



**From the SelectedWorks of Xiaofeng Wang**

---

June 2007

# Modeling Heterogeneity and Dependence for Analysis of Neuronal Data

Contact  
Author

Start Your Own  
SelectedWorks

Notify Me  
of New Work

---

Available at: <http://works.bepress.com/wang/13>

## Modeling heterogeneity and dependence for analysis of neuronal data

Xiaofeng Wang<sup>1,2,\*</sup>, Jiayang Sun<sup>3</sup>, Kenneth J. Gustafson<sup>4,5</sup> and Guang H. Yue<sup>6</sup>

<sup>1</sup>*Department of Quantitative Health Sciences, The Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195, U.S.A.*

<sup>2</sup>*School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, U.S.A.*

<sup>3</sup>*Department of Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, U.S.A.*

<sup>4</sup>*Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, U.S.A.*

<sup>5</sup>*Louis Stokes Cleveland VA Medical Center, 10701 East Boulevard, Cleveland, OH 44106, U.S.A.*

<sup>6</sup>*Department of Biomedical Engineering, The Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195, U.S.A.*

### SUMMARY

In this paper, we describe two types of neuroscience problems which challenge the typical statistical models assumed for analyzing neuronal data. This offers an opportunity for new modeling and statistical inference. In the first problem, the data are spatial neural counts which are often over-dispersed and spatially correlated so that a standard Poisson regression model is inadequate. In the second problem, the data are averaged electroencephalograph signals recorded during muscle fatigue, where a time series AR(1) regression model cannot fully capture all the variation and correlation structure in the data. It is shown that an additional parameter has to be included in the modeling of the correlation structure and that the role of the parameter differs from one channel to the other. We propose appropriate generalized models for these data, develop statistical procedures under the generalized models, and apply these procedures to the real data that motivated this paper. The effect of mis-specification of a correlation structure is also investigated. Copyright © 2007 John Wiley & Sons, Ltd.

**KEY WORDS:** spatial count data; EEG/EMG; neuroscience; negative-binomial model; heterogeneity; overdispersion; multiple testing

\*Correspondence to: Xiaofeng Wang, The Cleveland Clinic, 9500 Euclid Avenue/Wb4, Cleveland, OH 44195, U.S.A.

†E-mail: wangx6@ccf.org

Contract/grant sponsor: National Institutes of Health; contract/grant numbers: NS37400, HD40298

Contract/grant sponsor: Department of Defense; contract/grant number: DAMD17-01-1-0665

Contract/grant sponsor: National Science Foundation; contract/grant number: DMS-0308609

Contract/grant sponsor: Department of Veterans Affairs; contract/grant number: VA RR&D B3675R

## 1. INTRODUCTION

Neuroscience is a wide field of study that deals with the structure, function, and development of the nervous system. Advances in neuroscience tools and techniques have made it possible for researchers to examine the structure and function of the brain and spinal cord in greater detail than ever before. These tools and techniques include, for example, functional magnetic resonance imaging (fMRI), computerized electroencephalograph (EEG), and anatomical labeling techniques such as immediate early gene expression. These modern techniques may involve sophisticated instrumentation and equipment, which employ high-speed electronics and computers for data collection. The ever increasing amount of acquired neuroscience data requires improved statistical models and tools to extract information from neuroscience data in an efficient manner and to advance the efficacy of neuroscience-related studies.

In this paper, we will show two types of neuroscience problems which challenge the typical statistical models assumed for analyzing neuronal data, one with a discrete response variable and the other with a continuous response variable.

In problem one, we explore the three-dimensional topography of neurons in the spinal cord activated by reflex activity. The flexion withdrawal reflex automatically retracts the limb in response to noxious stimuli. This withdrawal reflex has been widely studied to better understand the organization of spinal motor systems [1]. Identification of neurons involved in specific behaviors, such as the flexion withdrawal reflex, is important to the understanding of the spatial organization of reflex systems. A variety of techniques may be used to label individual neurons, which may then be identified using microscopy. Currently, data presentation and analysis are most used often for two-dimensional qualitative comparisons. Statistical tools able to better identify anatomical compartments containing increased or decreased neural activity would substantially improve these studies. The statistical challenge in this type of study is how to model and analyze spatial count data with overdispersion and spatial correlation. The basis for the improved statistical analyses presented here was used to improve the analysis and presentation of the topography of spinal neurons [2].

In problem two, neuroscientists are concerned about brain activation during muscle fatigue. Increased fatigability occurs in every patient with muscle weakness, regardless of whether the weakness is due to a central or a peripheral neurological disorder [3]. The underlying mechanisms are not well understood, and there is a need to study fatigability systematically in neurology and rehabilitation. The behavior of the peripheral neuromuscular system during muscle fatigue has been studied extensively, but the role of the central nervous system in muscle fatigue is largely unknown. Without a good understanding of mechanisms of fatigue in health, an assessment of the mechanisms contributing to increased fatigability in neurological disorders is difficult. We investigated changes in brain activity during motor performance from non-fatigued to moderately fatigued to severely fatigued conditions in healthy volunteers. Handgrip force, surface electromyographic (EMG) signals of finger flexor and extensor muscles, and scalp electroencephalograms (EEG) were measured simultaneously [4]. The statistical challenge in this study is how to model multichannel force/EMG/EEG signals with appropriate covariance.

This paper is organized as follows. In Section 2, we describe the experiment background and data of the 'spinal neurons' problem, propose a spatial generalized linear model, and develop an algorithm for fitting the data under this generalized model and a test for testing the hypotheses of competing models. In Section 3, we present the experiment background and data of the 'muscle fatigue' problem and then explore adequate covariance structures for modeling such data in a

time-series regression analysis. The effect of mis-specification of the covariance structure in each of these two cases is investigated in the last subsections of Sections 2 and 3, respectively. In Section 4, we discuss the wider potential of the model-based approach for analyzing data in neuroscience.

## 2. SPATIAL MODEL FOR NEURAL COUNT DATA

### 2.1. Experiment and data

These experiments consisted of evoking a reflex in each of 10 cats, and then examining sections of the spinal cord to identify the location of neurons involved in this reflex [2]. The hindlimb flexion withdrawal reflex was evoked *via* electrical stimulation of two hindlimb nerves ( $n = 4$  animals at the tibial nerve and  $n = 4$  animals at the superficial peroneal nerve). Two animals were operated with unstimulated controls. The spinal cords were sectioned. Active spinal neurons were identified using immunocytochemical labeling of FOS, the product of the immediate early gene *c-FOS*, and examination of the serial cross-sections of the spinal cord. Table I contains the main variables and experimental design used in obtaining the relevant data from 10 animals. (A more detailed description of these experiments is reported separately in [2]. These main variables in Table I are sufficient to set a basis for our statistical development in the next subsections.)

*2.1.1. Three-dimensional anatomy.* The pertinent region of the spinal cord was divided into 240 anatomical compartments defined by three dimensions: 12 rostral/caudal sub-segments ( $Z$ ; along the spinal cord), 10 lamina ( $Y$ ; approximately top to bottom); and 2 sides ( $X$ ; right side (ipsilateral, stimulated) and left side (contralateral)). Anatomical views of the spinal cord are shown in Figure 1. Over 12 800 labeled neurons were categorized into these anatomical compartments.

The major goals of our statistical modeling and analyses for this problem are to: (1) determine significant experimental and analytical factors that affect the number of identified neurons; (2) determine spatial compartments where the number of labeled neurons is significantly greater in stimulated (treated) than in control animals; and (3) determine the relative significance (i.e.  $p$ -value) of the significant compartments.

### 2.2. Spatial models

The labeled cell data can be considered as spatial *lattice* count data as defined by Cressie [5], where the lattice data are observations from a random process observed over a countable collection of spatial regions. The count of active neurons at a particular location represents the count in the entire region indexed by that location.

Table I. Data variables.

|                                     |   |
|-------------------------------------|---|
| Animal                              | 1–10  |
| Nerve (treatment)                   | 1 = tibial, 2 = superficial peroneal, 3 = control |
| Anatomical location $s = (X, Y, Z)$ |   |
| Side ( $X$ )                        | 1 = right side (stimulated), 2 = left side        |
| Lamina ( $Y$ )                      | 1–10  |
| Longitudinal section ( $Z$ )        | 1–12  |
| Neurons (outcome)                   | Number of labeled neurons in the compartment      |

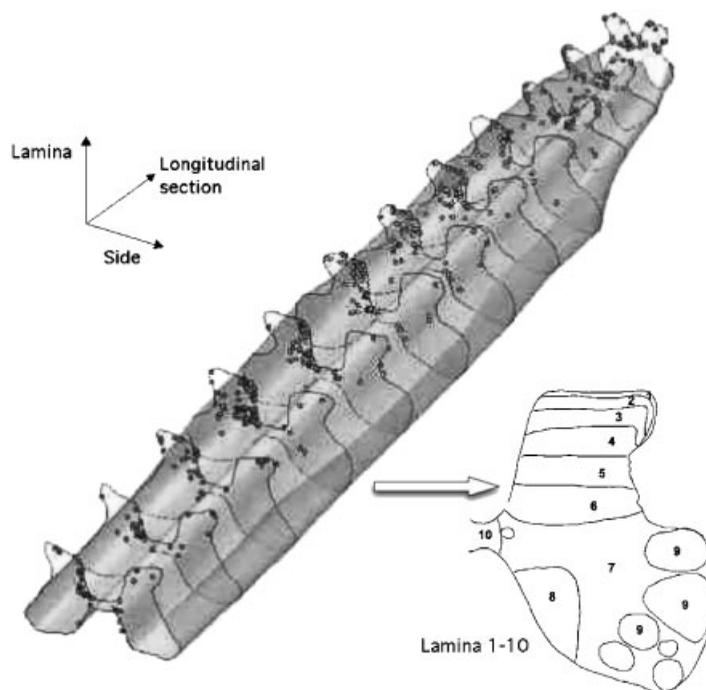


Figure 1. The three-dimensional anatomical views of the spinal cord and dimension definitions (see Table I). The figure is created using data adapted from [2].

As a natural guess, a reasonable model that fits the labeled cell data might be the Poisson regression model, in which the counts of labeled neurons are assumed to distribute independently according to the Poisson distributions, with means  $\mu = \mu(s)$  that vary according to spatial locations  $s$ . This Poisson regression model is a generalized linear model [6] with a link function  $g(\mu) = \log(\mu)$ , where

$$g(\mu) = \sum_{j=1}^q x_j \beta_j \quad (1)$$

The  $x$ 's are explanatory variables (including the spatial location  $s$  and treatment as shown in Table I) in equation (1), and  $\beta$ 's are unknown regression parameters to be estimated from the data. In particular, for the labeled cell data that are obtained on a regular grid (by 'section' and 'lamina'), a general decomposition might be

$$g(\mu_{klrt}) = \lambda + \text{section}_k + \text{lamina}_l + \text{nerve}_r + \text{side}_t$$

where {section} and {lamina} are fixed row and column effects,  $\lambda$  is the a fixed overall effect, and {nerve} and {side} represent fixed 'nerve' and 'side' effects as shown in Table I. However, our analysis under the Poisson regression model shows that there is a severe overdispersion; the Pearson chi square is about 5. This suggests that there is a greater variability among counts than would be expected for a Poisson regression model. This is expected, as a lattice count is the *sum* of

counts in its region, over which a general model is an inhomogeneous Poisson process. Hence, the lattice count is actually distributed as a mixture of Poisson distributions. The variance of a mixture distribution is often greater than the sum of variances of individual component distributions, and hence it leads to an overdispersion. This can be easily shown. Suppose that the distribution of  $Z$  is a finite mixture of  $m$  component distributions, with component means  $\mu_i$ , variances  $\sigma_i^2$ , and mixing parameters  $\alpha_i > 0$  for  $i = 1, \dots, m$ , such that  $\sum_{i=1}^m \alpha_i = 1$  then, obviously,

$$\text{Var}(Z) = \sum_{i=1}^m \alpha_i \left( \sigma_i^2 + \left( \mu_i - \sum_{j=1}^m \alpha_j \mu_j \right)^2 \right) \geq \sum_{i=1}^m \alpha_i \sigma_i^2 \tag{2}$$

where  $m$  approaches  $\infty$  as in the case of an inhomogeneous Poisson process, as long as the variance of  $Z$  is finite. We can apply Fubini's theorem to (2) to show that

$$\text{Var}(Z) \geq \int_C \sigma^2(s) \, d\pi(s) \tag{3}$$

Here  $\sigma^2(s)$  is the variance of the Poisson process at location  $s$  and  $\pi$  is the mixing distribution on compartment  $C$ . This type of overdispersion result (3) can also be shown and implied by other methods. For example, we can obtain (3) by applying Shaked's stochastic ordering [7] and constructing a uni-component density with mean  $\sum \alpha_i \mu_i$  and variance  $\sum \alpha_i \sigma_i^2$  in an application of the first un-numbered inequality [8, p. 31].

The *overdispersion* with respect to a Poisson distribution can be modeled naturally by integrating the Poisson distribution with respect to its conjugate distribution, the Gamma distribution. Then the resulting integrated distribution (i.e. marginal distribution in the Bayesian context) will be Negative binomial (NB). Indeed, as a referee pointed out, NB regression models have been proposed to deal with the extra-Poisson variation in previous studies, see; for example, [6, 9, 10].

More specifically, to lay out our specific model for this type of neuronal data and to have a basis for incorporating spatial correlation, later, let  $Y_i(\mathbf{s})$  be the count of cells,  $y_i(\mathbf{s})$  the realization of  $Y_i(\mathbf{s})$ , and  $\mathbf{x}_i(\mathbf{s})$  a  $p$ -dimensional vector of covariates, observed for subject  $i = 1, \dots, I$ , at a spatial location  $\mathbf{s}$ , where  $\mathbf{s} = (u, v)$  denotes the lattice coordinates of the spinal at which data may be obtained. Here  $\mathbf{s} = (u, v) = (1, 1), \dots, (U, V)$ , for example, represent particular lamina and slice in the labeled cell study. Using the matrix notation, for the  $i$ th subject,  $\mathbf{y}_i = (y_{i11}, \dots, y_{iU1}, \dots, y_{iUV})^t$  is the  $(UV) \times 1$  response vector and  $\mathbf{X}_i = (\mathbf{x}_{i11}, \dots, \mathbf{x}_{iU1}, \dots, \mathbf{x}_{iUV})^t$  is the  $(UV) \times p$  matrix of covariates.

To model spatial count data with overdispersion, we assume that, conditional on an unobserved random variable  $Z_{ij} = z_{ij}$  ( $j = 1, \dots, UV$ ),  $Y_{ij}$  has the Poisson distribution given by

$$p(y_{ij}|z_{ij}) = \frac{(\mu_{ij} z_{ij})^{y_{ij}} e^{-(\mu_{ij} z_{ij})}}{y_{ij}!} \tag{4}$$

with  $E(Y_{ij}|z_{ij}) = \text{Var}(Y_{ij}|z_{ij}) = \mu_{ij} z_{ij}$ .

Next, suppose that  $Z_{ij}$  has a Gamma distribution given by gamma( $1/\phi, \phi$ ), with a probability density function (pdf)

$$p(z) = \frac{1}{\Gamma(1/\phi)\phi^{1/\phi}} z^{1/\phi-1} e^{-z/\phi}$$

here we have abused and shall continue to abuse the notations a little by using  $p$  to denote all relevant pdfs.

It then follows easily by integration with respect to the mixing distribution that marginally  $Y_{ij}$  has the NB distribution given by  $\text{NB}(1/\phi, 1/(1 + \phi\mu_{ij}))$ , with a pdf

$$p(y_{ij}) = \int p(y|z)p(z) dz = \frac{\Gamma(y_{ij} + 1/\phi)}{\Gamma(1/\phi)\Gamma(y_{ij} + 1)} \frac{(\phi\mu_{ij})^{y_{ij}}}{(1 + \phi\mu_{ij})^{y_{ij} + 1/\phi}} \equiv f(y_{ij}; \mu_{ij}, \phi) \quad (5)$$

mean  $E(Y_{ij}) = \mu_{ij}$  and variance  $\text{Var}(Y_{ij}) = \mu_{ij} + \phi\mu_{ij}^2$ . The relationship of the response variable  $Y_{ij}$  with the covariates  $x_{ij}$  can still be modeled *via* the link function:  $E(Y_{ij}) = \mu_{ij} = \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta})$ , where  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of regression parameters and the inferences about  $\boldsymbol{\beta}$  can be made accordingly. See Section 2.4 for an application to the labeled cell data.

This NB model for the count data (without the knowledge of  $Z_{ij}$ ) is more general than the Poisson model, which is a special case of the NB model with  $\phi = 0$ . Note that the variance function is a quadratic function of the mean that allows us to capture the overdispersion in the data.

*Spatial correlation* (or non-zero covariance between data at different locations) is another source of variation that may be significant in the lattice count data. Typically, there are positive correlations between observations from spatially close sampling units. Failure to account for any non-ignorable spatial correlation usually leads to spuriously small standard errors of regression parameter estimates, and corresponding overstatement of the significance of regression effects. We consider two cases below.

*Case 1:* Spatial correlation disappears after the overdispersion within a spatial grid is controlled.

The model can be fitted by using the Newton–Raphson algorithm to maximize the log-likelihood function

$$L(\mathbf{y}, \boldsymbol{\mu}, \phi) = \sum_{i,j} \log(f(y_{ij}; \mu_{ij}, \phi)) \quad (6)$$

with respect to the regression parameters, where the sum is over all the observations and  $f$  is defined in (5). On the  $k$ th iteration, the algorithm updates the parameter vector  $\boldsymbol{\beta}_k$  with

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k - \mathbf{H}^{-1} \mathbf{w}$$

where  $\mathbf{H} = [h_{ij}] = [\partial^2 L / \partial \beta_i \partial \beta_j]$  is the Hessian (second derivative) matrix and  $\mathbf{w} = [w_j] = [\partial L / \partial \beta_j]$  is the first derivative vector of the log-likelihood function.

*Case 2:* Spatial correlations cannot be ignored after the overdispersion is controlled.

In order to model a spatial correlation, we adopt and generalize the method of generalized estimating equation (GEE) (see, e.g. [11, 12]). For the  $i$ th subject, let  $\mathbf{A}_i$  be the  $UV \times UV$  diagonal matrix with  $A_{ij} = \mu_{ij} + \phi\mu_{ij}^2$  as the  $j$ th diagonal element ( $j = 1, \dots, UV$ ). Also, let  $\mathbf{R}_i(\boldsymbol{\rho})$  be the  $UV \times UV$  ‘working’ correlation matrix, where the  $(j, j')$  element of  $\mathbf{R}_i(\boldsymbol{\rho})$  is an estimated correlation between  $y_{ij}$  and  $y_{ij'}$ , depending on a vector of unknown parameters  $\boldsymbol{\rho}$ . Thus, the covariance matrix for the  $i$ th subject is

$$\mathbf{V}_i(\boldsymbol{\rho}) = \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\rho}) \mathbf{A}_i^{1/2}$$

*Fitting algorithm in case 2:*

1. Compute an initial estimate of  $\beta$  with an NB model assuming independence as in Case 1. Specifically, we use a Newton–Raphson algorithm to maximize the log-likelihood function in (6) with respect to the regression parameters  $\beta$  and  $\phi$ .
2. Compute the empirical variogram of the *standardized Pearson residuals*

$$r_{ij} = (y_{ij} - \mu_{ij}) / \sqrt{\mu_{ij} + \phi \mu_{ij}^2}$$

from the fitted model to estimate the spatial correlation in the data. In the case where auto-correlation is only a function of distance, the empirical semi-variogram  $\gamma(d)$  of the residuals is estimated by

$$\gamma_i(d) = \frac{1}{2|N_i(d)|} \sum_{N_i(d)} (r_{is} - r_{is'})^2$$

where  $N_i(d)$  is the set of all pairs  $(s, s')$  with Euclidean distances  $d = |s - s'|$ ,  $|N_i(d)|$  is the number of distinct pairs in  $N_i(d)$ , and  $r_{is}$  and  $r_{is'}$  are residual values at spatial locations  $s$  and  $s'$  for animal  $i$ . In the following, we may suppress the subscript  $i$ . Of course, when  $\gamma_i$  are assumed same for all  $i$ , we shall estimate the common  $\gamma(d)$  by  $\bar{\gamma}(d)$ , the average of  $\gamma_i(d)$ 's.

3. Fit a parametric model for the semi-variogram. Different valid parametric families of models are discussed in, for example, [5]. For the labeled cell data, exploratory analysis showed that the exponential semi-variogram model

$$\gamma(d) = \gamma(d, \rho) = 1 - e^{-d/\rho}$$

fitted best in terms of minimizing least squares. It is then easy to write down the correlation matrix  $\mathbf{R}(\rho)$  based on the semi-variogram model:  $\mathbf{R}(\rho) \equiv R(d, \rho) = 1 - \gamma(d, \rho) = e^{-d/\rho}$  for the  $ij$ th component of the matrix that is  $d = |i - j|$  distance apart.

4. Use the GEEs or the score functions to estimate the parameters  $\beta$  and fit a variogram model iteratively, or use a penalized quasi-likelihood (PQL) estimation procedure described in [13]. We have applied both procedures for the labeled cell data, the results are very close as expected. The PQL algorithm is designed for generalized linear mixed models (GLMM), but in our case the model is simpler: it only contains non-independent errors and no random effects. The new GLIMMIX procedure in SAS 9.1 provides the function to fit the GLMM with NB distribution.

*2.3. Tests for overdispersion and spatial correlation*

*2.3.1. Overdispersion.* Under the NB model, the variance function  $\text{Var}(Y_{is}) = \mu_{is} + \phi \mu_{is}^2$ . Since the Poisson model is a special case of the NB model with  $\phi = 0$ , naturally we can test the overdispersion by testing the null hypothesis  $H_0 : \phi = 0$  against the alternative hypothesis  $H_1 : \phi > 0$ . A likelihood ratio statistic can be constructed by ‘(-2) multiplied by the difference between the fitted log-likelihoods under the Poisson and NB models’. This likelihood ratio test should be asymptotically equivalent to the score tests proposed by Dean and Lawless [14] and Dean [15]. See also [16] for applications to three different cases.

It should be noted that standard  $\chi^2$  approximation to the null distribution of these test statistics does not work well in small or moderate sample size situations. Either a correction to the test statistic or a high-order asymptotic approximation to the null distribution is necessary [15, 17]. Nevertheless, simulated critical values of the likelihood ratio test can be obtained easily without relying on either correction. That is what we choose to do. In fact, one can also decide which model is appropriate based on the Pearson  $\chi^2$  measure of overdispersion. See the data analysis section.

*2.3.2. Spatial correlation.* An examination of the existence of spatial correlation is needed before choosing a final model. The likelihood method, using full information of the data, will be more powerful than the method of GEE if the correlation is negligible. For this, we suggest to examine the Pearson residual  $r_{is} = (y_{is} - \mu_{is}) / \sqrt{\mu_{is} + \phi \mu_{is}^2}$ . We expect the correlation coefficient between  $r_{is}$  and  $r_{it}$ , estimated as

$$\frac{\sum_i r_{is} r_{it}}{\sqrt{\sum_i r_{is}^2} \sqrt{\sum_i r_{it}^2}}$$

to be close to zero, for  $s \neq t$  if there is no spatial correlation. In practice, these correlations are often sufficiently large so that a formal test is unnecessary. We can further investigate the sample variogram to confirm whether we have Case 1 or Case 2, and then model the correlation structure under Case 2.

#### 2.4. Simulation study and data analysis

In this section, we examine the effect of model mis-specification by a simulation study and apply our procedures to analyze the labeled cell data.

*2.4.1. Simulation.* In order to illustrate the importance of correctly modeling the heterogeneity structure, we conducted Monte Carlo experiments to evaluate the performance of the significance test about  $\beta$  under a correctly specified model (NB model) and a mis-specified model (Poisson model). In each simulation, NB pseudo-random numbers were generated to mimic the spatial lattice count data and then fitted with both a Poisson model and a NB model. The experiments were repeated 1000 times. Table II shows the results of the Monte Carlo experiments. The main entries in the table are the percentages that the level  $\alpha$  significance test was rejected out of these 1000 experiments, for two  $\alpha$  values 0.05 and 0.10, and three combinations of  $\mu_0$  and  $\phi$ . Here,  $\mu_0$  is a simplified version of  $\mu_{ij}$  in (4) and (5). As expected, the rejection frequencies (or type-I errors) resulting from the mis-specified model (Poisson model) are much larger than the expected level  $\alpha$ ; the results are especially bad for large  $\phi$ , the level 0.1 test wrongly rejected the null hypothesis more than 80 per cent of times when  $\mu_0 = 5$  and  $\phi = 5$ . On the other hand, the rejection frequencies under the correct model (NB model) are closer to  $\alpha$  and get closer to  $\alpha$  as the number of observations increases as they should be.

*2.4.2. Data analysis.* Table III displays the results of goodness of fit of the Poisson and NB models to the data. The fitted Poisson regression model shows a severe overdispersion with a Pearson  $\chi^2$  of 5.08. Pearson  $\chi^2$  is a standard measure of overdispersion, where a value of 1 indicates a perfect fit

Table II. Effect of model mis-specification.

|                       | $\alpha$ per cent | $H_0: \mu = \mu_0$ under Poisson model |        |         | $H_0: \mu = \mu_0$ under NB model |        |         |
|-----------------------|-------------------|--|--------|---------|-----------------------------------|--------|---------|
|                       |                   | 10 obs                                 | 50 obs | 100 obs | 10 obs                            | 50 obs | 100 obs |
| Case 1:               | 5                 | 23.3                                   | 22.3   | 6.2     | 11.2                              | 5.3    | 4.2     |
| $\mu_0 = 2, \phi = 1$ | 10                | 27.5                                   | 28.7   | 13.3    | 15.2                              | 9.3    | 8.8     |
| Case 2:               | 5                 | 56.8                                   | 46.1   | 45.8    | 10.3                              | 6.7    | 7.1     |
| $\mu_0 = 2, \phi = 5$ | 10                | 66.2                                   | 54.7   | 57.1    | 15.5                              | 11.1   | 10.3    |
| Case 3:               | 5                 | 73.7                                   | 65.1   | 37.7    | 25.4                              | 11.9   | 9.5     |
| $\mu_0 = 5, \phi = 5$ | 10                | 80.1                                   | 77.1   | 50.3    | 30.1                              | 19.9   | 12.8    |

Table III. Criteria for assessing goodness of fit.

| Criterion             | Poisson regression |           |          | NB regression with spatial correlation |         |          |
|-----------------------|--------------------|-----------|----------|--|---------|----------|
|                       | DF                 | Value     | Value/DF | DF                                     | Value   | Value/DF |
| Pearson $\chi^2$      | 6734               | 34 208.72 | 5.08     | 6734                                   | 8557.77 | 1.27     |
| $-2 \log$ -likelihood |                    | -90.36    |          |  | 4549.02 |          |

Table IV. Type III  $F$  tests of effects in the spatial model.

| Source  | DF | Pr> $F$ stat | Source          | DF | Pr> $F$ stat |
|---------|----|--------------|-----------------|----|--------------|
| Nerve   | 2  | 0.0278       | Nerve * lamina  | 18 | 0.0379       |
| Lamina  | 9  | <0.0001      | Nerve * section | 22 | 0.0461       |
| Side    | 1  | 0.0002       | Nerve * side    | 2  | 0.0007       |
| Section | 11 | 0.0213       |                 |    |              |

(no dispersion), greater than 1 indicates overdispersion, and less than 1 indicates underdispersion. Our NB regression model that accounts for both overdispersion and spatial correlation fitted the data well; it has a likelihood of 4549.02, which is much bigger than  $-90.36$ , the likelihood of the fitted Poisson model. The fitted NB model also produced almost no overdispersion (where Pearson  $\chi^2 = 1.27$ ).

Under our NB regression model, a sequence of Wald-type  $F$  tests [18] is used to assess the significance of each factor and their interactions in the model. Roughly speaking, a Wald-type  $F$  test checks for the significance of each factor and their interactions by controlling any other factors equal to or of lower degrees and orthogonal to any higher-order interactions that contain it. As shown in Table IV, all four factors, Nerve, Side ( $X$ ), Lamina ( $Y$ ), Section ( $Z$ ), and their interaction terms have  $p$ -values less than 0.0461, while the Lamina factor is the most significant factor ( $p < 0.0001$ ) that affects the number of labeled positive cells. This implies that the variation is the greatest across different laminar levels. A more detailed multiple comparison procedure can

be used to show which laminar level stands most different from the rest. This is what we are going to demonstrate in the next section.

### 2.5. Statistical mapping

An important and common question in the anatomical labeled cell studies is the identification of spatial compartments where the number of labeled cells is significantly different between animals under two different conditions: treated vs controlled; and treatment 1 vs treatment 2. This biological question of finding differential spatial compartments can be restated as a statistical problem in multiple hypothesis testing: simultaneously test all compartments to see which one rejects the null hypothesis of no difference between two groups of animals.

The procedure we are going to use may be called *simultaneous statistical mapping* (SSM). It is *simultaneous* because we will statistically control an overall error from testing all the compartments simultaneously. It is *statistical mapping* because the values of test statistics computed at each of the compartments lead to a statistical map, where the non-significant compartments may be set to zero, and significant compartments are highlighted with the observed values of the test statistics. Our SSM procedure is inspired from *statistical parameter mapping* (SPM) [19, 20], which is widely used in functional imaging analysis. The name SSM emphasizes the simultaneous nature of the tests. In our opinion, is more appropriate than SPM. SSM allows us to identify functionally specialized spatial compartments and is the most prevalent approach to characterizing functional anatomy and treatment-related changes.

Tests of the mean differences either between treated and control animals or between differently treated animals can be set up as tests of contrasts, one for each pair of the group conditions. The Wald test statistic [21] for testing  $\mathbf{L}'\boldsymbol{\beta} = 0$ , where  $\mathbf{L}$  is a contrast matrix, is defined by

$$S_w = (\mathbf{L}'\hat{\boldsymbol{\beta}})'(\mathbf{L}'\hat{\boldsymbol{\Sigma}}\mathbf{L})^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}}) \quad (7)$$

where  $\hat{\boldsymbol{\beta}}$  is the maximum likelihood estimate and  $\hat{\boldsymbol{\Sigma}}$  is its estimated covariance matrix. The Wald test statistics are computed by (7) for each compartment. The  $p$ -value of the test statistic at each compartment can be computed using the asymptotic distribution of  $S_w$  under the null hypothesis of no difference, which is chi square with  $r$  degrees of freedom, where  $r$  is the rank of  $\mathbf{L}$ .

In some imaging literature, a statistical mapping is a map of these test statistics or these pointwise  $p$ -values without adjusting for the multiplicity, or the simultaneous nature of these multiple tests.

Adjusting for the multiplicity is important. As the labeled cell experiment measures the cell count for hundreds of spatial compartments simultaneously, using a 0.05-level pointwise significance test for all compartments will lead to  $0.05 \times m$  compartments wrongly identified as significant even when the null hypothesis of no difference for any of the compartments is true and where  $m$  is the number of compartments. For example, if  $m = 500$ , there would be about 25 compartments wrongly declared to be significant if we used the pointwise critical value, that is, if we declared a significance for a compartment when its (pointwise)  $p$ -value is less than 0.05. This multiplicity error increases as the number of compartments increases. Thus, we need to perform a multiple testing procedure that controls an *overall* type-I error rate. Family-wise error rate (FWER) and false discovery rate (FDR) are the two most common overall type-I error rates. FWER is the probability of making at least one false-positive error. FDR, on the other hand, is the rate of false positives among all declared discoveries (significances) by a test. If FWER is also controlled by  $\alpha$ , then FDR is controlled by  $\alpha$ . Thus, FWER advocates a stronger control than FDR for an overall

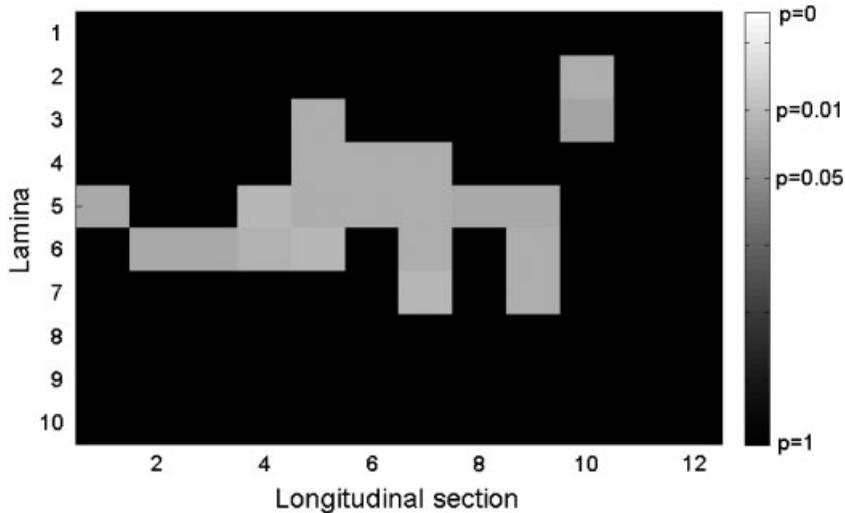


Figure 2. Statistical mapping in a labeled cell study. Anatomical locations that demonstrated a significantly greater number of labeled cells in stimulated animals than in control animals are identified. Data adapted from [2].

false positiveness. For this paper, we propose to use the Benjamini and Hochberg's procedure (BH procedure) [22] to control the FDR. This is equivalent to declaring that a compartment is significant if its  $p$ -value is less than the FDR 0.05 critical value  $p(0.05)$ . This 0.05-FDR critical value,  $p(0.05)$  is often much less than 0.05. A level- $\alpha$  FDR BH critical value is given by  $p(\alpha) = p_k$ , where  $k = \max\{i : p_i \leq \alpha \cdot i/m, i = 0, \dots, m\}$ ,  $p_0 = 0$ , and  $p_i$ 's are increasingly ordered  $p$ -values from the  $m$  compartments.

The BH procedure was initially proposed for multiple testing, in which individual test statistics are assumed to be independent of each other. However, the test statistics from different compartments can be correlated when data are spatially correlated. If the spatial correlation follows the correlation function  $e^{-d/\rho}$ , it is easy to show, as Benjamini and Yekutieli [23] did, that the correlation of the compartment test statistics satisfies the property of 'positive regression dependency on subsets' (PRDS). Hence, the BH procedure for independent tests still conservatively controls the overall FDR for PRDS-correlated multiple tests. Specifically, for our labeled cell data, our significance threshold for deciding which compartment is significantly different (between two treatments, or between a treatment and a control) is chosen with a BH-FDR controlling procedure that accounts for the multiplicity of tests. The resulting statistical map for the labeled cell study is shown in Figure 2. A model-based comparison (Wald test with FDR control,  $p < 0.05$  and FDR set to 0.05) was used to define anatomical locations where a significant change occurred (i.e. ones with  $p$ -value less than FDR  $p(0.05)$ ). The anatomical locations that demonstrated a significantly greater number of labeled cells in stimulated animals than in control animals represent specific populations of neurons involved in the reflex. These locations can be compared with the locations of other neural populations involved in similar reflexes to increase our understanding of the modular organization of neural circuits in the spinal cord. Therefore, Figure 2 demonstrates the utility of the statistical analysis to identify anatomically isolated areas of increased neural activity. These

areas seem to be isolated (or color being discontinued at boundaries of compartments) because we have analyzed the data compartment by compartment, which was the goal of this neuroscience study.

### 3. MODELING EEG/EMG SIGNALS IN A MUSCLE FATIGUE STUDY

#### 3.1. Experiment and data

Muscle fatigue has been studied for over a century, but almost no data are available to indicate how the brain perceives fatigue and modulates its signals to the fatiguing muscle. Although fMRI signals show clearly cortical activation patterns on two- or three-dimensional anatomical images, they tell little whether the signals reflect motor command- or sensory information processing-related cortical activities [24–26]. To address this issue, handgrip force, surface EMG signals of finger flexor and extensor muscles, and scalp electroencephalograms (EEG) were measured simultaneously in a fatigue handgrip task (Figure 3), in which each subject (total 8) performed 200 trials of maximal voluntary contraction (MVC) of handgrip. These 200 trials lasted 1400 s; each trial duration was 2 s and the inter-trial interval duration was 5 s.

The force and EMG data were processed using the Spike2 software package [4]. For the force data (sampled 100 points/s), the measured voltage signals were converted to force (N) using the calibration equation determined previously for the transducer used in the experiments [24]. Then, in each trial a mean value of the force was calculated from the data points that represented relatively stable force values (the ascending and descending portions of force values were excluded). These mean values were averaged again over 20 trials (out of 200) to obtain a total of 10 data points.

For the EMG data (sampled 2000 points/s), we first rectified all the recordings (flipping all negative data points to positive) and averaged all the data points for each 2-s trial to yield a mean value in each trial. Then, similarly as it was for the force data, we averaged values across 20 trials (out of 200) to also get 10 data points. Finally, the averaged force and EMG data were normalized to the corresponding maximal values, which were recorded at the beginning of the experiments and measured prior to the above-mentioned data analysis.

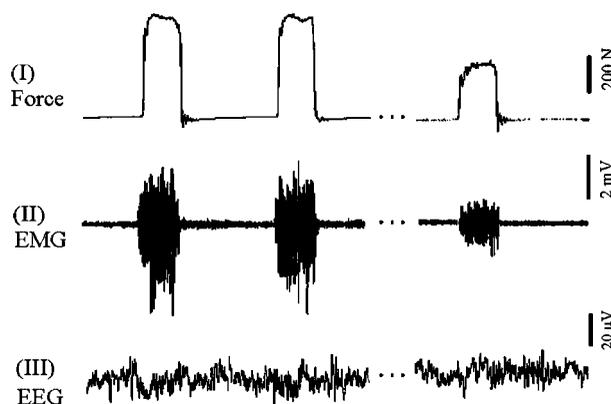


Figure 3. Simultaneous force, EMG, and EEG signals.

EEG signals were recorded from the scalp using a 64-channel NeuroSoft SYNAMPS system (version 4.2, NeuroScan, El Paso, TX, U.S.A.). The EEG data (sampled 250 points/s) consisted of two categories, movement-related cortical potential (MRCP) and power of frequency (POF). The MRCP was quantified by taking the highest MRCP value in each trial and then averaging the values across each of the consecutive 40 trials. The POF was calculated within a 750-ms window right before each contraction began (preparation phase) and a 2-s window (corresponding to the 2-s force or EMG) during the contraction (sustained phase) in each trial. The POF values for the preparation phase and that for the sustained phase were then averaged across each of the consecutive 40 trials. Because of relatively low signal-to-noise ratio, it needed to average a larger number of trials for the EEG data.

Note that although subjects were advised not to blink eyes, bite teeth, move head, or tense muscles other than those involved in the handgrip contractions, these activities occasionally occurred and the trials associated with these activities were excluded from further analysis. These trials were identified by visually inspecting all raw EEG data. On average, EEG data in  $35 \pm 4$  trials in each block were artifact free and were subjected to the subsequent analysis.

In summary, these data were longitudinal data collected for each of the eight individuals and could be considered as a multivariate time series. However, they differ from a typical, long time series in that they have been averaged over 20 time points in each of 10 blocks for the force and EMG data and over 40 points in each of five blocks for the EEG data, for each subject; thus, they consist of a large number of short series, collected at different time points. We shall show that, for some channels as a univariate longitudinal data set, fitting a first-order autoregressive AR(1) time series regression model to the data was inadequate. An additional parameter to model the correlation is needed in these channels and the role of the parameter differs from one channel to the other. On the other hand, if we treat these univariate series collectively as three multivariate series (each consisting of the same data type, such as EEG, or EMG, or force data, from multiple channels), our proposed general linear model (below) for multivariate repeated measures (over time) with a direct (Kronecker) product structure fits the data quite well.

### 3.2. Time series regression models

Most EEG/EMG signals are time series data. They can be fit reasonably well by the following time series regression model:

$$Y_{ijt} = \mathbf{x}'_{it} \boldsymbol{\beta} + u_{ijt}, \quad i = 1, \dots, M; \quad j = 1, \dots, N; \quad t = 1, \dots, T \quad (8)$$

where  $Y_{ijt}$  is the response variable, such as a force, or an EMG, or an EEG signal for the  $j$ th subject on the  $i$ th condition (e.g. performing a task, or receiving a stimuli or at rest) at the  $t$ th time period;  $\mathbf{x}_{it}$  denotes a non-stochastic regressor vector (e.g. a binary variable representing a task is performed or not at time  $t$  and a continuous variable denoting the age of the subject);  $\boldsymbol{\beta}$  is a vector of unknown parameters; and  $u_{ijt}$  is the random error. Again, let  $y_{ijt}$  be the realization of  $Y_{ijt}$  in (8).

A common assumption is that the process generating the regression errors is stationary [27], so the covariance of two errors depends only on their separation  $s$  in time:

$$C(u_t, u_{t+s}) = C(u_t, u_{t-s}) = \sigma^2 \rho_s$$

where  $\rho_s$  is the error autocorrelation at lag  $s$ . A usual model for the serial correlation of EMG/EEG signals assumes that the errors are AR(1) [27, 28]  $u_t = \rho u_{t-1} + \varepsilon_t$ , where the ‘primitive’ error terms  $\varepsilon_t$  are assumed to be mutually uncorrelated with  $E(\varepsilon_t) = 0$  and  $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ . Under this model,  $\rho_s = \rho^s$  and  $\sigma^2 = \sigma_\varepsilon^2 / (1 - \rho^2)$ . Using the generalized least-squares (GLS) method, we obtain an estimator of  $\beta$ :

$$\hat{\beta} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

where  $\mathbf{X}$  is the design matrix and  $\mathbf{V}$  is the estimated covariance matrix.

In our example, the force, EMG, and EEG data are longitudinal values collected from each of the eight individuals and were averaged over 20 or 40 time points in each of 10 or 5 blocks for each subject. So, the resulting regression errors from the blocked longitudinal data may not fit the AR(1) model or may not even be stationary in some channels. A generalization of the AR(1) model is the first-order heterogeneous autoregressive model, HAR(1). Its covariance of two errors has the form:

$$C(u_t, u_{t+s}) = C(u_t, u_{t-s}) = \sigma_t \sigma_{t+s} \rho_s$$

In some other channels, some EEG/EMG signals show that the current error depends on the random noises from the current and previous periods (rather than on the previous regressors). This leads to the first-order moving-average model, MA(1), or even an ARMA model if it also depends on the regressors.

A correct model specification is important. For example, mis-specification of covariance matrix will affect the estimation of variance of parameters, which will affect the test of hypotheses about relevant parameters. See Section 3.3 for some examples on the effects of covariance mis-specification. The most commonly used model selection criteria are Akaike Information Criterion (AIC) and Schwartz Bayesian Information Criterion (BIC). We shall use these criteria and a likelihood approach in Section 3.3 to choose adequate models for our force, EMG, and EEG data.

There is another challenging issue in modeling the muscle fatigue data. In the EMG and EEG signal recording, we often monitor signals from multiple locations. For example, EMG signals were recorded from the following three muscles: flexor digitorum superficialis (FDS), flexor digitorum profundus (FDP), and extensor digitorum (ED); EEG signals were recorded from 64 multiple channels. This leads to multivariate time series data. It is important to neuroscientists that one extracts information by combining the data from multiple channels.

Specifying direct Kronecker product structures of the covariance matrix in a general linear model is an attractive solution to model multivariate time series data [29]. Naturally, we can model these multivariate covariance structures by taking the Kronecker product of an unstructured matrix (that reflects the covariance across the multivariate observations at the same time point) with an additional structured covariance matrix (that reflects the covariance across the time for each fixed univariate series). For instance, to model three channels’ EEG signals, we consider the following error–covariance matrix based on the direct product of AR(1) covariance matrix and

an unstructured matrix:

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 \\ \sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3 \\ \sigma_3\sigma_1 & \sigma_3\sigma_2 & \sigma_3^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & & & \ddots & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}$$

It will be interesting to check whether the additional heterogeneity (to the AR(1) model) mentioned for some univariate series can be absorbed into the unstructured component covariance matrix in the multivariate case. This will be examined in the next section for our force, EMG, and EEG data.

### 3.3. Simulation study and data analysis

In this section, we examine the effect of covariance mis-specification by a simulation study and apply our generalized models to analyze the force/EMG/EEG data.

*3.3.1. Simulation.* As in Section 2, we perform a number of Monte Carlo experiments to evaluate the performance of the significance test about its mean  $\mu$  under a correctly specified model HAR(1) and a mis-specified model AR(1). We expect that the effect on the bias of an estimator of the mean is negligible, but that the effect on the variance of the estimator is large and hence it influences a significance test of the mean.

To mimic our short blocked series, we first generate  $N = 10$  samples of size  $T$  of the AR(1) process  $\{u_t, t = 1, 2, \dots, T\}$ :

$$(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(N)})$$

where each  $\mathbf{u}^{(i)}, i = 1, 2, \dots, N$ , represents a vector of  $T$  pseudo-random numbers from AR(1) with autocorrelation  $\rho = 0.8$ . Then, we generate the realization of the response variable  $\{y_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T\}$  by

$$y_{it} = \mu(t) + \sigma(t)u_{it}$$

We introduce a number of different departures from the AR(1) model by considering the following different functional forms of  $\mu(t)$  and  $\sigma(t)$ :

- Case 1:  $\mu(t) = 0, \sigma^2(t) = \sigma^2 + 0.05t$ .
- Case 2:  $\mu(t) = 0, \sigma^2(t) = \sigma^2 + 0.03t + 0.01t^2$ .
- Case 3:  $\mu(t) = 0 + 0.5t, \sigma^2(t) = \sigma^2 + 0.05t$ .

Data are then fitted by the best AR(1) model and the best HAR(1) model, respectively. We repeat the experiment 1000 times. Table V shows the results of the Monte Carlo experiments.

Table V. Simulation for mis-specification of time series models.

|        | $\alpha$ per cent | $H_0: \mu=0$ under AR(1) model |        |         | $H_0: \mu=0$ under HAR(1) model |        |         |
|--------|-------------------|--------------------------------|--------|---------|---------------------------------|--------|---------|
|        |                   | 10 obs                         | 50 obs | 100 obs | 10 obs                          | 50 obs | 100 obs |
| Case 1 | 5                 | 71.7                           | 32.8   | 25.2    | 8.2                             | 10.3   | 6.3     |
|        | 10                | 91.7                           | 70.3   | 33.2    | 12.3                            | 13.5   | 11.2    |
| Case 2 | 5                 | 90.8                           | 66.4   | 55.3    | 15.3                            | 11.2   | 9.3     |
|        | 10                | 95.0                           | 80.7   | 77.6    | 30.3                            | 22.1   | 15.7    |
| Case 3 | 5                 | 83.3                           | 55.5   | 32.1    | 22.4                            | 18.9   | 8.3     |
|        | 10                | 91.3                           | 63.5   | 43.1    | 27.2                            | 22.9   | 11.8    |

Table VI. Correlation specification based on the AIC criterion for the fatigue study.

|             |               |                   |               |               |
|-------------|---------------|-------------------|---------------|---------------|
| Signals     | Force         | EMG channel 1     | EMG channel 2 | EMG channel 3 |
| Correlation | HAR(1)        | AR(1)             | AR(1)         | AR(1)         |
| AIC         | -217.6        | -131.9            | -150.8        | -115.4        |
| Signals     | EEG channel 1 | EEG channel 2     | EEG channel 3 | EEG channel 4 |
| Correlation | ARMA(1,1)     | Compound symmetry | AR(1)         | AR(1)         |
| AIC         | 140.4         | 118.7             | 163.6         | 77.0          |

The main entries of the table are the rejection frequencies of the test for two different levels of significance (0.05 and 0.10). It is clear that the rejection frequencies are high when the covariance is mis-specified as AR(1), while the rejection frequencies are much smaller and approach the expected error rate  $\alpha$ , when the covariance is correctly specified by the HAR(1) model.

*3.3.2. Data analysis.* In the analysis of the muscle fatigue data, the first step in fitting an 'optimal' general linear model to each of the three categories of data (Force, EMG, EEG) is to select an appropriate covariance structure based on the AIC and BIC. The AIC and BIC were computed for both the multivariate time series and its individual univariate time series from each of the force, EMG, and EEG data sets. Briefly, some univariate sequences in the force, EMG, and EEG data can be considered from an AR(1) time series and others need an additional parameter to the AR(1) model to fit the data adequately (see Table VI); the corresponding multivariate time series (e.g. EMG signals of multiple channels) has a covariance that is a direct Kronecker product of an unstructured covariance matrix with an additional covariance matrix AR(1). Therefore, the additional parameter *is* absorbed into the unstructured component in the multivariate case. Table VI shows correlation specification based on the AIC criterion for some of the channels in the fatigue study, where EMG channel 1 is FDS, EMG channel 2 is FDP, EMG channel 3 is ED; and EEG channel 1 is C3, EEG channel 2 is C4, EEG channel 3 is FZ, EEG channel 4 is PZ. The detailed statistical analysis report can be found at <http://stat.case.edu/~xfwang/musclefatigue/>. Note that *compound symmetry* correlation in Table VI has correlation matrix with equal diagonal elements and equal off-diagonal elements [18, p. 107].

The second step of the analysis is, based on the covariance structures, to obtain the optimal general linear models and test whether the means by time blocks (e.g. B1–B10) and by variables

were constant. For example, for the EMG data based on unstructured AR(1) covariance, the means by blocks are significantly different from a constant ( $p < 0.0001$ ), while the EMG means of FDS, FDP, and ED channels were not significantly different ( $p = 0.06$ ).

The third step was carried out only if the means were significantly different in the second step. A pair-wise multiple comparison procedure (MCP, sometimes also referred to as mean separation tests [30]) is performed to identify which means were responsible for the difference. Using an MCP rather than a standard  $t$ -test is necessary to control the multiplicity issue arising from testing many hypotheses simultaneously, such as 45 pairs out of the 10-block data. In particular, a modified Tukey's MCP for correlated data was used for all pair-wise comparisons of GLS means of blocks.

The three-step analysis is conducted for each of the EMG, and EEG data, both as multivariate and as several univariate time series, and for the force data as univariate time series. In summary, with a global significance level at  $\alpha = 0.05$  (accounting for the overall multiplicity), our two primary hypotheses are confirmed statistically: (1) the brain is able to generate a constant, maximal command to drive the muscles despite significant muscle fatigue; and (2) the power of the EEG frequency (range from 0.5 to 35 Hz) would not change during the preparation/execution phase of the task, but would decrease during the holding phase of the contraction as muscle fatigue.

#### 4. DISCUSSION

In this paper, we addressed two types of neuroscience problems which challenge the typical statistical models assumed for analyzing neuronal data. The first problem dealt with the use of labeling techniques to identify active cells (neurons), a common experimental tool in neuroscience. Spatially compartmentalized neuronal data are readily obtained by biomedical scientists. However, few attempts have been made to address explicitly the statistical analysis of this type of spatial data. We provided a complete statistical procedure to identify the compartment or the location of cells of interests, spinal cord neurons. We proposed a spatial generalized linear model, from a hierarchical Poisson process coupled with a Gamma distributed mean (which marginally is a NB model), to account for the spatial correlation, inhomogeneity, and overdispersion in the labeled cell count data. The spatial relationship of these counts was captured by modeling the mean response by a log-quadratic function of four influential factors: Nerve, Lamina, Section, and Side. Our spatial generalized linear model is more general than the standard Poisson regression model which, does not account for either overdispersion or spatial correlation. The flexibility of our algorithm in modeling spatial correlation, overdispersion, and inhomogeneous (Poisson) process makes the spatial NB regression a good candidate for studying many neuroscience-related problems containing data obtained at multiple locations. The statistical mapping method in Section 2.5 is a powerful approach for characterizing the functional anatomy and identifying stimulation-related activity changes.

We also employed Benjamini and Hochberg multiple testing procedure [22] to control FDR at a prescribed level  $\alpha$ . It is reasonable to assume that the correlation in the labeled cell data is PRDS if there is any dependence. Research for developing FDR procedures in a general correlated case is open. As a future project, when we can effectively model the correlation structure from appropriate data, we plan to develop a simultaneous testing procedure using the Tube method as in [31, 32] for the NB regression to efficiently control FWER for all these compartments simultaneously. We expect that this new procedure that fully uses the correlation structure in the data will be more powerful than the ones that do not.

In the second problem, the data are averaged EEG/EMG signals recorded during muscle, fatigue where the AR(1) model cannot fully capture all the variation or correlation structure in the data. It is shown that for some channels with individual univariate series a additional parameter has to be included in the modeling of the correlation structure and the role of the parameter differs from one channel to the other. For the multivariate series, the correlation can be modeled by a direct product of an AR(1) covariance matrix with an unstructured covariance matrix. The latter covariance component, the unstructured covariance matrix, represents the inter-relationships between the multivariate observations. The study of the effect of model mis-specification shows that mis-specification of a correlation structure will lead to wrong statistical inferences.

#### ACKNOWLEDGEMENTS

This work was supported in part by NIH grants (NS37400 and HD40298), a Department of Defense grant (DAMD17-01-1-0665), an NSF grant (DMS-0308609), and Department of Veterans Affairs grant (VA RR&D B3675R). The authors would like to thank the two referees and the associate editor for their comments and suggestions on an earlier version of this manuscript.

#### REFERENCES

1. Schouenborg J. Modular organisation and spinal somatosensory imprinting. *Brain Research Reviews* 2002; **40**:80–91.
2. Gustafson KJ, Moffitt MA, Wang X, Sun J, Snyder S, Grill WM. Topography of spinal neurons active during the hindlimb withdrawal reflexes in the decerebrate cat. *Neuroscience* 2006; **141**:1983–1994.
3. McComas AJ, Miller RG, Gandevia SC. Fatigue brought on by malfunction of the central and peripheral nervous systems. *Advanced Experimental Medical Biology* 1995; **384**:495–512.
4. Liu JZ, Yao B, Siemionow V, Sahgal V, Wang X, Sun J, Yue GH. Fatigue induces greater brain signal reduction during sustained than preparation phase of maximal voluntary contraction. *Brain Research* 2005; **1057**:113–126.
5. Cressie NAC. *Statistics for Spatial Data*. Wiley: New York, 1993.
6. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall: New York, 1989.
7. Shaked M. On mixtures from exponential families. *Journal of the Royal Statistical Society, Series B* 1980; **42**:192–198.
8. Lindsay B. Mixture models: theory, geometry, and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 5, Institute of Mathematical Statistics, Hayward, California, 1995.
9. Margolin BH, Kaplan N, Zeiger E. Statistical analysis of the ames salmonella/microsome test. *Proceedings of the National Academy of Sciences USA*, vol. 78, 1981; 3779–3783.
10. Breslow NE. Extra-Poisson variation in log-linear models. *Applied Statistics* 1984; **33**:38–44.
11. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
12. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**: 121–130.
13. Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 1993; **48**:233–243.
14. Dean C, Lawless JF. Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association* 1989; **84**:467–472.
15. Dean C. Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association* 1992; **87**:451–457.
16. Collings BJ, Margolin BH. Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association* 1985; **80**:411–418.
17. Self SG, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 1987; **82**:605–610.
18. Davis CS. *Statistical Methods for the Analysis of Repeated Measurements*. Springer: New York, 2002.
19. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 1995; **2**:189–210.

20. Friston KJ. Introduction: experimental design and statistical parametric mapping. In *Human Brain Function* (2nd edn), Frackowiak RSJ *et al.* (eds). Academic Press: London, 2004.
21. Pawitan Y. In *All Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford University Press: New York, 2001.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 1995; **57**:289–300.
23. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 2001; **29**:1165–1188.
24. Liu JZ, Dai TH, Sahgal V, Brown RW, Yue GH. Nonlinear cortical modulation of muscle fatigue: a functional MRI study. *Brain Research* 2002; **957**:320–329.
25. Liu JZ, Shan ZY, Zhang LD, Sahgal V, Brown RW, Yue GH. Human brain activation during sustained and intermittent submaximal fatigue muscle contractions: an fMRI study. *Journal of Neurophysiology* 2003; **90**: 300–312.
26. Liu JZ, Zhang LD, Yao B, Sahgal V, Yue GH. Fatigue induced by intermittent maximal voluntary contractions is associated with significant losses in muscle output but limited reductions in functional MRI-measured brain activation level. *Brain Research* 2005; **1040**:44–54.
27. Chatfield C. *The Analysis of Time Series: An Introduction* (6th edn). CRC Press: New York, 2003.
28. Goel V, Brambrink AM, Baykal A, Koehler RC, Hanley DF, Thakor NV. Dominant frequency analysis of EEG reveals brains response during injury and recovery. *IEEE Transactions on Biomedical Engineering* 1996; **43**:1083–1192.
29. Galecki AT. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics—Theory and Methods* 1994; **23**:3105–3119.
30. Hsu JC. *Multiple Comparisons: Theory and Methods*. CRC Press: New York, 1999.
31. Sun J, Loader C, McCormick WP. Confidence bands in generalized linear models. *Annals of Statistics* 2000; **28**:429–460.
32. Sun J. Multiple comparisons for a large number of parameters. *Biometrical Journal* 2001; **43**:627–643.