

## MODELLING MULTILEVEL SPATIAL BEHAVIOUR IN BINARY-MARK MUSCLE FIBRE CONFIGURATIONS

BY TILMAN M. DAVIES, MATTHEW R. SCHOFIELD, JON CORNWALL AND  
PHILIP W. SHEARD

*University of Otago*

The functional properties of skeletal muscles depend on the spatial arrangements of fast and slow muscle fibre types. Qualitative assessment of muscle configurations suggest that muscle disease and normal ageing are associated with visible changes in the spatial pattern, though a lack of statistical modelling hinders our ability to formally assess such trends. We design a nested Gaussian conditional autoregressive (CAR) model to quantify spatial features of dichotomously marked muscle fibre networks and implement it within a Bayesian framework. Our model is applied to data from a human skeletal muscle and results reveal spatial variation at multiple levels across the muscle. The model provides the foundation for future research in describing the extent of change to normal muscle fibre type parameters under experimental or pathological conditions.

**1. Introduction.** The ability of skeletal muscle to develop tension to move or to resist the movement of joints is fundamental to the ability of animals to interact with their environment. In humans, collections of individual muscle fibres are arranged into bundles called *fascicles*, and fascicles in turn are grouped together into the functional units we call muscles. The individual muscle fibres can be broadly categorised into *slow twitch* (“type I”) or *fast twitch* (“type II”) based on their relative speed of contraction.

Fibre type composition and distribution is crucial in the determination of muscle function (Cornwall and Kennedy (2015), Faulkner et al. (2007), Lexell, Taylor and Sjöström (1988), Miller, Woodley and Cornwall (2016)) with departures from “typical” spatial distributions likely to reflect age- and disease-related imperatives. Age-related alterations in the spatial distribution of muscle fibre types are thought to be due to changes in usage pattern, and these might arise by change to lifestyle or as a consequence of remodelling of nerve-muscle connections resulting from selective loss of nerves (Aare et al. (2016), Mosole et al. (2014)). The specific biological drivers for selective nerve loss in disease states or in normal ageing remain largely unknown, though consequential changes to muscle fibres is a well-established and conspicuous feature (Cornwall and Sheard (2012), Webster et al. (1988)). Assessment of muscle samples from individuals with diseases such as amyotrophic lateral sclerosis, type 2 diabetes and Duchenne muscular dystrophy

---

Received March 2018; revised September 2018.

*Key words and phrases.* Gaussian process, hierarchical model, Bayesian inference, physiology.

shows diagnostically important clustering of fibre types (Oberbach et al. (2006), Webster et al. (1988)). The spatial behaviours can be complex—in the instance of Duchenne muscular dystrophy, for example, there is a tendency for muscle fibres of the same type to cluster across the whole muscle as well as at the edges of individual fascicles (cf. Webster et al. (1988)).

As a result of these observations, fibre-type spatial distributions have been widely acknowledged as varying in some disease states and with age. Despite the potential importance of this conjecture with respect to understanding the pathogenesis of neuromuscular conditions and both normal and abnormal age-related changes, the spatial patterns formed by configurations of fast- and slow-twitch fibres have proved difficult to quantify—owed in part to a lack of relevant statistical tools. As a consequence relatively few formal inferences have been made concerning the manifestation of fibre-type distributions within or between muscle fascicles (Cornwall and Sheard (2012)) and, in turn, the ways in which any variation might be systematic or predictable.

Previous work has largely been limited to subjective visual inspections, as well as qualitative, semiquantitative and exploratory test-based approaches (see the overview in Davies, Cornwall and Sheard (2013)) with no development of robust statistical models that enable precise quantification of spatial distributions of fibre types across whole muscles or within muscle fascicles. Such a model would provide a tool to undertake research addressing basic biological principles, including how skeletal muscles change as we age and how neuromuscular pathologies manifest and progress.

Our objective is to develop a modelling framework to better understand the spatial behaviour of such fibre-type observations. We propose using thresholded hidden Gaussian conditional autoregressions to describe the binary response, following Weir and Pettitt (1999, 2000) and Pettitt, Weir and Hart (2002). These are part of a class of latent probit regression models (Chib and Greenberg (1998)). Our main methodological contribution lies in *nesting* a pair of latent Gaussian conditional autoregressive (CAR) processes, allowing us to describe spatial variation separately at the fibre and fascicle levels. The former describes how fibres vary within a fascicle, while the latter simultaneously allows the parameters describing this behaviour to vary spatially across the muscle. Sufficient flexibility exists in this framework for the incorporation of spatial covariate information, as we shall demonstrate with a measure of fibre and fascicle proximity to local boundaries.

The remainder of the article is structured as follows. In Section 2 we describe the type of data to be analysed, including origin and preprocessing, and present a novel data set to be analysed. Our proposed model is in Section 3 with model fitting outlined in Section 4. The results of application to the real data are given in Section 5. In Section 6 we provide concluding remarks on the extensibility of the modelling framework for coping with multiple muscles or subjects as well as additional examples of related research.

## 2. Data collection and processing.

**2.1. Dissection and immunohistochemistry.** While the scrutiny of muscle tissue can be performed for smaller biopsies taken from live individuals, larger sections (including cross-sections of entire muscles) are difficult to acquire and are most commonly accessed through the use of bodies donated to medical science. In either case identification of fast- and slow-twitch fibres is achieved by taking transverse sections through the specimen in such a way as to ensure all fibres within the muscle are represented in the section and by then processing the section using *immunohistochemistry*, where the two fibre types react differently to a specific chemical treatment—leaving one type darkly stained and the other unstained. The result is photographed through a microscope.

Figure 1 shows such a photograph which will serve as the focus of our analysis; fast-twitch fibres are stained dark. It represents a cross-section of an entire muscle taken from the neck of an elderly cadaver and processed in accordance with the institutional ethical requirements of the University of Otago and the New Zealand Human Tissue Act ([New Zealand Govt. \(2008\)](#)). A consequence of the immunohistochemistry procedure, observable in Figure 1, is the potential lateral displacement of some fascicles from their immediate neighbours. Noteworthy is the fact that there would be no such spacing between adjacent regions of fibres *in situ*, and so we ignore these large clefts in the geometrical identification of the neighbour networks (Section 2.2).

Image processing software ([Rueden et al. \(2017\)](#)) is subsequently used to identify individual fibres as planar coordinates and their corresponding type. Fascicles over the entire muscle were then identified independently by two experts in muscle biology and morphology based on the visual assessment of fibre arrangement and the identification of cleavage planes. These planes denote the region between adjacent fascicles, and their presence is commonly used to assist in the identification of contiguous bundles of fibres.

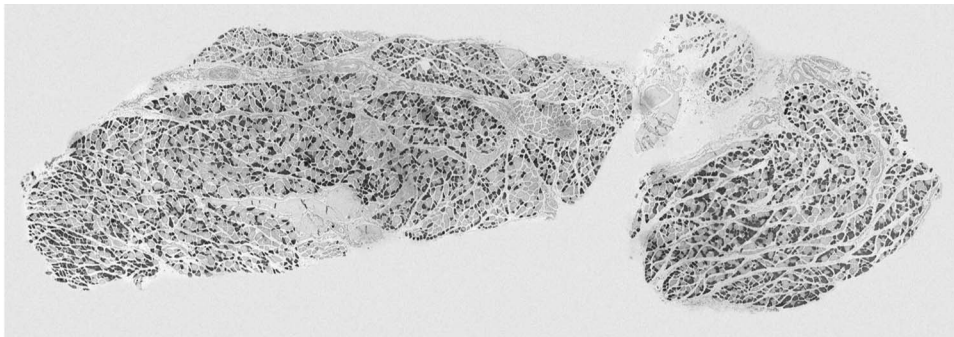


FIG. 1. Muscle cross-section following immunohistochemistry to identify slow-twitch fibres (light—unstained) and fast-twitch fibres (dark—stained).

**2.2. Computational geometry.** The next step is to obtain sensible geometric representations of the shapes of the fascicles. These constructs are important in our analysis for two reasons. First, they will be used to truncate the triangulations of each bundle of fibres so the estimated neighbour network in each case respects the irregular shape of the corresponding fascicle. Second, and perhaps more importantly, we must be able to compute for each fibre some measure of “distance from centre” which inherently requires knowledge of the boundary of the observation set.

In [Davies, Cornwall and Sheard \(2013\)](#) the three example fibre bundles are well approximated by simple convex hulls. This is inappropriate in the current analysis because many fascicles and indeed the overall muscle do not form convex geometric shapes. Instead, we turn to a lesser-known generalisation of the convex hull, namely the  $\alpha$ -shape ([Edelsbrunner et al. \(1983\)](#); see also [Pateiro-López and Rodríguez-Casal \(2010\)](#)), which is capable of producing geometric approximations that do respect concavity. To identify the neighbour networks, we use *Delaunay* triangulations (see, e.g., [Okabe et al. \(2000\)](#)); these are computed for each bundle of fibres for which there is precedent in the literature with respect to muscle data ([Davies, Cornwall and Sheard \(2013\)](#), [Fonseca et al. \(2003\)](#), [Pernuš \(1988\)](#)). Using the geometric centroids of each  $\alpha$ -shape, these steps are repeated at the fascicle level to identify the fascicle-to-fascicle neighbour network across the entire muscle. The interested reader is referred to Section 1 of Supplement A ([Davies et al. \(2019\)](#)) for further details on these geometric techniques.

The topmost image of Figure 2 shows the entire collection of 6246 fibres of our data as well as the neighbour network for each fibre bundle, presenting the first spatial layer of the hierarchical relationship we aim to model. The bottom image shows the second spatial layer of the hierarchy. This is provided as the muscle-wide  $\alpha$ -shape as well as all 199 fascicle-specific  $\alpha$ -shapes used to define the networks visible in the previous image. The fascicle centroids are marked off, and the muscle-wide fascicle neighbour network is superimposed. Hereinafter, for simplicity we will refer to type I, slow twitch fibres as “light” and type II, fast twitch fibres as “dark.”

It remains important to acknowledge that any geometric representation of either the fibre or fascicle neighbourhood structure is approximate, and we offer some additional discussion on this issue in Supplement A ([Davies et al. \(2019\)](#)). Informal sensitivity analysis, which involved manipulating the extent to which fibre neighbours were identified close to fascicle boundaries as well as altering the fascicle-level neighbour network to instead follow distance-based rules, had minimal effect on the final results for the data set studied here.

**3. Data and model.** We propose to model the data through nested CAR processes. This allows the binary responses to be informed at both the fibre level (according to spatial positioning relative to other fibres within the fascicle) and at

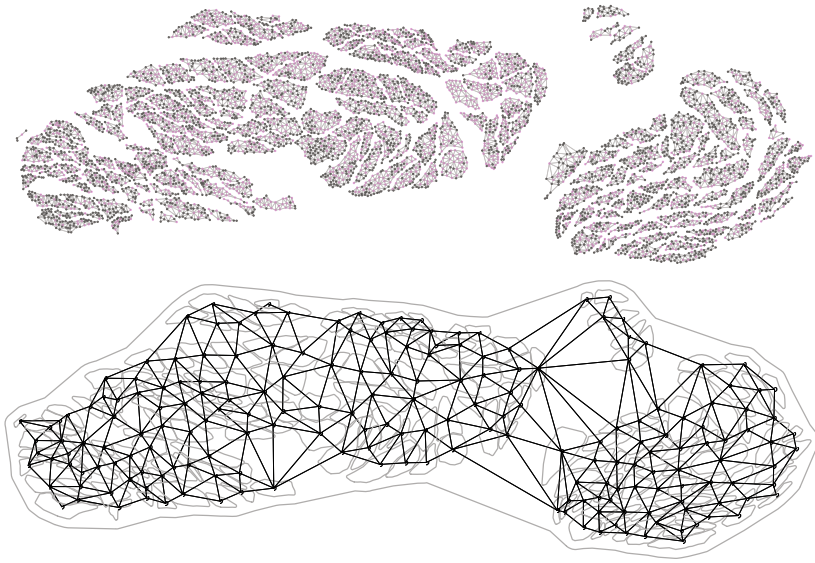


FIG. 2. Multilevel spatial structure of the muscle fibres and fascicles. Top: Individual fibres in their respective Delaunay neighbour networks (grey lines); “dark” fibres are black; “light” fibres are pink. Bottom:  $\alpha$ -shapes as the boundaries for the overall muscle and the fascicles (grey polygons), and the muscle-wide fascicle neighbour network as a Delaunay triangulation of the geometric centroids of the fascicle polygons (black lines).

the fascicle level (according to the spatial positioning of the fascicle within the muscle).

The fibre-level model uses the approach of Pettitt, Weir and Hart (2002) for CAR modelling of binary data. We extend their approach through inclusion of spatially referenced covariates. The covariate effects vary by fascicle and are hierarchically modelled with a further CAR process. The hierarchical model is a special case of the linear model of coregionalisation (Banerjee, Carlin and Gelfand (2015), Banerjee et al. (2008), Finley, Banerjee and Gelfand (2015), Gelfand et al. (2004), Ren and Banerjee (2013), Wackernagel (1998)).

The two CAR processes inform different aspects of the scientific application. The fascicle-level CAR process describes how the effects of the spatially referenced covariates vary throughout the muscle. This is of direct interest in understanding muscle function. The fibre-level CAR process describes the spatial variation of the observations after accounting for covariates. This is of interest insofar as it helps us to assess the extent to which the covariates fail to describe spatial patterns present in the observed data.

**3.1. Observation process.** Assume a data set comprising  $N$  individual muscle fibres. These are spatially arranged in bundles of  $K$  disjoint fascicles, with  $n_k$  fibres in fascicle  $k$ ,  $k \in \{1, \dots, K\}$ , such that  $N = n_1 + n_2 + \dots + n_K$ .

Each of the  $N$  fibres has a corresponding binary mark; let  $y_{ik} = 1$  if fibre  $i$  of fascicle  $k$  is dark with  $y_{ik} = 0$  for light. The vector  $\mathbf{y}_k = (y_{1k}, \dots, y_{n_k k})$  denotes the observations from fascicle  $k$  with the  $N$ -vector

$$(3.1) \quad \begin{aligned} \mathbf{y} &= [y_{11}, \dots, y_{n_1 1}, y_{12}, \dots, y_{n_2 2}, \dots, y_{1K}, \dots, y_{n_K K}]^T \\ &= [\mathbf{y}_1^T, \dots, \mathbf{y}_K^T]^T \end{aligned}$$

giving the observations across the entire muscle.

We commence design of our multilevel probit model by following [Chib and Greenberg \(1998\)](#), where the binary process is described in terms of a latent normal variable, thresholded at zero. We let  $z_{ik}$  denote a latent continuous variable for fibre  $i$  in fascicle  $k$ . The vector  $\mathbf{z}_k = (z_{1k}, \dots, z_{n_k k})$  denotes the values of the latent variable for the observations from fascicle  $k$  with the  $N$ -vector  $[\mathbf{z}_1^T, \dots, \mathbf{z}_K^T]^T$  giving the variables across the entire muscle in the same way as  $\mathbf{y}$ . The thresholding leads to a deterministic relationship between  $y_{ik}$  and  $z_{ik}$ :

$$(3.2) \quad \begin{aligned} p(y_{ik} = h | z_{ik}) &= \begin{cases} 1 & \text{if } I(z_{ik} > 0) = h, \\ 0 & \text{otherwise,} \end{cases} \\ i &\in \{1, \dots, n_k\}, k \in \{1, \dots, K\}, h \in \{0, 1\}, \\ p(\mathbf{y} | \mathbf{z}) &= \prod_{k=1}^K \prod_{i=1}^{n_k} p(y_{ik} | z_{ik}). \end{aligned}$$

In the following sections we outline the multilevel model for  $\mathbf{z}$ .

**3.2. Latent fibre-level process.** The aim is to describe how the type of fibre varies spatially within a fascicle, particularly with respect to proximity to the edge of the fascicle. We do this by considering the inverse distance of each fibre from fascicle edge as a covariate. This is found by first measuring  $\varepsilon_{ik}$ , the Euclidean distance in  $\mathbb{R}^2$  of the  $i$ th fibre in fascicle  $k$  to the nearest boundary of the  $\alpha$ -shape of its owning fascicle. We then invert this, yielding a metric that increases with distance from the middle of the fascicle, and standardise. To this end we define

$$\delta_{ik} = \frac{d_{ik} - \bar{d}_k}{s_{d_k}},$$

where  $d_{ik} = \varepsilon_{ik}^{-1}$  and  $\bar{d}_k$  and  $s_{d_k}$  are the mean and standard deviation respectively of  $d_{1k}, \dots, d_{n_k k}$ . The vector  $\boldsymbol{\delta}_k = [\delta_{1k}, \dots, \delta_{n_k k}]$  gives these standardised distances for all fibres in fascicle  $k$ , with the  $N$ -vector

$$(3.3) \quad \boldsymbol{\delta} = [\boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T, \dots, \boldsymbol{\delta}_K^T]^T$$

giving the distances for all fibres within the muscle.



There might exist other unmeasured drivers of the observed spatial behaviour. To capture potentially important residual effects, we consider a CAR error structure for the fibres in each fascicle. The neighbour network is represented using

$$\omega_{ijk} = \begin{cases} 1 & \text{if } i \text{ is a Delaunay neighbour of } j \text{ within fascicle } k, \\ 0 & \text{otherwise;} \end{cases}$$

$i, j \in \{1, \dots, n_k\}; k \in \{1, \dots, K\}$ . This allows us to consider the neighbour structure at the fascicle and muscle level,

(3.4)

$$\mathbf{\Omega}_k = \begin{bmatrix} 0 & \omega_{12k} & \dots & \omega_{1n_k k} \\ \omega_{21k} & 0 & \dots & \omega_{2n_k k} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n_k 1k} & \omega_{n_k 2k} & \dots & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_2 & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Omega}_K \end{bmatrix},$$

where  $\mathbf{0}$  denotes a matrix of zeros of arbitrary size. For any given  $k$ ,  $\omega_{iik} = 0 \forall i$ , so that all  $\mathbf{\Omega}_k$ 's and hence  $\mathbf{\Omega}$  have zeros on the main diagonal. This neighbour network is represented in the top of Figure 2.

We account for the conditional autoregressive spatial dependence in fascicle  $k$  through the specification of a precision matrix  $\mathbf{Q}_k(\rho)$ . Using similar notation as in Cressie (1993) and Pettitt, Weir and Hart (2002), we let

$$(3.5) \quad \mathbf{Q}_k(\rho) = \sigma^{-2} \mathbf{M}_k(\rho)^{-1} \{ \mathbf{I}_{n_k} - \mathbf{C}_k(\rho) \}; \quad \sigma^{-2} > 0, -\infty < \rho < \infty,$$

where  $\sigma^{-2}$  is the scalar, universal precision,  $\rho$  is a scalar, universal spatial dependence parameter, and  $\mathbf{I}_{n_k}$  is the  $n_k \times n_k$  identity matrix. Spatial dependence, based directly on the fibre neighbour network, is incorporated through the  $n_k \times n_k$  matrix  $\mathbf{C}_k(\rho)$ . The form of  $\mathbf{C}_k(\rho)$  requires presence of the corresponding matrix  $\mathbf{M}_k(\rho)$  in (3.5) to ensure positive definiteness of the covariance matrix itself. Due to the thresholding at zero, the parameter  $\sigma^{-2}$  is set to 1 to ensure identifiability; we suppress it in further equations.

The quantity  $\mathbf{M}_k(\rho) = \text{diag}[m_{1k}, \dots, m_{n_k k}]$  is a diagonal matrix with elements

$$m_{ik} = (1 + |\rho| \omega_{ik}^*)^{-1},$$

where  $\omega_{ik}^*$  is the number of neighbours of the  $i$ th fibre in the  $k$ th fascicle

$$(3.6) \quad \omega_{ik}^* = \sum_{j=1}^{n_k} \omega_{ijk}; \quad i, j \in \{1, \dots, n_k\}, k \in \{1, \dots, K\}.$$

This is equivalent to summing over the  $i$ th row or column of  $\mathbf{\Omega}_k$  in (3.4).

The matrix  $\mathbf{C}_k(\rho)$  is defined as

$$\mathbf{C}_k(\rho) = \begin{bmatrix} 0 & c_{12k} & \dots & c_{1n_k k} \\ c_{21k} & 0 & \dots & c_{2n_k k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n_k 1k} & c_{n_k 2k} & \dots & 0 \end{bmatrix} \quad \text{with } c_{ijk} = \rho \omega_{ijk} m_{ik}; i \neq j.$$

Defining  $\beta_{0k}$  as the fascicle-specific intercept and  $\beta_{1k}$  as a fascicle-specific effect of distance from edge, the corresponding model for  $\mathbf{z}_k$  is

$$p(\mathbf{z}_k | \boldsymbol{\beta}, \rho) = \text{MVN}(\mathbf{1}_{n_k} \beta_{0k} + \boldsymbol{\delta}_k \beta_{1k}, \mathbf{Q}_k(\rho)); \quad k \in \{1, \dots, K\},$$

where  $\mathbf{1}_{n_k}$  is a  $n_k$ -vector of ones and we define the collections  $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0K})^\top$ ,  $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{1K})^\top$ , and  $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_1^\top)^\top$ . The notation  $\text{MVN}(\mathbf{a}, \mathbf{B})$  denotes a multivariate normal distribution with mean  $\mathbf{a}$  and precision matrix  $\mathbf{B}$ .

The model  $p(\mathbf{z}_k | \boldsymbol{\beta}, \rho)$  is equivalent to the CAR model described in [Pettitt, Weir and Hart \(2002\)](#), with a linear model describing the mean. They remark upon several appealing properties of the above form of  $\mathbf{Q}_k(\rho)$ :

- The denominator terms  $m_{ik}$  in  $\mathbf{C}_k(\rho)$  and  $\mathbf{M}_k(\rho)$  allow  $\rho$  to be defined for all real values on  $(-\infty, +\infty)$ . A simpler version of these definitions, used by [Weir and Pettitt \(2000\)](#) for their analysis of Finnish toad data, involves removal of the  $(1 + |\rho| \omega_{ik}^*)^{-1}$  from the  $\mathbf{C}_k(\rho)$  (i.e., the denominator is set to 1). This leads to  $\mathbf{C}_k(\rho) = \rho \boldsymbol{\Omega}_k$  and removes the need for  $\mathbf{M}_k(\rho)$  in (3.5). However, the covariance matrix is only positive definite on a union of intervals determined by the eigenvalues of  $\boldsymbol{\Omega}_k$ ; hence  $\rho$  would only be permitted to take on particular values that correspond to intervals defined by the data observed. Our preference is to work with  $\mathbf{Q}_k(\rho)$  described above, so that the spatial dependency parameter  $\rho$  is comparable between data sets.
- The determinant  $|\mathbf{Q}_k(\rho)|$  can be computed efficiently, as described in Section 4. Such strategies were generalised by [Jin, Carlin and Banerjee \(2005\)](#) and allow for model fitting of fascicles with many fibres.

The joint model for  $\mathbf{z}$  is given as  $p(\mathbf{z} | \boldsymbol{\beta}, \rho) = \prod_{k=1}^K p(\mathbf{z}_k | \boldsymbol{\beta}, \rho)$ , due to the conditional independence between fascicles. This can also be written in a “global” form convenient for describing the fitting algorithms in Section 4 as

$$(3.7) \quad p(\mathbf{z} | \boldsymbol{\beta}, \rho) = \text{MVN}(\mathbf{A}_1 \boldsymbol{\beta}_0 + \mathbf{A}_\delta \boldsymbol{\beta}_1, \mathbf{Q}(\rho)),$$

where  $\mathbf{A}_1$  and  $\mathbf{A}_\delta$  are the appropriate  $n \times K$  block diagonal matrices

$$\mathbf{A}_1 = \text{diag}[\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_K}] \quad \text{and} \quad \mathbf{A}_\delta = \text{diag}[\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_K].$$

The global precision matrix

$$(3.8) \quad \mathbf{Q}(\rho) = \mathbf{M}(\rho)^{-1} \{ \mathbf{I}_N - \mathbf{C}(\rho) \}; \quad -\infty < \rho < \infty$$

is found by taking block diagonals of the fascicle-specific matrices:

$$(3.9) \quad \begin{aligned} \mathbf{M}(\rho) &= \text{diag}[\mathbf{M}_1(\rho), \dots, \mathbf{M}_K(\rho)]; \\ \mathbf{C}(\rho) &= \text{diag}[\mathbf{C}_1(\rho), \dots, \mathbf{C}_K(\rho)]. \end{aligned}$$

These model descriptions make clear that conditional on  $\boldsymbol{\beta}$  and  $\rho$  there is no covariation between fibres of different fascicles. We allow the elements of  $\boldsymbol{\beta}$  to co-vary as we describe in the next section.



**3.3. Fascicle-level process.** It is expected that the fascicle-specific parameters  $\beta_{0k}$  and  $\beta_{1k}$  will vary spatially across the muscle. As with the fibre-specific model we allow these parameters to depend on the standardised inverse distance from fascicle centre to muscle edge,  $\Delta_k = \frac{D_k - \bar{D}}{S_D}$ , where  $D_k$  is the Euclidean distance from centre of fascicle  $k$  to the edge of the muscle (obtained in the same way as earlier, now by inversion of the shortest distance from fascicle centroid to the muscle boundary), and  $\bar{D}$  and  $S_D$  are the sample mean and standard deviation of the  $D_1, \dots, D_K$ . The muscle-wide collection is denoted  $\mathbf{\Delta} = [\Delta_1, \Delta_2, \dots, \Delta_K]^T$ .

It is conceivable that the lateral displacement of some fascicles as a by-product of the immunohistochemistry procedure noted in Section 2.1 will have some effect on these measurements. For the purposes of the current analysis, we assume any such effects are negligible, supported in part by the use of the geometric centroid of each fascicle to calculate the nearest-edge lengths, coupled with the fact the muscle region is itself approximated by an  $\alpha$ -shape.

To account for spatial structure, we define the muscle-wide neighbour network illustrated on the bottom of Figure 2. As before we find the Delaunay neighbours; using the geometric centroids of the  $K$   $\alpha$ -shapes as the positional reference points,

$$(3.10) \quad \eta_{k\ell} = \begin{cases} 1 & \text{if fascicle } k \text{ is a neighbour of } \ell, \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\boldsymbol{\eta} = \begin{bmatrix} 0 & \eta_{12} & \dots & \eta_{1K} \\ \eta_{21} & 0 & \dots & \eta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{K1} & \eta_{K2} & \dots & 0 \end{bmatrix}$$

is the  $K \times K$  muscle-wide neighbourhood matrix.

We account for the spatial dependence across the muscle through specification of a  $K \times K$  precision matrix  $\mathbf{R}(\xi)$  that is of the same form as each  $\mathbf{Q}_k(\rho)$  in (3.5). Specifically, let

$$(3.11) \quad \mathbf{R}(\xi) = \mathbf{L}(\xi)^{-1} \{ \mathbf{I}_K - \mathbf{G}(\xi) \}; \quad -\infty < \xi < \infty,$$

where  $\mathbf{L}(\xi)$  is a diagonal matrix  $\mathbf{L}(\xi) = \text{diag}[l_1, \dots, l_K]$  with elements  $l_k = (1 + |\xi| \eta_k^*)^{-1}$  with  $\eta_k^* = \sum_{\ell=1}^K \eta_{k\ell}$  being the number of fascicle neighbours for fascicle  $k$ . Thus we have

$$(3.12) \quad \mathbf{G}_\xi = \begin{bmatrix} 0 & g_{12} & \dots & g_{1K} \\ g_{21} & 0 & \dots & g_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{K1} & g_{K2} & \dots & 0 \end{bmatrix} \quad \text{with } g_{k\ell} = \xi \eta_{k\ell} l_k; k \neq \ell.$$

We now define the  $K$ -vector  $\mathbf{u}$ ,

$$(3.13) \quad p(\mathbf{u}|\xi) = \text{MVN}(\mathbf{0}_K, \mathbf{R}(\xi)),$$

where  $\mathbf{0}_K$  is a  $K$ -vector of zeros. These variables contribute to the model for  $\beta_0$  and  $\beta_1$ ,

$$(3.14) \quad p(\beta_l | \theta_\beta, \mathbf{u}) = \text{MVN}(\alpha_l \mathbf{1}_K + \gamma_l \Delta + v_l \mathbf{u}, \tau_l \mathbf{I}_K); \quad l \in \{0, 1\},$$

where  $\alpha_l$  describes the intercept;  $\gamma_l$  is the effect on  $\beta_l$  of distance from fascicle centre to muscle edge;  $v_l$  gives the weightings of spatial variable  $\mathbf{u}$  on  $\beta_l$ , and  $\tau_l$  is the precision of  $\beta_l$ . In practice we model in terms of  $\sigma_l = \tau_l^{-0.5}$ ,  $l \in \{0, 1\}$  so that  $\theta_\beta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1, v_0, v_1, \sigma_0, \sigma_1)^\top$  is the collection of parameters used to describe  $\beta$ .

The variable  $\mathbf{u}$  is a common spatial process underlying both  $\beta_0$  and  $\beta_1$ . This reflects our prior belief that  $\beta_{0k}$  and  $\beta_{1k}$  will be correlated; a fascicle with a large positive intercept (a high proportion of dark fibres at mean distance from centre) is unlikely to have a large positive effect of distance. The model for  $\beta$  in (3.14) can be thought of as factor analysis, where the single factor varies spatially according to a CAR model. Note that the spatial variables  $\mathbf{u}$  and the parameters  $v_0$  and  $v_1$  can jointly change sign without affecting the model fit. To avoid sign switching problems, we postprocess the posterior draws to ensure  $v_0 > 0$ . If this involves changing the sign of  $v_0$ , we also reverse the sign of  $v_1$  and  $\mathbf{u}$ .

Equation (3.14) also highlights the aforementioned links to multivariate spatial coregionalisation models. Taking  $(\beta_{0k}, \beta_{1k})$  to be a bivariate response at each fascicle location,  $\mathbf{u}$  represents the key ingredient of a shared, spatially correlated field separately affecting the component-specific processes by way of  $v_0$  and  $v_1$  (cf.  $\mathbf{v}(\mathbf{s})$  in equation (7.20) of Banerjee, Carlin and Gelfand (2015)).

Finally, it is prudent to consider the potential for spatial confounding in mixed-effect spatial regression models, whereby fixed predictors co-vary alongside the spatial random effect (Hodges and Reich (2010)). In our specification of the between-fascicle process, we assume any such confounding is negligible given the single spatial covariate of “distance from muscle centre,”  $\Delta$ , is unlikely to be highly correlated with the Delaunay network defining spatial dependence in the random effect  $\mathbf{u}$ .

**4. Implementation.** We fit the multilevel model described above using Bayesian methods. Denoting the full collection of parameters as  $\theta = (\theta_\beta^\top, \rho, \xi)^\top$ , the joint posterior is

$$(4.1) \quad p(\mathbf{z}, \mathbf{u}, \beta, \theta | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{z}) p(\mathbf{z} | \beta, \rho) \prod_{l \in \{0, 1\}} p(\beta_l | \theta_\beta, \mathbf{u}) p(\mathbf{u} | \xi) p(\theta),$$

where the model terms are defined above and  $p(\theta)$  is the joint prior distribution. We assume parameters are independent a priori with

$$\begin{aligned} p(\sigma_l) &= t_{(0, \infty)}(0, b_\sigma, 3); & p(\alpha_l) &= \text{N}(0, b_\alpha); & p(\gamma_l) &= \text{N}(0, b_\gamma); \\ p(v_l) &= \text{N}(0, b_v); & p(\rho) &= \text{N}(0, b_\rho); & p(\xi) &= \text{N}(0, b_\xi), \end{aligned}$$

where  $l \in \{0, 1\}$ ;  $t_{(h_1, h_2)}(a, b, c)$  denotes a truncated t-distribution on the interval  $(h_1, h_2)$  with centre  $a$ , scale  $b^{-1}$  and  $c$  degrees of freedom, and  $N(a, b)$  denotes a univariate normal with mean  $a$  and precision  $b$ . We choose the hyperparameters  $b_{(\cdot)}$  so that we have a combination of weakly informative and vague priors as detailed in Section 2 of Supplement A (Davies et al. (2019)).

It is challenging to efficiently sample from the joint posterior distribution in (4.1). Separately updating all unknowns in turn leads to a Markov chain with impractically poor mixing. There is high posterior correlation between parameters. The parameters  $\sigma_l, l \in \{0, 1\}$  have extended sojourns near 0, leading to little movement in the corresponding parameters  $\beta_l$ . We explored the use of hierarchical centring approaches, and the problem persisted. Similarly, there is high dependence between  $\mathbf{u}$  and  $\xi$ . To overcome these problems, we use a scheme where we jointly update  $\sigma_l$  and  $\beta_l$ , and  $\mathbf{u}$  and  $\xi$ . We sample all other variables separately in turn from their full conditional densities if of known form, or using a random-walk Metropolis–Hastings algorithm otherwise. Pilot runs of several thousand iterations identify suitable values for the tuning parameters required for the Metropolis–Hastings proposals. An alternative is to consider adaptive tuning methods; see, for example, Andrieu and Thoms (2008). In what follows we outline the joint updates along with the full conditional densities for the other variables.

4.1. *Joint update of  $\sigma_l$  and  $\beta_l$ .* We jointly update  $\sigma_l$  and  $\beta_l$  together using a Metropolis step. We make use of a known full conditional density for  $\beta_l$  in determining the proposal density  $J(\cdot|\cdot)$ ,

$$J(\sigma_l^*, \beta_l^* | \theta_l^{(t)}) = J(\sigma_l^* | \sigma_l^{(t)}) p(\beta_l^* | \sigma_l^*, \dots), \quad l \in \{0, 1\},$$

where  $p(\beta_l^* | \sigma_l^*, \dots)$  is the full conditional distribution for  $\beta_l$ ,

$$p(\beta_l | \cdot) = \frac{p(\mathbf{z} | \beta, \rho) p(\beta_l | \theta_\beta, \mathbf{u})}{Z(\sigma_l)} = \text{MVN}(\Lambda_l^{-1} m_l, \Lambda_l),$$

with  $\Lambda_0 = \mathbf{A}_1^\top \mathbf{Q}(\rho) \mathbf{A}_1 + \tau_0 \mathbf{I}_K$ ,

$$m_0 = \mathbf{A}_1^\top \mathbf{Q}(\rho) (\mathbf{z} - \mathbf{A}_\delta \beta_1) + \tau_0 \mathbf{I}_K (\alpha_0 \mathbf{1}_K + \gamma_0 \Delta + \nu_0 \mathbf{u}),$$

and  $\Lambda_1 = \mathbf{A}_\delta^\top \mathbf{Q}(\rho) \mathbf{A}_\delta + \tau_1 \mathbf{I}_K$ ,

$$m_1 = \mathbf{A}_\delta^\top \mathbf{Q}(\rho) (\mathbf{z} - \mathbf{A}_1 \beta_0) + \tau_1 \mathbf{I}_K (\alpha_1 \mathbf{1}_K + \gamma_1 \Delta + \nu_1 \mathbf{u}).$$

We use the notation  $Z(\sigma_l)$  to make it clear that the normalising constant of this full conditional distribution depends on  $\sigma_l$  as well as other parameters not involved in this joint update. This choice of proposal leads to cancellation in the Metropolis acceptance probability,  $r$ ,

$$r = \frac{Z(\sigma_l^*) p(\sigma_l^*)}{Z(\sigma_l^{(t)}) p(\sigma_l^{(t)})} \frac{J(\sigma_l^{(t)} | \sigma_l^*)}{J(\sigma_l^* | \sigma_l^{(t)})},$$

such that it is a function of  $\sigma_l$  but not  $\beta_l$ . The proposal  $J(\sigma_l^*|\sigma_l^{(t)})$  is chosen to be a truncated normal distribution (truncated at zero) with a location parameter given by the current value  $\sigma_l^{(t)}$ .

4.2. *Full conditional distribution for latent variable  $\mathbf{z}$ .* The full conditional distribution for  $z_{ik}$  is  $p(z_{ik}|\cdot) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\boldsymbol{\beta}, \rho)$ . This has a truncated normal distribution

$$p(z_{ik}|\cdot) = \begin{cases} N_{[0,\infty)}(\bar{z}_{ik}, (1 + |\rho|\omega_{ik}^*)), & y_{ik} > 0, \\ N_{(-\infty,0)}(\bar{z}_{ik}, (1 + |\rho|\omega_{ik}^*)), & y_{ik} \leq 0, \end{cases}$$

where

$$\bar{z}_{ik} = \beta_{0k} + \beta_{1k}\delta_{ik} + \frac{\rho}{1 + |\rho|\omega_{ik}^*} \sum_{j=1}^{n_k} \omega_{ijk}(z_{jk} - \beta_{0k} - \beta_{1k}\delta_{jk}),$$

and  $N_{(h_1, h_2)}(a, b)$  is the truncated normal distribution on the interval  $(h_1, h_2)$  with mean  $a$  and precision  $b$ .

4.3. *Full conditional distribution for spatial parameter  $\rho$ .* The full conditional distribution for  $\rho$  is  $p(\rho|\cdot) \propto p(\mathbf{z}|\boldsymbol{\beta}, \rho)p(\rho)$  which is not of known form. Evaluating the proportional full conditional distribution presents a computational challenge as we require the determinant of  $\mathbf{Q}(\rho)$ , the global precision matrix. However, we can largely avoid this challenge by recognising that we can rewrite (3.8) in the following way:

$$\begin{aligned} \mathbf{Q}(\rho) &= \mathbf{I}_N + |\rho|\boldsymbol{\omega}^* - \rho\boldsymbol{\Omega} \\ &= \begin{cases} \mathbf{I}_N - \rho(\boldsymbol{\Omega} - \boldsymbol{\omega}^*), & \rho > 0, \\ \mathbf{I}_N, & \rho = 0, \\ \mathbf{I}_N - \rho(\boldsymbol{\Omega} + \boldsymbol{\omega}^*), & \rho < 0, \end{cases} \end{aligned}$$

where  $\boldsymbol{\Omega}$  is given in (3.4) and  $\boldsymbol{\omega}^* = \text{diag}[\omega_{11}^*, \dots, \omega_{n_1 1}^*, \dots, \omega_{1K}^*, \dots, \omega_{n_K K}^*]$  is a diagonal matrix of the  $\omega_{ij}^*$  values from (3.6) arranged in the natural order corresponding to  $\mathbf{C}(\rho)$  and  $\mathbf{M}(\rho)$  in (3.9). The above is equivalent to equation (2.6) in Pettitt, Weir and Hart (2002), and is important because it allows the determinant of  $\mathbf{Q}(\rho)$  to be found for any  $\rho$  based on precomputed eigenvalues of the matrices  $\boldsymbol{\Omega} - \boldsymbol{\omega}^*$  and  $\boldsymbol{\Omega} + \boldsymbol{\omega}^*$ . Thus, we avoid the heavy computational burden at each iteration of the algorithm, allowing for computation in large network settings.

4.4. *Joint update of  $\mathbf{u}$  and  $\xi$ .* We jointly update  $\xi$  and  $\mathbf{u}$  together using a Metropolis step as with  $\sigma_l$  and  $\beta_l$  above. The proposal density  $J(\cdot|\cdot)$  is

$$J(\xi^*, \mathbf{u}^*|\boldsymbol{\theta}_l^{(t)}) = J(\xi^*|\xi^{(t)})p(\mathbf{u}^*|\xi^*, \dots), \quad l \in \{0, 1\},$$

where  $p(\mathbf{u}^*|\xi^*, \dots)$  is the full conditional distribution for  $\mathbf{u}$ ,

$$p(\mathbf{u}|\cdot) = \frac{p(\boldsymbol{\beta}_0|\boldsymbol{\theta}_\beta, \mathbf{u}), p(\boldsymbol{\beta}_1|\boldsymbol{\theta}_\beta, \mathbf{u})p(\mathbf{u}|\xi)}{Z(\xi)} = \text{MVN}(\boldsymbol{\Gamma}^{-1}\mathbf{a}, \boldsymbol{\Gamma})$$

with  $\boldsymbol{\Gamma} = \mathbf{R}_\xi + \sum_{l \in \{0,1\}} \tau_l v_l^2 \mathbf{I}_K$  and  $\mathbf{a} = \sum_{l \in \{0,1\}} \tau_l v_l (\boldsymbol{\beta}_l - \alpha_l \mathbf{1}_K - \gamma_l \boldsymbol{\Delta})$ . The Metropolis acceptance probability,  $r$ , is  $r = \frac{Z(\xi^*)p(\xi^*)}{Z(\xi^{(t)})p(\xi^{(t)})} \frac{J(\xi^{(t)}|\xi^*)}{J(\xi^*|\xi^{(t)})}$ , that is, a function of  $\xi$  but not  $\mathbf{u}$ . The proposal  $J(\xi^*|\xi^{(t)})$  is a normal distribution centred on the current value  $\xi^{(t)}$  and truncated at zero.

**4.5. Remaining full conditional distributions.** The full conditional distributions for  $\alpha_l$ ,  $\gamma_l$  and  $v_l$  (indexed by  $l \in \{0, 1\}$ ) are

$$p(\alpha_l|\cdot) = \text{N}(\tau_l \psi_{\alpha_l}^{-1} (\boldsymbol{\beta}_l - \gamma_l \boldsymbol{\Delta} - v_l \mathbf{u})^\top \mathbf{1}_K, \psi_{\alpha_l}) \quad \text{where } \psi_{\alpha_l} = \tau_l K + b_\alpha;$$

$$p(\gamma_l|\cdot) = \text{N}(\tau_l \psi_{\gamma_l}^{-1} (\boldsymbol{\beta}_l - \alpha_l \mathbf{1}_K - v_l \mathbf{u})^\top \boldsymbol{\Delta}, \psi_{\gamma_l}) \quad \text{where } \psi_{\gamma_l} = \tau_l \boldsymbol{\Delta}^\top \boldsymbol{\Delta} + b_\gamma;$$

and

$$p(v_l|\cdot) = \text{N}(\tau_l \psi_{v_l}^{-1} (\boldsymbol{\beta}_l - \alpha_l \mathbf{1}_K - \gamma_l \boldsymbol{\Delta})^\top \mathbf{u}, \psi_{v_l}) \quad \text{where } \psi_{v_l} = \tau_l \mathbf{u}^\top \mathbf{u} + b_v.$$

**5. Results.** We fit the above model to the neck muscle data introduced in Section 2. Care must be taken with the starting value of  $\mathbf{z}$ , where the sign of all components must match the observed binary marks in  $\mathbf{y}$ . For the inverse cumulative distribution function of the standard normal,  $\Phi^{-1}$ , we set  $z_{ik}^{(0)} = \beta_{0k}^{(0)} = \Phi^{-1}(n_k^{-1} \sum_{j=1}^{n_k} y_{jk}) \forall i \in \{1, \dots, n_k\}; k \in \{1, \dots, K\}$ , with any instances of  $\pm\infty$  replaced accordingly by  $\pm 4$ . We run four parallel Markov chains with varying starting values. Each chain is run for 220,000 iterations, discarding the first 20,000 as burn-in and retaining every 20th value thereafter to reduce storage requirements.

Fascicle image plots for  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\beta}_1$  showing the posterior expectations are presented in Figure 3 based on the pooled results of the four chains. We provide a similar plot showing their posterior standard deviations in Section 3 of Supplement A (Davies et al. (2019)), where we also detail interactive 3-dimensional versions of Figure 3 (in which a third axis reflects posterior standard deviation); these are accessible online.

Pooled posterior summaries are provided in Table 1 for all scalar parameters. Plots and other diagnostics we used to assess convergence appear in Supplement A (Davies et al. (2019)), and the reader is provided with the data set and all necessary R code (R Core Team (2018)) to run the algorithm in Supplement B (Davies et al. (2019)).

The components of  $\boldsymbol{\beta}_0$  describe the relative dark/light count for each fascicle. There are two distinct spatial pockets of fascicles with a larger proportion of dark fibres in the left and rightmost extremes of the muscle, though we also note a middle corridor of “even” and “greater light” ratios. This agrees with a visual

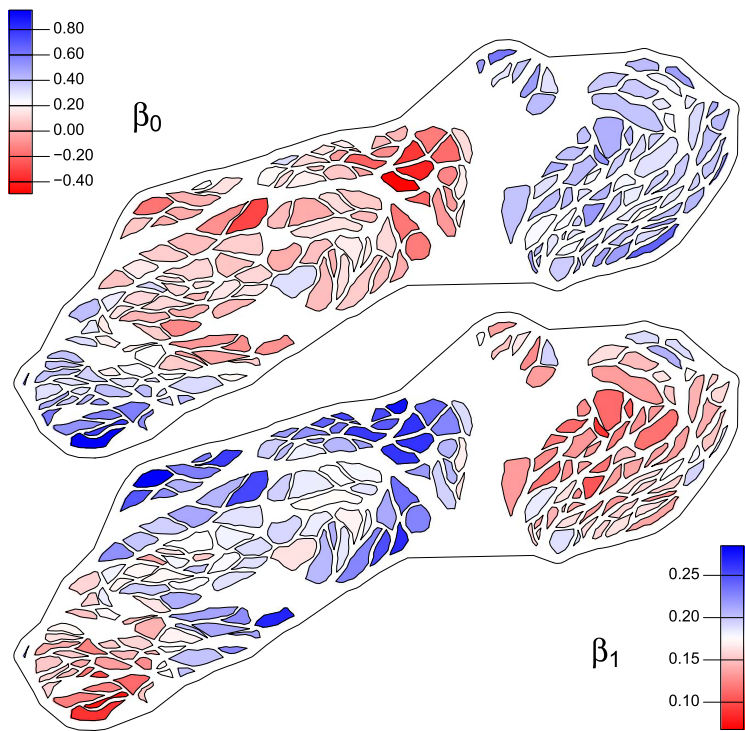


FIG. 3. Posterior expectations of all components of  $\beta_0$  (top) and  $\beta_1$  (bottom). The colour bands in each case are centred with white on the corresponding  $\alpha_1$  posterior mean.

TABLE 1  
Posterior summaries of scalar parameters. With the exception of  $\sigma_0$  and  $\sigma_1$ , bold entries depict those parameters whose corresponding credible interval excludes zero

Parameter	Post. expectation	95% cred. int.
$\rho$	0.0013	(−0.0243, 0.0276)
$\alpha_0$	<b>0.2300</b>	<b>(0.0336, 0.4369)</b>
$\gamma_0$	<b>0.0635</b>	<b>(0.0071, 0.1188)</b>
$\nu_0$	<b>1.342</b>	<b>(0.820, 2.009)</b>
$\sigma_0 (\tau_0^{-1/2})$	0.0607	(0.0027, 0.1479)
$\alpha_1$	<b>0.1764</b>	<b>(0.1215, 0.2269)</b>
$\gamma_1$	0.0134	(−0.0245, 0.0508)
$\nu_1$	<b>−0.2464</b>	<b>(−0.5221, −0.0219)</b>
$\sigma_1 (\tau_1^{-1/2})$	0.0402	(0.0017, 0.1030)
$\xi$	<b>14.99</b>	<b>(5.24, 28.61)</b>



inspection of the data in Figure 2 and provides evidence of nonuniform fibre type distributions across the muscle section.

The parameter  $\gamma_0$  describes how the  $\beta_{0k}$ 's are associated with distance from muscle centre. Its 95% credible interval excludes zero and provides weak evidence that the relative count of dark fibres *increases* in fascicles closer to the muscle edge.

Each  $\beta_{1k}$  describes the effect of distance from fascicle centre on the relative abundance of dark fibres for the  $k$ th fascicle. Overall positivity of these components suggests a tendency across the muscle of dark fibres within a given fascicle to become more abundant as we near the boundary of its owning fascicle. That said, this behaviour appears weaker—or at least harder to detect—in the aforementioned leftmost and rightmost pockets which tend to “own” more dark fibres to begin with. Elsewhere, as indicated by Figure 3, there is a stronger tendency of dark fibres to be positioned nearer the boundary for those fascicles with more even (and greater light) relative counts of the two fibre types. In terms of the statistical model, this aligns with a conjecture made in Section 3.3. Scientifically, this provides an interesting example of a previously unquantified phenomenon which has received little attention in the physiological literature.

Unlike the mean structure of  $\beta_0$ , however, the posterior of the slope parameter  $\gamma_1$  is not readily distinguishable from zero. Thus, while there is evidence the total amount of dark fibres in a given fascicle increases with fascicle proximity to muscle edge, the results suggest the magnitude of the effect of relative dark-to-light abundance as we move from centre to edge of a given fascicle is not itself affected by the position of the fascicle with respect to the edge of the muscle. This offers an interesting insight into the biological drivers for this process which indicates both across-muscle and within-muscle factors influencing fibre-type proportionality.

Strong positive spatial correlation is evident for the components within each of  $\beta_0$  and  $\beta_1$ , leading to the aforementioned pockets of similar spatial trends. These effects are due to the common spatial field  $\mathbf{u}$ ; the dependence parameter of which  $\xi$  exhibits a clear positively centred posterior. The opposite signs of the spatial offset coefficients  $\nu_0$  and  $\nu_1$ , accompanied in both instances by posterior credible intervals excluding zero, lead to the rough “inversion” of magnitudes of the  $(\beta_{0k}, \beta_{1k})$  pairs across the muscle.

A final remark concerns the fibre-type spatial dependence parameter  $\rho$ . With an associated interval straddling zero, this suggests that there does not remain any detectable *residual* systematic behaviour in the spatial configuration of the binary-mark fibres. This provides an important point of difference with previous work. In both Venema (1993) and Davies, Cornwall and Sheard (2013), detection of like-fibre spatial interaction via such a correlation parameter was the end goal, thereby merely allowing a researcher to search for evidence against random, independent dark/light configurations. In the current model, however, this spatial dependence parameter is a term only intended to capture unexplained spatial structure in the binary observations following adjustment by  $\beta_{0k}$  and  $\beta_{1k}$  for any given fascicle  $k$ . The inability to detect a nonzero  $\rho$  for these data means we have been able to

explain said spatial structure by the regression of relative dark/light counts with respect to distance from fascicle centre.

**6. Discussion.** Using a latent nested Gaussian CAR process we have demonstrated how one might effectively model dichotomously marked skeletal muscle fibre configurations. The proposed framework is specifically designed in such a way that the natural hierarchy of the spatial structure is itself used to inform various parameters taken to drive the binary patterning, thus providing quantifiable measures of the inherent spatial relationships, both within and between fascicles, across an entire muscle.

To illustrate the above, we analysed a cross-section of a human neck muscle, offering a novel impression of how fast- and slow-twitch fibres are arranged in their respective fascicles. Incorporation of deterministic predictors of the inherent spatial patterning, namely distance from fascicle and muscle centre, showed that we could partially explain the configuration at hand; flagged in particular by the corresponding lack of detection of residual spatial dependence at the fibre level. As far as we are aware this relationship has not been demonstrated elsewhere by any targeted statistical model and thus forms an important conclusion for wider consideration by the physiological research community.

From a statistical standpoint, the flexibility of the Gaussian CAR framework allows for very natural extensions. Our next step is to consider such processes for multiple muscles simultaneously, either by examining samples taken from multiple individuals, from multiple locations within a given individual or indeed both. For example, to answer age- or disease-related questions, we would envisage collecting such muscle data for multiple individuals by first identifying appropriate strata. The model could then be extended to incorporate a third “person” level, thereby allowing us to factor in the disease status or age of the muscle samples and allow joint estimation of relevant parameters in the presence of all potentially important factors. Advancements to fitting strategies for Bayesian probit regression models (Durante (2018)) offer promising options for improving sampling algorithms in these more complicated settings.

The class of statistical model discussed here can be applied to similar data sets in the physical and biological sciences. Studies in spatial ecology, for example, often deal with data that are hierarchically clustered in space (see, e.g., Takashina et al. (2018)). Nesting spatially dependent CAR processes would thus be useful for describing both local and global fluctuations in relevant characteristics of plant or animal populations. For similar physiological data sets as studied in this work, test-based tools (such as those we term “count statistics” in Davies, Cornwall and Sheard (2013)) are popular in their simplicity and, as such, their independent development continues today (for a recent example see Kelly et al. (2018)). However, such techniques applied in isolation lack sufficient flexibility to provide formal inferences regarding joint, interacting and/or latent unobserved drivers of systematic

spatial behaviours. Indeed, sophisticated hierarchical models offer a rich alternative for applications in the wider biological research arena where relatively simple summary statistics and test-based methods are employed (see an example and commentary in Davies, Sheard and Cornwall (2016), Makino, Funayama and Ikegaya (2016)). Investigations that marry targeted statistical modelling solutions with the pursuit of novel scientific hypotheses have the potential to provide unique insights into these kinds of complex biological processes.

**Acknowledgements.** The authors acknowledge M. Pollard for assistance in imaging and data collection. Two anonymous referees and an Associate Editor are thanked for their constructive feedback.

## SUPPLEMENTARY MATERIAL

**Geometry, Diagnostics, and  $\beta$  Posterior Plots** (DOI: [10.1214/18-AOAS1214SUPPA](https://doi.org/10.1214/18-AOAS1214SUPPA); .pdf). A PDF document providing additional examples of the geometrical treatment of the muscle data, various diagnostic plots of the MCMC, and additional graphics related to the sampled posteriors of  $\beta$ .

**R Code and Data Files** (DOI: [10.1214/18-AOAS1214SUPPB](https://doi.org/10.1214/18-AOAS1214SUPPB); .zip). A .zip archive containing R code and data necessary to repeat the analysis. Readers should open `muscle.R` and follow the instructions therein.

## REFERENCES

- AARE, S., SPENDIFF, S., VUDA, M., ELKRIEF, D., PEREZ, A., WU, Q., MAYAKI, D., HUSSAIN, S. N. A., HETTER, S. and HEPPLER, R. T. (2016). Failed reinnervation in ageing skeletal muscle. *Skeletal Muscle* **6** 29.
- ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. [MR2461882](#)
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. [MR3362184](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. [MR2523906](#)
- CHIB, S. and GREENBERG, E. (1998). Analysis of multivariate probit models. *Biometrika* **85** 347–361.
- CORNWALL, J. and KENNEDY, E. (2015). Fibre types of the anterior and lateral cervical muscles. *Eur. Spine J.* **24** 1986–1991.
- CORNWALL, J. and SHEARD, P. W. (2012). Spatial aggregation of type II fibres in human cervical muscles. *Clin. Anat.* **25** 533.
- CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. [MR1239641](#)
- DAVIES, T. M., CORNWALL, J. and SHEARD, P. W. (2013). Modelling dichotomously marked muscle fibre configurations. *Stat. Med.* **32** 4240–4258. [MR3118352](#)
- DAVIES, T. M., SHEARD, P. W. and CORNWALL, J. (2016). Comment on Makino et al. and observations on spatial modeling. *Anat. Sci. Int.* **91** 423–424.

- DAVIES, T. M., SCHOFIELD, M. R., CORNWALL, J. and SHEARD, P. W. (2019). Supplement to “Modelling multilevel spatial behaviour in binary-mark muscle fibre configurations.” DOI:10.1214/18-AOAS1214SUPPA, DOI:10.1214/18-AOAS1214SUPPB.
- DURANTE, D. (2018). Conjugate Bayes for probit regression via unified skew-normals. Available at [arXiv:1802.09565](https://arxiv.org/abs/1802.09565).
- EDELSBRUNNER, H., KIRKPATRICK, D. G. and SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory* **29** 551–559. [MR0713690](#)
- FAULKNER, J. A., LARKIN, L. M., CLAFLIN, D. R. and BROOKS, S. V. (2007). Age-related changes in the structure and function of skeletal muscles. *Clin. Exp. Pharmacol. Physiol.* **34** 1091–1096.
- FINLEY, A. O., BANERJEE, S. and GELFAND, A. E. (2015). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *J. Stat. Softw.* **63** 1–28.
- FONSECA, S., WILSON, I. J., HORGAN, G. W. and MALTIN, C. A. (2003). Slow fiber cluster pattern in pig longissimus thoracis muscle: Implications for myogenesis. *J. Anim. Sci.* **81** 973–983.
- GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. and SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST* **13** 263–312. [MR2154003](#)
- HODGES, J. S. and REICH, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *Amer. Statist.* **64** 325–334. [MR2758564](#)
- JIN, X., CARLIN, B. P. and BANERJEE, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61** 950–961. [MR2216188](#)
- KELLY, N. A., HAMMOND, K. G., STEC, M. J., BICKEL, C. S., WINDHAM, S. T., TUGGLE, S. C. and BAMMAN, M. M. (2018). Quantification and characterization of grouped type I myofibers in human aging. *Muscle Nerve* **57** E52–E59.
- LEXELL, J., TAYLOR, C. C. and SJÖSTRÖM, M. (1988). What is the cause of the ageing atrophy? Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15- to 83-year-old men. *J. Neurol. Sci.* **84** 275–294.
- MAKINO, K., FUNAYAMA, K. and IKEGAYA, Y. (2016). Spatial clusters of constitutively active neurons in mouse visual cortex. *Anat. Sci. Int.* **91** 188–195.
- MILLER, A., WOODLEY, S. and CORNWALL, J. (2016). Fibre types of the longus capitis and longus colli muscles in elderly females. *Anat. Sci. Int.* **91** 163–168.
- MOSOLE, S., CARRARO, U., KERN, H., LOEFLER, S., FRUHMANN, H., VOGELAUER, M., BURGGRAF, S., MAYR, W., KRENN, M., PATERNOSTRO-SLUGA, T. et al. (2014). Long-term high-level exercise promotes muscle reinnervation with age. *J. Neuropathol. Exp. Neurol.* **73** 284–294.
- NEW ZEALAND GOVT. (2008). Human Tissue Act. Public Act No 28, Wellington.
- OBERBACH, A., BOSSENZ, Y., LEHMANN, S., NIEBAUER, J., ADAMS, V., PASCHKE, R., SCHÖN, M. R., BLÜHER, M. and PUNKT, K. (2006). Altered fiber distribution and fiber-specific glycolytic and oxidative enzyme activity in skeletal muscle of patients with type 2 diabetes. *Diabetes Care* **29** 895–900.
- OKABE, A., BOOTS, B., SUGIHARA, K. and CHIU, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. [MR1770006](#)
- PATEIRO-LÓPEZ, B. and RODRÍGUEZ-CASAL, A. (2010). Generalizing the convex hull of a sample: The R package alphahull. *J. Stat. Softw.* **34** 1–28.
- PERNUŠ, F. (1988). The Delaunay triangulation and the shape hull as tools in muscle fibre analysis. *Pattern Recogn. Lett.* **8** 197–202.
- PETTITT, A. N., WEIR, I. S. and HART, A. G. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Stat. Comput.* **12** 353–367. [MR1951708](#)

- R CORE TEAM (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- REN, Q. and BANERJEE, S. (2013). Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. *Biometrics* **69** 19–30. [MR3058048](#)
- RUEDEN, C. T., SCHINDELIN, J., HINER, M. C., DEZONIA, B. E., WALTER, A. E., ARENA, E. T. and ELICEIRI, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinform.* **18** 529.
- TAKASHINA, N., BEGER, M., KUSUMOTO, B., RATHNAYAKE, S. and POSSINGHAM, H. P. (2018). A theory for ecological survey methods to map individual distributions. *Theor. Ecol.* **11** 213–223.
- VENEMA, H. W. (1993). Estimation of the parameters of a binary Markov random field on a graph with application to fibre type distributions in a muscle cross-section. *IMA J. Math. Appl. Med. Biol.* **10** 115–133.
- WACKERNAGEL, H. (1998). *Multivariate Geostatistics: An Introduction with Applications*, 2nd ed. Springer, New York.
- WEBSTER, C., SILBERSTEIN, L., HAYS, A. P. and BLAU, H. M. (1988). Fast muscle fibers are preferentially affected in Duchenne muscular dystrophy. *Cell* **52** 503–513.
- WEIR, I. S. and PETTITT, A. N. (1999). Spatial modelling for binary data using a hidden conditional autoregressive Gaussian process: A multivariate extension of the probit model. *Stat. Comput.* **9** 77–86.
- WEIR, I. S. and PETTITT, A. N. (2000). Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **49** 473–484. [MR1824553](#)

T. M. DAVIES  
 M. R. SCHOFIELD  
 DEPARTMENT OF MATHEMATICS AND STATISTICS  
 UNIVERSITY OF OTAGO  
 PO BOX 56, DUNEDIN 9054  
 NEW ZEALAND  
 E-MAIL: [tdavies@maths.otago.ac.nz](mailto:tdavies@maths.otago.ac.nz)  
[mschofield@maths.otago.ac.nz](mailto:mschofield@maths.otago.ac.nz)

J. CORNWALL  
 CENTRE FOR EARLY LEARNING IN MEDICINE  
 OTAGO MEDICAL SCHOOL  
 UNIVERSITY OF OTAGO  
 PO BOX 56, DUNEDIN 9054  
 NEW ZEALAND  
 E-MAIL: [jon.cornwall@otago.ac.nz](mailto:jon.cornwall@otago.ac.nz)

P. W. SHEARD  
 DEPARTMENT OF PHYSIOLOGY  
 UNIVERSITY OF OTAGO  
 PO BOX 56, DUNEDIN 9054  
 NEW ZEALAND  
 E-MAIL: [phil.sheard@otago.ac.nz](mailto:phil.sheard@otago.ac.nz)  
 URL: <http://phsl.otago.ac.nz/peoplelab.php?lab=22>