture of two normals or a normal and a point mass is used, a prior must still be placed on the binary variables γ_{jk}. The simplest prior is

$$\Pr(\gamma_{jk} = 1) = 1 - \Pr(\gamma_{jk} = 0) = \pi_{jk}, \qquad \gamma_{jk} \text{ independent } \forall j,k \tag{5}$$

The only article not using such a prior is Crouse, Nowak, and Baraniuk (1998), which instead introduces dependence among the different γ_{jk}, either at similar locations or similar resolutions.

One article related to Figure 1(b), but with an important difference is Holmes and Denison (1998). In this case an infinite number of scale mixtures are used rather than two discrete choices.

3 A comparative look at current practice

In this section we give details on the different prior structures as indicated in Tables 1 and 2. Techniques using only one additional hierarchical level on the prior structure of θ (i.e. as in Figure 1(a)) are considered first, followed by the majority of papers, which place at least two levels of prior on θ (as in

	Prior			
Paper	θ	σ	γ	shrinkage
Chipman et.al. (1997)	N-N	fixed	indep.	shrink
Crouse et.al. (1998)	N-N	fixed	depend.	shrink
Abramovich et.al. (1998)	P-N	fixed	indep.	thresh
Clyde & George (1998)	P-N	fixed	indep.	shrink
Clyde et.al. (1998)	P-N	IG	indep.	shrink
Vidakovic (1998)	P-T	exp	indep.	thresh
Holmes & Denison (1998)	N*	IG	—	shrink/thresh

TABLE 2. Comparison of techniques using a mixture prior on wavelet coefficients. Abbreviations are N=Normal, P=point mass, T=student's T, IG=Inverse Gamma. * Holmes and Denison employ a continuous mixture of normal, rather than two mixture elements. Prior independence is still assumed.

Figure 1(b)). The use of differing notations in the original prior necessitates some translation; we use the notation of Section 2.

The papers by Ruggeri and Vidakovic (1995) and Ruggeri (1999) are particularly interesting because they describe structures for priors on wavelet coefficients that differ from the variants on the normal linear model that most authors use. Here, the authors offer a fairly comprehensive catalog of possible distributions for the wavelet coefficients dependent only on a location parameter (the scale parameter is assumed to be known), and suggest priors corresponding to each of the choices. They consider a number of combinations of location parameter priors for θ (listed first) with distributional models for d (listed second): double exponential/normal, normal/normal, double exponential/double exponential, t/t (both with the same and different degrees of freedom), normal/double exponential, t/double exponential, and t/normal. The authors do not focus on how the hyperparameters of the prior should be chosen, other than suggesting empirical Bayes methods.

Vannucci and Corradi (1997, this volume) consider a model with an inverse gamma prior on σ and a multivariate normal prior on θ. They elaborate on an idea introduced in Vidakovic and Muller (1995) that the prior covariance Σ be non-diagonal, i.e. coefficients are correlated. The correlation structure is considered both for coefficients within the same resolution level j and coefficients in different levels. Coefficients are arranged in such a fashion that the covariance matrix can be expressed as a diagonal band outside which correlations are zero. Within the band, the largest correlations are between coefficients that are close in location and scale. The covariance matrix is specified in terms of two hyperparameters, simplifying the elicitation process. Values of these hyperparameters are chosen either by empirical Bayes methods, or by making these parameters a component of the model and simulating their posterior distribution via MCMC. An alternate method for introducing dependence among wavelet coefficients is considered by Crouse et. al. (1998). The two approaches are compared in the description of Crouse et. al. below.

The remaining techniques considered use two or more hierarchical levels of prior on the wavelet coefficients θ. The majority considered use either a mixture of two normals or a mixture of a normal and a point mass. First, we outline methods that use a mixture of two normal distributions for a prior on θ. The main distinction between Chipman, Kolaczyk and McCulloch (1997) and Crouse et. al. (1998) is that the former assume independence in a prior for γ as in (5), while Crouse et. al. introduce dependencies.

In Chipman et. al. (1998), a hierarchical prior (4) is placed on the wavelet coefficients θ and the noise level σ is assumed known. The wavelet coefficients prior is a scale mixture of two normals, each with mean 0, and variances w_{jk}^2 and $v_{jk}^2 \gg w_{jk}^2$. Independence prior (5) on mixture indicators γ is used. Conditional on γ, wavelet coefficients are assumed independent in the prior.

The hyperparameters to be chosen are thus $\sigma, v_{jk}, w_{jk}, \pi_{jk}$. Hyperparameters indexed by jk are assumed to vary with resolution j, but be fixed across location k. This yields level-dependent shrinkage, similar to SureShrink (Donoho and Johnstone 1995). Each of hyperparameters are given an interpretation that allows default values to be estimated from the data. While fixing σ may be theoretically unappealing, this assumption enables rapid closed form calculation of posterior means for wavelet coefficients. The interpretation of π_j is the percentage of coefficients at a resolution level which are expected to be non-negligible. Prior standard deviation w_j represents the magnitude of a negligible coefficient, while v_j gives the largest magnitude of a significant coefficient.

These hyperparameters are not directly elicited. Instead, associations are made between the interpretations given above and

1. The magnitude of an insignificant change in the response Y
2. The magnitude of a substantial change in a given wavelet basis function, and
3. The percent of coefficients at a given resolution level that are above a noise threshold of Donoho, Johnstone, Kerkyacharian and Picard (1995)

Crouse et. al. (1998) consider the same specification of prior for θ as Chipman et. al. (1997) above, except for the innovation of considering dependent priors on the mixture parameters γ_{jk}. Crouse et. al. consider two types of Markov dependence among coefficients. In both cases, the γ_{jk} are arranged in a grid corresponding to the resolution and location of the corresponding wavelet coefficient. The dependency is Markov in the sense that whether or not a specific γ_{jk} is nonzero will depend on the state of its immediate neighbors. These neighbors could either be indicators corresponding to coefficients at the same resolution and location $k \pm 1$, or at an analogous location and at resolution $j \pm 1$. Links within a resolution level are referred to as a "hidden Markov chain", while links across resolutions are a "hidden Markov tree". In essence, if a certain coefficient is significant, then its neighbors (however they are defined) are also likely to be significant. There is an important distinction between this induced correlation and that of Vannucci and Corradi (1997, this volume), where correlation is directly induced on the coefficients via Σ. When correlations are assumed at the γ level, significant coefficients will make neighbors more likely to be significant, but no restrictions are placed on the direction of the significance. A significant coefficient is more likely to have significant neighbors, but they are equally likely to be large with the same or opposite sign. Introducing positive correlations on the coefficients θ themselves via Σ says quite another thing. If a coefficient is large, then a neighboring coefficient is also likely to be large and of the same sign. If the signal to be modeled oscillates rapidly, neighboring wavelet coefficients may be large but of opposite sign, making dependence introduced via γ more appropriate.

The majority of the other mixture-coefficient prior papers rely on a mixture of a point mass and a normal. Abramovich et. al. (1998), Clyde and George (1998), and Clyde, Parmigiani and Vidakovic (1998) all consider such priors. The first two are quite similar, while the last places a prior on σ rather than assuming a fixed value.

In Abramovich et. al. (1998) and Abramovich and Sapatinas (this volume), prior (4) for θ is used, with $w_{ij} = 0$, yielding a mixture of a point mass at zero and a $N(0, v_j^2)$. Independence prior (5) is placed on γ. It is assumed that the prior distribution (i.e., the choice of the hyperparameters v_j^2, π_j) is the same for any given resolution level j, and that the noise level, σ, is known or a good estimate is available. This is a structure similar to that found in Clyde et. al. (1998), and can be viewed as an extreme case of the formulation given in Chipman et. al. (1997). These hyperparameters are assumed to have structure such that $v_j^2 = 2^{-aj}C_1$ and $\pi_j = \min(1, 2^{-Bj}C_2)$, where C_1, C_2, a and B are assumed to be non-negative constants. Some interpretation of these constants is given to explain how they might be derived, particularly in the context of decision-theoretic thresholding rules. For example, in the discrete wavelet transform, this prior structure will correspond to the universal threshold of Donoho & Johnston (1994) when $a = B = 0$, and letting $C_1 \to \infty$, $C_2 \to 0$, such that as n increases, $\sqrt{C_1}/C_2\sigma n \to 1$. A nice intuitive interpretation of the parameters is also given, in that the prior expected number of non-zero wavelet coefficients on the jth level is $2^{j(1-B)}C_2$. Thus, if $B > 1$, the expected number of non-zero coefficients is finite, and if $B = 0$, this implies a prior belief that all coefficients on all levels have the same probability of being non-zero. If $B = 1$, this assumes that the expected number of non-zero wavelet coefficients is the same on each level. An interesting contribution made by these authors is to connect the parameters a and B, and the Besov space parameters

(s, p), so that if a particular Besov space were chosen to represent prior beliefs, a, B could then be numerically derived. In particular, the authors recommend that a, B be chosen based on prior knowledge about regularity properties of the unknown function be studied. For the parameters C_1 and C_2, the authors propose an empirical-Bayes type procedure that relates C_1 and C_2 to the *VisuShrink* threshold of Donoho and Johnstone (1994). In practice, the authors recommend a "standard" choice of $a = 0.5, B = 1$. Part of the novelty of the approach is the idea that by simulating observations from different Besov spaces, one can elicit the correct space from which to choose the prior. One suggestion made for future research is to consider incorporation of a dependence structure between the wavelet coefficients. A weakness of this approach is that while B has a nice interpretation, the parameters C_1, C_2 don't have good intrinsic interpretability, and so any elicitation of these parameters would be very difficult.

Clyde and George (1998) employ the usual mixture of a point mass and normal distribution as a prior on the wavelet coefficients (i.e. $w_{jk} = 0$ in (4)). They propose an empirical Bayes procedure based on conditional likelihood approximations given the MAD estimate of σ proposed by Donoho et. al. (1995). This doesn't differ a great deal from the approach of Abramovich et. al. (1998).

In Clyde et. al (1998), priors are specified for both the wavelet coefficients θ and the noise standard deviation σ. For θ, (4) is used with $w_{jk} = 0$ and $v_{jk}^2 = c_{jk}\sigma^2$, yielding a mixture of a point mass at zero and a $N(0, c_{jk}\sigma^2)$. Prior independence of different coefficients is assumed. The parameter $\sqrt{c_{jk}}$ indicates the magnitude (relative to σ) of a large (or significant) coefficient at scale j and location k. This parameter is chosen to be constant across scale and location in practice (i.e. $c_{jk} = c$). An inverse gamma prior with parameters ν and λ is placed on σ^2.

An interpretation of c is obtained via linkages to information criteria given in George and Foster (1997). Different choice of c correspond to BIC, AIC, or the risk inflation criterion of Foster and George (1994) and Donoho and Johnstone (1994). The elements of γ are assumed independent. The mixing probability π_{jk} in (5) is chosen to vary across scale j but not location k. The interpretation of π_j is similar to that of Chipman et. al. (1997), although no automatic choices or elicitation recommendations are made.

In Vidakovic (1998), priors similar to both Figure 1(a) and Figure 1(b) are explored. For Figure 1(a), an exponential prior is placed on σ^2 (rather than the inverse gamma used in many other papers considered here), and a scaled t-distribution on θ_{jk}. Prior parameters to be selected are the prior mean of σ^2, the prior variance of θ, and the degrees of freedom for the t prior on θ. No recommendations regarding elicitation are made, although empirical default values are proposed. Shrinkage is achieved with this model via the posterior mean (i.e. Bayes rule). No coefficients are shrunk exactly to zero (i.e. thresholded). To achieve thresholding, the prior on θ_{jk} is modified to be a mixture of a scaled t and a point mass at 0. No choice is given for the mixture probabilities π_{jk}. This differs slightly from (4) as employed by Abramovich et. al. (1998), Clyde and George (1998), and Clyde et. al. (1998), which mix a point mass with a normal. Thresholding is obtained via the Bayes factor.

Holmes and Denison (1998) also assume $\theta \stackrel{iid}{\sim} N(0, v)$ as in Figure 1(a), and an inverse gamma prior on σ. The difference is that rather than choosing v (they call this λ^{-1}), a hyperprior is placed upon it. This is similar to the other approaches that use a scale mixture of normals, but with the distinction that the prior on the variance is a continuous prior, rather than a prior on two different values. That is, rather than using mixture of two normals, Holmes and Denison use an infinite mixture of normal with different variances. In this way, the paper more closely resembles the approaches using a mixture of two normals or a normal and point mass.

They note that this gives an intuitive interpretation to each component v_i^{-1}, which is to shrink the classical least squares estimate of θ by $(1 + \sigma^2 v_i^{-1})^{-1}$. Because the conjugate inverse-gamma or Wishart prior for v^{-1} would do little to aid elicitation, the authors turn to the fact that beliefs are likely to be most available on the complexity and smoothness on the underlying signal that is being reconstructed. Following the result of Hastie & Tibshirani (1990) on identifying the degrees of freedom of a linear smoother, the authors derive the prior

$$v_i^{-1} \mid \sigma^2 \propto \exp(-c \sum_{i=1}^{N} (1 + \sigma^2 v_i^{-1})^{-1}), \tag{6}$$

where c is a constant for penalizing model complexity. The hyperparameter c is shown to have a direct relationship with the log model probability, so that $c = \{0, 1, 0.5 \log n, \log n\}$ correspond to well-known model choice criteria (Bayes Factor, AIC, BIC, RIC respectively).

This degrees of freedom connection to model complexity allows the elicitation of the prior distribution to capture the effect of both the number of coefficients in the model and the extent of the prior shrinkage on them.

4 Future work and conclusions

One component of the wavelet models upon which priors have not been placed is the basis (X in Figures 1(a) and 1(b)). All articles reviewed here have conditioned on the choice of wavelet basis. Choice of an appropriate basis is something of an art form, making formal elicitation difficult. Wavelet bases are often grouped together, and indexed by an integer parameter within each group. For example, Daubechies (1988) has a compactly supported family of wavelets indexed by an integer N, with larger values of N yielding smoother basis functions. Priors might be placed on an indexing parameter within a family of basis functions, or on a range of different families. The first problem seems more approachable in terms of elicitation, provided that a degree of smoothness could be quantified and elicited. Yau and Kohn (1999) consider priors on the basis for purposes of model averaging, but place equal priors weight on each of four bases.

The possibilities for priors and their elicitation will grow as the field of wavelets grows. Interesting possibilities include wavelet packets, a means of constructing a new basis out of several individual bases, and methods that shift either the original data or the wavelet basis.

References

Abramovich, F. and Sapatinas, T. (1999) "Bayesian Approach to Wavelet Decompositions and Shrinkage", this volume.

Abramovich, F., Sapatinas, T., and Silverman, B.W. (1998) "Wavelet Thresholding via a Bayesian Approach", *Journal of the Royal Statistical Society, Series B*, 60, 725–750.

Chipman, H., Kolaczyk, E. and McCulloch, R. (1997) "Adaptive Bayesian Wavelet Shrinkage", *Journal of the American Statistical Association*, 92, 1413–1421.

Clyde, M., and George, E.I. (1998) "Robust Empirical Bayes Estimation in Wavelets", ISDS Discussion Paper 98-21, Duke University, Institute of Statistics and Decision Sciences

Clyde, M., Parmigiani, G., Vidakovic, B. (1998). "Multiple Shrinkage and Subset Selection in Wavelets", Biometrika 85, 391-402.

Crouse, M., Nowak, R. and Baraniuk, R. (1998). "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models". *IEEE Trans. Signal Processing*, 46, 886–902.

Craig, P.S., Goldstein, M., Seheult, A.H., and Smith, J.A. (1998) "Constructing Partial Prior Specifications for Models of Complex Physical Systems", (with discussion) *The Statistician*, 47, 37–54.

Daubechies, I. (1988) "Orthonormal Bases of Compactly Supported Wavelets" *Communications in Pure and Applied Mathematics*, 41, 909-996

Donoho, D. L. and Johnstone, I.M. (1994) "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, 81, 425–455.

Donoho, D. L. and Johnstone, I.M. (1995) "Adapting to Unknown Smoothness via Wavelet Shrinkage", *Journal of the American Statistical Association*, 90, 1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995) "Wavelet Shrinkage: Asymptopia?" (with discussion), *Journal of the Royal Statistical Society, Series B*, 57, 310-370.

Foster, D. and George, E. (1994) "The Risk Inflation Criterion for Multiple Regression", *Annals of Statistics* 22, 1947–75.

George, E. and Foster, D. (1997) "Empirical Bayes Variable Selection", Technical Report, University of Texas at Austin.

Holmes, C.C., and Denison, D.G.T. (1998) "Bayesian Wavelet Analysis with a Model Complexity Prior", in *Bayesian Statistics 6*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, Oxford University Press.

Kadane, J.B., and Wolfson, L.J. (1998) "Experiences in Elicitation" (with discussion), *The Statistican*, 47, 3–20.

Kass, R.E. and Wasserman, L. (1997) "The Selection of Prior Distribution by Formal Rules", *Journal of the American Statistical Association*, 96, 1343–1370.

O'Hagan, A. (1998) "Eliciting Expert Beliefs in Substantive Practical Applications", (with discussion) *The Statistician*, 47, 21–36.

Ruggeri, F. (1999) "Robust Bayesian and Bayesian Decision Theoretic Wavelet Shrinkage", this volume.

Ruggeri, F. and Vidakovic, B., (1995) "A Bayesian Decision Theoretic Approach to Wavelet Thresholding", ISDS Discussion Paper 95-35, Duke University, Institute of Statistics and Decision Sciences

Vannucci, M. and Corradi, F. (1997) "Some Findings on the Covariance Structure of Wavelet Coefficients: Theory and Methods in a Bayesian Perspective" *Journal of the Royal Statistical Society, Series B*, provisionally accepted.

Vannucci, M. and Corradi, F. (1999) "Modeling Dependence in the Wavelet Domain", this volume.

Vidakovic, B. (1998) "Nonlinear Shrinkage With Bayes Rules and Bayes Factors", *Journal of the American Statistical Association*, 93, 173–179.

Vidakovic, B. and Muller, P. (1995) "Wavelet Shrinkage with Affine Bayes Rules with Applications", Technical Report DP-95-36, ISDS, Duke University.

Yau, P. and Kohn, R. (1999) "Wavelet Nonparametric Regression Using Basis Averaging", this volume.