

Jeffrey S. MORRIS, Marina VANNUCCI, Philip J. BROWN, and Raymond J. CARROLL

We thank the reviewers for their interesting discussion and insightful comments and suggestions. As issues differ among the three discussions, we respond to each one separately.

1. RESPONSE TO KOLACZYK

We thank Eric for his kind comments and insightful investigation relating the regularization and variance component parameters of our model to the Besov space of the functions. His results are interesting and deep, and we hope the readers take the time to understand his results and consider their implications. In what follows we attempt to further interpret his theorem in light of our results. In the process we highlight two important features of our approach. First, although our shrinkage prior is only present at the top level of the hierarchy, our method provides some smoothing for functions at all hierarchical levels. Second, through the variance components, the data themselves help modulate the amount of smoothing and shrinkage to the mean at each hierarchical level. The chosen regularization parameters interact with the variance components to determine the degree of smoothing done at the various levels.

Table 1 relates the notation in his theorems to the notation in our paper. Also note that our notation for the wavelet scales differs. We define the basis functions as $\psi^{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$, while Kolaczyk uses $\psi^{j,k}(x) = 2^{j/2} \times \psi(2^jx - k)$, meaning that our level $j = 1$ refers to the finest wavelet level, and his $j = 1$ refers to the coarsest level. If J is the number of wavelet coefficient levels obtained, Kolaczyk's j is the same as our $J - j + 1$.

The theorem cited from Abramovich, Sapatinas, and Silverman (1998) (Result 1) states that the Besov space in which the function is assumed to belong depends on the choice of the regularization parameters τ_j^2 , the variances of the nonzero wavelet coefficients at the various wavelet resolution levels. The underlying Besov space should affect the amount of smoothing done in estimation, because different Besov spaces allow varying degrees of smoothness. Kolaczyk states that "Increasing α_0 for fixed p , for example, increases the range of smoothness allowed for $g_{0,0}$." This agrees with our intuition and with our empirical results, because larger α_0 imply smaller τ_j^2 , which causes wavelet coefficients to be shrunk more strongly toward 0, leading to more regularization and, thus, more smoothness in the resulting estimates.

Kolaczyk extends these results to our hierarchical model (5)–(9) in Result 2, which states roughly that the Besov space of the crypt-level functions g_{abc} depends on an interaction among the regularization parameters and the variance components at the various hierarchical levels, with the dominating variance components playing the major role. This does not surprise us, because it seems natural that the characteristics of the observed functions at the various levels should also play a role in determining the underlying Besov space of the true functions and, thus, the amount of smoothing done in the function estimation. In fact, it seems to be a positive aspect of our model that the data themselves automatically have a part in regulating the smoothness of our estimates. We will return to this point later, but first apply Result 2 to our dataset.

We estimate α_m from Result 2 by simply taking the slope from a linear regression of $\log_2(\sigma_{m,j}^2)$ on j . Figure 1 contains the scatterplots and fitted lines for $m = 0, 1$, and 2 for rats fed a diet with fish or corn oil. As Kolaczyk mentions, the crypt-level variation is the dominant factor, with $\alpha_2 < \alpha_0 < \alpha_1$ for both diets. If robust regression is used to fit the crypt-level variance components, this inequality is even more extreme, with $\alpha_2 = 1.61$ and 1.54 for diets of fish and corn oil, respectively. However, as he also points out, these parameters may result in the "broken boundary" phenomenon he describes. One way to describe this phenomenon is to say that the underlying smoothness of the true functions is constrained by different variance components, depending on the L^p space in which the functions are assumed to reside. We are not sure of the practical consequences of this, except to say that the variance components interact in subtle ways to determine the underlying Besov space of the functions. This in itself does not seem objectionable to us.

Looking more closely at Figure 1, we see noticeable curvature in the log-linear scatterplots for the crypt-level variance components for both diets. If a quadratic term is introduced to these models, it is statistically significant only for the crypt-level variance components ($p = .006$ and $p = .005$). The fact that this curvature is present for both diets, which contain observations from independent rats, leads us to question the assumption in Result 2 that the variance components decay exponentially in j for all m . We are not sure if this nonlinearity is unique to our dataset, related to the spiky nature of the data, or characteristic of the hierarchical model in general. This would be an interesting question to investigate. Also, Result 2 deals specifically with the function space of g_{abc} , the crypt-level functions. It would also be interesting to see if it is possible to prove like results for the functions at the other hierarchical levels g_{ab} or g_a .

We agree with Kolaczyk that one should not draw causative inferences from subtle observations such as the larger-than-expected crypt-level variance components at levels 3–5, and in this article we tried to avoid such terminology. We merely wanted to point out that these wavelet levels seem to coincide with the peaks that we see in the observed crypt profiles, and further proceeded to offer a possible biological explanation as to why such peaks would be characteristic of these data. We agree that subtle results such as these are certainly not strong enough to support causal inference, but we do believe they raise an interesting biological hypothesis about how MGMT operates (not previously introduced to our knowledge) that can be tested in future experiments.

The hierarchical nature of our model has interesting and useful implications for understanding how the variance components modulate the amount of smoothing done in the estimation of functions at the various hierarchical levels, as well as

Table 1. Relating the Notation of Kolaczyk to Morris et al.

| Kolaczyk | Morris et al. | Interpretation in Morris et al. |
|----------------------|--------------------------|-------------------------------------------------------------|
| $g_{0:0}(x)$ | $g_a(t)$ | Mean-level function |
| $g_{0:1}(x)$ | $g_{ab}(t)$ | Rat-level function |
| $g_{0:2}(x)$ | $g_{abc}(t)$ | Crypt-level function |
| j | $J - j + 1$ | Index for wavelet level |
| $\theta_{0:0}^{j,k}$ | $\theta_a^{J-j+1,k}$ | Mean-level wavelet coefficient |
| $\theta_{0:1}^{j,k}$ | $\theta_{ab}^{J-j+1,k}$ | Rat-level wavelet coefficient |
| $\theta_{0:2}^{j,k}$ | $\theta_{abc}^{J-j+1,k}$ | Crypt-level wavelet coefficient |
| $\sigma_{1,j}^2$ | $\sigma_{1,J-j+1}^2$ | Rat-level variance component |
| $\sigma_{2,j}^2$ | $\sigma_{2,J-j+1}^2$ | Crypt-level variance component |
| π_j | p_j | Prior probability of nonzero mean-level wavelet coefficient |
| $\sigma_{0,j}^2$ | τ_j^2 | Prior variance of nonzero mean-level wavelet coefficients |

the amount of borrowing of strength across experimental units. Consider the classical hierarchical model:

$$Y_{ij} \sim N(\mu_i, \sigma^2), \quad (1)$$

$$\mu_i \sim N(\theta, \xi^2). \quad (2)$$

Note that the posterior mean estimator of μ_i is given by $(\bar{Y}_i \xi^2 + \theta \sigma^2) / (\xi^2 + \sigma^2)$. This can be written as a weighted average whereby the empirical estimator \bar{Y}_i is shrunken linearly toward the prior mean θ . The amount of shrinkage is determined by

the relative magnitudes of the variance components, with more shrinkage as $\sigma^2 \gg \xi^2$.

These same principles apply to our multilevel hierarchical model (5)–(9), and help explain the degree of shrinkage to the mean as well as smoothing inherent in estimation of the mean-, rat-, and crypt-level functions. Because $\sigma_e^2 \ll \sigma_{1,j}^2$, the crypt-level coefficients $\theta_{abc}^{j,k}$ are not shrunken strongly toward the corresponding rat means $\theta_{ab}^{j,k}$, which indicates that the estimate of each crypt-level function g_{abc} does not borrow much

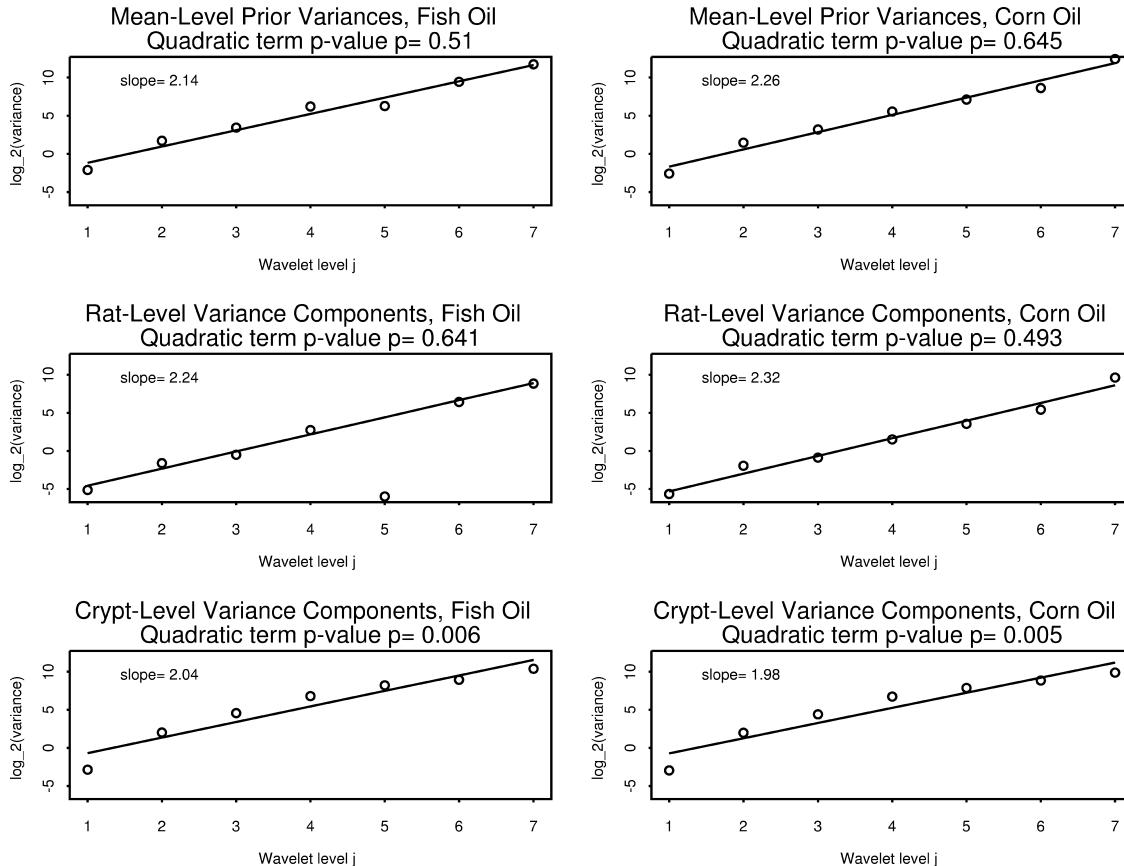


Figure 1. Plot of $\log_2(\text{VC})$ Versus Wavelet Level for Mean-, Rat-, and Crypt-Level Variance Components for Rats Fed Diets of Fish or Corn Oil, Along With the Linear Regression Fits. Robust regression was used for the rat-level variance component for the rats fed fish oil to minimize the effect of the obvious outlier. Note that the mean- and rat-level variances appear to follow the log-linear trend, but there is curvature in the plot for the crypt-level variance components for both diets. When a quadratic term is added to the model, it is statistically significant for the crypt-level variance components but not for the rat- or mean-level variance components for both diets.

strength from other crypts from that rat. The rat-level coefficients $\theta_{ab}^{j,k}$ are shrunk more strongly toward their mean $\theta_a^{j,k}$ because $\sigma_{1,j}^2 > \sigma_{2,j}^2$ for all j . For the mean-level coefficients, the amount of shrinkage toward the prior mean of 0 depends strongly on $\gamma_a^{j,k}$. When it is 1, there is little shrinkage because $\sigma_{2,j}^2 \ll \tau_j^2$. When $\gamma_a^{j,k} = 0$, then $\theta_a^{j,k}$ is thresholded, which can be viewed as degenerate shrinkage to 0 brought about by a prior variance that is effectively 0. As explained in the main article, Bayesian model averaging over $\gamma_a^{j,k}$ induces nonlinear shrinkage toward 0, whereby large coefficients experience little shrinkage, whereas smaller ones are accelerated strongly toward 0.

In the context of wavelet regression, the amount of shrinkage toward 0 determines the amount of smoothing done in function estimation. It is easy to see that this leads to regularization of the mean-level functions g_a , but this smoothing also “trickles down” to other levels of the hierarchy as modulated by the variance components. Because the rat-level coefficients $\theta_{ab}^{j,k}$ are shrunk strongly toward their means $\theta_a^{j,k}$, which are shrunk differentially toward 0, indirectly the $\theta_{ab}^{j,k}$ are also shrunk differentially toward 0. As a result, the estimates of g_{ab} are not only shrunk toward their mean functions g_a , but also visibly

smoothed. Figure 2 contains the rat-level function estimates for all rats at the 12-hour time point, and demonstrates both shrinkage to the mean and smoothing effects.

2. RESPONSE TO WU AND HUANG

We thank Colin and Jianhua for their enthusiasm regarding our work, and we applaud their effort to explore generalizations of our approach applied to a broader class of clustered functional data. They introduce an interesting model that can accommodate both time-varying and time-invariant covariates and the between-curve correlation imposed by the hierarchical model. The essence of this approach involves writing out the design matrices corresponding to the wavelet bases, choosing a subset of basis functions and specifying a covariance structure for them, and then fitting the hierarchical model using maximum likelihood. They mention that this model has been applied to splines and can, in principle, be used for other bases, as well.

Their proposed manner of handling the wavelet bases is fundamentally different from ours: They explicitly write out the design matrices for the wavelet bases, whereas we use the DWT to compute the wavelet coefficients for the observed functions and then model in the wavelet space. An advantage of their approach is that it can naturally handle unequally sampled data,

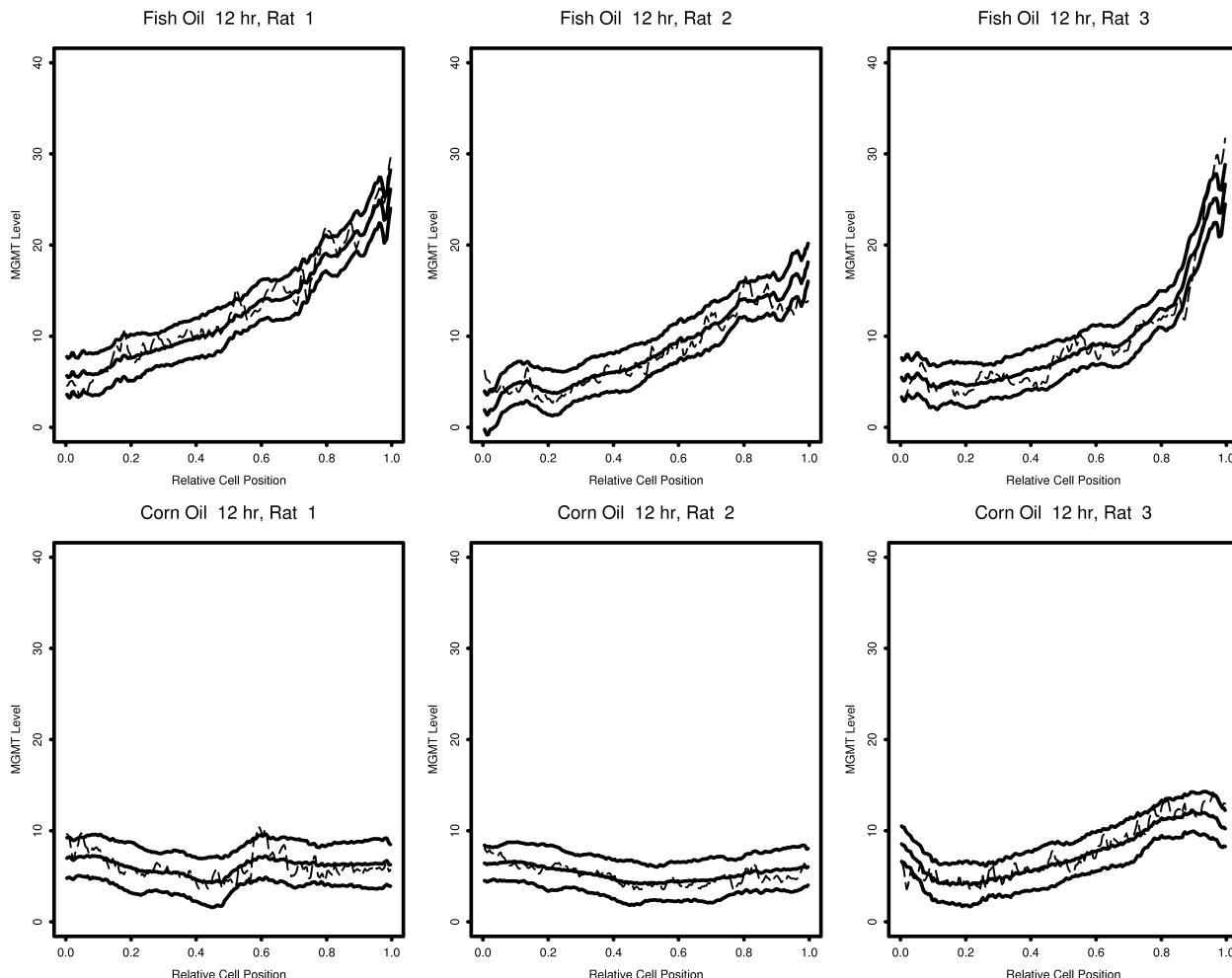


Figure 2. Posterior Mean MGMT Functions for Individual Rats in the 12-hour Group, Distal Colon, With 90% Posterior Pointwise Bounds. The dotted lines indicate the empirical estimates of the rat functions obtained by simply averaging the observed profiles over the rats' sampled crypts. Note the shrinkage toward the mean and smoothing inherent in our estimation of the rat curves.

whereas an advantage of ours is that it is much quicker to compute. Also, they propose a frequentist approach whereas ours is Bayesian, but as we point out in the article, the only informative priors in our approach are the shrinkage parameters, which regulate the amount of smoothing done in function estimation, so this should not have a strong effect on the substantive results.

As Wu and Huang mention, there are a number of practical issues that must be resolved before this model or any model for hierarchical functional data can be implemented. We highlight some of these issues here, and clarify how our approach handles them. The first is how to decide which wavelet basis functions are included in the model. This issue is crucial because this selection is the vehicle by which smoothing is introduced into the function estimation. We perform this selection in a Bayesian fashion via the mixture prior we place on the treatment-level coefficients. To us, building the selection directly into the modeling is the most natural approach and automatically accounts for the variability introduced by the selection when performing inference. An alternative approach is to derive appropriate thresholding rules, which would require further theoretical work because the single-function results in the literature do not immediately transfer to the hierarchical setting.

Another challenge is deciding how to model the covariance of the basis functions. Given the large number of basis functions present in the model for hierarchical functional data, it is not feasible to model their covariance in an unstructured fashion. Parsimonious alternatives include selecting a small number of wavelet coefficients, then assuming unstructured covariance or imposing a structure on the covariance matrix. We take the latter approach, and for us, parsimony is achieved by assuming independence among the wavelet coefficients for a given function. This assumption is commonly made in wavelet regression, and is heuristically supported by the whitening property and orthogonality of the wavelet transform. It engenders computational feasibility because it allows the wavelet-space model to be fit marginally, that is, one wavelet coefficient at a time.

We note that independence in the wavelet space does not imply independence of the observations within the function, but rather allows varying degrees of autocorrelation in the data space. Generally speaking, more heteroscedasticity across wavelet levels leads to more autocorrelation in the data space. It is possible to convert covariance matrices back and forth between the data and wavelet space, so we can see the data-space covariance matrix induced by our diagonal covariance matrix in the wavelet space. Given a wavelet-space covariance matrix Σ^* , the data-space covariance matrix is $\Sigma = W\Sigma^*W'$, where W is the DWT matrix. The results of Vannucci and Corradi (1999) can also be used to do this conversion, which amounts to performing a two-dimensional IDWT on Σ^* when Σ^* is diagonal. We consider this independence assumption a convenient and reasonably flexible working model, but feel that it would be a valuable exercise to consider other approaches relaxing this assumption.

Model fitting is another issue that needs to be addressed. How can the model be fit in a computationally feasible way, and will the model be fit using frequentist or Bayesian methods? The C++ code we have developed to perform the MCMC runs at the rate of more than 500 iterations per minute for our application, which consists of 738 functions sampled at 256 points

each, and the chain appears to converge quite rapidly. We have not investigated frequentist approaches, but it seems feasible to use an EM algorithm or Newton–Raphson to obtain maximum likelihood or REML estimates for our model, as long as the shrinkage prior is treated as part of the model. We cannot comment at this time on the computational feasibility of other approaches, but it would, no doubt, highly depend on how the selection of basis functions is done and what covariance structure is assumed on them.

In our opinion the key advantages of our Bayesian approach are inferential. As previously mentioned it naturally takes into account more sources of variability when performing inference, such as the variability in estimating the covariance parameters and the variability in selecting basis functions to include in the model. Other approaches would require first selecting sets of basis functions, then obtaining inference conditional on this set. It is possible that the selection process introduces considerable variability into the estimation procedure, so conditional inference may lead to underestimation of the estimators' variability. Second, given the posterior samples, we have a natural way of performing inference, joint or marginal, on any function of the model parameters. This inference is conditional on the sample sizes of our study, so does not rely on asymptotic assumptions. This is important for our application with only three rats (clusters) per treatment group, because any asymptotic result based on the number of clusters is not likely to hold.

There are many difficult issues in modeling and fitting hierarchical functional data, and certainly our approach is not the only one that can be taken. We thank Wu and Huang for introducing some interesting ideas and bringing to light many of these issues. We have focused our discussion on wavelet-based approaches. We have no doubt that like methods using smoother bases such as B splines would be appropriate and effective for data with smoother characteristics, but we refrain from discussing these here. As Wu and Huang noted, our choice of the wavelet basis for our application was motivated by the spiky nature of the observed crypt-level curves, which we did not feel could be modeled effectively in such a joint model using smoother bases such as splines.

3. RESPONSE TO CLYDE AND GEORGE

We thank Merlise and Ed for their kind comments and specific suggestions for choosing the regularization parameters and performing inference. We find their suggestions to be interesting and very appropriate. Here we respond to their suggestions, and also discuss some model diagnostic issues that they have raised.

They first suggest allowing the regularization parameters to differ by treatment group and choosing them automatically using either an empirical Bayes or a fully Bayesian approach. We agree that it is appropriate to allow differing degrees of smoothness in the treatment functions, especially when a data-dependent method is used for choosing the regularization parameters. They present an adaptation of their empirical Bayes method (Clyde and George 1999, 2000) to the hierarchical functional context. We applied this method to the MGMT data, obtaining MLE's for the regularization parameters for each diet–time combination; see Table 2. As they suggest, we substituted MOM estimates for the variance components. We also

Table 2. Empirical Bayes Estimates of Shrinkage Parameters for Each Diet-Time Combination

| Parameter | Diet | (hour) | Scale | Time | | | | | |
|------------|------|--------|-------|------|-------|------|------|------|------|
| | | | | 6 | 5 | 4 | 3 | 2 | j |
| p_{aj} | Fish | 0 | 1.000 | .060 | .584 | .310 | .234 | .611 | .326 |
| | Fish | 3 | 1.000 | .175 | .070 | .718 | .157 | .818 | .750 |
| | Fish | 6 | 1.000 | .041 | .975 | .151 | .525 | .118 | .292 |
| | Fish | 9 | 1.000 | .447 | .223 | .268 | .399 | .210 | .593 |
| | Fish | 12 | 1.000 | .096 | .854 | .063 | .142 | .234 | .161 |
| | Corn | 0 | 1.000 | .587 | .109 | .702 | .766 | .597 | .398 |
| | Corn | 3 | 1.000 | .865 | 1.000 | .863 | .688 | .316 | .777 |
| | Corn | 6 | 1.000 | .080 | .773 | .100 | .449 | .226 | .079 |
| | Corn | 9 | 1.000 | .087 | .472 | .702 | .589 | .474 | .185 |
| | Corn | 12 | 1.000 | .046 | .056 | .044 | .120 | .102 | .183 |
| T_{aj}^2 | Fish | 0 | 35.68 | .00 | .93 | .22 | .11 | .91 | .22 |
| | Fish | 3 | 47.22 | .07 | .00 | .78 | .00 | 1.14 | 1.00 |
| | Fish | 6 | 29.96 | .00 | 2.19 | .00 | .68 | .00 | .21 |
| | Fish | 9 | 30.94 | .27 | .08 | .15 | .49 | .08 | .47 |
| | Fish | 12 | 41.86 | .00 | 1.11 | .00 | .00 | .13 | .00 |
| | Corn | 0 | 30.37 | .54 | .00 | .87 | 1.52 | 1.27 | .32 |
| | Corn | 3 | 31.72 | .89 | 4.23 | 1.49 | 1.74 | .41 | 1.43 |
| | Corn | 6 | 30.78 | .00 | 1.04 | .00 | .42 | .19 | .00 |
| | Corn | 9 | 37.97 | .00 | .38 | .96 | .42 | .40 | .05 |
| | Corn | 12 | 17.31 | .00 | .00 | .00 | .00 | .00 | .03 |

generated plots of the posterior mean curves for each diet-time combination, given these choices of the regularization parameters (not shown), and compared them with the plots using fixed values of the regularization parameters (Fig. 3 in the main article). We saw no substantive differences between the two sets of plots, but they differed slightly in the amount of smoothing done at some of the treatment levels. The empirical Bayes method appears to give adaptive smoothing, allowing different levels of smoothness for the different treatment mean functions based on the characteristics of their data.

Not surprisingly, we find that estimates of p_j and T_j^2 are larger for treatment groups with mean curves that demonstrate more local features. For example, the unsmoothed mean curve for the corn oil, 3-hour group contains several large bumps. The empirical Bayes estimates for many of the p_j are large for this treatment, and, as a result, the bumps are largely retained in the smoothed estimate. On the other hand, no local features are evident in the corn oil, 12-hour group, and $p_j \approx 0$ for all wavelet levels. This smooth function can be well represented using just the scaling coefficients. We also explored a fully Bayesian method with diffuse priors for the regularization parameters and obtained similar results.

Note that the empirical Bayes estimates of T_j^2 were very small for all wavelet levels. From our discussion in Section 4.2 of the article, this means that all wavelet coefficients experience a large degree of shrinkage, even those that are large in magnitude. This linear shrinkage results in attenuation of local features. This may be the “right thing” to do here, because it appears that the underlying mean functions are quite smooth, and maybe we do not want to believe that local features in the curves are real. However, in a case where the true mean function contains local features, this type of attenuation may be a problem.

Clyde and George also suggested computing Bayes factors for statistical inference on the treatment functions. This is an excellent suggestion because, as they note, the tractability of the marginal distributions of our model when the variance components are known opens up the possibility of computing the

Bayes factors for hypotheses of interest without too much trouble.

They also bring up the issue of model misspecification, and ask if any diagnostic is available to check whether the independence assumption in the wavelet space is reasonable. One quick and easy way to investigate this assumption is to plot the method of moments estimates of the correlation matrices at the various hierarchical levels in the data and wavelet spaces and see whether the DWT appears to sufficiently decorrelate the data. A wavelet-space correlation matrix that is close to diagonal indicates that the independence assumption is reasonable. For example, consider the crypt-level covariance matrix Σ_1 , with elements $\Sigma_{1,(i,i')} = \text{cov}\{Y_{abc}(t_i), Y_{abc}(t_{i'})\}$. An easily computed method of moments estimator for this matrix is given by $\widehat{\Sigma}_{1,\text{MOM}}$, with the (i, i') element given by $\{\sum_{a,b} (C_{ab} - 1)\}^{-1} \sum_{a,b,c} \{Y_{abc}(t_i) - \bar{Y}_{ab \cdot}(t_i)\} \{Y_{abc}(t_{i'}) - \bar{Y}_{ab \cdot}(t_{i'})\}$. The corresponding correlation matrix can be found by multiplying each element $\widehat{\Sigma}_{1,\text{MOM},(i,i')}$ by $\{\widehat{\Sigma}_{1,\text{MOM},(i,i)} \widehat{\Sigma}_{1,\text{MOM},(i',i')}\}^{-1/2}$. Correlation matrices at the other hierarchical levels can be estimated likewise. The corresponding wavelet correlation matrices can be computed using the same formulas, with the double subscript (j, k) substituted for i .

Figure 3 contains image plots corresponding to the MOM estimates of the crypt- and rat-level correlation surfaces in both the data space and the wavelet space. The plot is gray scaled such that a correlation of +1 is white, -1 is black, and 0 is gray. Note that the crypt-level correlation matrix in the data space is close to diagonal, indicating weak autocorrelation between measurements within a given crypt. The corresponding wavelet-space correlation matrix shows evidence of being decorrelated. There is a much stronger degree of autocorrelation evident in the rat-level curves, which is largely eliminated when moving to the wavelet space. With our choice of wavelet basis, the independence assumption appears to be a reasonable working model for these data. Because different wavelet bases will lead to different decorrelation properties, these plots can be used to select a wavelet basis that accommodates the independence assumption for a given dataset. They can also be used to

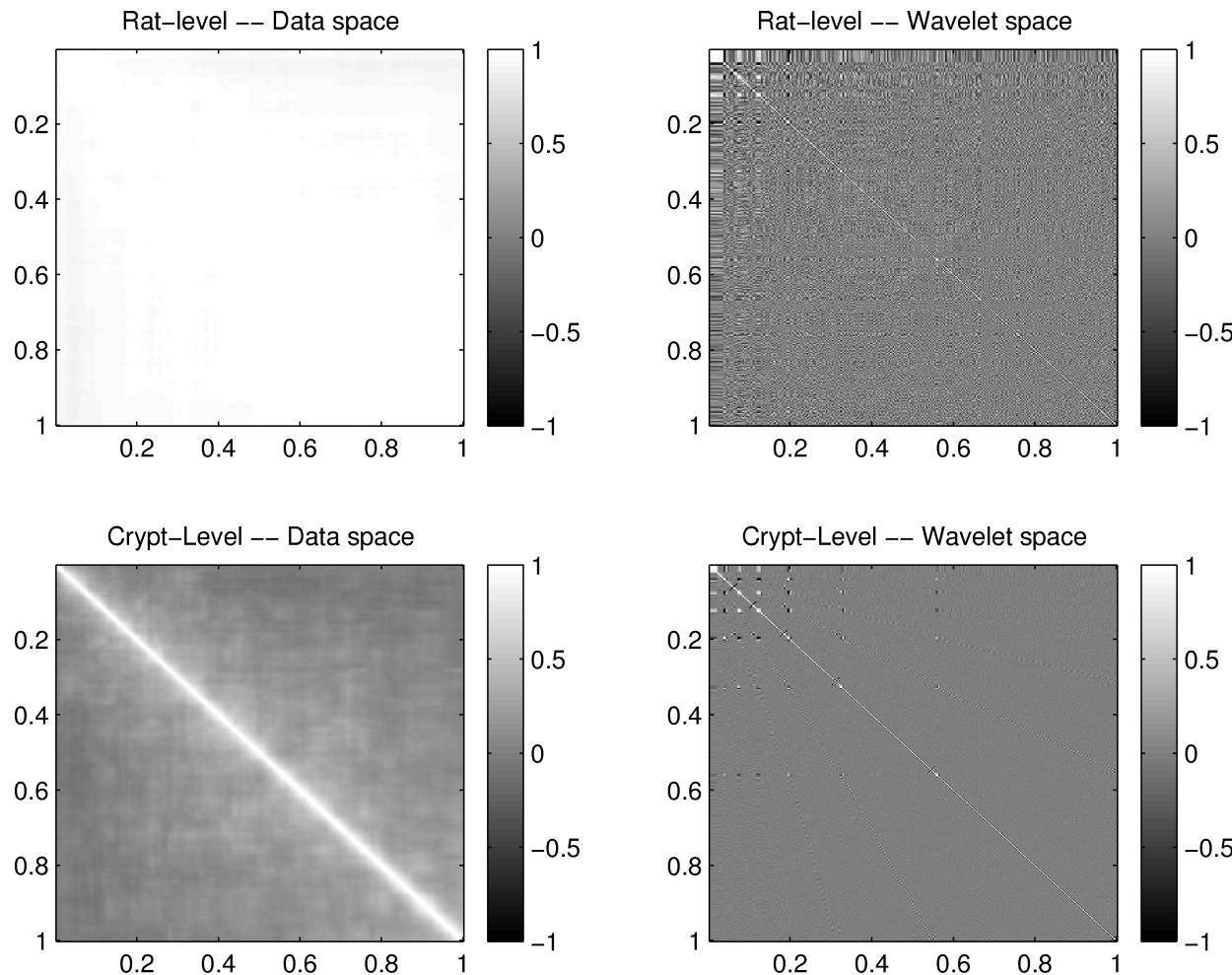


Figure 3. Method of Moments Estimates of Within-Curve Correlation Matrices in Data and Wavelet Spaces at the Rat and Crypt Hierarchical Levels. Note the decorrelation in the wavelet space compared to the data space.

suggest alternative covariance structures for the wavelet coefficients if one does not wish to assume independence.

Note that these correlation matrices reflect the characteristics of the functions at the various hierarchical levels. The fact that the crypt-level data are inherently local, characterized by discrete peaks, is reflected in the low degree of autocorrelation present. The rat-level curves, being relatively smooth and flat (see Fig. 2), demonstrate a large degree of autocorrelation. If the functions were perfectly flat, each entry in the correlation matrix would be 1.

Clyde and George also raise a concern that the variability in cell size and position in the crypt may lead to some oversmoothing. Morris et al. (2002) demonstrated that the distal colon cells are not uniformly distributed along the crypt wall, and the size, position, or both of the cells varies from crypt to crypt. We do not expect this misalignment problem to prevent us from picking out key features and treatment effects in the data, but it is a potential source of attenuation in the function estimation, and so it may be worthwhile to consider methods for aligning the functions. Morris et al. (2002) modeled the distribution of relative cell depths within a crypt as the order statistics of a Beta, and they built a measurement error step into their MCMC to model outcomes as functions of the relative crypt depth when cell position measurements were unavailable. It may be possible to

add a similar step to adjust for the misalignment problem here, but there are issues that would need to be worked out. Repeated transformation of the functional domain would lead to computational difficulties in our framework, because the DWT's for each crypt would need to be recomputed each time it is rescaled. Additionally, assay limitations in the MGMT data prevent us from actually seeing the cell delineations, so any attempt at re-alignment would involve inferences on what constitutes a cell, introducing considerable error.

[Received XXXX. Revised XXXX.]

ADDITIONAL REFERENCE

- Abramovich, F., Sapatinas, T., and Silverman, B. W. (1998), "Wavelet Thresholding via a Bayesian Approach," *Journal of the Royal Statistical Society, Ser. B*, 60, 725–749.
- Clyde, M., and George, E. I. (1999), "Empirical Bayes Estimation in Wavelet Nonparametric Regression," in *Bayesian Inference in Wavelet Based Models*, eds. P. Muller and B. Vidakovic, New York: Springer-Verlag, pp. 309–322.
- (2000), "Flexible Empirical Bayes Estimation for Wavelets," *Journal of the Royal Statistical Society, Ser. B*, 60, 681–698.
- Morris, J. S., Wang, N., Lupton, J. R., Chapkin, R. S., Turner, N. D., Hong, M. Y., and Carroll, R. J. (2002), "A Bayesian Analysis of Colonic Crypt Structure and Coordinated Response to Carcinogen Exposure Incorporating Missing Crypts," *Biostatistics*, 3, 529–546.
- Vannucci, M., and Corradi, F. (1999), "Covariance Structure of Wavelet Coefficients: Theory and Models in a Bayesian Perspective," *Journal of the Royal Statistical Society, Ser. B*, 61, 971–986.