Supplemental material on: Bayesian Emulation and Calibration of a Stochastic Computer Model of Mitochondrial DNA Deletions in Substantia Nigra Neurons

D. A. Henderson, R. J. Boys, K. J. Krishnan, C. Lawless and D. J. Wilkinson

Abstract

This manuscript contains supplemental material on the paper "Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in substantia nigra neurons".

1 Comments on hazards in the biological model

The hazard rates corresponding to the five chemical reactions listed in equation (1) in the paper are

$$h_1(\boldsymbol{Y}, c_1) = c_1 Y_1, \qquad h_2(\boldsymbol{Y}, c_2) = c_2 \left(\frac{Y_1}{Y_1 + Y_2}\right), \qquad h_3(\boldsymbol{Y}, c_3) = c_3 Y_1,$$
$$h_4(\boldsymbol{Y}, c_4) = c_4 \left(\frac{Y_2}{Y_1 + Y_2}\right), \qquad h_5(\boldsymbol{Y}, c_5) = c_5 Y_2.$$

Each cell is initiated (at birth) with a population of 1000 normal mtDNA. The synthesis rates, c_2 and c_4, are set such that the cell tries to maintain homeostasis in its mtDNA copy number by always trying to keep the number of mtDNA at the initial level of 1000. The idea here is that if the total number of mtDNA molecules in the cell at any time is equal to the initial amount, that is, $Y_1 + Y_2 = 1000$, then each molecule type is synthesised at the same hazard rate as molecules are degraded. Any reduction in total mtDNA copy number causes a disproportionate increase in the synthesis rate for all molecule types thereby maintaining equilibrium. So we take $c_2 = 1000c_3$ and $c_4 = 1000c_5$. Additionally we make the simplifying assumption that the degradation rates for the normal and mutated mtDNA are equal, that is $c_5 = c_3$. Thus the reaction hazard rates become

$$h_1(\boldsymbol{Y}, c_1) = c_1 Y_1, \qquad h_2(\boldsymbol{Y}, c_3) = \frac{1000 c_3 Y_1}{Y_1 + Y_2}, \qquad h_3(\boldsymbol{Y}, c_3) = c_3 Y_1,$$
$$h_4(\boldsymbol{Y}, c_3) = \frac{1000 c_3 Y_2}{Y_1 + Y_2}, \qquad h_5(\boldsymbol{Y}, c_3) = c_3 Y_2,$$

as given in equation (2) in the paper.

2 Further details of the model for the deletion accumulation data

In this section we describe in more depth the dataset on deletion accumulation and how we model it.

The proportion of mtDNA deletions in each sample of 25 neurons is measured indirectly using one of two real-time polymerase chain reaction (RT-PCR) experimental techniques. In very simple terms, a

PCR experiment starts off with an initial unknown quantity of DNA, Y, and the experiment proceeds by doubling the quantity of DNA in cycles. The output is a measurement of the number of cycles C_t until a prespecified threshold level T is exceeded, as measured in terms of fluorescence when subjected to a laser. In theory the relationship between these quantities is $T = Y \times 2^{C_t}$ and so the initial quantity of DNA can be estimated from $Y = T \times 2^{-C_t}$.

The mtDNA molecule is circular and approximately 16kb in length. It has a major arc (approximately 11kb) and a minor arc (approximately 5kb). Deletions are only assumed to occur in the major arc, not in the minor arc. Two genes are targeted in the PCR experiments: ND1 from the minor arc and ND4 which resides on the major arc. For a given sample, let Y_{ND1} and Y_{ND4} denote the number of copies of mtDNA with the ND1 gene and ND4 gene present (respectively) and assume that deletions will only affect ND4 and not ND1. Then there are a total of Y_{ND1} copies of mtDNA in the sample and only $Y_{ND1} - Y_{ND4}$ copies of mtDNA with deletions (because the deleted copies of mtDNA in ND4 do not show up). Therefore we can estimate the proportion of mtDNA with deletions by

$$p = \frac{Y_{ND1} - Y_{ND4}}{Y_{ND1}} = 1 - \frac{Y_{ND4}}{Y_{ND1}}.$$

The first experimental technique is called the simplex method (this corresponds to technique 2 in our notation). The DNA from the sample of 25 neurons is split evenly into two samples. On one sample RT-PCR is run targeting the ND1 gene, and on the other sample RT-PCR is run targeting the ND4 gene. The other experimental technique is called the multiplex method and this involves running RT-PCR on both genes simultaneously from the combined sample of neurons.

Focussing on the simplex method for the purposes of exposition, the rationale behind the use of the RT-PCR method for determining the proportion of deletions is as follows. First we assume that there are Y_{ND1} and Y_{ND4} copies of ND1 and ND4 in the initial samples respectively and that the threshold fluorescence level T is reached after C_{ND1} cycles for the ND1 probe and after C_{ND4} cycles for the ND4 probe. Hence

$$T = 2^{C_{\text{ND1}}} Y_{\text{ND1}} \quad \text{and} \quad T = 2^{C_{\text{ND4}}} Y_{\text{ND4}}.$$

Therefore,

$$Y_{\text{ND1}} = \frac{T}{2^{C_{\text{ND1}}}} \quad \text{and} \quad Y_{\text{ND4}} = \frac{T}{2^{C_{\text{ND4}}}},$$

and plugging these values into the expression for the proportion of deletions gives

$$p = 1 - \frac{Y_{\text{ND4}}}{Y_{\text{ND1}}} = 1 - \frac{T/2^{C_{\text{ND4}}}}{T/2^{C_{\text{ND1}}}} = 1 - 2^{-(C_{\text{ND4}} - C_{\text{ND1}})} = 1 - 2^{-\Delta C_t},$$

where $\Delta C_t = C_{\text{ND4}} - C_{\text{ND1}}$. Rearranging the above expression gives

$$\Delta C_t = -\log_2(1 - p),$$

and for a given sample i this corresponds to the quantity y_i in the paper.

The measurement error model is derived from the following basic assumptions. The quantities that are actually measured in the PCR experiments are the number of cycles till fluorescence is reached. As noted in the paper, Larionov *et al.* (2005) suggest that a normal distribution adequately models these measurements. Specifically, we assume that the observed numbers of cycles are biased and noisy measurements of the true numbers of cycles, that is

$$C_{ND1}^{\text{obs}} \sim N(C_{ND1} + b_s, \sigma_s^2) \quad \text{and} \quad C_{ND4}^{\text{obs}} \sim N(C_{ND4} + b_s, \sigma_s^2).$$

Here C_{ND1}^{obs} and C_{ND4}^{obs} are the observed values of C_{ND1} and C_{ND4}, respectively. We assume a fixed bias denoted b_s and a fixed variance denoted σ_s^2. The subscript s is used to denote the simplex method. We

further assume that C_{ND1}^{obs} and C_{ND4}^{obs} are independent since they are obtained from separate samples. It follows that the observed value of ΔC_t has a normal distribution,

$$\Delta C_t^{\text{obs}} = C_{ND4}^{\text{obs}} - C_{ND1}^{\text{obs}} \sim N(C_{ND4} + b_s - C_{ND1} - b_s, 2\sigma_s^2),$$

that is

$$\Delta C_t^{\text{obs}} \sim N(\Delta C_t, 2\sigma_s^2).$$

For a particular sample i the above equation is the measurement error model given in the paper since ΔC_t^{obs} corresponds to z_i, ΔC_t corresponds to y_i, and $2\sigma_s^2$ corresponds to ϕ_2^{-1}. Therefore even when we allow for systematic bias in the measurement process we end up with a measurement error model which does not include a bias term explicitly.

Both experimental techniques are based on this so-called "ΔC_t" method. The main difference between the techniques is that in the multiplex method we start off with twice as many copies of ND1 and ND4 as in the simplex method (as the sample has not been halved). This means that the threshold should be reached in one fewer cycle for each probe. The multiplex method should in theory reduce the sources of uncertainty in the experimental procedure. However, it contains an element of "competition" between the two probes as they compete for the limited resource which is the laser. We model this element of competition through the following bivariate normal structure

$$\begin{pmatrix} C_{\text{ND1}}^{M,\text{obs}} \\ C_{\text{ND4}}^{M,\text{obs}} \end{pmatrix} \sim N\left(\begin{pmatrix} C_{\text{ND1}} - 1 + b_m \\ C_{\text{ND4}} - 1 + b_m \end{pmatrix}, \begin{pmatrix} \sigma_m^2 & \rho\sigma_m^2 \\ \rho\sigma_m^2 & \sigma_m^2 \end{pmatrix} \right).$$

The mean vector of the observations results from the fact that there is twice as much DNA in the multiplex sample as in the simplex sample and this corresponds to one fewer doubling cycle. We assume a systematic bias of b_m, which may be different to that from the simplex procedure. We also assume a fixed variance of σ_m^2 and a correlation of ρ. It might be thought that competition would lead to a negative correlation between the observations and hence $\rho < 0$.

From the above bivariate normal model we have that

$$\Delta C_t^{M,\text{obs}} = C_{\text{ND4}}^{M,\text{obs}} - C_{\text{ND1}}^{M,\text{obs}} \sim N\left(C_{\text{ND4}} - 1 + b_m - C_{\text{ND1}} + 1 - b_m, 2(1-\rho)\sigma_m^2\right),$$

that is,

$$\Delta C_t^{M,\text{obs}} \sim N\left(\Delta C_t, 2(1-\rho)\sigma_m^2\right),$$

which has the same mean as, but a different variance from, ΔC_t^{obs}. If we further assume that the individual measurement error variances are the same in the two procedures, that is $\sigma_s^2 = \sigma_m^2$, then the only difference in the variances of the observed ΔC_t measurements is caused by the correlation in the samples ρ. In terms of the notation used in the paper, we have $\phi_1^{-1} = 2(1-\rho)\sigma_m^2$. Note that we do not make the assumption that $\sigma_s^2 = \sigma_m^2$ in the paper and so we have not explicitly estimated the correlation ρ by $\rho = 1 - \phi_2/\phi_1$.

Thus allowing for different levels of bias in the two sets of measurements results in a measurement error model which has the same mean in the two experimental techniques, but possibly different variances. The model does account for technique-specific biases in the measurements but these biases cancel out in the derivation of the measurement error model. Because the technique-specific biases in the measurements cancel out in the derivation of the measurement error model, any differences between the means of the two experimental techniques that may be suggested by the data from Figure 1 in the paper should be due to random variation according to our model. We have therefore not included an additional technique-specific bias term into the model.

The data on deletion accumulation are tabulated in Table 1. For example, there are 10 observations for individual 1, three made using technique 1 ("multiplex"), the other seven made using technique 2 ("simplex").

3 Details of MCMC scheme for exact Bayesian calibration

In Section 3.2 of the paper, we state that it is possible to sample from the posterior density of the calibration parameters $\boldsymbol{\theta}$, the measurement error parameters $\boldsymbol{\phi}$ and the latent data $\boldsymbol{y}$ using a Markov chain Monte Carlo scheme, despite the fact that the density of the latent data can not be computed directly. The details are as follows.

The joint posterior density is given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\phi}),$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)^{\mathrm{T}}$. We sample from $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x})$ using a Metropolis-Hastings within Gibbs procedure. This entails sampling each unknown quantity or block of related quantities from their full conditional distribution using a Metropolis-Hastings transition kernel when direct sampling is not possible. We split the scheme into two blocks of full conditionals: $\boldsymbol{\phi}|\cdots$ and $\boldsymbol{\theta}, \boldsymbol{y}|\cdots$, where '$\cdots$' denotes all the other variables in the model. Because of our choice of gamma distribution for the precisions we can sample directly from the full conditional distribution $\boldsymbol{\phi}|\cdots$. The full conditional distribution $\boldsymbol{\theta}, \boldsymbol{y}|\cdots$ is not available in closed form and so a Metropolis-Hastings step is used. Specifically, at iteration $k+1$,

1. sample $\phi_j^{(k)}|\boldsymbol{y}, \boldsymbol{z} \sim \text{Gamma}(C_j, D_j)$, for $j = 1, 2$, where

$$C_j = a_{\phi_j} + \frac{1}{2}\sum_{i=1}^{n} \mathbb{I}(o_i = j), \qquad D_j = b_{\phi_j} + \frac{1}{2}\sum_{i=1}^{n} \left\{\mathbb{I}(o_i = j)(z_i - y_i)^2\right\},$$

   and $\mathbb{I}(x)$ is the indicator function which equals 1 if x is true and equals 0 otherwise.

2. (a) generate a candidate value $\tilde{\boldsymbol{\theta}}$ from the distribution with transition kernel $q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$;
   (b) generate a candidate vector of latent data $\tilde{\boldsymbol{y}}$ from the stochastic kinetic model using the candidate parameter $\tilde{\boldsymbol{\theta}}$. This proposal has density $\pi(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{\theta}}, \boldsymbol{x})$.
   (c) Set $\boldsymbol{\theta}^{(k+1)} = \tilde{\boldsymbol{\theta}}$, $\boldsymbol{y}^{(k+1)} = \tilde{\boldsymbol{y}}$ with probability $\min(1, A)$, otherwise retain the current values by setting $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)}$, $\boldsymbol{y}^{(k+1)} = \boldsymbol{y}^{(k)}$. The acceptance ratio is given by

$$A = \frac{\pi(\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta})}\frac{\pi(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{\theta}}, \boldsymbol{x})}{\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})}\frac{\pi(\boldsymbol{z}|\tilde{\boldsymbol{y}}, \boldsymbol{\phi})}{\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\phi})}\frac{q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})}{q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})}\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})}{\pi(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{\theta}}, \boldsymbol{x})} = \frac{\pi(\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta})}\frac{\pi(\boldsymbol{z}|\tilde{\boldsymbol{y}}, \boldsymbol{\phi})}{\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\phi})}\frac{q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})}{q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})}.$$

Note that proposing the latent data $\boldsymbol{y}$ by using a sample from the computer model means that all terms involving the density of the latent data $\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$ cancel out of the acceptance ratio. This cancellation means that computation of $\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$ (which we assume to be prohibitive or impossible) is not required as part of the MCMC scheme. So, in theory, we can always sample from the exact posterior distribution provided that we can sample from the distribution of the latent data, even when we cannot compute the density of the latent data. Note also that we can update $\boldsymbol{y}$ from its full conditional distribution (that is, not jointly with $\boldsymbol{\theta}$) by using independence proposals from the computer model. In fact, the conditional independence in the latent data means that we can sample each y_i from its full conditional. However, we cannot sample from the full conditional for $\boldsymbol{\theta}$ because of the presence of the density $\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$ in the acceptance ratio, and so a joint update of $(\boldsymbol{\theta}, \boldsymbol{y})$ is essential.

4 Further details on experimental design

This section gives more details of the simulation design we adopted for fitting the emulator in Section 4.3 in the paper.

The n_D-point simulation design we have used is in fact the union of three designs. The first component design is a 2^4-factorial design on the extreme points of the input space. We take the extreme points for θ_1 and θ_2 to be the 0.0005-quantile and 0.9995-quantile of their normal prior distributions; the extreme points for θ_3 are 0.5 and 1, which are the limits of its uniform prior distribution; and the extreme points of x were taken to be 1 and 110. A major consideration with this design is that we would like the emulation model to be a useful surrogate for the simulation model across the whole prior parameter space, as well as being able to predict well for a realistic range of ages. The extreme points for θ_1 and θ_2 contain 99.9% of the prior probability for each parameter marginally. It is unlikely that we will need predictions for ages over 110 years, since so few people survive past that age.

The second component design is the Cartesian product of an 8-point Latin hypercube sample for $\boldsymbol{\theta}$, and the 13 unique ages at which we have observations

$$\boldsymbol{x}^\star = (19, 20, 32, 42, 44, 51, 56, 72, 75, 77, 81, 89, 91)^{\mathrm{T}}.$$

Designs based on Latin hypercubes were popularised by McKay *et al.* (1979) and have become widely used throughout the computer experiments literature. They have good space-filling properties whilst maintaining uniform coverage of the univariate inputs; see Santner *et al.* (2003) for further details. Our choice of a Cartesian product of these x values and a Latin hypercube sample is motivated by Kennedy and O'Hagan (2001) who comment that it is intuitively reasonable to include the x values at which we have experimental data into the simulation design.

The third component design consists of a sample of 130 points from the prior distribution for $\boldsymbol{\theta}$. The corresponding 130 x-inputs are sampled uniformly from the integers $\{1, 2, \ldots, 110\}$. This design aims to give good coverage over the support of the prior distribution. The subsequent large number of inter-point distances may be beneficial when estimating Gaussian process parameters, as pointed out by Rougier (2001).

5 Integrated emulation and calibration

Given the model for the emulator and the simulation training data, we can construct an integrated Bayesian model for joint emulation and calibration. The joint posterior density function corresponding to the integrated model is

$$\begin{aligned}\pi(&\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi}_\eta, \boldsymbol{\psi}_\eta, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\eta}_y, \boldsymbol{\xi}_y, \boldsymbol{\rho} | \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{D}^\star, \boldsymbol{W}) \\ &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})\pi(\boldsymbol{\eta}|\boldsymbol{\psi}_\eta, \boldsymbol{D}^\star)\pi(\boldsymbol{\xi}|\boldsymbol{\psi}_\xi, \boldsymbol{D}^\star)\pi(\boldsymbol{\eta}_y|\boldsymbol{\eta}, \boldsymbol{\psi}_\eta, \boldsymbol{\theta}, \boldsymbol{x}^\star, \boldsymbol{D}^\star) \\ &\quad \times \pi(\boldsymbol{\xi}_y|\boldsymbol{\xi}, \boldsymbol{\psi}_\xi, \boldsymbol{\theta}, \boldsymbol{x}^\star, \boldsymbol{D}^\star)\pi(\boldsymbol{\rho}|\boldsymbol{\eta}_y, \boldsymbol{\xi}_y, \boldsymbol{x})\pi(\boldsymbol{W}|\boldsymbol{\eta}, \boldsymbol{y})\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\phi}), \end{aligned} \tag{1}$$

in which the uniform joint densities of the GP parameters $\boldsymbol{\psi}_\eta$ and $\boldsymbol{\psi}_\xi$ are subsumed into the proportionality sign.

We use the notation $\eta_j = \eta(\boldsymbol{u}_j)$ and $\xi_j = \xi(\boldsymbol{u}_j)$ for $j = 1, \ldots, n_{D^\star}$, and therefore $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{n_{D^\star}})^{\mathrm{T}}$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{n_{D^\star}})^{\mathrm{T}}$. Here $n_{D\star}$ is the number of design points at which we have training data. The density of the latent means and the latent log standard deviations is multivariate normal (from the definition of a GP), with

$$\begin{aligned}\pi(\boldsymbol{\eta}|\boldsymbol{\psi}_\eta, \boldsymbol{D}^\star) &= \frac{1}{(2\pi)^{n_{D^\star}/2}|\boldsymbol{\Sigma}_{\boldsymbol{\eta}}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu}_{\boldsymbol{\eta}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta} - \boldsymbol{\mu}_{\boldsymbol{\eta}})\right\} \\ \pi(\boldsymbol{\xi}|\boldsymbol{\psi}_\xi, \boldsymbol{D}^\star) &= \frac{1}{(2\pi)^{n_{D^\star}/2}|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}})\right\},\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{\eta}}$, is a vector of length $n_{D^\star}$ with each element equal to $\psi_{\eta,1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is an $n_{D^\star} \times n_{D^\star}$ positive definite covariance matrix with (i, j)th element $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}(i, j) = c_\eta(\boldsymbol{u}_i, \boldsymbol{u}_j)$; and similarly for ξ replacing η.

The vectors of 'predictive' means and log standard deviations corresponding to the unique ages at which we have observations, $\boldsymbol{x}^\star$, are denoted by $\boldsymbol{\eta}_y = (\eta_{y,1}, \ldots, \eta_{y,n_{x^\star}})^{\mathrm{T}}$ and $\boldsymbol{\xi}_y = (\xi_{y,1}, \ldots, \xi_{y,n_{x^\star}})^{\mathrm{T}}$, respectively, where $n_{x^\star} = 13$ is the number of unique ages. The conditional densities for these predictive means and log standard deviations are derived from standard multivariate normal distribution theory and the properties of the relevant Gaussian process prior. For example,

$$\boldsymbol{\eta}_y | \boldsymbol{\eta}, \boldsymbol{\psi}_\eta, \boldsymbol{\theta}, \boldsymbol{x}^\star, \boldsymbol{D}^\star \sim N_{n_{x^\star}}(\boldsymbol{\mu}_{\boldsymbol{\eta}_y|\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_y|\boldsymbol{\eta}}),$$

that is $n_{x^\star}$-dimensional multivariate normal with conditional mean vector

$$\boldsymbol{\mu}_{\boldsymbol{\eta}_y|\boldsymbol{\eta}} = \boldsymbol{\mu}_{\boldsymbol{\eta}_y} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}_y,\boldsymbol{\eta}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}_{\boldsymbol{\eta}}),$$

and conditional covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}_y|\boldsymbol{\eta}} = \boldsymbol{\Sigma}_{\boldsymbol{\eta}_y} - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_y,\boldsymbol{\eta}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\eta}_y,\boldsymbol{\eta}}^{\mathrm{T}}.$$

Here, $\boldsymbol{\mu}_{\boldsymbol{\eta}_y}$ is a vector of length $n_{x^\star}$ with each element equal to $\psi_{\eta,1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_y}$ is an $n_{x^\star} \times n_{x^\star}$ positive definite covariance matrix with (i,j)th element

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}_y}(i,j) = c_\eta\{(\boldsymbol{\theta}^{\mathrm{T}}, x_i^\star)^{\mathrm{T}}, (\boldsymbol{\theta}^{\mathrm{T}}, x_j^\star)^{\mathrm{T}}\},$$

where $x_i^\star$ is the ith element of $\boldsymbol{x}^\star$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_y,\boldsymbol{\eta}}$ is an $n_{x^\star} \times n_{D^\star}$ matrix with (i,j)th element

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}_y,\boldsymbol{\eta}}(i,j) = c_\eta\{(\boldsymbol{\theta}^{\mathrm{T}}, x_i^\star)^{\mathrm{T}}, \boldsymbol{u}_j\}.$$

The structure of the density for $\boldsymbol{\xi}_y$ is identical to that of $\boldsymbol{\eta}_y$ but with an obvious change of notation.

The density of the latent data $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^{\mathrm{T}}$ is simply the product of $n = 90$ independent normal densities with means and log standard deviations selected from the appropriate predictive means and log standard deviations, namely

$$\pi(\boldsymbol{\rho} | \boldsymbol{\eta}_y, \boldsymbol{\xi}_y, \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\exp(\xi_{y,I(x_i)})} \exp\left\{ -\frac{(\rho_i - \eta_{y,I(x_i)})^2}{2\exp(2\xi_{y,I(x_i)})} \right\},$$

where the function $I(x_i) = \{j : x_i = x_j^*\}$ is needed to map individual ages x_i to the unique ages $\boldsymbol{x}^\star$.

The density of the emulator training data is the product of independent normal densities,

$$\pi(\boldsymbol{W} | \boldsymbol{\eta}, \boldsymbol{\xi}) = \prod_{j=1}^{n_{D^\star}} \prod_{k=1}^{n_j} \frac{1}{\sqrt{2\pi}\exp(\xi_j)} \exp\left\{ -\frac{(w_{jk} - \eta_j)^2}{2\exp(2\xi_j)} \right\}.$$

The densities of the experimental data $\pi(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{\phi})$, the calibration parameters $\pi(\boldsymbol{\theta})$ and the measurement error parameters $\pi(\boldsymbol{\phi})$ follow from the definition of the Bayesian model in Section 3.1 of the paper.

A Metropolis-Hastings within Gibbs scheme can be used to sample values from the joint posterior distribution (1). Each component value or vector of values is sampled in turn conditional on the current sampled values of the other components via an appropriate Metropolis-Hastings transition kernel. We take the transition kernels to be the appropriate dimension multivariate normal density centered on the current sampled value when a Gibbs update is not available. At each iteration of the scheme, the emulation model is constructed by augmenting the design matrix $\boldsymbol{D}^\star$ with the input corresponding to the current value of $\boldsymbol{\theta}$. This feature should lead to an increase in the accuracy of the emulator in regions of the input space with high posterior support.

However, sampling from the posterior distribution is fairly challenging, and this is due in no small part to the large number of often highly correlated unknown quantities. Calculations on large matrices are

required throughout the algorithm and this leads to the algorithm being slow to run. Together with the long run lengths that are required to sample from such a complex posterior distribution this effectively prohibits the fitting of the full integrated model in this form.

Separating the emulation stage from the calibration stage breaks down the inference problem into two more computationally feasible tasks, albeit at the expense of a lack of shared information and 'borrowing of strength' between the two tasks. In particular, since posterior information about likely values of $\boldsymbol{\theta}$ is no longer included in the fitting of the emulation model, the emulator is likely to be a slightly less accurate model for the simulator in regions of the input space with high posterior support. However, the drawback of this potential small reduction in accuracy is more than compensated by the computational benefits of teating the emulation and calibration tasks separately.

6 Emulator simulation

The role of the fitted emulation model is to act as a fast substitute for the simulation model in the calibration scheme and for studying output in predictive *in silico* experiments. Both tasks require predictive simulations from the fitted emulation model, as outlined below.

Suppose we require simulations at n_v input configurations $\boldsymbol{u}_j^v$, for $j = 1, \ldots, n_v$. Let $\boldsymbol{D}_v$ denote the prediction design matrix, with jth row equal to $\boldsymbol{u}_j^v$. A sample of K values from the posterior predictive distribution of the emulator at these inputs can be obtained as follows. For $k = 1, \ldots, K$

1. sample $\boldsymbol{\psi}_\eta^{(k)}, \boldsymbol{\psi}_\xi^{(k)} | \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\xi}}$ from the joint posterior distribution (equation (5) in the paper);

2. sample $\boldsymbol{\eta}_v^{(k)} | \widehat{\boldsymbol{\eta}}, \boldsymbol{\psi}_\eta^{(k)}, \boldsymbol{D}_v \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\eta}_v|\widehat{\boldsymbol{\eta}}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_v|\widehat{\boldsymbol{\eta}}}^{(k)}\right)$;

3. sample $\boldsymbol{\xi}_v^{(k)} | \widehat{\boldsymbol{\xi}}, \boldsymbol{\psi}_\xi^{(k)}, \boldsymbol{D}_v \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_v|\widehat{\boldsymbol{\xi}}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_v|\widehat{\boldsymbol{\xi}}}^{(k)}\right)$;

4. sample $\rho_{v,j}^{(k)} | \eta_{v,j}^{(k)}, \xi_{v,j}^{(k)} \sim N\left(\eta_{v,j}^{(k)}, \exp\{2\xi_{v,j}^{(k)}\}\right), \quad j = 1, \ldots, n_v.$

This procedure returns K sampled values of the logit transformed proportion of deletions in 25 cells from $\rho_{v,j} | \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\xi}}, \boldsymbol{D}_v$, for $j = 1, \ldots, n_v$. In the above, $\boldsymbol{\eta}_v = (\eta_{v,1}, \ldots, \eta_{v,n_v})^{\mathrm{T}}$ and $\boldsymbol{\xi}_v = (\xi_{v,1}, \ldots, \xi_{v,n_v})^{\mathrm{T}}$. In step 2, standard multivariate normal distribution theory and the properties of the relevant GP give the mean vector and covariance matrix as

$$\boldsymbol{\mu}_{\boldsymbol{\eta}_v|\widehat{\boldsymbol{\eta}}}^{(k)} = \boldsymbol{\mu}_{\boldsymbol{\eta}_v}^{(k)} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}_v,\widehat{\boldsymbol{\eta}}}^{(k)} \left(\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\eta}}}^{(k)} + \exp(\psi_{\eta,7}^{(k)}) \boldsymbol{I}\right)^{-1} (\widehat{\boldsymbol{\eta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\eta}}}^{(k)}),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v|\widehat{\boldsymbol{\eta}}}^{(k)} = \boldsymbol{\Sigma}_{\boldsymbol{\eta}_v}^{(k)} - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_v,\widehat{\boldsymbol{\eta}}}^{(k)} \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{(k)} + \exp(\psi_{\eta,7}^{(k)}) \boldsymbol{I}\right)^{-1} \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v,\widehat{\boldsymbol{\eta}}}^{(k)}\right)^{\mathrm{T}},$$

where $\boldsymbol{\mu}_{\boldsymbol{\eta}_v}^{(k)}$ is an n_v–vector with each element equal to $\psi_{\eta,1}^{(k)}$, $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v}^{(k)}$ is an $n_v \times n_v$ positive definite covariance matrix with (i,j)th element $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v}^{(k)}(i,j) = c_\eta(\boldsymbol{u}_i^v, \boldsymbol{u}_j^v)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v,\widehat{\boldsymbol{\eta}}}^{(k)}$ is an $n_v \times n_{D^{\star\star}}$ matrix with (i,j)th element $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_v,\widehat{\boldsymbol{\eta}}}^{(k)}(i,j) = c_\eta(\boldsymbol{u}_i^v, \boldsymbol{u}_j)$. Similar results hold for the mean vector and covariance matrix in step 3.

7 External validation: neuron survival data

The data used for external validation are taken from Table 2(A) in Fearnley and Lees (1991) and comprise the number of neurons surviving in samples from the caudal substantia nigra of 36 individuals without

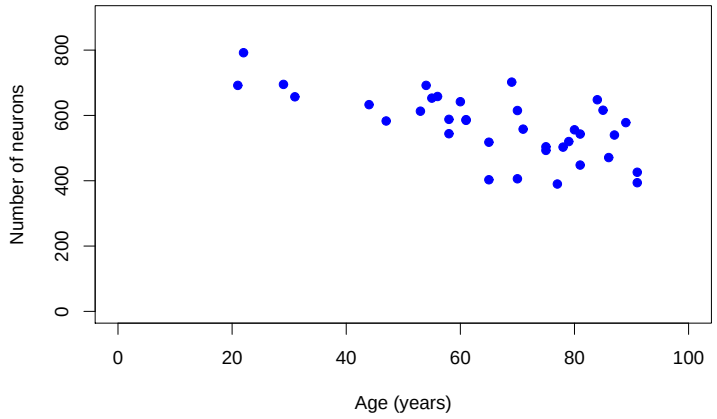

Figure 1: Neuron survival data. Observed number of neurons in the substantia nigra of 36 individuals of varying ages.

Parkinson's disease. Figure 1 shows a scatterplot of number of neurons surviving against the age of these individuals and indicates that the number of neurons probably decreases with age. Note that the data used in this paper are a corrected version of that in Fearnley and Lees (1991): the total number of surviving neurons for the 22-year-old individual is 792, not 692. These (corrected) data are tabulated in Table 2.

We now formulate the observational model for these data and link it with the biological model under scrutiny.

7.1 Bayesian model

Suppose that individual i_s has $y^s_{i_s}$ neurons surviving in the region of their substantia nigra from which the sample was taken ($i_s = 1, \ldots, 36$). However, due to observational error, only $z^s_{i_s}$ neurons are recorded as present. Assuming that each neuron is observed independently with probability ϕ_s, the observational model is binomial, that is

$$z^s_{i_s} \sim \text{Bin}(y^s_{i_s}, \phi_s).$$

We represent our uncertainty about the observation probability parameter ϕ_s through a beta distribution

$$\phi_s \sim \text{Beta}(a_{\phi_s}, b_{\phi_s})$$

whose parameters are chosen to reflect the belief that the measurement process is reasonably accurate ($a_{\phi_s} = 90$, $b_{\phi_s} = 10$; $E(\phi_s) = 0.9$).

We model the number of neurons in the substantia nigra of the i_sth individual, $y^s_{i_s}$, using the stochastic kinetic model of mtDNA deletion and cell death described in Section 2 of the paper. Recall that the marginal distribution of $y^s_{i_s}$ is analytically intractable and that, although it is relatively straightforward to sample from this distribution, such simulation can be time-consuming. To obtain a single realisation of the true number of neurons $y^s_{i_s}$ for individual i_s using the calibration parameter $\boldsymbol{\theta}$, we must simulate N_{i_s} independent realisations from the computer model, one for each substantia nigra neuron that individual i starts out with at birth. The N_{i_s} neurons are simulated until time $x^s_{i_s}$ years. If at any time up to $x^s_{i_s}$ years, the proportion of deletions in a neuron attains or exceeds the lethal threshold θ_3 then that neuron is deemed not to have survived. The number of neurons out of N_{i_s} that survive till time $x^s_{i_s}$ is a realisation of $y^s_{i_s}$.

In order to complete the model description, we need plausible values for the N_{i_s}, the number of neurons at birth in the region of the caudal substantia nigra used in the Fearnley and Lees study. Their study assumes that each individual starts out with the same number of neurons, so that $N_{i_s} = N$ for all i_s. They then estimate N to be $N^* = 795$ by fitting a straight line to the data $(\boldsymbol{x}^s, \boldsymbol{z}^s)$ using ordinary least squares. In the absence of any other information regarding neuron numbers at birth, we model our beliefs about its value by using a fairly diffuse Poisson distribution (with mean $N^* = 795$).

7.2 Emulating the survival output

Obtaining values for the $y^s_{i_s}$ from the simulation model is slow, typically taking tens of minutes. This is due to having to trace the outcome of a large number of individual neurons, and here we have (roughly) a 32-fold increase in the number of neurons compared to the earlier analysis. Therefore, we again turn to the computational advantages of using an emulator.

The key output for the simulation model at input $\boldsymbol{u} = (\boldsymbol{\theta}^T, x)^T$ is the binary outcome describing whether the neuron has survived until age x (where $S(\boldsymbol{u}) = 1$ denotes survival). Let $p^s(\boldsymbol{u})$ denote the probability of survival using input $\boldsymbol{u}$. As cells are simulated independently, it follows that the number of surviving cells in a sample of size N is

$$y^s(\boldsymbol{u}) \sim \text{Bin}(N, p^s(\boldsymbol{u})).$$

This simple parametric model for the output from the simulation model can be used to construct an emulator by assuming the survival probability is a smooth function of the inputs $\boldsymbol{u}$. As in Section 4.2 of the paper, we work on an (unconstrained) logit scale and use a Gaussian process prior with parameters $\boldsymbol{\psi}_{\rho^s}$ to model $\rho^s(\cdot) = \text{logit}\{p^s(\cdot)\}$. For reasons of brevity, we omit the details on fitting the emulator but follow essentially the same procedure as described in Sections 5.1 to 5.3 of the paper. For instance, we reduce the complexity of the model substantially by fixing the $n_D = 250$ logit-transformed survival probabilities, $\boldsymbol{\rho}^s$, at their estimated values from the simulator output. The reduced and simplified emulation model was fitted using MCMC in a manner analogous to that described in Section 5.2 of the paper. The fitted model for $\rho^s(\cdot)$, which we denote by $\widehat{\rho}^s(\cdot)$, is based on Monte Carlo estimates of the posterior means of $\boldsymbol{\psi}_{\rho^s}$ and on the conditional mean of the Gaussian process, as was the case for the fitted models for the simulation output on deletion accumulation. We note that these simplifications in the emulation model had little effect in terms of its predictive accuracy.

7.3 Model validation

The Bayesian model has the following hierarchical structure:

$$\begin{aligned}
z^s_{i_s} | y^s_{i_s}, \phi_s &\sim \text{Bin}(y^s_{i_s}, \phi_s), \qquad i_s = 1, \ldots, 36 \\
y^s_{i_s} | N, \boldsymbol{\theta}, x^s_{i_s} &\sim \text{Bin}(N, \text{expit}[\widehat{\rho}^s\{(\boldsymbol{\theta}^{\mathrm{T}}, x^s_{i_s})^{\mathrm{T}}\}]), \qquad i_s = 1, \ldots, 36 \\
\phi_s &\sim \text{Beta}(a_{\phi_s}, b_{\phi_s}), \\
N &\sim \text{Poisson}(N^\star),
\end{aligned}$$

where $\widehat{\rho}^s(\cdot)$ is the fitted emulation model described in Section 7.2, $a_{\phi_s} = 90$, $b_{\phi_s} = 10$ and $N^\star = 795$.

External validation (or out-of-sample prediction) is performed by simulating from the predictive distribution of $z^s_{\text{pred}} | \boldsymbol{z}, x^s_{i_s}$, that is, from the posterior predictive distribution of $z^s_{i_s}$ for individual i_s calculated using the posterior distribution of the parameters $\boldsymbol{\theta}$ from the dataset on deletion accumulation. Realisations are obtained by first sampling from the posterior distribution whose density is given by equation (6) in the paper and then, for each individual, simulating from the Bayesian model above. Specifically, for $k = 1, \ldots, K$:

- sample $\boldsymbol{\theta}^{(k)}|\boldsymbol{z}$, a sample from the posterior distribution;
- sample $N^{(k)} \sim \text{Poisson}(N^\star)$;
- sample $\phi_s^{(k)} \sim \text{Beta}(a_{\phi_s}, b_{\phi_s})$;
- sample $y_{i_s}^{s(k)}|N^{(k)}, \boldsymbol{\theta}^{(k)}, x_{i_s}^s \sim \text{Bin}(N^{(k)}, \text{expit}\{\widehat{\rho}_s[\{(\boldsymbol{\theta}^{(k)})^{\mathrm{T}}, x_{i_s}^s\}^{\mathrm{T}}]\}), \qquad i_s = 1, \ldots, 36;$
- sample $z_{i_s}^{s(k)}|y_{i_s}^{s(k)}, \phi_s^{(k)} \sim \text{Bin}(y_{i_s}^{s(k)}, \phi_s^{(k)}), \qquad i_s = 1, \ldots, 36.$

Here, sampling from the posterior distribution is achieved by using the converged and thinned output reported in Section 6.1 of the paper. Thus the predictive sample for each individual consists of $K = 5000$ (essentially) uncorrelated values. Figure 5 in the paper displays summaries of these predictive distributions.

References

Fearnley, J. M. and Lees, A. J. (1991) Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain*, **114**, 2283–2301.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society, Series* B, **63**, 425–464.

Larionov, A., Krause, A. and Miller, W. (2005) A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*, **6**, 1–16.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.

Rougier, J. (2001) Comment on "Bayesian calibration of computer models" by Kennedy and O'Hagan. *Journal of the Royal Statistical Society, Series* B, **63**, 453.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*. New York: Springer.

Table 1: Measurements of deletion accumulation data in 15 individuals. Listed for each sample i is the experimental technique, o_i, the individual on whom the measurement is taken, $\mathcal{I}_i$, the age of the individual, x_i, and the RT-PCR measurement of deletion accumulation, z_i.

i	o_i	$\mathcal{I}_i$	x_i	z_i	i	o_i	$\mathcal{I}_i$	x_i	z_i	i	o_i	$\mathcal{I}_i$	x_i	z_i
1	1	1	19	-0.14343	31	2	4	42	1.1	61	2	10	75	1.19
2	1	1	19	-0.1393	32	2	4	42	0.513	62	2	10	75	1.18
3	1	1	19	-0.01207	33	2	4	42	0.919	63	2	10	75	0.25
4	1	2	20	-0.13376	34	2	5	44	1.676	64	2	11	75	1.052
5	1	2	20	0.21	35	2	5	44	1.078	65	2	11	75	0.896
6	1	2	20	0.15786	36	2	5	44	0.928	66	2	11	75	1.211
7	1	3	32	-0.38716	37	2	5	44	1.645	67	2	11	75	0.95
8	1	3	32	-0.00574	38	2	6	51	1.021	68	2	11	75	1.54
9	1	3	32	0.27586	39	2	6	51	1.142	69	2	12	77	1.397
10	1	5	44	0.51987	40	2	6	51	1.403	70	2	12	77	1.856
11	1	5	44	0.68656	41	2	6	51	0.871	71	2	12	77	2.426
12	1	5	44	1.124	42	2	6	51	0.13	72	2	12	77	1.898
13	1	7	51	-0.23206	43	2	7	51	1.25	73	2	12	77	0.655
14	1	7	51	0.2217	44	2	7	51	-0.918	74	2	13	81	1.481
15	1	7	51	0.24086	45	2	7	51	0.512	75	2	13	81	1.21
16	2	1	19	0.625	46	2	7	51	1.02	76	2	13	81	0.79
17	2	1	19	0.105	47	2	8	56	0.558	77	2	14	89	2.269
18	2	1	19	1.421	48	2	8	56	0.897	78	2	14	89	1.993
19	2	1	19	1.177	49	2	8	56	0.539	79	2	14	89	0.693
20	2	1	19	0.88	50	2	9	72	1.573	80	2	14	89	2.716
21	2	1	19	1.03	51	2	9	72	1.133	81	2	14	89	1.42
22	2	1	19	0.49	52	2	9	72	1.184	82	2	14	89	2.06
23	2	2	20	0.413	53	2	9	72	0.985	83	2	14	89	2.37
24	2	2	20	0.611	54	2	9	72	2.38	84	2	15	91	1.719
25	2	2	20	0.564	55	2	9	72	1.69	85	2	15	91	2.055
26	2	3	32	0.806	56	2	9	72	1.49	86	2	15	91	1.995
27	2	3	32	1.061	57	2	10	75	1.712	87	2	15	91	2.564
28	2	3	32	0.989	58	2	10	75	1.243	88	2	15	91	2.128
29	2	3	32	1.138	59	2	10	75	1.035	89	2	15	91	2.24
30	2	4	42	0.965	60	2	10	75	1.044	90	2	15	91	1.36

Table 2: Neuron survival data. The rows are: individual, age in years, and observed number of neurons.

i_s	1	2	3	4	5	6	7	8	9	10	11	12
$x^s_{i_s}$	21	22	29	31	44	47	53	54	55	56	58	58
$z^s_{i_s}$	692	792	695	657	633	583	613	692	653	658	588	544
i_s	13	14	15	16	17	18	19	20	21	22	23	24
$x^s_{i_s}$	60	61	61	65	65	69	70	70	71	75	75	77
$z^s_{i_s}$	642	587	585	403	518	702	406	615	558	493	504	390
i_s	25	26	27	28	29	30	31	32	33	34	35	36
$x^s_{i_s}$	78	79	80	81	81	84	85	86	87	89	91	91
$z^s_{i_s}$	503	520	556	543	448	648	616	471	540	578	426	394