

# Supplementary Material for Morris and Carroll (2005)

This document contains supplementary material for the following paper:

Morris, JS and Carroll, RJ (2005). Wavelet-Based Functional Mixed Models.

The supplementary material includes:

1. Description of Metropolis-Hastings Method for Covariance Parameters
2. Demonstration of Adaptive Regularization of Random Effect Functions
3. Trace Plots from the MCMC

## 1 Metropolis-Hastings Method for Covariance Parameters

Here we outline the details of the Metropolis-Hastings procedure we use to update the covariance parameters  $\boldsymbol{\Omega}_P, \boldsymbol{\Omega}_R, \boldsymbol{\Omega}_Q$ , and  $\boldsymbol{\Omega}_S$ .

First, the covariance parameters  $q_{jk}^*$  and  $s_{jk}^*$  of  $\boldsymbol{\Omega}_Q$  and  $\boldsymbol{\Omega}_S$  are updated as a block for each  $j = 1, \dots, J$  and  $k = 1, \dots, K_j$ , as follows. The objective function is

$$f(q_{jk}^*, s_{jk}^* | \mathbf{d}_{jk}, \boldsymbol{\beta}_{jk}^*, \boldsymbol{\Omega}_P, \boldsymbol{\Omega}_R) \propto |\Sigma_{jk}|^{-1/2} [\exp\{-1/2(\mathbf{d}_{jk} - X\boldsymbol{\beta}_{jk}^*)' \Sigma_{jk}^{-1} (\mathbf{d}_{jk} - X\boldsymbol{\beta}_{jk}^*)\}] f(q_{jk}^*, s_{jk}^*), \quad (1)$$

where recall  $\Sigma_{jk} = ZP(\boldsymbol{\Omega}_P)Z' * q_{jk}^* + R(\boldsymbol{\Omega}_R) * s_{jk}^*$ . If these parameters are only indexed by  $j$  (see Section ??), then this objective function takes the product of these terms over  $k = 1, \dots, K_j$ .

Candidate values  $(q1)_{jk}$  and  $(s1)_{jk}$  are sampled from a independent truncated normal proposal densities centered at the previous values  $(q0)_{jk}$  and  $(s0)_{jk}$  with proposal variances  $(vq)_{jk}$  and  $(vs)_{jk}$ , respectively. The density is truncated at 0 to ensure positivity of the variance components. Thus, the proposal density is given by

$$p\{(q1)_{jk}, (s1)_{jk} | (q0)_{jk}, (s0)_{jk}\} \propto \Phi\{-(q0)_{jk} / \sqrt{(vq)_{jk}}\} \{(vq)_{jk}^{-1/2}\} \exp[-\{(q1)_{jk} - (q0)_{jk}\}^2 / 2(vq)_{jk}] * \Phi\{-(s0)_{jk} / \sqrt{(vs)_{jk}}\} \{(vs)_{jk}^{-1/2}\} \exp[-\{(s1)_{jk} - (s0)_{jk}\}^2 / 2(vs)_{jk}],$$

where  $\Phi(x)$  is the normal cdf evaluated at  $x$ .

The sampling proceeds as follows for each  $j$  and  $k$  at each iteration of the MCMC.

1. Generate proposal values  $(q1)_{jk}$  and  $(s1)_{jk}$  from  $\text{TrN}\{(q0)_{jk}, (vq)_{jk}\}$  and  $\text{TrN}\{(s0)_{jk}, (vs)_{jk}\}$ , respectively, where  $\text{TrN}$  refers to a normal truncated at 0 and  $(q0)_{jk}$  and  $(s0)_{jk}$  are the values of  $q_{jk}^*$  and  $s_{jk}^*$  from the previous iteration.
2. Compute  $a_{jk}$ , given by

$$a_{jk} = \frac{f((q1)_{jk}, (s1)_{jk} | \mathbf{d}_{jk}, \boldsymbol{\beta}_{jk}^*, \boldsymbol{\Omega}_P, \boldsymbol{\Omega}_R) p((q0)_{jk}, (s0)_{jk} | (q1)_{jk}, (s1)_{jk})}{f((q0)_{jk}, (s0)_{jk} | \mathbf{d}_{jk}, \boldsymbol{\beta}_{jk}^*, \boldsymbol{\Omega}_P, \boldsymbol{\Omega}_R) p((q1)_{jk}, (s1)_{jk} | (q0)_{jk}, (s0)_{jk})}$$

3. Generate  $u$  from a  $U(0,1)$ , if  $u < a_{jk}$ , then let  $q_{jk}^* = (q1)_{jk}$  and  $s_{jk}^* = (s1)_{jk}$  for the current iteration, otherwise  $q_{jk}^* = (q0)_{jk}$  and  $s_{jk}^* = (s0)_{jk}$ .

It is important to carefully choose the proposal variances  $(vq)_{jk}$  and  $(vs)_{jk}$  since if they are too large, the acceptance probabilities will be low and the sampler inefficient, while if too small, the acceptance probabilities will be too high and the sampler may not adequately explore the space of the objective function. Often trial and error is used to come up with proposal variances with good properties. This is infeasible here because of the very large number of parameters, so we instead determine them automatically from the data. We compute rough estimates of the asymptotic variances for the maximum likelihood estimators of  $q_{jk}^*$  and  $s_{jk}^*$ , conditional on initial estimates of  $\beta_{jk}^*$ . Since the amount of variability in estimating the MLE should be a reasonable ballpark estimate for the variance of the objective function given weak prior information, this should be a good starting point. In order to err on one side of taking slightly larger jumps, we multiplied these estimates by a factor of 1.5 to get the proposal variances. We found that this strategy led to reasonable acceptance probabilities for all of our motivating examples.

Similar steps can be followed to update the parameters contained in  $\Omega_P$  and  $\Omega_R$ .

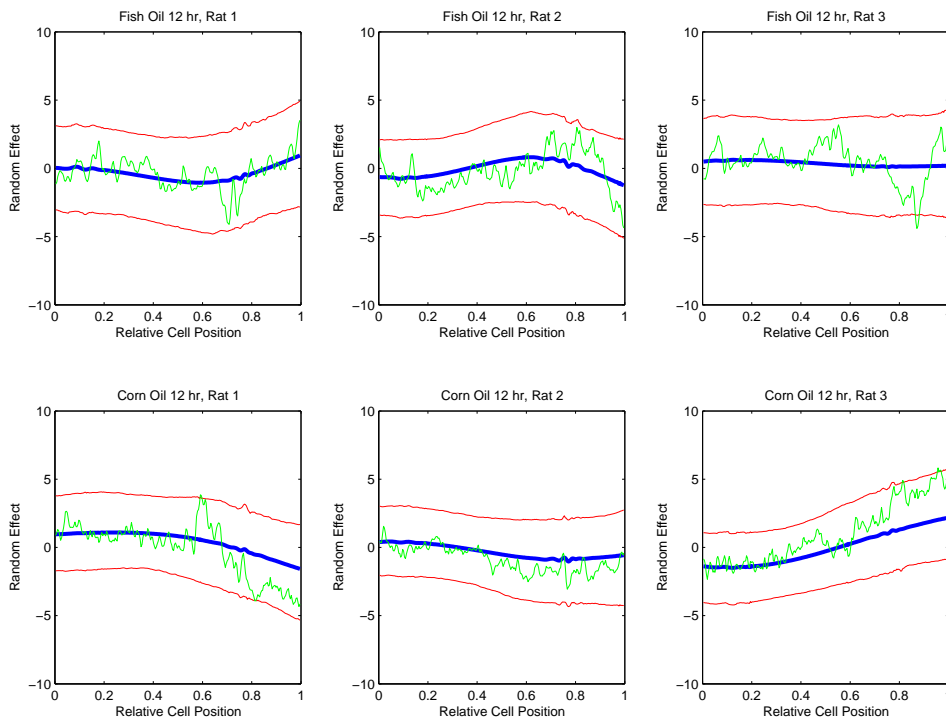


Figure 1: *Estimation of Rat-Level Random Effect Functions*: Plots of the rat-level random effect functions for the rats sacrificed at the 12 hour time point in the example. The green line is the raw (non-regularized) estimate of the random effect function, the blue line is the regularized posterior mean estimate from our method, and the red lines are the 90% posterior credible intervals.

## 2 Adaptive Regularization of Random Effect Functions

In Section 4.3 of Morris and Carroll (2005), we claim that our framework is adaptive enough to handle spikiness in the fixed or random effect functions. Here we illustrate this point. In the example presented in the paper, the fixed effect functions and rat-level random effects were both quite smooth, but the crypt-level residual deviation functions were very spiky. Figure 1 contains a plot of the posterior mean and 90% pointwise posterior credible bands for the rat-level random effects at the 12 hour time point. Note how the wavelet-based method is able to accommodate smooth random effects when suggested by the data. Also note how our estimation procedure is not simply "smoothing" the individual random effect functions, as is commonly done in other methods, but rather shrinks the individual functions towards the group mean function as one would expect for random effects. This is most evident for the corn oil rats near the top of the crypt (relative cell position 1).

To illustrate that this framework is adaptive enough to also handle spiky random effects, consider an example in mass spectrometry proteomics. This application yields functional data characterized by many spikes roughly corresponding to versions of proteins present in the bio-

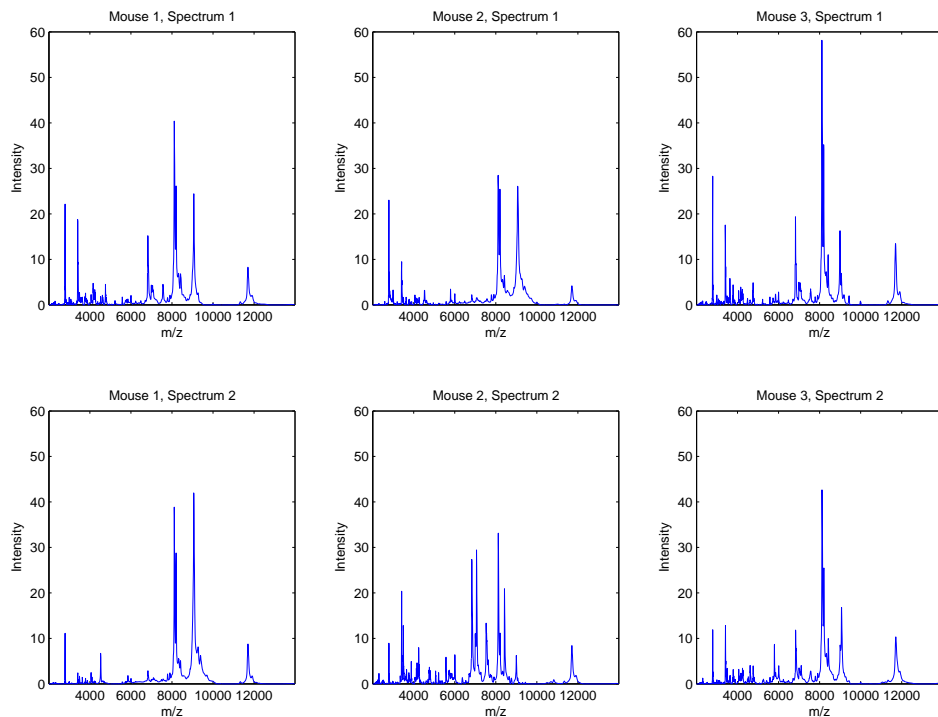


Figure 2: *SELDI mass spectra*: Plot of SELDI spectra for first 3 mice, with the top obtained using the low laser intensity scan and the bottom using the high laser intensity scan.

logical sample. Figure 2 contains a plot of six spectra from a study performed at MD Anderson Cancer Center. These data contain spectra from 16 mice, each injected with one of two cell lines into one of two organs. For each mouse, spectra were obtained from a SELDI mass spectrometry instrument with two laser intensity settings, low and high, yielding a total of 32 spectra. Figure 2 contains the two spectra from the first 3 mice. This example is not contained in the Morris and Carroll (2005) paper, but is included here to illustrate this point. We applied our wavelet-based functional mixed model to these data with 5 fixed effect functions: an overall mean, cell-line main effect function, organ main effect function, cell-line by organ interaction function, and laser intensity effect function, plus 16 independent random effect functions, one per mouse.

Figure 3 contains a plot of the posterior mean for the latter four fixed effect functions along with 90% pointwise posterior credible bands. Note how these fixed effect functions are very spiky, which is natural since we would expect that any effects would be specific to certain proteins, yielding spiky effect functions. In this example, the mouse-level random effect functions are also very spiky. Figure 4 contains a plot of posterior means for the random effect functions for the first 3 mice, as well as the corresponding mouse-specific fitted mean functions (posterior predictive mean functions). Note how the method is able to accommodate spiky random effect functions when the data suggests it is necessary.

The reason for the ability of this approach to handle the spikiness is that, unlike any other

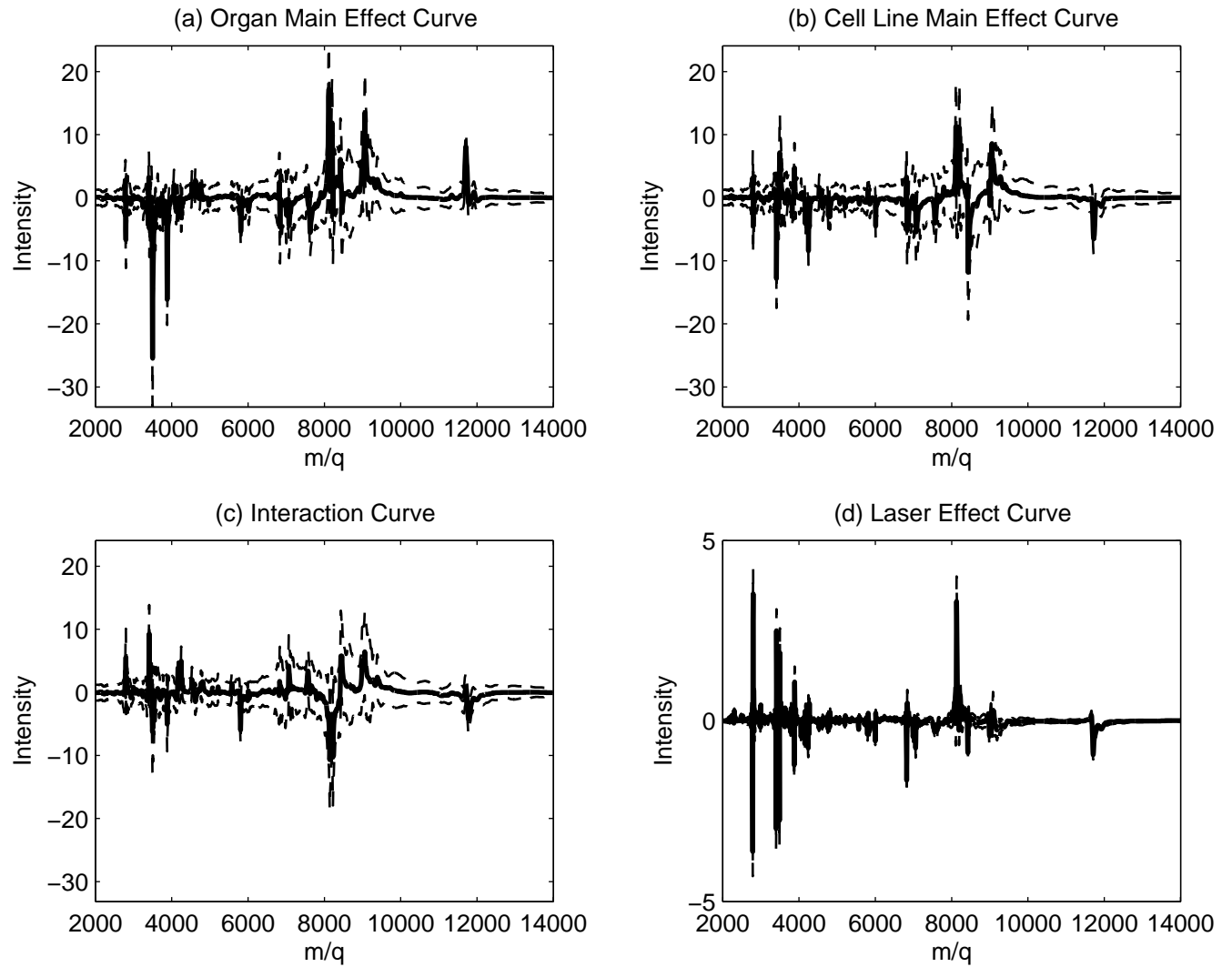


Figure 3: *Fixed Effect Functions*: Posterior mean (regularized) estimates of fixed effect estimates, with 90% pointwise posterior credible bands.

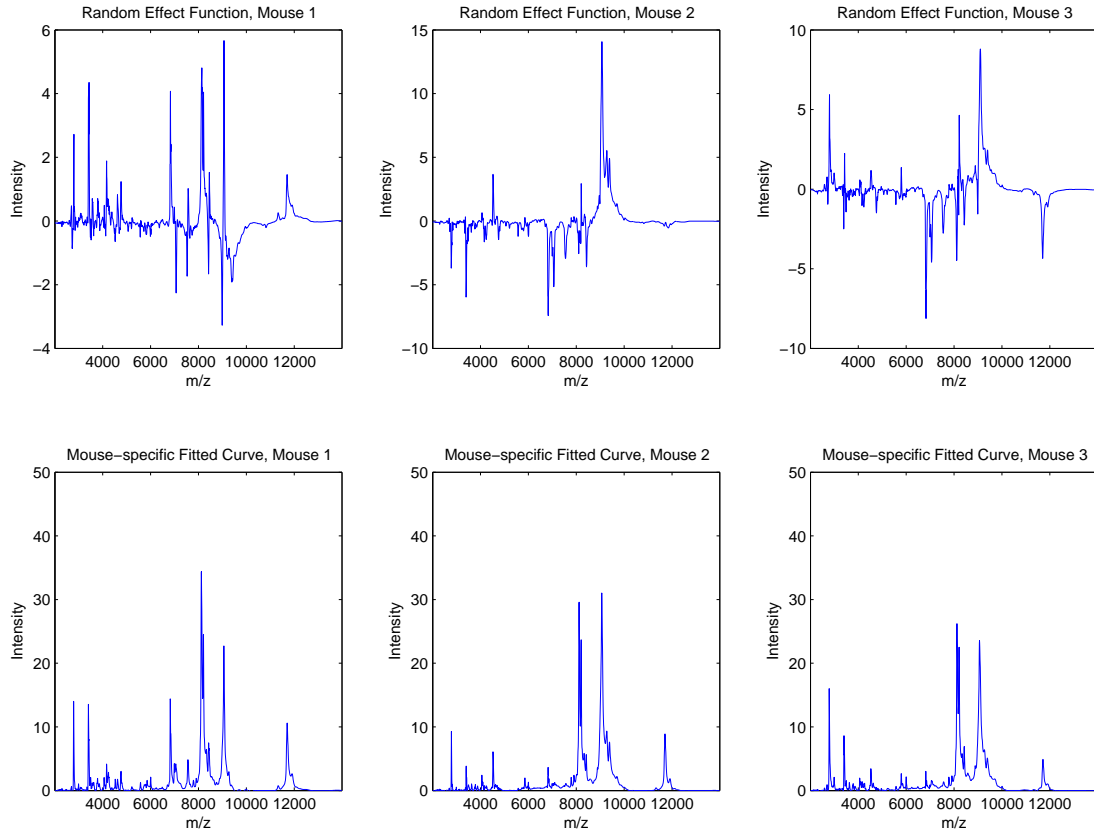


Figure 4: *Mouse random effect functions*: Posterior mean (regularized) estimates of mouse random effect functions for first 3 mice (top), and corresponding mouse-specific fitted mean functions (posterior predictive mean functions, bottom).

applications of wavelet regression in existing literature, our method allows *separate variance components for each wavelet coefficient*. That is, the diagonal elements of  $Q^*$  are all allowed to be different, across both scales  $j$  and locations  $k$ , as indicated by the notation  $q_{jk}$ . These variance components serve dual interrelated purposes: to model the curve-to-curve variability in that wavelet coefficient, and to serve as a regularization parameter for the random effect functions. Thus, in our modeling we effectively have *separate regularization parameters for each wavelet coefficient*. As described below, this leads to adaptive smoothing of these wavelet coefficients. This phenomenon is unique to the multiple function case, since it is not feasible in the single function nonparametric regression setting to have separate smoothing parameters for each coefficient. It is possible in the replicated functional setting because we are able to borrow strength across functions. Because these smoothing parameters are also variance components, it is straightforward to estimate them from the data, and even better than that, our Bayesian approach automatically takes their uncertainty into account when performing inference.

The fact that we estimate these regularization parameters for the random effect functions from the data allows the regularization to adapt to the features of the data. Here is a brief description of how this works. Wavelet coefficients that tend to be important for representing even a small number of random effect functions will have relatively large subject-level variance components. These large variances will lead to less shrinkage of these coefficients, and thus the features represented by these wavelet coefficients will tend to be preserved in the regularized random effect function estimates. Wavelet coefficients that are unimportant for representing the random effect functions will be close to zero, leading to a small variance, strong shrinkage, and a regularization of these features. In the mass spectrometry example, the wavelet coefficients corresponding to the strong spikes in the random effect functions will have large variance components, while wavelet coefficients corresponding to regions that are flat spots or very small bumps will be strongly shrunken and thus regularized.

Heuristically, wavelet shrinkage leads to "adaptive regularization" because it is *selective* in how it shrinks the wavelet coefficients towards zero – smaller coefficients are more strongly shrunken than the larger ones. In the classical single function wavelet regression setting and for our fixed effect functions (for which we have no replicate curves we are willing to assume come from a common distribution), this selective shrinkage is accomplished by thresholding or application of the spike-slab prior (i.e. the mixture of point mass at zero and Gaussian). For our random effect functions, it is accomplished a different way by having different variances for each wavelet coefficient, but the effect is similar, making it able to handle spiky random effect functions as evidenced by Figure 2.

Note that global shrinkage and non-adaptive estimation would result from a Gaussian prior on wavelet coefficients if a common variance was used for all wavelet coefficients, or even for all wavelet coefficients within a given resolution level  $j$  (which is what almost all other wavelet regression methods do by only allowing the variance components to vary by  $j$  but not location  $k$ ). However, by allowing different variances by  $j$  and  $k$ , we are able to handle the spiky random effects quite well. Again, since Gaussian shrinkage is regulated by the variance, allowing different variances for each wavelet coefficient allows adaptive shrinkage of the wavelet coefficients towards zero, which tends to preserve features corresponding to strong spikes.

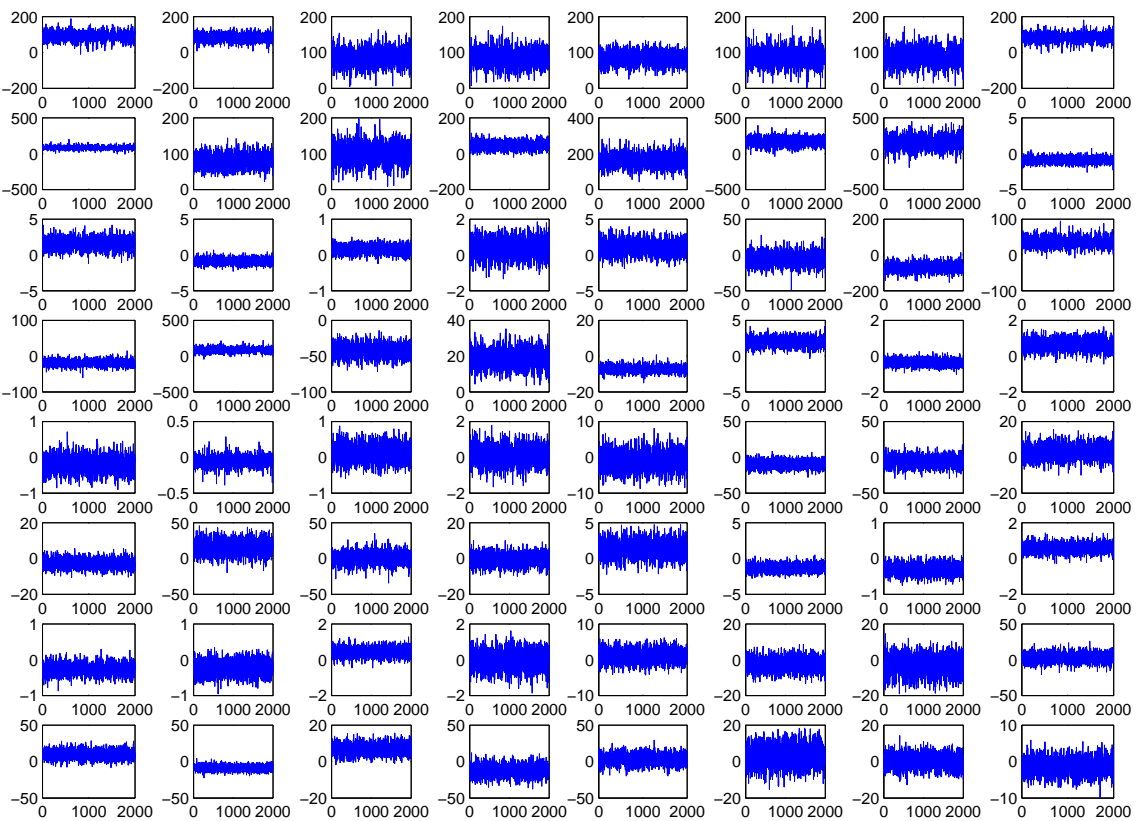


Figure 5: *Trace plot of subset of wavelet coefficients for 1<sup>st</sup> fixed effect function.* Trace plots of 64 wavelet coefficients for the first fixed effect function, the mean MGMT profiles for rats in the fish oil, time 0 group. There are over 3000 wavelet coefficients for the 14 fixed effect functions altogether. These are the 64 lowest frequency coefficients, which account for a vast majority of the variability in the functions (99.79% of the total energy). These trace plots suggest the chain has converged and is mixing nicely.

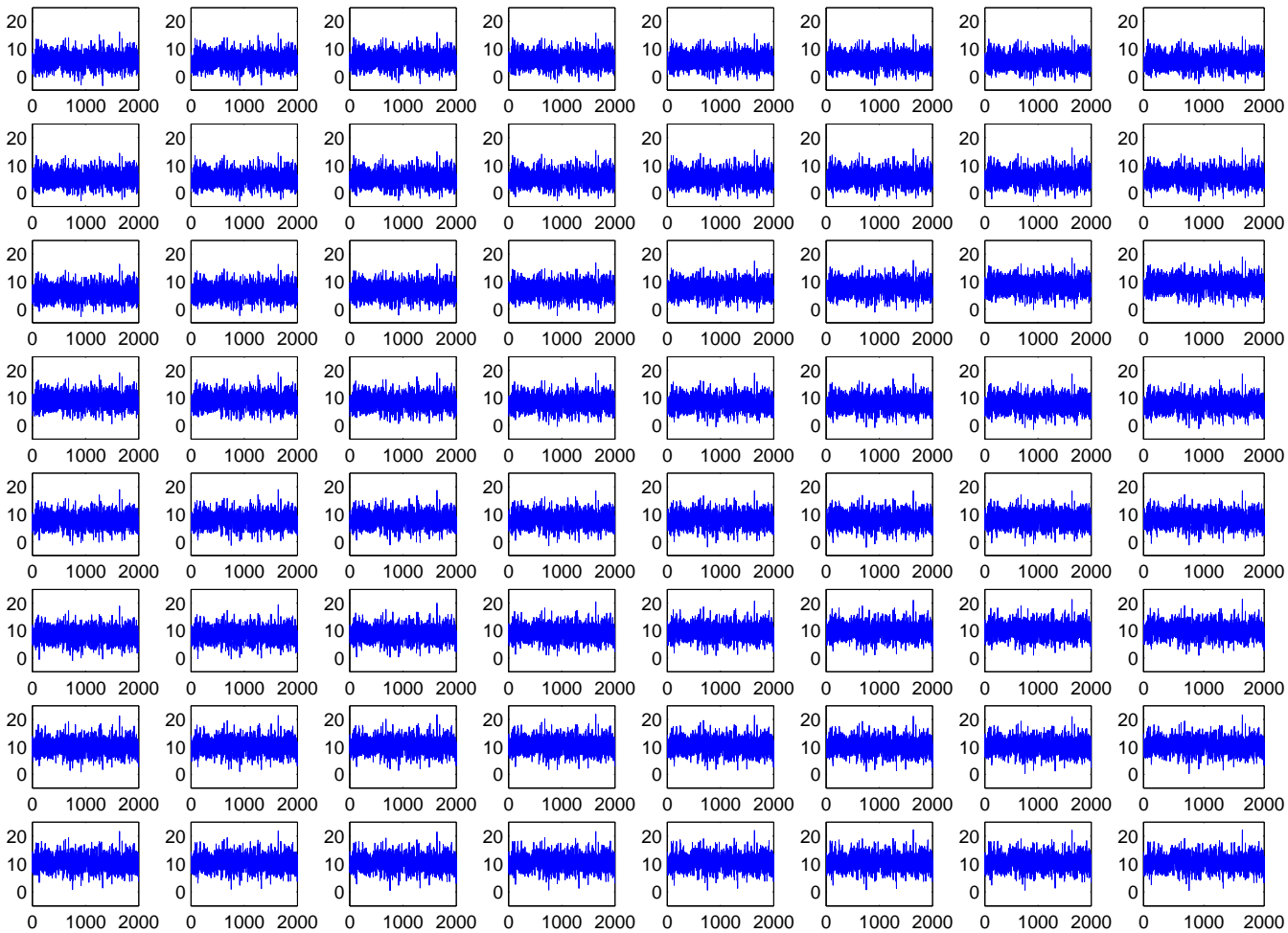


Figure 6: *Trace plot of 1<sup>st</sup> fixed effect function at each of 64 grid points on the interval  $(0, 0.25)$ .*  
 These trace plots suggest again, that the chain has converged and is mixing nicely.

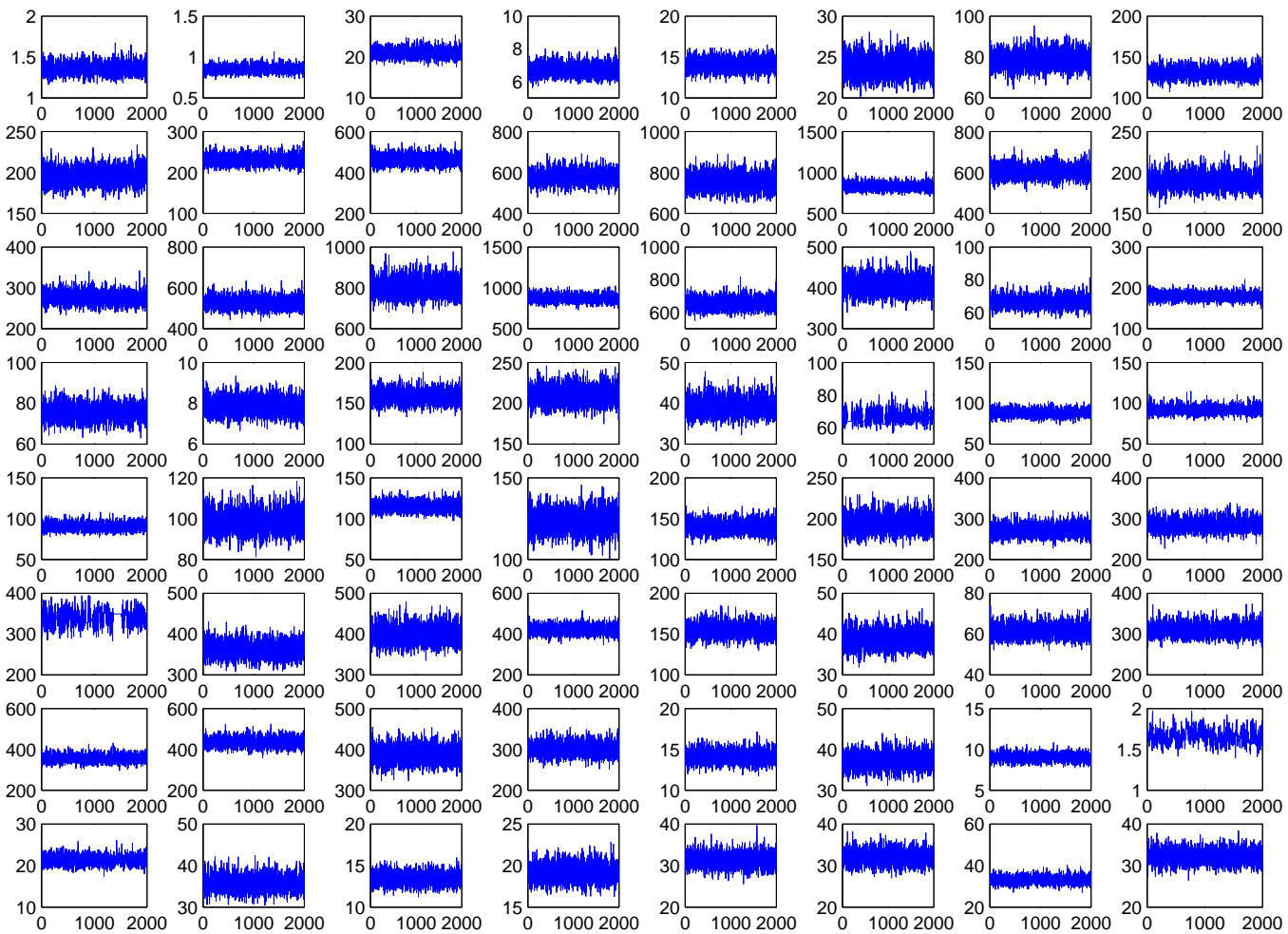


Figure 7: *Trace plot of subset of variance components.* Trace plot of the 64 lowest frequency curve-to-curve wavelet space variance components in the model. These 64 coefficients account for 99.65% of the total energy in terms of the curve-to-curve variability. This plot contains just 64 of the over 500 variance components in the model. These trace plots suggest the chain has converged and is mixing nicely.