
© Copyright 2007 Blackwell Publishing
Motor unit number estimation using reversible jump Markov chain Monte Carlo

P.G. Ridall†
School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia.

A.N. Pettitt
School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia.

N. Friel
Department of Statistics, University of Glasgow, U.K.

P. A. McCombe
Royal Brisbane and Women’s Hospital and School of Medicine, University of Queensland, Brisbane, Queensland, Australia.

R. D. Henderson
Royal Brisbane and Women’s Hospital and School of Medicine, University of Queensland, Brisbane, Queensland, Australia.

Summary. We present an application of reversible jump Markov chain Monte Carlo (RJMCMC) from the field of neurophysiology where we seek to estimate the number of motor units within a single muscle. Such an estimate is needed for monitoring the progression of neuro-muscular diseases such as amyotrophic lateral sclerosis (ALS). Our data consist of action potentials recorded from the surface of a muscle in response to stimuli of different intensities applied to the nerve supplying the muscle. During the gradual increase in stimulus intensity from threshold to supramaximal, all motor units are progressively excited. However, at any given submaximal stimulus intensity, the number of units that are excited is variable, because of random fluctuations in axonal excitability. Furthermore, the individual motor unit action potentials exhibit variability. To account for these biological properties, Ridall et al. (2006) developed a model of motor unit activation capable of describing the response where the number of motor units, \( N \), is fixed. The purpose of this paper is to extend that model so that the possible number of motor units, \( N \), is a stochastic variable. In this paper we illustrate the elements of our model, show that the results are reproducible and show that our model can measure the decline in motor unit numbers during the course of ALS. Our method holds promise of being useful in the study of neurogenic diseases.

Keywords: alternation, amyotrophic lateral sclerosis (ALS), reversible jump, MCMC, motor neurone disease (MND), motor unit number estimation (MUNE).

1. Introduction

Motor units are responsible for the contraction of muscles. Each motor unit consists of an anterior horn cell in the spinal cord, a motor axon running within the peripheral nerve, its

†Address for correspondence: P.G. Ridall, School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, 4001, Australia.
E-mail: g.ridall@qut.edu.au
terminal branches and the muscle fibres supplied by that axon. One anterior horn cell is capable of supplying many muscle fibres. The true number of motor units within a single muscle is not known but in degenerative neuromuscular diseases, these numbers decrease with time. In Amyotrophic lateral sclerosis (ALS), clinical assessment of muscle weakness is inadequate as a measure of disease progression because the loss of motor units is masked by a process known as collateral sprouting (Bjornskov et al., 1984). The aim of this paper is develop a methodology to obtain the posterior distribution of \( N \), the number of motor units supplying a muscle. In this way progression of the disease can be monitored.

Research into motor unit number estimation has spanned over 30 years. Methods have ranged from the original incremental technique (McComas et al., 1971), multiple point stimulation (Doherty and Brown, 1993), the “MUESA” method (Slawnych et al., 1996), spike triggered averaging (Bromberg, 1993), to the so-called Poisson or statistical method Daube (1995) and the recent, almost identical method of Blok et al. (2005) which assumes a binomial distribution instead of a Poisson. The incremental method of McComas et al. (1971) is thought to overestimate motor unit numbers because it fails to account for alternation (Stein and Yang, 1990), which is the firing of different combinations of motor units in response to a stimulus. This occurs because of the variability of the threshold of motor axons, such that the response to a stimulus is probabilistic. If \( n \) motor units are firing some of the time (this excludes motor units that never fire at that stimulus or motor units that are firing all of the time at that stimulus intensity) then there are \( 2^n - 1 \) possible increments of potential each one consisting of a different combination of motor units. This means that the number of distinct observed increments is usually greater than \( n \). Thus estimates of the motor unit numbers based on the counting of increments will be upwardly biased.

The Poisson statistical method allows for alternation in its estimate of motor unit numbers, but makes two assumptions. Firstly it assumes that all single motor unit action potentials are identical in size. Secondly it assumes that the number of motor units \( n \) that are firing stochastically (not all of the time) at a fixed stimulus has a Poisson distribution. In our model we make assumptions that better reflect the physiological mechanisms involved in the stimulus and response of motor units.

Our data collection protocol consists of measuring the magnitude of the muscle action potentials evoked by an electrical stimuli applied at the nerve. The response, known as compound muscle action potential (CMAP), is the summation of all the single motor unit action potentials (MUAPs) that are evoked by a given stimulus. In Ridall et al. (2006) we constructed a model that describes this response by accounting for individual unit properties: the mean and precision of the threshold of a single motor unit and the mean single MUAP size. These parameters are modelled as random effects. A large numbers of models conditioned on \( N \) were constructed and the model selected was the value of \( N \) that minimised the Bayes information criterion (BIC). The BIC was used to find an approximation to the log of the marginal posterior probability, \( \log p(N|y) \), the number of remaining motor units. However this approximation is asymptotic and its accuracy unknown. Furthermore, running multiple models to convergence is time consuming.

In this paper we construct an enlarged model where \( N \) is used as a stochastic variable rather than a fixed quantity. In this way we are able to obtain a posterior probability distribution which can be used to quantify the uncertainty of \( N \). In order to traverse model space, we make use of the theory of trans-dimensional modelling outlined in Green (1995). In a standard mixture model, each observation is allocated to one component of the mixture. In contrast with our model each observation is allocated a subset of motor units. This subset depends on a covariate, the magnitude of the stimulus level. Our model
is similar to the change-point model discussed in Green (1995) in that there is an ordering of one of our parameters.

A feature of our RJMCMC is that we use two different representations of the latent variable, that is used to indicate which motor units are firing for a given stimulus. Fixed dimension MCMC updates, where the number of remaining motor units, \( N \), is held constant, are efficiently carried out by using a Gaussian latent variable to represent the threshold. A unit fires if and only if the stimulus exceeds that threshold. Alternately, for moves that result in a change in \( N \), a binary latent variable is used to represent whether a unit is firing or not. This is so that reallocation of each observation to the subset of the parameters can be done by Gibbs sampling. A second characteristic of our RJMCMC is that stochastic moves for the allocation variable are required both for increasing the value of \( N \) and decreasing its value.

In Section 2 we discuss the data collection process and present some clinical details of our patients. In Section 3 we present considerations leading to a formulation of our model. In Section 4 we give the details of the RJMCMC that we use to infer \( N \). In Section 5 we present our results, in Section 6 a discussion and in Section 7 our conclusions. Appendices provide details of the RJMCMC algorithm.

2. Electrophysiological techniques

In our electrophysiological studies, a nerve is repeatedly electrically stimulated at an intensity and frequency that can be controlled and the response recorded by electrodes taped to the appropriate muscle group. We used a Viking IV EMG machine, modified to allow collection of large trains of stimuli of varying intensity. The evoked muscular response is known as the CMAP and is recorded as both CMAP area and amplitude.

Our protocol for data collection resembles that of McComas et al. (1971) in that a response to a graded stimulus is used but, unlike McComas et al. (1971), we use stimulus levels covering the whole range and we do not make subjective judgments of the location of increments in potential. For our study we use data from two patients whose clinical details are given in Table 1. Note that for each of our studies, the ulnar nerve was stimulated. However, we have also carried out successful studies with stimulation of the median and peroneal nerves.

Two studies, each comprising 400 observations, are taken from Patient 1 on the same day. These are used to illustrate both our fixed dimensional model and also the reproducibility of our method of MUNE. For Patient 2, nine studies of 550 observations each were carried out over an eighteen month period and these are used to show how our method can be used to quantify decline in motor unit numbers over the course of the disease, from the time when the muscle was of normal strength to when there was very little strength. The data for this a patient are shown in Figure 1. Note the appearance of increasingly large discontinuities in the curves as the disease progresses and how the maximum CMAP decreases with time.

3. Background to the fixed \( N \) model

Here we present a brief summary of the biological considerations which leads to a formulation of our model. A good non-mathematical account can be found in Benarroch et al. (1999).
Fig. 1. The figure shows a series of nine stimulus-response curves collected on patient 2 over an approximate eighteen month period. The horizontal axes show the stimulus in mA and the vertical axes show the CMAP area in $\mu V ms$. The dates are on which the studies were conducted are shown in the title for each study. Note the appearance of increasingly large discontinuities in the curves as the disease progresses. Note also how the maximum CMAP decreases with time.

Table 1. The table summarises the clinical details of two patients with ALS or a closely related disease. Note that all recordings were taken from the *abductor digiti minimi* (ADM) muscle in the hand upon stimulation of the ulnar nerve in the wrist.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Age</th>
<th>Nerve</th>
<th>Muscle</th>
<th>Clinical symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>80</td>
<td>Left</td>
<td>ADM</td>
<td>Severe upper and lower limb weakness.</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>66</td>
<td>Right</td>
<td>ADM</td>
<td>Severe lower limb weakness and progressive lower limb weakness.</td>
</tr>
</tbody>
</table>
We divide the description of the biology into the three components:

(i) the depolarisation of the axon membrane;
(ii) the single motor unit action potentials (MUAPs);
(iii) the summation of the single MUAPs to give the CMAP.

We illustrate the basic concepts by referring to Figure 2, which displays an analysis of a data set taken from patient 1.

The depolarisation of the axon membrane: This mechanism is well known and follows from the work of Hodgkin and Huxley (1952). At rest, sodium and potassium ion pumps maintain a constant potential across the membrane. An action potential only occurs if the applied stimulus exceeds the threshold of the axon at a particular instance of time causing depolarisation. This threshold for firing is thought not to be precise but variable over a range (Verveen, 1960) leading to a major source of stochasticity in the model. Furthermore, the threshold can be described by a Gaussian distribution (Bergmans, 1970), (Bostock et al., 1998), (Bruce et al., 1999). Upon the depolarisation of the axon membrane, sodium ions rush across the membrane into the axon and an action potential is transmitted along the axon. This is followed by a smaller movement of potassium ions in the reverse direction. Within a few milliseconds (the refractory period), the movement of ions is reversed and the original resting membrane potential is restored. In our studies this occurs well before the next stimulus arrives at either 1000 or 500 msec (depending on the frequency of stimulation) after the previous stimulus.

We describe the threshold of a particular unit by two unique excitability parameters, the threshold mean or the stimulus at which a motor unit has a 50% chance of firing and the threshold precision. These parameters are thought to be related to the number and type of open ion channels (Hales et al., 2004) and can be formulated as random effects in our model. The probability of each unit firing can be described by a sigmoidal cumulative Gaussian excitability curve (Stein and Yang, 1990). The shape and location of these curves are determined by the excitability parameters just described and are shown in Figure 2 (1B). The overlapping of these excitability curves, as in regions one and two of Figure 2 (1B), leads to an effect known as alternation. This means that for a given stimulus differing combinations of motor units fire leading to significant variability in the CMAP. In region one of Figure 2 (1B), five units are alternating, four are alternating in region two, three in region three and two in region four. In region one, where the model suggests five units, there are a maximum number of \( n = 2^5 - 1 = 31 \) different ways that at least one of these five units can fire.

The single MUAPs: The potential is propagated along the axons, through the terminal branches to the motor synapses where acetyl choline is released. This causes depolarisation of the muscle membrane and the generation of a motor unit action potential (MUAP) which is the summation of the action potentials of all the muscle fibres of that motor unit. Because the size of the MUAP varies between motor units (Feinstein et al., 1955) we describe the between unit variation of single MUAPs by a hyper-distribution. There is also variability within motor units, which may be due to stochastic events at the terminal branches and at the neuromuscular junction. Motor
unit variability is increased in ALS, and it is important to take this into account in the modelling. This is a major difference between our model and the Poisson method of Daube (1995), which does not allow for variation in single MUAP sizes.

The summation of the single MUAPs to give the CMAP: At the surface of the skin, an electrode records the CMAP. This is the sum of all the single MUAPs that fire at that instance as well as a component of baseline noise. Following Day and Hulliger (2001) we assume that the potentials are additive, although this is strictly the case only for rectified signals. The expected CMAPs from every sixth data point are shown as bars in Figure 2(1C) with the contributing single MUAPs shown as stripes within the bars, showing that the assumption gives an adequate description of the data. Figure 2 (2C) shows the expected CMAP area every second observation in region 1. Notice the large number of combinations of the five single MUAPs that can be activated but that this is considerably fewer than the maximum number possible, \( n = 31 \).

We now give a mathematical description of each of the above components.

3.1. Details of the statistical model

Let \( N \) be the number of motor units and let \( k \) represent a given motor unit where \( k = 1, 2, \ldots, N \). Let \( y_t \) be the measurement of CMAP area recorded at stimulus \( S_t \), for \( t = 1, 2, \ldots, T \), where \( T \) is the number of measurements taken. We denote all the observations of CMAP by \( y = \{y_1, y_2, \ldots, y_T\} \) and all the stimulus values by \( S = \{S_1, S_2, \ldots, S_T\} \).

3.1.1. The depolarisation of the axon membrane

We assume that the depolarisation of the axon membrane or the firing of the motor unit is independent both of the firing of other motor units and of whether the unit has fired at previous instances of time.

The firing or depolarisation of unit \( k \) at time \( t \) can be denoted by either a binary latent variable, \( s_{k,t} \), the firing state (either on or off), or a Gaussian latent variable, \( \tau_{k,t} \) the threshold. Unit \( k \) fires if and only if its threshold, \( \tau_{k,t} \), is exceeded by stimulus, \( S_t \), i.e.

\[
s_{k,t} = 1(S_t > \tau_{k,t}), \tag{1}
\]

where \( 1() \) is the indicator function.

Although the stimulus is assumed to be known, the threshold is assumed to vary randomly over a range defined by the normal distribution:

\[
\tau_{k,t} \sim N \left( m_k, \frac{1}{\delta_k^2} \right) . \tag{2}
\]

The range of the threshold for each unit is defined by two excitability parameters, \( m_k \) and \( \delta_k^2 \). The first is the mean threshold, \( m_k \), which is the stimulus at which a motor unit has a 50\% probability of firing and determines the horizontal location of the excitability curves shown in Figure 2 (1B) or (2B).
Fig. 2. This is an illustration of the implementation of our fixed \( N \) model on a data set taken from patient 1 whose clinical details are given in Table 1. In this model \( N = 14 \). From top left we have: Panel (1A) is the data set consisting of CMAP area, \( y_t \) plotted against the applied stimulus, \( S_t \). Note that the data has been divided into 4 regions. Region 1 shows the series of roughly horizontal rows of points that appear when the stimulus is close to \( S = 46mA \). The second, third and fourth regions show similar sets of lines when the stimulus is close to \( S = 52mA \), \( S = 56mA \) and \( S = 61mA \) respectively. Panel (2A) is an enlargement of region 1. Panel (1B) shows the excitability curves that describe the probability of each unit firing as a function of stimulus. The vertical axis is the probability of the applied stimulus (assumed fixed) exceeding the stochastic threshold, (2). These curves also split into four groups according to the portion of the data they describe. The numbers of overlapping curves describing each region are 5,4,3 and 2 respectively. An enlargement of the 5 excitability curves describing region 1 is shown in (2B). Note that in region 2 of the data, units 1-5 are all firing with near certainty. The mean threshold, \( m_1 \), is defined as the stimulus for which a unit has a 50 per cent probability of firing. The first three, \( m_1, m_2, m_3 \), are labelled in (2B). Panel (1C) shows the estimated total expected CMAP area, \( \mu_t \), of every sixth point in the data set as the stimulus increases. The shaded areas within each bar of the expected CMAP show the contributing single MUAPs. Again an enlargement describing region 1 is shown in Panel (2C). The observed CMAP area is marked by a ‘x’ symbol for a comparison with respect to the expected.
In setting a prior for \( \mathbf{m} \) we require one that gives economical values for the number of units, \( N \). We use a prior that penalises small spacings between the ordered values of \( m_k \). We apply the transformation

\[
s_j = m_j - m_{j-1}, \quad j = 1, 2, \ldots, N + 1
\]

for the ordered \( m_k \) where \( m_0 = \min(S) \) and \( m_{N+1} = \max(S) \) and, noting that the Jacobian for the transformation is 1, we assume a linearly transformed Dirichlet distribution on the spacings, \( \mathbf{s} = \{s_1, s_2, \ldots, s_{N+1}\} \), and transforming back to \( \mathbf{m} \) we obtain:

\[
p(\mathbf{m}) \propto \prod_{j=1}^{N+1} (m_j - m_{j-1})^{K-1}.
\]

For \( K = 1, 2, 3, \ldots \), this prior for \( \mathbf{m} \) is equivalent to taking the components of \( \mathbf{m} \) equal to the \( K \)th, \((2K)\)th, \( \ldots \) order statistics for a random sample of size \( K(N+1)-1 \) from the uniform distribution on the interval \((m_0, m_{N+1})\). This generalises the prior of Green (1995). The parameter \( K \) in (3) can be used to control the ‘repulsion’ between the values of \( m_k \) with \( K = 1 \) implying no ‘repulsion’. Note that \( \lim_{K \to \infty} \frac{s_i}{m_{N+1} - m_0} \to \frac{1}{N+1} \). We therefore suggest \( K \) should be reasonably small. In our analysis we set \( K = 2 \).

The second excitability parameter, \( \delta_k^2 \), is the precision of the threshold of that unit and determines the steepness of the corresponding excitability curve. The threshold precision parameters, \( \delta_k^2 \), are allocated Gamma random effects:

\[
\delta_k^2 \sim \text{Gamma}(\alpha_\delta, \beta_\delta)
\]

where \( \alpha_\delta \) and \( \beta_\delta \) assume non-informative uniform priors on the interval \((0, 10)\). We experimented with different values for the upper bounds of \( \alpha_\delta \) and \( \beta_\delta \) so that the posterior distribution of these parameters was unchanged by the choice of upper bound. The value of 10 used in the analysis achieved this requirement.

We denote all the parameters used in this section by

\[
\Theta_z = \{\mathbf{m}, \delta, \alpha_\delta, \beta_\delta\}.
\]

### 3.1.2. The single motor unit action potentials

Each single MUAPs is assumed to be independently distributed around the mean for that unit, \( \mu_k \), with a common variance \( \sigma^2 \). The single MUAPs, \( \mu_k \), are allocated Gamma distributions truncated on the left by \( \mu_{\min} \)

\[
\mu_k \sim \text{Gamma}(\alpha_{\mu}, \beta_{\mu}) \mathbf{1} (\mu_k > \mu_{\min}).
\]

The minimum size of a single MUAP, \( \mu_{\min} \), is not known and there has been considerable speculation about its size, (Henderson et al., 2006), (Bromberg, 2003). We experimented with the setting of this parameter and used the value that gave the best reproducibility in MUNE. This was a setting of \( \mu_{\min} = 100 \mu V \cdot ms \) and is similar to the value found by Shefner et al. (2004) although in that study the size was expressed as amplitude rather than area of the potential.
In our data analysis we allocated $k$ random effects by allowing $\alpha_\mu$ and $\beta_\mu$ to vary. These parameters are given uniform priors: $\alpha_\mu \sim U(0, 5)$ and $\beta_\mu^{-1} \sim U(0, 1000)$. Similar comments to that following (4) apply to the choice of these upper bounds.

We also allocate $\sigma^2$, a measure of the within unit variability, a non-informative inverse Gamma prior, $\sigma^2 \sim \text{Inv.Gamma}(\alpha_3, \beta_3)$ with $\alpha_3 = \beta_3 = 0.001$.

### 3.1.3. The summation of the single MUAPs to give the CMAP

Using the assumption of additivity, the expected CMAP will be the sum of expected single MUAPs and the expected level of baseline noise where the baseline noise is normally distributed around its mean, $\mu_b$, with variance $\sigma_b^2$. Similarly, the variance of the CMAP area will be the sum of the variances of those units that fire together with a component known as the baseline variance. However, in our historical studies non-Gaussian distributions are evident. Thus to account for long tailed and non-Gaussian variation the marginal distribution of $y$ about its expected value is taken to be a $t$-distribution with $2\gamma$ degrees of freedom. This is achieved by letting:

$$
(y_t \mid \tau_t, \mu, \mu_b, \sigma_b, \sigma, q_t) \sim N(\mu_t^T, V_t),
$$

$$
\mu_t^T = \mu_b + \sum_{k=1}^{N} s_{k,t} \mu_k
$$

(7)

$$
V_t = \frac{\sigma^2 n_t + \sigma_b^2}{q_t}, \quad n_t = \sum_{k=1}^{N} s_{k,t},
$$

(8)

$$
q_t \sim \text{Gamma}(\gamma, \gamma).
$$

(9)

where $s_{k,t}$ is given by (1). In our studies the parameter $\gamma$ of (9) was set to 2 giving $y$ marginally a $t$-distribution with 4 degrees of freedom.

The priors for the mean and variance of the baseline noise, $\mu_b$ and $\sigma_b^2$ respectively, are set using our historical studies. The correct setting of $\mu_b$ is crucial because if it is set too low then baseline noise can be confused with one small unit firing and $\sigma_b^2$ is underestimated. Many of our studies start with approximately 30-50 observations where the stimulus is so low that we can safely assume that no units are firing. To set our priors for $\mu_b$, we used the summaries of 21 studies. They had an average mean of $49(\pm 35) \mu Vms$ where the standard deviation (s.d.) is shown in brackets. For $\mu_b$ we use a Gamma prior with the same mean but twice this standard deviation. Thus $\mu_b$ is allocated a Gamma prior with $\alpha_1 = (\frac{\text{mean}}{2s.d.})^2 \approx 0.5$ and $\beta_1 = \frac{\text{mean}}{2s.d.} \approx 0.01$.

In the same 21 studies, we calculated the variances which yielded a mean of $1.743 \times 10^3(\pm 2.425 \times 10^3)$, with the standard deviation shown in brackets. In a similar way we allocate $\sigma_b^2$ an inverse Gamma prior with the same mean and four times our standard deviation; an inverse Gamma prior with $\alpha_4 = 2 + (\frac{\text{mean}}{4s.d.})^2 \approx 2.03$ and $\beta_4 = 1.743 \times 10^3(\alpha_4 - 1) \approx 1800$.

We denote all the parameters used here by

$$
\Theta_y = \{\mu, \mu_b, \sigma, \sigma_b, q, \alpha_\mu, \beta_\mu\}.
$$

(10)
4. The trans-dimensional Markov chain.

We now describe a Markov chain where $N$ can be used as a stochastic variable in an enlarged model which incorporates varying dimensional parameter spaces. We describe two general types of moves in the MCMC updates: the fixed $N$ moves which do not result in any change in the dimension of the model; and the varying $N$ moves which increase or decrease the value of $N$ by one. For the fixed $N$ moves, we find it convenient to use a continuous latent variable (the threshold, $\tau_{k,t}$) to indicate which units are firing. With varying $N$ moves, on the other hand, it is necessary to reallocate the firing state of each unit to each observation. This is more easily done if the latent state is binary when it can be done by carrying out Gibbs sampling. Therefore, for the purpose of varying $N$ moves, we re-express the latent Gaussian random variable, $\tau$, as a binary random variable, $s$, using (1). This binary latent variable has a prior distribution given by

$$p_{k,t} = p(s_{k,t} = 1 \mid m, \delta) = \Phi(\delta_k(S_t - m_k)) .$$

The terms in the above are described in Section 3.1.1.

The reverse reexpression of binary latent variables as Gaussian latent variables is now described. If the trans-dimensional proposal is accepted we re-express the newly created binary latent variables as Gaussian random variables so that the next fixed $N$ move can be made. To achieve this, and only for those units in the proposal, the continuous latent variable, $\tau$, is sampled from the truncated Gaussian distributions:

$$\tau_{k,t} \sim \left\{ \begin{array}{ll} \mathcal{N}(m_k, \frac{1}{\pi^k}) \mathbf{1}(\tau_{k,t} < S_t), & s_{k,t} = 1, \\ \mathcal{N}(m_k, \frac{1}{\pi^k}) \mathbf{1}(\tau_{k,t} > S_t), & s_{k,t} = 0, \end{array} \right.$$  

where the top line of (12) is the normal distribution truncated on the right by $S_t$ and the bottom line of (12) is the normal distribution truncated on the left by $S_t$.

4.1. The probability model

All the parameters (apart from $N$) are denoted by $\Theta = \{\Theta_z, \Theta_y\}$. where $\Theta_z$ is given by (5) and $\Theta_y$ is given by (10). The probability model is shown both in the directed acyclic graph in Figure 3 and mathematically below as

$$p(y, s, \Theta, N) = p(y \mid \Theta_y, s)p(s \mid \Theta_z, S, N)p(\Theta_y \mid N)p(\Theta_z \mid N)p(N)$$
where

\[ p(y | \Theta_y, s) = \prod_{t=1}^{T} \frac{e^{-\frac{(y_t - \mu_t)^2}{2 \sigma_t^2}}}{\sqrt{2\pi \sigma_t}}, \tag{14} \]

\[ p(s | \Theta_s, S, N) = \prod_{t=1}^{T} \prod_{k=1}^{N} p_{k,t}^{s_{k,t}} (1 - p_{k,t})^{1-s_{k,t}}, \tag{15} \]

\[ p(\Theta_y | N) = p(\sigma^2) p(\sigma^2) p(\mu_k) p(\mu_k) \prod_{k=1}^{N} p(\mu_k), \tag{16} \]

\[ p(\Theta_s | N) = p(\alpha) p(\beta) p(m) \prod_{k=1}^{N} p(\delta^2_k), \tag{17} \]

\[ p(N) \propto \begin{cases} 1, & N \in \{N_{\min}, \ldots, N_{\max}\} \\ 0, & \text{otherwise.} \end{cases} \tag{18} \]

Equation (14) is the likelihood of the observations conditional on the binary latent variable, \( s_{k,t} \). The components of this expression, \( \mu_t \) and \( \sigma_t \), are given in Section 3.1.3. Equation (15) is the distribution of the latent variable where \( p_{k,t} \), given by (11), denotes the probability of a unit firing at an instance of time. Equation (16) describes the priors and the random effects for those parameters that describe the observations given the latent variable, where the random effects, \( \mu_k \), are described in (6). Equation (17) describes the priors and the random effects for the parameters describing the latent variable where the random effects and \( \delta^2_k \), are described by (4). The prior for the number of units, (18), is given a discrete uniform prior with \( N_{\max} \) and \( N_{\min} \) set appropriately.

4.2. The types of moves

For cross model jumps, in our posterior simulation, we found that convergence was assisted by using moves of two types. The types of moves we use are listed below:

i) Update the unknowns without changing the dimension \( N \) of the model.

ii) Increase \( N \) by one (a birth) by taking two adjacent units, when ordered by the values of the mean threshold parameter, and replacing them by three adjacent units or the reverse move (a death).

iii) Increase \( N \) by one by taking one unit and replacing it by two adjacent units or the reverse move.

4.2.1. Fixed \( N \) move

These moves are standard Metropolis-Hastings within Gibbs sampling and are described in detail in Appendix A.

4.2.2. Varying \( N \) moves

We broadly discuss the ingredients of a trans-dimensional move following Green (1995). For all our move types, we consider increasing or decreasing values of \( N \) only by one. The
Fig. 3. The DAG represents the probability model, equation (13), implemented in this paper. The round nodes represent stochastic variables and the rectangular nodes represent known data. The binary latent variable is denoted by $s_k$, and the Gaussian latent variable by $\tau_{k,t}$. Note the parameters shown can be partitioned into two subsets: $\Theta = \{\Theta_z, \Theta_y\}$ where $\{m, \delta^2\} \subset \Theta_z$ given by (5) are parameters that describe the latent variable and $\{\mu, \mu_k, \sigma^2, \sigma_z^2, q\} \subset \Theta_y$ given by (10) are parameters that describe the data. The priors have been omitted to improve the appearance of the diagram. Note that as $N$ changes the dimension of all the variables on the upper plate also change.
change in the parameters and latent variable is given by:

\[ \{ N, u, \Theta^N, s^N \} \overset{N \rightarrow N+1}{\underset{N \rightarrow N+1}{\rightarrow}} \{ N + 1, \Theta^{N+1}, s^{N+1} \} \]

where \( u \) is a variable generated to achieve dimension matching in the continuous parameters (Green, 1995). Having proposed a new model, the probability of acceptance, \( A \), consists of three components: the posterior ratio, the proposal ratio and the Jacobian, giving

\[ A = \min[1, \text{(posterior ratio)} \times \text{(proposal ratio)} \times |\text{Jacobian}|]. \quad (19) \]

We discuss each of these terms briefly. Firstly the posterior ratio can be obtained from the probability model. We propose moves that change the dimension of the parameters \( m, \delta \) and \( \mu \) and the latent variable \( s \). Only the components involving these terms will be needed in the posterior ratio. The posterior distribution of all unknowns \( x = \{ \Theta, s, N \} \) is \( p(x | y) \propto p(y, s, \Theta, N) \) where \( p(y, s, \Theta, N) \) is given by (13).

Secondly, we consider the problem of creating proposals and subsequently calculating the proposal ratio. We determine different move types, \( M \), where the move type is selected from a discrete probability distribution chosen to achieve good mixing. Following Section 3.3 of Green (1995), the proposal ratio in (19) involves calculating the ratio of the product of the probability of changing the dimension and selecting the move type, the probability density of proposing the continuous parameters, \( \Theta \), and the probability of allocating of parameters to observations which, in our case, are represented by the binary latent variable \( s \). For the upward move or birth, \( N \rightarrow N + 1 \), the proposal probability is given by

\[
\text{proposal ratio} = \frac{q_{N \rightarrow N+1}^{N+1} \cdot p(u)}{q_{N \rightarrow N+1}^N \cdot p(u)}, \quad (20)
\]

where \( r_{M}^{N \rightarrow N+1} \) is the probability of proposing to increase \( N \) by one given that move type \( M \) has been selected. The expression \( q_{N \rightarrow N+1}^{N+1} \) is the probability of proposing new discrete parameters in the \( N + 1 \) model space and \( p(u) \) is the probability of choosing the random numbers needed to generate the extra continuous parameters in the \( N + 1 \) space. Similarly \( q_{M}^{N \rightarrow N+1} \) and \( r_{M}^{N \rightarrow N+1} \) are defined for the downward move of death \( N + 1 \rightarrow N \). The general form of the bijection on the continuous parameters is

\[
\begin{pmatrix}
\mu_N^N & u_1 \\
\delta_N^N & u_2 \\
m_N^N & u_3
\end{pmatrix}
\overset{\text{birth}}{\Rightarrow}
\begin{pmatrix}
\mu_{N+1}^{N+1} \\
\delta_{N+1}^{N+1} \\
m_{N+1}^{N+1}
\end{pmatrix},
\]

where the \( u_i \in u \) are random numbers used to create the extra dimension in the upward move (or birth) and to absorb a dimension in the reverse move. We discuss schemes where the transformation is given by (21). In these schemes most of the parameters of (21) are left unchanged. We now discuss two different schemes for increasing or decreasing the dimension by one. In the first, which is the more complex, two units are selected and replaced by three and the reverse of this and labelled moves 2-3 in Appendix B. The second scheme involves replacing one unit by two units and the reverse of this. These are labelled moves 4 and 5 in Appendix C.

The proposals for \( \mu \) must conserve their sum, \( \sum_i \mu_i \). Our motivation for moves 2-3 for the first scheme (shown in Appendix B) came from an inspection of the MCMC output of
Table 2. Here we compare two ways of moving down a dimension in our trans-dimensional model. The first move involves replacing three units (where two of the three are single MUAPS are nearly equal) by two units. The second involves replacing two units by one unit. The left-hand table shows the values of $\mu_i^5$ from the model that incorporates alternation and $\mu_i^6$ from the model that does not. Note that $\mu_i^5 + \mu_s^5 + \mu_s^6 \approx \mu_i^4 + \mu_s^4$, $\mu_i^4 \approx \mu_s^6 \approx \mu_i^3$ and $\mu_s^5 \approx \mu_s^2 - \mu_i^4$. Moves of this kind are incorporated in move types 2 and 3 and shown in Appendix B. The right-hand table compares $\mu_i^5$ from the 5 unit model and $\mu_i^4$ from the 4 unit model. Here we note $\mu_i^5 = \mu_i^3 + \mu_i^4$. Moves capable of describing such a transformation between these two models kind are incorporated in move types 4 and 5 and shown in Appendix C. The data are also analysed in Ridall et al. (2006) where MUNE using the BIC yields a most probable estimate of $N = 5$. Note, the units are in scaled CMAP area to facilitate the arithmetic.

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i^5$</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>$\mu_i^6$</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>9</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i^4$</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>$\mu_i^5$</td>
<td>20</td>
<td>20</td>
<td>24</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

two fixed $N$ models using the approach of Ridall et al. (2006) †. The data was taken from the left ulnar nerve of patient 3 whose clinical symptoms are described in Table 1. The left-hand panel of Table 2 is taken from realisations of the MCMC output for the “alternation model” for 5 units and the “no alternation” model for 6 units. If a shift from the 6 to the 5 unit model was to occur, units 3, 4 and 5 in the 6 unit model would be re-expressed by units 4 and 5 of the 5 unit model. The first three units would remain unchanged. Note $\mu_i^6 \approx \mu_s^6 \approx \mu_i^4$ and $\mu_s^6 \approx \mu_s^2 - \mu_i^4$. The first two units and the last unit are left unchanged. Appendix B moves preserve approximate equalities of this type. In the right-hand panel of Table 2 we show the motivation for the second simpler scheme shown in Appendix C. Here units 3 and 4 in the 5 unit model can be replaced by just one unit in the smaller model. The first two units and the last unit are unchanged. Here $\mu_i^5 + \mu_i^5 \approx \mu_i^3$. The proposals for $m$ and $\delta$ between these two classes of move types also differ subtly.

Not only does the proposal have to be considered for the continuous parameters but also for all the states (represented by the discrete latent variable) which indicate which units are firing

$$(s_{1:N,t}^N)^{\text{birth}} = (s_{1:N+1,t}^{N+1})^{\text{death}} \quad t = 1, 2, \ldots, T.$$ 

Most of the states are unchanged by the transformation. The remaining states, signifying which units are firing, are sampled from the full conditionals using Gibbs sampling. In the most complex scheme for each observation at instance $t$, two adjacent states are discarded in the lower dimensional model and reallocated to one of $2^3 = 8$ possibilities per observation in the higher dimensional model using Gibbs sampling. In the reverse direction the eight states for three units (for every observation) are replaced by one of $2^2 = 4$ possible states in the lower dimensional model. Note that the reallocation problem for MUNE is a more complex problem than it is for standard mixtures. In a split move for mixtures, the observations

†In the analysis of Ridall et al. (2006), the BIC indicated that the $N = 5$ model was the most probable model. The choice of the $N = 5$ model was later verified by the reversible jump methodology presented in this paper. However, the details of the RJMCMC analysis for this data set are not described in this paper.
belonging to one component are reallocated to one of the enlarged group of components using Gibbs sampling (Richardson and Green, 1997). In the reverse direction the merge move is deterministic. On the other hand, for the MUNE problem reallocation is done for every observation and rather than being deterministic, the allocation into lower dimensional model space is stochastic.

The last component of the acceptance probability given by (19) is the Jacobian for the bijection (21) which must be calculated for each move type. The details for this are given in Appendices B and C.

5. Results of our RJMCMC

In this section we analyse the data described in Section 2. The MCMC was implemented in Matlab. Multiple Markov chains were run on each data set with an equal number of iterations of the RJMCMC used for burn-in and recording the posterior distribution. To show convergence we use three methods. The first is a visual check to ensure that the each chain converges to the same distribution. The second is a visual check of the trace plots of $N$ to ensure the within chain variation is constant. The third is a formal method that tests for these kinds of convergence based on a three way ANOVA and is discussed in Section 5.1.1.

We now illustrate our model on two patients. For patients 1 we repeat the data collection process a short time later with the stimulating electrodes removed and repositioned differently. For patient 2 the data were collected at intervals over a period of about eighteen months. We use these data, firstly to show our technique can be used to measure progression of the disease and, secondly to show that our technique can be used for MUNE when $N$ has a moderate size.

5.1. Patient 1: A replicated study

Figure 4 Panels (1A)-(2A) show data collected from patient 1. His clinical details are given in Table 1. The recordings were taken in one session with a change in the position of the stimulating electrodes but with no change in the position of the recording electrodes. The trace plots Figure 4 (1B) and (2B) suggest convergence of the chain after a burn-in of 100,000 observations. Figure 4 (1C) and (2C) show the two posterior distributions recorded from 100,000 iterations after a similar burn-in period. A summary of the posterior distribution is given in Table 4. We note that there is a small difference in the posterior distributions for the two recordings but the modes are equal.

5.1.1. A formal test for convergence

In this section we demonstrate a formal test for convergence of our RJMCMC using the MCMC output from the first study on patient 1. Brooks and Giudici (1998) suggest a three way ANOVA on a continuous variable for assessing RJMCMC convergence diagnostics. However in our model the primary quantity of interest is $N$ the dimension of the model which is a count. We therefore conducted an analog of a three way ANOVA for a multinomial generalised linear model (GLM) on the counts in each category over the values that $N$ was distributed: $N \in \{12, \ldots, 18\}$. These counts from the MCMC were thinned at a rate of
Fig. 4. Panels (1A) and (2A) show two data sets collected from patient 1 with the recording electrode fixed but the position of the stimulating electrode altered. Panels (1B) and (2B) show the corresponding trace plots from the output of the RJMCMC after a burn-in of 100000 iterations. Panels (1C) and (2C) show the posterior distributions of the corresponding data sets.
Table 3. The table shows a three way ANOVA from a Poisson GLM

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Δ Dev</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between chain</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>category</td>
<td>6</td>
<td>481.23</td>
<td></td>
</tr>
<tr>
<td>chain</td>
<td>2</td>
<td>6.34 × 10⁻³</td>
<td>1.00</td>
</tr>
<tr>
<td>chain:category</td>
<td>12</td>
<td>13.39</td>
<td>0.34</td>
</tr>
<tr>
<td><strong>Within chain</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>segment</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>category:segment</td>
<td>6</td>
<td>7.12</td>
<td>0.31</td>
</tr>
<tr>
<td>chain:segment</td>
<td>2</td>
<td>0.07</td>
<td>0.97</td>
</tr>
<tr>
<td>chain:category:segment</td>
<td>12</td>
<td>4.01</td>
<td>0.98</td>
</tr>
</tbody>
</table>

one in ten thousand so that the independence assumption could reasonably be used. The three factors are the three chains, the two categories of segment of the MCMC, segment 1: iterations 1 to 500,000 and segment 2: iterations 500,001 to 1,000,000, with the third factor being seven categories of $N$.

In order to test that the three chains give the same distribution for the posterior of $N$, we refer to the deviance difference for the category×chain interaction term for the between chain ANOVA shown in Table 3 to $\chi^2_{12}$, obtaining a $p$-value of 0.34 suggesting no lack of support for the chains giving the same distribution of $N$.

In order to investigate convergence we compare the deviance difference for the three-way interaction, category×chain×segment in the within chain ANOVA (Table 3) and refer the value to $\chi^2_{12}$ obtaining a $p$-value of 0.98 suggesting no lack of support for convergence of the chains.

5.2. Patient 2: A serial study

Figure 1 Panels (1A)-(4A) show data collected from patient 2. His clinical details are given in Table 1. We use the study to demonstrate the decline in numbers over a period of about eighteen months. Figure 5 shows the corresponding posterior distributions which are summarised in Table 4. They show a slow decline in motor unit numbers. The corresponding trace plots for $N$ are shown in Figure 6 for one million observations after a similar burn-in time and convergence would appear to be satisfied.

6. Discussion

We have constructed a model using assumptions based on the current understanding of motor unit biology. In this paper we have demonstrated how the model can be used to obtain reproducible and reasonable estimates of motor unit numbers for muscles with up to about seventy motor units. In addition, our model can also be used to retrieve useful information about the population of units such as the distribution of excitability properties and the distribution of single MUAP sizes.
Fig. 5. The marginal posterior distributions, $p(N|y)$, of the data sets of Patient 2 displayed in Figure 1 are displayed. These were estimated from a run of a million iterations after a similar burn-in.

Table 4. A summary of the posterior distribution for each of the data sets used in this paper. It shows the dates on which the recordings were made, the mode of the posterior, $p(N|y)$, a 95% credible interval for the same posterior. The last column is the number of observations in the data set.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Date</th>
<th>Mode</th>
<th>95% C.I.</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dec 04</td>
<td>14</td>
<td>13 17</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Dec 04</td>
<td>14</td>
<td>14 16</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>May 04</td>
<td>76</td>
<td>69 84</td>
<td>562</td>
</tr>
<tr>
<td></td>
<td>Jan 05</td>
<td>56</td>
<td>53 61</td>
<td>2817</td>
</tr>
<tr>
<td></td>
<td>Feb 05</td>
<td>50</td>
<td>44 57</td>
<td>1241</td>
</tr>
<tr>
<td></td>
<td>April 05</td>
<td>43</td>
<td>39 49</td>
<td>1149</td>
</tr>
<tr>
<td></td>
<td>May 05</td>
<td>40</td>
<td>36 44</td>
<td>689</td>
</tr>
<tr>
<td></td>
<td>June 05</td>
<td>41</td>
<td>37 44</td>
<td>1176</td>
</tr>
<tr>
<td></td>
<td>July 05</td>
<td>38</td>
<td>33 42</td>
<td>773</td>
</tr>
<tr>
<td></td>
<td>Sept 05</td>
<td>28</td>
<td>26 32</td>
<td>1293</td>
</tr>
<tr>
<td></td>
<td>Nov 05</td>
<td>13</td>
<td>12 15</td>
<td>839</td>
</tr>
</tbody>
</table>
Fig. 6. The trace plots from the RJMCMC performed on the data displayed in Figure 1 taken from 1000000 iterations after a similar burn-in time.
The data collection protocol plays an important role in MUNE. For example a sufficient number of observations need to be collected for a stimulus intensity sufficiently low that it can be assumed that no units are firing. This provides information for $\mu_b$, the mean of the baseline noise, and $\sigma_b^2$, the variance of the baseline noise. At the moment our observations are collected manually but we are planning to improve our data collection protocol and automate the data collection process.

As the number of motor units increases, the rate of convergence reduces because of an increasing amount of missing information (See, for example, Liu (2001)). For these studies, it is therefore necessary to increase the number of observations. In a normal patient, it is thought that there are more than 100 motor units in muscle groups such as the abductor digiti minimi (Shefner, 2001). Estimates like this would lead, in our formulation of the model to the use of more than three hundred parameters and a large number of latent variables. Ways of simplifying the model so that the number of unknowns is reduced, need to be investigated.

The acceptance rate of the cross model jumps has been improved by a factor of about 5-10 by adopting the t-distribution for the distribution of the observations about their expected values. Other improvements should be considered. Possible ways of increasing the acceptance rate of RJMCMC are to use the ideas of Brooks et al. (2003). Firstly, we need to optimise the tuning parameters and secondly, to explore the use of correlated random variables for our proposals for the cross model jumps. We are also looking at simulated tempering where parallel chains run at higher temperatures which display faster mixing can be used for proposals for the RJMCMC (Liu, 2001).

We have used both informative and non-informative priors. Where there is substantial information about a parameter value, either from the literature or from historical studies, we have used that information in the choice of prior. Where there is little information we have used non-informative priors and investigated the sensitivity of the posterior to such choices.

6.1. Sensitivity analysis

We have already discussed the priors for $\mu_{\text{min}}$ in Section 3.1.1. Setting of $\mu_{\text{min}} = 0$ leads to a small increase in the modal estimate of $N$. In Section 3.1.2 we discussed the informative prior used for $\mu_b$ which if set too low causes an overestimate of MUNE. In the same section we discussed $\sigma_b^2$ which also has a mildly informative prior.

Here we examine the assumption of the size of the tail of the distribution of the observations about their expected values. The parameter that controls this is $\gamma$ where $2\gamma$ is the degrees of freedom of the t-distribution. To examine the sensitivity of $\gamma$ we used data set (1), the data shown in Figure 2 (1A) and data set (2), the data shown in Figure 1 (E). We used $\gamma = 2$ in our analysis but here we set $\gamma$ to 1.5 and 10 and compare our estimates of the posterior distribution of $N$. The left-hand side of Table 5 shows the relevant part of our sensitivity analysis for this parameter. Increasing the value of parameter $\gamma$ in our data set (2), where $N$ was moderately large, did result in an increased value of $\hat{N}$, the posterior modal estimate of $N$. This is because, for $\gamma$ taking a higher value, outlying observations or groups of observations can be explained by extra units. With $\gamma$ taking a low value, outlying observations can be explained with existing units. Data analysis (Henderson et al., 2006), suggests a long-tailed non-normal distribution supporting a t-distribution with a small number of degrees of freedom. The model could be extended to allow $\gamma$ to be an unknown parameter.
Table 5. The table shows the results for MUNE of sensitivity to various values of the parameters $\gamma$ and $K$. Here $2\gamma$, the degrees of freedom of the t-distribution, is a measure of the size of the tail of the distribution. Also $K$, the parameter of the Dirichlet distribution, governs the penalty on the size of the spacing between the values of $m$. The data sets used are shown in the table as (1),(2) and (3). Data set (1) is shown in Figure 2 (1A) and data set(2) is shown in Figure 1 (E).

<table>
<thead>
<tr>
<th>Data set</th>
<th>$K$</th>
<th>$\gamma$</th>
<th>95% C.I.</th>
<th>Mode</th>
<th>Data set</th>
<th>$K$</th>
<th>$\gamma$</th>
<th>95% C.I.</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2</td>
<td>1</td>
<td>13 18</td>
<td>14</td>
<td>(1)</td>
<td>2</td>
<td>1</td>
<td>13 18</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>13 17</td>
<td>14</td>
<td></td>
<td>2</td>
<td>2</td>
<td>13 17</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>13 17</td>
<td>14</td>
<td></td>
<td>3</td>
<td>2</td>
<td>13 16</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10</td>
<td>13 17</td>
<td>14</td>
<td></td>
<td>6</td>
<td>2</td>
<td>12 15</td>
<td>13</td>
</tr>
<tr>
<td>(2)</td>
<td>2</td>
<td>1</td>
<td>32 44</td>
<td>38</td>
<td>(2)</td>
<td>2</td>
<td>1</td>
<td>50 65</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>36 44</td>
<td>40</td>
<td></td>
<td>2</td>
<td>2</td>
<td>36 44</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>36 44</td>
<td>40</td>
<td></td>
<td>3</td>
<td>2</td>
<td>36 44</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10</td>
<td>36 43</td>
<td>40</td>
<td></td>
<td>6</td>
<td>2</td>
<td>31 29</td>
<td>34</td>
</tr>
</tbody>
</table>

Another important setting is the degree of repulsion, $K$, used for the prior of $m$. This was discussed in Section 3.1.1. In our model $K$ was set to 2. Here we repeat an experiment similar to the one described in the last paragraph except $K$ is set to 1,3 and 6. We use the data shown in Figure 2 (2A) and in Figure 1 (E). For data set (1) shown in Table 5 there is little sensitivity to the modal estimate $\hat{N}$ to changes in $K$. For data set (2) unrestricted spacing, or a setting of $K = 1$, results in a much higher value of $\hat{N}$ and an increased right tail in the posterior distribution for $N$. In a similar manner $K = 6$ results in lower values for $\hat{N}$ with less variation in the posterior. However the model is slightly sensitive to changes in values of $K = 2$ to $K = 3$. Larger values of $K$ essentially space out the values of $m$ and allow for little variation in spacing and overlapping of excitability curves. We find this implausible.

We examine the settings for the priors for the hyper-parameters $\delta_3$ and $\beta_3$ which prescribe the parameter $\delta_3^2$ (See Equation (4)). The hyper-priors for $\delta_3$ and $\beta_3$ were taken uniform distributions on $(0, 10)$. Varying this uniform distribution from $U(0, 5)$ to $U(0, 20)$ had very little impact on the posterior distribution of $N$. Similar comments apply to varying the hyper-priors of $\alpha_\mu$ and $\beta_\mu$ (See Equation (6)): for $\alpha_\mu$ the prior is changed from $U(0, 5)$ to $U(0, 10)$ and $\beta_\mu^{-1}$ from $U(0, 1000)$ to $U(0, 500)$.

Finally we examine the setting of $\alpha_3$ and $\beta_3$ which are the priors which control $\sigma^2$, the within unit variability. We let both of these parameters be equal to $10^{-6}$, $10^{-2}$ and $10^{-1}$ but found our estimate of $\hat{N}$ robust to these choices. If $\sigma^2$ were small then the value of $\beta_3 = 10^{-3}$ would be informative but typically $\sigma^2 > 1$ and the results for the full conditional of $\sigma^2$ given in Appendix A demonstrate the uninformative nature of the value of $10^{-3}$ for $\beta_3$ in this case.

7. Conclusion

MUNE is an important problem that has received considerable attention over a 35 year period but with no satisfactory and universally accepted method emerging (Shefner, 2001). Using current knowledge, in Ridall et al. (2006), we produced a set of assumptions to construct a Bayesian stochastic model where each unit was allocated a single MUAP size, two excitability parameters and an associated latent variable. In that work, a number of
models were run with $N$ being fixed. In the present work, the variable of interest, $N$, is a stochastic variable. We determine the marginal posterior distribution of the number of motor units using the approach of Green (1995). In order to improve the convergence properties of the chain, we have used a dual representation of the latent variable. For fixed $N$ moves, the latent variable is Gaussian whereas for variable $N$ moves, it is binary. The binary latent variable is necessary for allocating observations to units that are firing and is done by sampling from the full conditionals.

In this paper we have demonstrated the usefulness of our model by describing the rate of loss of units of a patient with rapidly progressing ALS. At the present time the only means of monitoring the progress of patients with neuromuscular diseases such as ALS is through medical examination and the observation of clinical symptoms such as muscle strength. We are not yet at the stage where MUNE using our method can be conducted routinely on all patients regardless of the presence of disease but we are optimistic that this goal is in sight.

Acknowledgements

We wish to thank the staff and patients of Royal Brisbane and Women’s Hospital for their full cooperation without which this research would not be possible and also to the editors and referees for their helpful comments on earlier drafts. The research of the first three authors was supported by Australian Research Council grants.

References


Appendix A
The full conditionals for within model MCMC. (Move type $M = 1$)

We use Metropolis-Hastings within Gibbs sampling.

$\tau_t \mid \ldots$

Let $\tau_t = \{\tau_{1,t}, \tau_{1,t}, \ldots, \tau_{1,t}\}$. Since $p(\tau_t \mid \ldots) \propto p(\tau_t \mid \Theta_t)p(y_t | \tau_t, \Theta_t, S_t)$ at each time instance, $t$, a proposal $\tau_t \rightarrow \tilde{\tau}_t$ is made from the term on the left above, the prior distribution:

$\tilde{\tau}_t \sim N(m, 1/\delta^2),$

and accept with probability calculated from the term on the right.

$\mu_b \mid \ldots$

We use (7) and (8) for our definition of $\mu_t^T$, $n_t$ and $V_t$. Since

$$p(\mu_b \mid \ldots) \propto p(\mu_b) \prod_{t=1}^T N(y_t; \mu_t^T, V_t)$$

$$\propto p(\mu_b) N\left(\mu_b; \frac{\sum_{t=1}^T y_t - \sum_{k=1}^N \mu_k s_{k,t}}{\sum_{t=1}^T 1/V_t}, \frac{1}{\sum_{t=1}^T 1/V_t}\right),$$

a proposal can be made from the term on the right and the acceptance probability calculated using the term on the left.

$\mu_k \mid \ldots$

We let $\tilde{\mu}_{k,t} = \mu_t^T - s_{k,t} \mu_k$. Since

$$p(\mu_k \mid \ldots) \propto p(\mu_k) N\left(\mu_k; \frac{\sum_{t=1}^T y_t - \sum_{k=1}^N \tilde{\mu}_{k,t}}{\sum_{t=1}^T 1/V_t}, \frac{1}{\sum_{t=1}^T 1/V_t}\right),$$

a proposal can be made from the term on the right and the acceptance probability calculated using the term on the left and given by (6).
The mode, \( \hat{\sigma}^2 \), of the full conditional is found by numerically solving for \( \sigma^2 \) in the equation

\[
\frac{\partial \log p(\sigma^2 | \ldots)}{\partial \sigma^2} = -\frac{1}{2} \sum_{t=1}^{T} \frac{n_t}{q_t V_t} + \frac{1}{2} \sum_{t=1}^{T} \frac{n_t(y_t - \mu_t^T)^2}{q_t V_t^2} + \frac{\beta_3}{\sigma^4} - \frac{\alpha_3 + 1}{\sigma^2} = 0,
\]

and approximating the variance by \( \hat{V}(\hat{\sigma}^2) = (-H(\hat{\sigma}^2))^{-1} \) at that mode where

\[
H(\hat{\sigma}^2) = 2 \sum_{t=1}^{T} \frac{n_t^2}{q_t^2 V_t^2} - \sum_{t=1}^{T} \frac{n_t^2(y_t - \mu_t^T)^2}{V_t^4 q_t^2} - \frac{2\beta_3}{\sigma^6} + \frac{\alpha_3 + 1}{\sigma^4} \bigg|_{\hat{\sigma}^2}.
\]

We make a proposal from the Gamma distribution Gamma(\( \alpha, \beta \)), with the same mode and variance as the full conditional where the parameters \( \alpha \) and \( \beta \) can be found by solving the simultaneous equations \( \frac{\alpha - 1}{\beta} = \hat{\sigma}^2 \) and \( \frac{\alpha}{\beta^2} = (-H(\hat{\sigma}^2))^{-1} \).

\[ \sigma_b^2 | \ldots \]

We use a similar scheme to that used for \( \sigma^2 \). The mode, \( \hat{\sigma}_b^2 \), is found by numerically solving for \( \sigma_b^2 \) in the equation

\[
\frac{\partial \log p(\mu, \mu_b, \sigma_b, \sigma | y)}{\partial \sigma_b^2} = -\frac{1}{2} \sum_{t=1}^{T} \frac{1}{q_t V_t} + \frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - \mu_t^T)^2}{q_t V_t^2} + \frac{\beta_4}{\sigma_b^4} - \frac{\alpha_4 + 1}{\sigma_b^2} = 0
\]

and approximating the variance by \( \hat{V}(\hat{\sigma}_b^2) = (-H(\hat{\sigma}_b^2))^{-1} \) at that mode where

\[
H(\hat{\sigma}_b^2) = \left( 2 \sum_{t=1}^{T} \frac{1}{q_t^2 V_t^2} - \sum_{t=1}^{T} \frac{(y_t - \mu_t^T)^2}{q_t^4 V_t^4} - \frac{2\beta_4}{\sigma_b^6} + \frac{\alpha_4 + 1}{\sigma_b^4} \right) \bigg|_{\hat{\sigma}_b^2}.
\]

We make a proposal from the Gamma distribution with the same mode and variance as the full conditional using the same method as above.

\[ \delta^2 | \ldots \]

For \( k = 1, 2, \ldots, N \) we sample \( \delta_k^2 \) from its full conditional:

\[
(\delta_k^2 | \ldots) \sim \text{Gamma} \left( \frac{T}{2} + \alpha_5, \frac{\sum_{t=1}^{T} (\tau_{k,t} - m_k)^2}{2} + \beta_5 \right).
\]

\[ m | \ldots \]

For \( k = 1, 2, \ldots, N \) we sample from the doubly truncated normal distribution

\[
\tilde{m}_k \sim N \left( \frac{\sum_{t=1}^{T} \tau_{k,t}}{T}, \frac{1}{T \delta_k^2} \right) 1(m_{k-1} < \tilde{m}_k < m_{k+1}),
\]

and use (3) to calculate the acceptance ratio.
We make a bivariate normal proposal from

\[
\begin{pmatrix}
\hat{\alpha}_\delta \\
\hat{\beta}_\delta 
\end{pmatrix} \sim N\left( \begin{pmatrix}
\hat{\alpha}_\delta \\
\hat{\beta}_\delta 
\end{pmatrix}, V \right).
\]

where \(\hat{\alpha}_\delta, \hat{\beta}_\delta\) are the maximum likelihood estimates of the parameters of a Gamma distribution fit to \(\delta^2\) and \(V\) is four times the inverse of Fisher’s information matrix for that fit.

We use a similar scheme to that used to update \(\alpha_\mu, \beta_\mu\).

For \(t = 1, 2, \ldots, T\) we sample \(q_t\) from its full conditional

\[
(q_t \mid \ldots) \sim \text{Gamma}(\gamma + \frac{1}{2}, \gamma + \frac{(y_t - \mu^T_t)^2}{2(n_t \sigma^2 + \sigma^2_b)}).
\]

Appendix B
Details of the varying \(N\) move types \(M = 2, 3\)

These moves consist of using 3 units to describe what was described by two units and the reverse move. For the birth move, we randomly select \(i \in \{2, 3, \ldots, N\}\) with equal probability and propose replacing units \(i - 1\) and \(i\) in \(N\) model space by three units, \(i - 1, i\) and \(i + 1\), in \(N + 1\) space. For the death move, in the reverse direction, \(N + 1 \rightarrow N\), (assuming there are \(N + 1\) units), we select \(i \in \{2, 3, \ldots, N\}\) and replace units \(i - 1, i\) and \(i + 1\) by two units. Note that \(p_{\text{birth}}(i) = p_{\text{death}}(i)\). The general form of the bijection on the continuous parameters is

\[
\begin{pmatrix}
\mu^N_{i-1;i} \\
\delta^N_{i-1;i} \\
m^N_{i-1;i} 
\end{pmatrix} \begin{pmatrix}
u_1 \\
u_2 \\
u_3 
\end{pmatrix} \begin{pmatrix}
\mu^{N+1}_{i-1;i+1} \\
\delta^{N+1}_{i-1;i+1} \\
m^{N+1}_{i-1;i+1} 
\end{pmatrix}.
\]

The random variables \(u_1, u_2\) are randomly generated from univariate distributions. \(u_1\) is simulated from the normal distribution, \(u_1 \sim N(0, \sigma^2_1)\), where \(\sigma^2_1\) is a tuning parameter. \(u_2\) is simulated from the Gamma distribution, \(u_2 \sim G(\nu, \nu)\) where \(\nu\) is a tuning parameter. The variable \(u_3\) is simulated from the uniform distribution: \(u_3 \sim U(0, 1)\). The corresponding transformation on the discrete variables is

\[
\begin{pmatrix}
\mu^N_{1:i-2} \\
\delta^N_{1:i-2} \\
m^N_{1:i-2} 
\end{pmatrix} \begin{pmatrix}
u_1 \\
u_2 \\
u_3 
\end{pmatrix} \begin{pmatrix}
\mu^{N+1}_{i-1;i+1} \\
\delta^{N+1}_{i-1;i+1} \\
m^{N+1}_{i-1;i+1} 
\end{pmatrix}.
\]

with the units before and after remaining unchanged but relabelled. For instance in the forward move, \(\mu^N_{1:i-2}\) will be relabelled as \(\mu^{N+1}_{1:i-2}\) and when \(\mu^N_{i-1;i+1}\) is inserted \(\mu^N_{i+1:N}\) will...
be relabelled as $\mu_{i+2:N+1}^{N+1}$ with corresponding changes in the labels for $\delta^{N+1}$, $m^{N+1}$ and $s^{N+1}$.

(a) The excitability parameters, $m$ and $\delta$.

The proposals from $N$ model space to $N+1$ model space for the excitability parameters are

\[
\begin{pmatrix}
\delta_{i-1}^{N+1} & \delta_i^{N+1} & \delta_{i+1}^{N+1} \\
m_{i-1}^{N+1} & m_i^{N+1} & m_{i+1}^{N+1}
\end{pmatrix}
:=
\begin{pmatrix}
\delta_{i-1}^{N} & \delta_i^{N+1} & \delta_{i+1}^{N} \\
m_{i-1}^{N} & m_i^{N} + u_2(m_i^{N} - m_{i-1}^{N}) & m_{i+1}^{N}
\end{pmatrix}.
\]

The inverse of this transformation from $N+1$ model space to $N$ model space for the excitability parameters, is

\[
\begin{pmatrix}
\delta_{i-1}^{N} & \delta_i^{N} & u_2 \\
m_{i-1}^{N} & m_i^{N} & u_3
\end{pmatrix}
:=
\begin{pmatrix}
\delta_{i-1}^{N+1} & \delta_i^{N+1} & \delta_{i+1}^{N+1} \\
m_{i-1}^{N+1} & m_i^{N+1} + u_3(m_i^{N+1} - m_{i+1}^{N+1}) & m_{i+1}^{N+1}
\end{pmatrix}.
\]

(b) The single MUAPS. The increments in area parameters, $\mu$.

The single MUAPS must be transformed, for both the $N \rightarrow N+1$ move and the $N+1 \rightarrow N$ move in such a way that their sum is conserved, that is $\sum_{j=i-1}^{i+1} \mu_j^N = \sum_{j=i-1}^{i+1} \mu_j^{N+1}$. If this constraint is not applied then the rest of the $\mu$s would have to be rescaled. We consider two possible moves, $M = 2, 3$, represented by the rows of the matrix below, designed for switching between the alternation and the no alternation models. For each of these there are six possible permutations.

\[
\begin{pmatrix}
\mu_{i-1}^{N+1} & \mu_i^{N+1} & \mu_{i+1}^{N+1} \\
\end{pmatrix}
:=
\begin{pmatrix}
\mu_{i-1}^{N} & \mu_i^{N} - \mu_{i-1}^{N} + u_1 & \mu_{i-1}^{N} - u_1 \\
\mu_i^{N} & \mu_{i+1}^{N} - \mu_i^{N} + u_1 & \mu_{i+1}^{N} - u_1
\end{pmatrix}.
\]

If the proposal does not satisfy the prior condition: $\mu_{i-1}^{N+1}, \mu_i^{N+1}, \mu_{i+1}^{N+1} > \mu_{\text{min}}$, it is rejected. For the inverse transformation we select $i \in \{2, 3, \ldots, N\}$ and propose replacing units $i-1$, $i$ and $i+1$ by two units. The transformation from $N+1$ model space to $N$ model space for the increments is

\[
\begin{pmatrix}
\mu_{i-1}^{N} & \mu_i^{N} \\
\end{pmatrix}
:=
\begin{pmatrix}
\mu_{i-1}^{N+1} + \mu_{i+1}^{N+1} & \mu_{i-1}^{N+1} - \mu_{i+1}^{N+1} \\
\mu_i^{N+1} + \mu_{i+1}^{N+1} & \mu_i^{N+1} - \mu_{i+1}^{N+1}
\end{pmatrix}.
\]

The right-hand column of these matrices show that if $\mu_{i-1}^{N+1}$ and $\mu_{i+1}^{N+1}$ differ greatly in magnitude then the move will be rejected with high probability.

(c) The discrete variable, $s$

To reallocate the new units to each observation, we propose discarding two of the allocations for each observation but keep the rest as shown in (23). Having new values of $\delta^{N+1}$, $\mu^{N+1}$ and $m^{N+1}$, the three new states $s_{i-1;i+1:t}^{N+1}$ for each observation are
selected by block Gibbs sampling from the full conditionals. $s_{t-1:i+1,t}^{N+1} \sim p(s_{t-1:i+1,t}^{N+1} \mid \ldots)$ where $s_{t-1:i+1,t}^{N+1} \in \{(0,0),(0,1),\ldots,(1,1,1)\}$ and

$$p(s_{t-1:i+1,t}^{N+1} \mid \ldots) \propto p(s_{t-1:i+1,t}^{N+1}, \delta_{t-1:i+1,t}^{N+1}, m_{t-1:i+1,t}^{N+1}) p(y_t \mid \Theta_y^{N+1}, s^{N+1})$$

(24)

and $p(s_{t-1:i+1,t}^{N+1}, \delta_{t-1:i+1,t}^{N+1}, m_{t-1:i+1,t}^{N+1}) = \prod_{j=t-1}^{t+1} p_{j,t}(s_{j,t}^{N+1} \mid \ldots)$ is given by (11). The proposal probability, needed to calculate the acceptance ratio, is given by $q^{N \rightarrow N+1} = \prod_{t=1}^{T} p_t(s_{t-1:t}^{N+1} \mid \ldots)$ where $p(s_{t-1:t}^{N+1} \mid \ldots)$ are the normalised full conditionals given in (24). The reverse proposal from three to two units must also be considered: $q^{N+1 \rightarrow N} = \prod_{t=1}^{T} p_t(s_{t-1:t}^{N} \mid \ldots)$ where

$$p(s_{t-1:t}^{N} \mid \ldots) \propto p(s_{t-1:t}^{N}, \delta_{t-1:t}^{N}, m_{t-1:t}^{N}) p(y_t \mid \Theta_y^{N}, s^{N})$$

and $s_{t-1:t}^{N} \in \{(0,0),(0,1),\ldots,(1,1,1)\}$. Finally, the birth move $\{x^N,u\} \rightarrow x^{N+1}$ is accepted with probability $A(x^N,x^{N+1})$ where $x^N = \{N,\Theta_y^N, s^N\}$ and

$$A(x^N,x^{N+1}) = \min \left(1, \frac{p(x^{N+1} \mid y) q^{N+1 \rightarrow N} r^{N+1 \rightarrow N} J}{p(x^N \mid y) q^{N \rightarrow N+1} r^{N \rightarrow N+1} \prod_{i=1}^{3} p(u_i)} \right),$$

(25)

where $r$ is described below equation (20) and where

$$J = \left| \frac{\partial \Theta^{N+1}}{\partial (u, \Theta^N)} \right|$$

$$= (m_i^N - m_{i-1}^N) \frac{\delta_i^N + \delta_{i-1}^N}{2}$$

$$= (m_{i+1}^N - m_{i-1}^N) \frac{\delta_{i+1}^N + \delta_{i-1}^N}{2}.$$ 

The acceptance ratio for the reverse of this move: the $N+1 \rightarrow N$ move, is $B(x^{N+1},x^N) = \min \left(1, \frac{1}{B(x^N,x^{N+1})} \right)$ where $B(x^N,x^{N+1})$ is given by the second term in brackets on the right-hand side of (25).

If the proposal is accepted the newly created binary random variables are re-expressed as Gaussian latent variables using (12).

**Appendix C**
**Details of the varying $N$ move types ($M = 4, 5$)**

Replace one unit by two units or the reverse move.

There are two moves of this type. For the birth move, we randomly select with equal probability $i \in \{1,2,\ldots,N\}$ and propose replacing that unit in $N$ model space by two units $i$ and $i+1$ in $N+1$ space. In the reverse direction (assuming there are $N$ units), we select
with equal probability \( i \in \{1, 2, \ldots, N - 1\} \) and replace units \( i \) and \( i + 1 \) by one unit. The general form of the bijection on the continuous parameters is

\[
\begin{pmatrix}
\mu_i^N \\
\delta_i^N \\
m_i^N
\end{pmatrix}
\begin{pmatrix}
\mu_i^{N+1} \\
\delta_i^{N+1} \\
m_i^{N+1}
\end{pmatrix}
\begin{pmatrix}
\mu_i^N \\
\delta_i^N \\
m_i^N
\end{pmatrix}
\]

birth

death

In the birth move, a new location and shape parameter are introduced and the area of one unit is shared among two others. The contribution of one unit to the observation is reallocated to two units.

(a) The excitability parameters, \( \delta, m \).

For \( \delta \) we propose two moves, \( M = 6, 7 \), which we depict by the rows of the matrix below.

\[
\begin{pmatrix}
\delta_i^{N+1} \\
\delta_i^{N+1}
\end{pmatrix}
\begin{pmatrix}
\delta_i^N \\
\delta_i^N
\end{pmatrix}
+ u_5 \delta_i^N \\
+ u_5 \delta_i^N
\]

and \( u_5 \sim \text{Gamma}(\nu, \nu) \) and \( \nu \) is a tuning parameter. For the reverse move from \( N + 1 \) to \( N \) model space (assuming there are \( N + 1 \) units):

\[
\begin{pmatrix}
\delta_i^N \\
\delta_i^N
\end{pmatrix}
+ u_5 \delta_i^N \\
+ u_5 \delta_i^N
\]

For the location parameters we propose two moves, \( M = 6, 7 \), which we depict by the rows of the matrix below.

\[
\begin{pmatrix}
m_i^{N+1} \\
m_i^{N+1}
\end{pmatrix}
\begin{pmatrix}
m_i^N \\
m_i^N
\end{pmatrix}
+ u_6 (m_i^N - m_i^{N-1}) \\
+ u_6 (m_i^N - m_i^{N-1})
\]

and where \( u_6 \sim U(0, 1) \). The reverse moves are

\[
\begin{pmatrix}
m_i^N \\
m_i^N
\end{pmatrix}
\begin{pmatrix}
m_i^{N+1} \\
m_i^{N+1}
\end{pmatrix}
+ m_i^N - m_i^{N+1} \\
+ m_i^N - m_i^{N+1}
\]

where \( m_i^{N+1} = \min(S_i) \) and \( m_i^{N+1} = \max(S_i) \)

(b) The single MUAPs, \( \mu \)

The transformation from \( N \) model space to \( N + 1 \) model space on \( \mu \) is:

\[
\begin{pmatrix}
\mu_i^{N+1} \\
\mu_i^{N+1}
\end{pmatrix}
\begin{pmatrix}
\mu_i^N \\
\mu_i^N
\end{pmatrix}
+ u_4 (\mu_i^N - \mu_{\min}) \\
+ u_4 (\mu_i^N - \mu_{\min})
\]

where \( u_4 \sim U(0, 1) \). Equation (27) was motivated by observing that in comparing \( N \) to \( N + 1 \) dimensional models one unit was shared among two others. The right-hand panel of Table 2 compares \( \mu_i^3 \) from the 5 unit model and \( \mu_i^4 \) from the 4 unit model. Note that \( \mu_i^3 \equiv \mu_i^3 + \mu_i^3 \). The reverse move from \( N + 1 \) to \( N \) model space is:

\[
\begin{pmatrix}
\mu_i^N \\
\mu_i^N
\end{pmatrix}
\begin{pmatrix}
\mu_i^{N+1} \\
\mu_i^{N+1}
\end{pmatrix}
+ \mu_i^{N+1} - \mu_{\min}
\]

\[
+ \mu_i^{N+1} - \mu_{\min}
\]
(c) The discrete variables, $s$

For the transformation $N \rightarrow N + 1$ one unit is reallocated to two units. This and the reverse move are given by

$$
\begin{pmatrix}
  s_i^{N,t} \\
  s_{i+1,t}^{N+1}
\end{pmatrix}
\xrightarrow{\text{birth}}
\begin{pmatrix}
  s_i^{N,t} \\
  s_{i+1,t}^{N+1}
\end{pmatrix}, \quad t = 1, 2, \ldots, T.
$$

The allocation is done by Gibbs sampling in the same manner as for the previous move. In the forward move, four possible reallocations are selected for each observation and in the reverse direction only two.

The acceptance ratio for the forward move $N \rightarrow N + 1$ given by (25) where for move $M = 4$

$$
J = \left| \frac{\partial \Theta^{N+1}}{\partial (u, \Theta^{N})} \right|
= \delta_i^N (m_i^N - m_{i-1}^N)(\mu_i^N - \mu_{\min})
= \delta_{i+1}^{N+1} (m_{i+1}^{N+1} - m_{i-1}^{N+1})(\mu_i^{N+1} + \mu_{i+1}^{N+1} - \mu_{\min}),
$$

and for move $M = 5$

$$
J = \delta_i^N (m_{i+1}^N - m_i^N)(\mu_i^N - \mu_{\min})
= \delta_{i+1}^{N+1} (m_{i+2}^{N+1} - m_i^{N+1})(\mu_i^{N+1} + \mu_{i+1}^{N+1} - \mu_{\min}).
$$