Supplement to "Mechanical heterogeneities in the subendothelial matrix develop with age and decrease with exercise"

The additive model

To assess the age and exercise effects on arterial stiffness, we estimated the following additive model: for mouse i and spatial location $\boldsymbol{s} = (s_x, s_y)$ of the elastic modulus map,

$$Y_i(\boldsymbol{s}) = \boldsymbol{x}_i'\boldsymbol{\beta} + f_i(\boldsymbol{s}) + \epsilon_i(\boldsymbol{s}) \tag{1}$$

where $Y_i(\boldsymbol{s})$ is the square-root-transformed elastic modulus, $\boldsymbol{x}_i$ is the vector of predictors, $\boldsymbol{\beta}$ is the corresponding vector of unknown regression coefficients to be estimated, $f_i(\boldsymbol{s})$ is the unknown spatial variability term to be estimated, and $\epsilon_i(\boldsymbol{s})$ is the independent and identically distributed measurement error, which is assumed to be normally distributed. The square-root transformation of the elastic modulus is necessary to satisfy the normality assumption for the measurement errors, $\epsilon_i(\boldsymbol{s})$. The vector of predictors, $\boldsymbol{x}_i$, contains indicators for the age and exercise groups of each mouse, and does not depend on the spatial location. The regression coefficients, $\boldsymbol{\beta}$, may be interpreted as in a standard linear regression model, and inference may be conducted on the group effects. The spatial locations, $\boldsymbol{s}$, are not directly comparable across mice: while the AFM measurements were recorded on an 11×11 point grid within a $100\mu\text{m} \times 100\mu\text{m}$ area of each artery, the precise locations within each artery cannot be made identical across mice. As a result, the spatial terms, $f_i(\boldsymbol{s})$, must be mouse-specific, and an ideal estimator of $f_i(\boldsymbol{s})$ should be invariant to rotations of the spatial coordinate system for $\boldsymbol{s}$. The *thin plate spline* satisfies this and other optimality and smoothness properties, and is easy to implement within the `mgcv` package in `R`.

For each spatial effects term, $f_i(\boldsymbol{s})$, the thin plate spline produces an accompanying *smoothing parameter*. Consider a simplified setting in which we wish to model a response, y_i, as a smooth function of a (one-dimensional) predictor, x_i: $y_i = f(x_i) + \epsilon_i$, where ϵ_i is a random error term. The thin plate spline is the estimator of f which minimizes the penalized least squares criterion

$$\sum_i (y_i - f(x_i))^2 + \lambda J_1(f) \tag{2}$$

where $J_1(f) = \int [f''(x)]^2 \, dx$ penalizes functions with large (in magnitude) second derivatives, i.e., rougher functions. The smoothing parameter, $\lambda > 0$, controls the trade-off between the first term, which measures goodness-of-fit, and the second term, which measures roughness. As $\lambda \to 0$, the first term dominates, and f approaches a rough interpolation of the data, y_i; as $\lambda \to \infty$, the second term dominates, and f approaches a linear function. More details can be found in (Wood, 2006).

Under model (1), the penalized least squares criterion (2) generalizes to

$$\sum_{i,\boldsymbol{s}} \left(Y_i(\boldsymbol{s}) - \boldsymbol{x}_i'\boldsymbol{\beta} - f_i(\boldsymbol{s})\right)^2 + \sum_i \lambda_i J_2(f_i) \tag{3}$$

where

$$J_2(f) = \int\int \left(\frac{\partial^2 f(s_x, s_y)}{\partial s_x^2}\right)^2 + \left(\frac{\partial^2 f(s_x, s_y)}{\partial s_x \partial s_y}\right)^2 + \left(\frac{\partial^2 f(s_x, s_y)}{\partial s_y^2}\right)^2 \, ds_x \, ds_y,$$

which extends J_1 using partial derivatives to measure roughness in both spatial directions s_x and s_y. The criterion (3) is minimized jointly over $\boldsymbol{\beta}$ and f_i for all mice i. The smoothing parameters, λ_i, corresponding to each mouse-specific spatial term, f_i, are estimated from the data using generalized cross validation for computational efficiency, but other procedures produced similar results.

The bootstrap

To compare the spatial heterogeneity of the age and exercise groups, we computed the within-group sample means of the log-smoothing parameters, $\log \lambda_i$. The (natural) logarithm transformation stabilizes the variance of λ_i, and produces more reasonable comparisons.

Since the sampling distribution of $\log \lambda_i$ is unknown, we approximate it using the bootstrap (Efron and Tibshirani, 1994). The bootstrap uses a re-sampling procedure to estimate the sampling distribution based on the empirical distribution function of the data, and is broadly applicable.

Let $N = 22$ be the number of mice. The bootstrap algorithm is the following:

For $b = 1, \ldots, B = 1,000$ bootstrap simulations,

1. Sample $N = 22$ mice *with replacement* from the data to form the bootstrap data, $\left\{Y_i^{(b)}(\boldsymbol{s})\right\}$;
2. Estimate model (1) by solving (3) with the bootstrap data, $\left\{Y_i^{(b)}(\boldsymbol{s})\right\}$;
3. Compute the within-group sample means of the bootstrap log-smoothing parameters, $\left\{\log \lambda_i^{(b)}\right\}$.

The approximate sampling distributions of the within-group means of the log-smoothing parameters may then be compared using the bootstrap simulations. Note that we re-sample at the mouse level to preserve the within-mouse spatial structures, which is essential for measuring spatial heterogeneity.