Web-based supplementary materials for "**Bayesian log-Gaussian Cox process regression: with applications to meta-analysis of neuroimaging working memory studies**" by P. Samartsidis, C.R. Eickhoff, S.B. Eickhoff, T.D. Wager, L. Feldman Barrett, S. Atzil, T.D. Johnson and T.E. Nichols

1 Gradient expressions for the LGCP

Let $\beta_k \equiv \mu_k$ $(k = 0, \ldots, K)$. The log-posterior, up to a normalising constant is given by:

$$\ell\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \{\boldsymbol{\gamma}_k\}_{k=1}^{K^*} \mid \cdot\right) \propto \sum_{i=1}^{I}\left[-\sum_{j=1}^{V} A_{v_j}\lambda_i(v_j) + \sum_{j=1}^{V} \mathbf{1}_{v_j \in \mathbf{x}_i} \log \lambda_i(v_j)\right] + \log \text{priors}, \quad (1)$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)^\top$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)^\top$, $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_{K^*})^\top$, $\boldsymbol{\rho} = (\rho_0, \ldots, \rho_{K^*})^\top$, $A_{v_j} = \mathbf{1}_{v_j \in \mathcal{B}}$ and the intensity function at each voxel v_j is defined as:

$$\lambda_i(v_j) = \alpha_i \exp\left(\sum_{k=0}^{K} \beta_k z_{ik}\right) \exp\left(\sum_{k=0}^{K^*} \sigma_k \left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik}\right) \quad (2)$$

We now calculate the derivatives with respect to the parameters of interest.

1.1 Partial derivatives with respect to β_l

We have that:

$$\begin{aligned}\frac{\partial \log \lambda_i(v_j)}{\partial \beta_l} &= \frac{\partial}{\partial \beta_l}\log \alpha_i + \frac{\partial}{\partial \beta_l}\sum_{k=0}^{K}\beta_k z_{ik} + \frac{\partial}{\partial \beta_l}\sum_{k=0}^{K^*}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\ &= z_{il}. \end{aligned} \quad (3)$$

As a result:

$$\begin{aligned}\frac{\partial \ell\left(\beta_l \mid \cdot\right)}{\partial \beta_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial \beta_l}\lambda_i(v_j) - \mathbf{1}_{v_j \in \mathbf{x}_i}\frac{\partial}{\partial \beta_l}\log \lambda_i(v_j)\right] + \frac{\partial}{\partial \beta_l}\log \pi(\beta_l) \\ &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j \in \mathbf{x}_i} z_{il}\right] - \frac{\partial}{\partial \beta_l}\frac{\beta_l^2}{2\tau^2} \\ &= -\sum_{j=1}^{V}\sum_{i=1}^{I} A_{v_j}\lambda_i(v_j)z_{il} + \sum_{i=1}^{I} n_i z_{il} - \frac{\beta_l}{\tau^2}, \end{aligned} \quad (4)$$

where n_i is the total number of foci in study i.

1.2 Partial derivatives with respect to σ_l

We have that:

$$\begin{aligned}\frac{\partial \log \lambda_i(v_j)}{\partial \sigma_l} &= \frac{\partial}{\partial \sigma_l}\log \alpha_i + \frac{\partial}{\partial \sigma_l}\sum_{k=0}^{K}\beta_k z_{ik} + \frac{\partial}{\partial \sigma_l}\sum_{k=0}^{K^*}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\ &= \left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il}. \end{aligned} \tag{5}$$

Therefore:

$$\begin{aligned}\frac{\partial \ell\left(\sigma_l \mid \cdot\right)}{\partial \sigma_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial \sigma_l}\lambda_i(v_j) - \mathbf{1}_{v_j\in\mathbf{x}_i}\frac{\partial}{\partial \sigma_l}\log\lambda_i(v_j)\right] + \frac{\partial}{\partial \sigma_l}\log\pi(\sigma_l) \\ &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{vj}\lambda_i(v_j)\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il}\right] - \frac{\partial}{\partial \sigma_l}\frac{\sigma_l^2}{2\tau^2} \\ &= -\sum_{j=1}^{V}\sum_{i=1}^{I}\left[\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j\left(A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right)\right] - \frac{\sigma_l}{\tau^2} \\ &= -\sum_{j=1}^{V}\left[\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right]\right] - \frac{\sigma_l}{\tau^2}. \end{aligned}$$

1.3 Partial derivatives with respect to ρ_l

Again:

$$\begin{aligned}\frac{\partial \log \lambda_i(v_j)}{\partial \rho_l} &= \frac{\partial}{\partial \rho_l}\log \alpha_i + \frac{\partial}{\partial \rho_l}\sum_{k=0}^{K^*}\beta_k z_{ik} + \frac{\partial}{\partial \rho_l}\sum_{k=0}^{K}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\ &= \sigma_l\frac{\partial}{\partial \rho_l}\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il}. \end{aligned} \tag{6}$$

For ease of exposition we complete the derivation for the one-dimensional case; however, similar arguments can be used when $\mathcal{B} \subset \mathbb{R}^3$. Matrices $\mathbf{R}_l$ are circulant and so, the matrix-vector product $\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l$ can be found using the discrete Fourier transform as

$$\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l = \mathbf{F}\boldsymbol{\Phi}_l^{1/2}\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l, \tag{7}$$

where $\boldsymbol{\Phi}_l$ are the diagonal matrices containing the eigenvalues of $\mathbf{R}_l$ and $\mathbf{F}$ is the matrix of eigenvectors. In Equation (7), the only term depending on ρ_l is $\boldsymbol{\Phi}_l$ and, hence:

$$\frac{\partial}{\partial \rho_l}\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l = \mathbf{F}\frac{\partial}{\partial \rho_l}\boldsymbol{\Phi}_l^{1/2}\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l \tag{8}$$

We know that $\boldsymbol{\Phi}_l = \mathrm{diag}\left\{\phi_{l_0},\ldots,\phi_{l_{V-1}}\right\}$, where for $k = 0,\ldots,V-1$ we have that:

$$\phi_{l_k} = \sum_{j=0}^{V-1}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right), \tag{9}$$

where ι is the imaginary unit. Now it is straightforward to see that for $k = 0, \ldots, V-1$:

$$\begin{aligned}
\frac{\partial}{\partial \rho_l} \phi_{l_k}^{1/2} &= \frac{\partial}{\partial \rho_l} \sqrt{\sum_{j=0}^{V-1} \exp\left(-\rho_l ||v_0, v_j||^{\delta_l}\right) \exp\left(-\frac{2\pi\iota kj}{V}\right)} \\
&= \frac{\frac{\partial}{\partial \rho_l} \sum_{j=0}^{V-1} \exp\left(-\rho_l ||v_0, v_j||^{\delta_l}\right) \exp\left(-\frac{2\pi\iota kj}{V}\right)}{2\sqrt{\sum_{j=0}^{V-1} \exp\left(-\rho_l ||v_0, v_j||^{\delta_l}\right) \exp\left(-\frac{2\pi\iota kj}{V}\right)}} \\
&= \frac{-\sum_{j=1}^{V-1} d\left(v_0, v_j\right)^{\delta_l} \exp\left(-\rho_l ||v_0, v_j||^{\delta_l}\right) \exp\left(-\frac{2\pi\iota kj}{V}\right)}{2\phi_{l_k}^{1/2}} \\
&= -\frac{1}{2} \frac{\psi_{l_k}}{\phi_{l_k}^{1/2}}, \qquad (10)
\end{aligned}$$

where ψ_{l_k} can be viewed as the k-th eigenvalue of the of a circulant matrix $\mathbf{S}_l$ with base $\mathbf{s}_l = \left[||v_0, v_0||^{\delta_l} \exp\left(-\rho_l ||v_0, v_0||^{\delta_l}\right), \ldots, ||v_0, v_{V-1}||^{\delta_l} \exp\left(-\rho_l ||v_0, v_{V-1}||^{\delta_l}\right)\right]$ and $\mathbf{S}_l = \mathbf{F}\boldsymbol{\Psi}_l\mathbf{F}^{\mathrm{H}}$, $\boldsymbol{\Psi}_l = \mathrm{diag}\left\{\psi_{l_0}, \ldots, \psi_{l_{V-1}}\right\}$. Overall we see that:

$$\begin{aligned}
\frac{\partial}{\partial \rho_l} \mathbf{R}_l^{1/2} \boldsymbol{\gamma}_l &= \mathbf{F} \frac{\partial}{\partial \rho_l} \boldsymbol{\Phi}_l^{1/2} \mathbf{F}^{\mathrm{H}} \boldsymbol{\gamma}_l \\
&= -\frac{1}{2} \mathbf{F} \left[\boldsymbol{\Psi}_l \oslash \boldsymbol{\Phi}^{1/2}\right] \mathbf{F}^{\mathrm{H}} \boldsymbol{\gamma}_l \\
&= -\frac{1}{2} \mathbf{Q}_l \boldsymbol{\gamma}_l, \qquad (11)
\end{aligned}$$

where $\oslash$ stands for element wise division. Combining Equations (6) and (11), we find that:

$$\frac{\partial \log \lambda_i(v_j)}{\partial \rho_l} = -\frac{1}{2} \sigma_l \left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j z_{il}. \qquad (12)$$

So:

$$\begin{aligned}
\frac{\partial \ell\left(\rho_l \mid \cdot\right)}{\partial \rho_l} &= -\sum_{i=1}^{I} \sum_{j=1}^{V} \left[A_{v_j} \frac{\partial}{\partial \rho_l} \lambda_i(v_j) - \mathbf{1}_{v_j \in \mathbf{x}_i} \frac{\partial}{\partial \rho_l} \log \lambda_i(v_j)\right] + \frac{\partial}{\partial \rho_l} \log \pi(\rho_l) \\
&= -\sum_{i=1}^{I} \sum_{j=1}^{V} \left[A_{vj} \lambda_i(v_j) \left(-\frac{1}{2}\right) \sigma_l \left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j z_{il} - \mathbf{1}_{v_j \in \mathbf{x}_i} \left(-\frac{1}{2}\right) \sigma_l \left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j z_{il}\right] \\
&\quad -\frac{\partial}{\partial \rho_l} \mathbf{1}_{\rho_l \in [\rho_{\mathrm{low}}, \rho_{\mathrm{upp}}]} \\
&= \frac{\sigma_l}{2} \sum_{j=1}^{V} \sum_{i=1}^{I} \left[\left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j \left(A_{v_j} \lambda_i(v_j) z_{il} - \mathbf{1}_{v_j \in \mathbf{x}_i} z_{il}\right)\right] \\
&= \frac{\sigma_l}{2} \sum_{j=1}^{V} \left[\left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j \sum_{i=1}^{I} \left[A_{v_j} \lambda_i(v_j) z_{il} - \mathbf{1}_{v_j \in \mathbf{x}_i} z_{il}\right]\right].
\end{aligned}$$

1.4 Partial derivatives with respect to $\boldsymbol{\gamma}_l$

Finally:

$$\begin{aligned}\frac{\partial \log \lambda_i(v_j)}{\partial \boldsymbol{\gamma}_l} &= \frac{\partial}{\partial \boldsymbol{\gamma}_l}\log \alpha_i + \frac{\partial}{\partial \boldsymbol{\gamma}_l}\sum_{k=0}^{K^*}\beta_k z_{ik} + \frac{\partial}{\partial \boldsymbol{\gamma}_l}\sum_{k=0}^{K}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\ &= \sigma_l \mathbf{r}_{l_j} z_{il}, \end{aligned} \tag{13}$$

where $\mathbf{r}_{l_j}$ is the j-th row of the matrix $\mathbf{R}_l^{1/2}$. Now we can see that:

$$\begin{aligned}\frac{\partial \ell\left(\boldsymbol{\gamma}_l \mid \cdot\right)}{\partial \boldsymbol{\gamma}_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial \boldsymbol{\gamma}_l}\lambda_i(v_j) - \mathbf{1}_{v_j\in\mathbf{x}_i}\frac{\partial}{\partial \boldsymbol{\gamma}_l}\log\lambda_i(v_j)\right] + \frac{\partial}{\partial \boldsymbol{\gamma}_l}\log\pi(\boldsymbol{\gamma}_l) \\ &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\lambda_i(v_j)\sigma_l\mathbf{r}_{l_j}z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}\sigma_l\mathbf{r}_{l_j}z_{il}\right] - \frac{\partial}{\partial \boldsymbol{\gamma}_l}\frac{\boldsymbol{\gamma}_l^{\mathrm{T}}\boldsymbol{\gamma}_l}{2} \\ &= -\sigma_l\sum_{j=1}^{V}\left[\mathbf{r}_{l_j}\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right]\right] - \boldsymbol{\gamma}_l \\ &= -\sigma_l\sum_{j=1}^{V}\left[\mathbf{r}_{l_j}\mathbf{c}_{l_j}\right] - \boldsymbol{\gamma}_l \\ &= -\sigma_l\left(\mathbf{R}_l^{1/2}\right)^{\top}\mathbf{c}_l - \boldsymbol{\gamma}_l \\ &= -\sigma_l\mathbf{R}_l^{1/2}\mathbf{c}_l - \boldsymbol{\gamma}_l, \end{aligned} \tag{14}$$

since $\mathbf{R}$ is a nested block circulant matrix, where $\mathbf{c}_l$ are V-vectors with elements $\mathbf{c}_{l_j} = \sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x_i}}z_{il}\right]$.

1.5 Random effects updates

Let J be the total number of publications from which the I studies in the meta-analysis have been retrieved, and let $\mathcal{C}_j$ the set of studies retrieved from paper j ($j = 1, \ldots, J$). For all $j = 1, \ldots, J$ we have that

$$\begin{aligned}\pi(\alpha_j|\cdot) &\propto \pi(\alpha_j)\prod_{i\in\mathcal{C}_j}\pi(\mathbf{x}_i|\lambda_i) \\ &\propto \alpha_j^{\kappa-1}\exp\left\{-\alpha_j\kappa\right\}\exp\left\{-\alpha_j\sum_{i\in\mathcal{C}_j}\sum_{v=1}\lambda_{iv}^*\right\}\alpha_j^{\sum_{i\in\mathcal{C}_j}n_i} \\ &= \alpha_j^{\sum_{i\in\mathcal{C}_j}n_i+\kappa-1}\exp\left\{-\alpha_j\left(\sum_{i\in\mathcal{C}_j}\sum_{v=1}\lambda_{iv}^*+\kappa\right)\right\}. \end{aligned}$$

Therefore, we draw α_j from a $\mathcal{G}\left(\sum_{i\in\mathcal{C}_j}n_i+\kappa, \sum_{i\in\mathcal{C}_j}\sum_{v=1}\lambda_{iv}^*+\kappa\right)$ distribution.

2 Simulation studies

In order evaluate the performance of the proposed HMC algorithm to sample from the posterior distribution of the latent GPs, we consider two simulation setups. In the first we draw samples directly from the log-Gaussian Cox process model, whereas in the second we create synthetic studies based on a different model to assess its robustness to model misspecification. For consistency, all processes are defined on the same brain atlas used in the application of Section 5, consisting of 216,040 2mm^3 cubic voxels. The average number of foci per simulated dataset is kept low (mean number of foci per study is 5) to resemble the sparsity of points observed in real CBMA data. Finally, the total number of studies is fixed to 200 in both analyses, similar to the sample sizes available in real applications (Kang et al., 2011, for example).

2.1 Setup 1

In this setting we simulate 200 studies, with two spatially varying covariates that account for the mean of two groups of studies, and two non-spatially varying covariates. For $i = 1, \ldots, 200$ we set:

$$\lambda_{iv} = \exp\left\{\sum_{k=1}^{2}\left(\mu_k + \sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_v\right)z_{ik} + \sum_{i=3}^{4}\beta_k z_{ik}\right\}, \tag{15}$$

where $z_{i1} \sim \text{Bernoulli}(0.5)$, $z_{i2} = 1 - z_{i1}$, $z_{i3} \sim \text{Uniform}[-1, 1]$ and $z_{i4} \sim \text{Bernoulli}(0.5)$. Note that this parametrisation of the covariates implies existence of two types of studies, say type 1 and 2, with different spatially varying means and the effect of one continuous and one categorical covariate. The expected total number of foci is 3.99 and 4.16 for studies of type 1 and 2 respectively. We draw $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ from their $\mathcal{N}_V(0, \mathbf{I})$ prior and fix the values of the scalar parameters shown in Table 1. We run the HMC algorithm of Section 4 for 10,000 iterations, discarding the first 4,000 as a burn-in and save every 6 iterations for a total of 1,000 saved posterior samples. This took roughly 14 hours on an NVIDIA Tesla K20c GPU card.

Results are summarised in Table 1 and Figure 3. In Table 1 we see that the scalar parameters are estimated accurately despite the sparsity of points in the realisations. The 95% credible intervals contain the true values of all the parameters in the setup. Traceplots for the parameters σ_1, σ_2, ρ_1 and ρ_2 can be found in Figure 1, whereas trace plots for μ_1, μ_2, β_3 and β_4 can be found in Figure 2. The red lines indicate the true parameter values.

For $z_{i3} = z_{i4} = 0$, the median expected number of points is 3.97 (95% CI [3.84,4.10])for type 1 and 4.61 (95% CI [4.46,4.78]) for type 2. These values are very similar to the values we observe in the simulated dataset, that is 3.98 for type 1 and 4.53 for type 2. This indicates that our model does a good job fitting the data. The shape of the latent Gaussian processes $\mu_k + \sigma_k \mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k$ is generally captured for both types as can be seen in Figure 3. In particular, we can see that the maxima in the true and estimated images

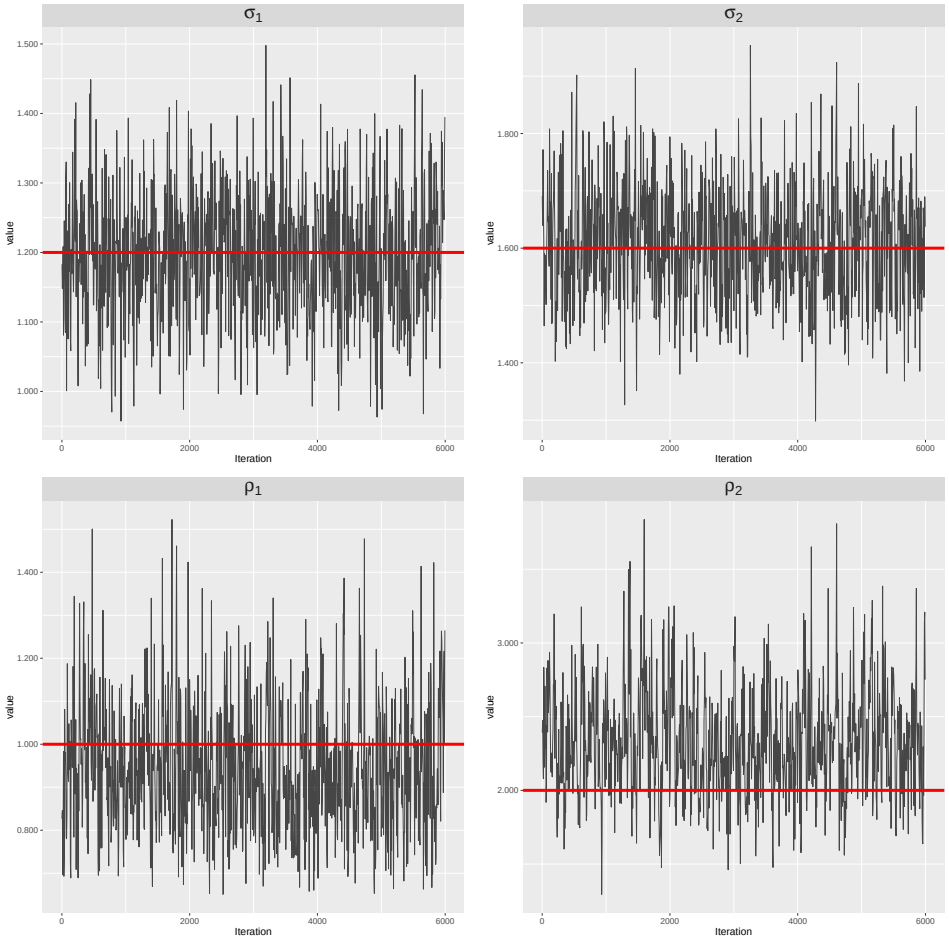

Figure 1: Posterior traceplots for the scalar parameters of the LGCP model used to fit the data of Section 2.1. Top row: standard deviations. Bottom row: correlation decay parameters ($\times 100$). The true values are indicated by the solid red lines.

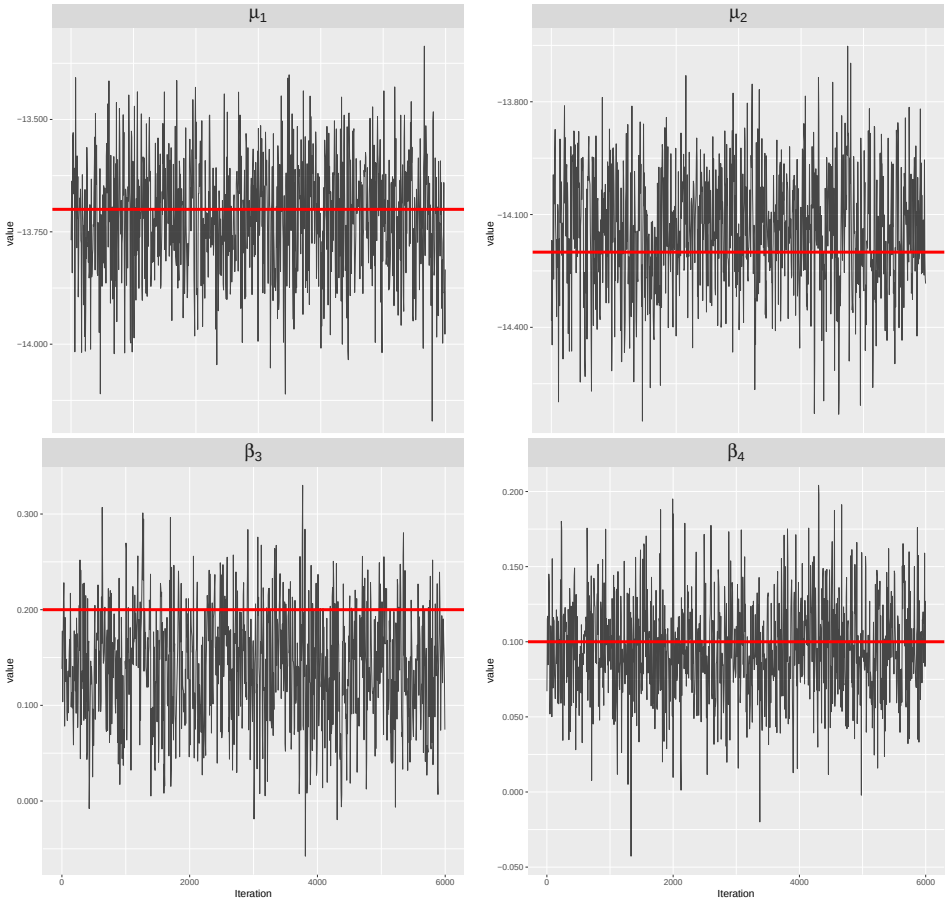

Figure 2: Posterior traceplots for the scalar parameters of the LGCP model used to fit the data of Section 2.1. Top row: overall latent process means. Bottom row: regression coefficients for covariates z_3 and z_4. The true values are indicated by the solid red lines.

appear roughly in the same locations. The same cannot be said about the other values but this is expected given the dearth of information in regions of low intensity.

Table 1: Posterior summaries of the scalar parameters of the LGCP model, fit to the simulated data of Section 2.1. Results are based on 1,000 posterior draws. The values for the correlation parameters ρ_1, ρ_2 are multiplied by 100. The values for β_1 and β_2 are multiplied by 10.

Parameter	**True Value**	**Posterior median**	95% **credible interval**
μ_1	-13.7	-13.72	-13.99 , -13.48
μ_2	-14.2	-14.14	-14.47 , -13.86
σ_1	1.2	1.19	1.01 , 1.38
σ_2	1.6	1.61	1.43 , 1.81
ρ_1	1	0.93	0.69 , 1.27
ρ_2	2	2.30	1.69 , 3.15
β_3	2	1.44	0.22 , 2.52
β_4	1	0.95	0.32 , 1.65

2.2 Setup 2

In this setup we create datasets with a pattern of points that follows brain structures of interest. Again there are two types of studies, say type 1 and type 2. For each study i, $i = 1, \ldots, 200$, we generate the total number of points from a Negative Binomial distribution with mean $\mu = 6 + 2z_{i3} - \mathbf{1}_{\{z_{i4}=0\}} + \mathbf{1}_{\{z_{i4}=1\}}$ and variance $\mu^2/20$. For the covariates, $z_{i3} \sim \text{Uni}[-1, 1]$ and $z_{i4} \sim \text{Bernoulli}(0.5)$. Once we know the exact number of foci per study, we assign the study uniformly at random to one of the 2 types and the distribute its foci as follows. For type 1, foci appear systematically in the following regions: each focus can be observed in the right amygdala (B_R) with probability 55%, the orbifrontal cortex (B_C) with probability 30% or anywhere else in the brain with probability 15%. The configuration for type 2 differs in that most of the points will go to the left amygdala (B_L) instead of the right amygdala. If a focus is assigned to one of the three broad regions, the exact location has a uniform distribution over the region. In the fourth column of Figure 4 the regions in red and blue correspond to the left and right amygdala respectively while the orbifrontal cortex is coloured in green.

HMC is run for 10,000 iterations, discarding the 4,000 first as a burn-in and saving every 6 to obtain a total of 1,000 samples from the posterior. The run took approximately 15 hours on a Tesla K20c GPU card.

Results are shown in Figure 4 where in the first two columns we see median posterior log-intensities for the two types, in different axial slices. In both cases, we find that the regions with the highest intensities are the amygdalae and that the orbifrontal cortex is a region of high posterior intensity as well. The median expected number of points is 5.81 for type 1 (95% CI [5.36,6,32]) and 6.45 for type 2 (95% CI [5.97,6.97]). The observed values are 6.27 and 6.73 respectively.

Conditional on there being exactly one focus, we can estimate the probability that this focus appears in any subset $B \subseteq \mathcal{B}$ as $\int_B \lambda(\xi)\, d\xi / \int_{\mathcal{B}} \lambda(\xi)\, d\xi$. Using the posterior

draws obtained from the HMC algorithm, we can obtain the posterior distribution of any such quantity. For our simulated type 1 data we find that the median posterior probability of observing a focus in the right amygdala (B_R) is 0.43 (95% CI [0.40,0.48]). For type 2, the probability of observing a focus in the left amygdala (B_L) is 0.42 (95% CI [0.39,0.46]). For the orbifrontal cortex (B_C) the median posterior probabilities are 0.25 for type 1 and 0.23 for type 2, with 95% credible intervals $[0.22, 0.28]$ and $[0.20, 0.26]$ respectively. We therefore see that the model underestimates the probabilities for B_R, B_L and B_C. This bias can be attributed to the smoothness that is imposed by our parameter δ thus leading to increases intensities just outside these regions as well as regions where noise foci appear.

An interesting question one may ask is which are the regions of the brain that are activated by one type or the other, but not both. To answer this, one can construct the mean standardised posterior difference map computed as the ratio of the posterior mean of the difference $(\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v$, to the posterior standard deviation of that difference: $\frac{(\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v}{\mathrm{sd}\left((\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v\right)}$. Extreme negative or positive values are evidence of differences between the two types. We show the difference map in the third column of Figure 4. As we see, the model distinguishes the the two types in the amygdala but the differences are small in the rest of the brain.

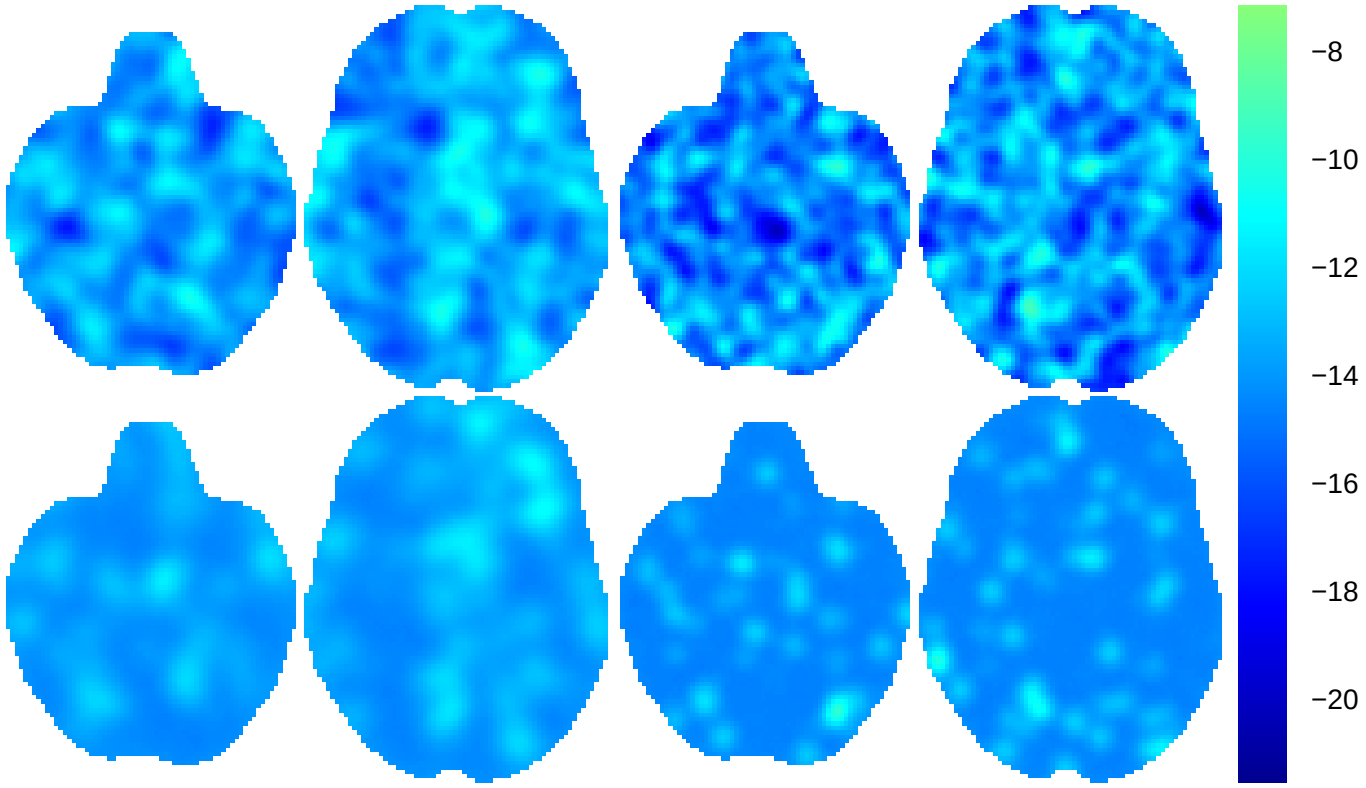

Figure 3: Some true (top row) and estimated (bottom row) latent Gaussian processes for type 1 (columns 1 and 2) and type 2 (columns 3 and 4) in the simulation setup 1 of Section 2.1. Columns 1 and 3 correspond to axial slice $z = -22$; columns 2 and 4 correspond to axial slice $z = 4$. While they may appear dissimilar at first, observe that the most intense regions of the true and estimated intensity match up; in this punishingly sparse setting (mean ≈ 4 foci per 3D image), the less intense regions have too few points to learn the intensity.

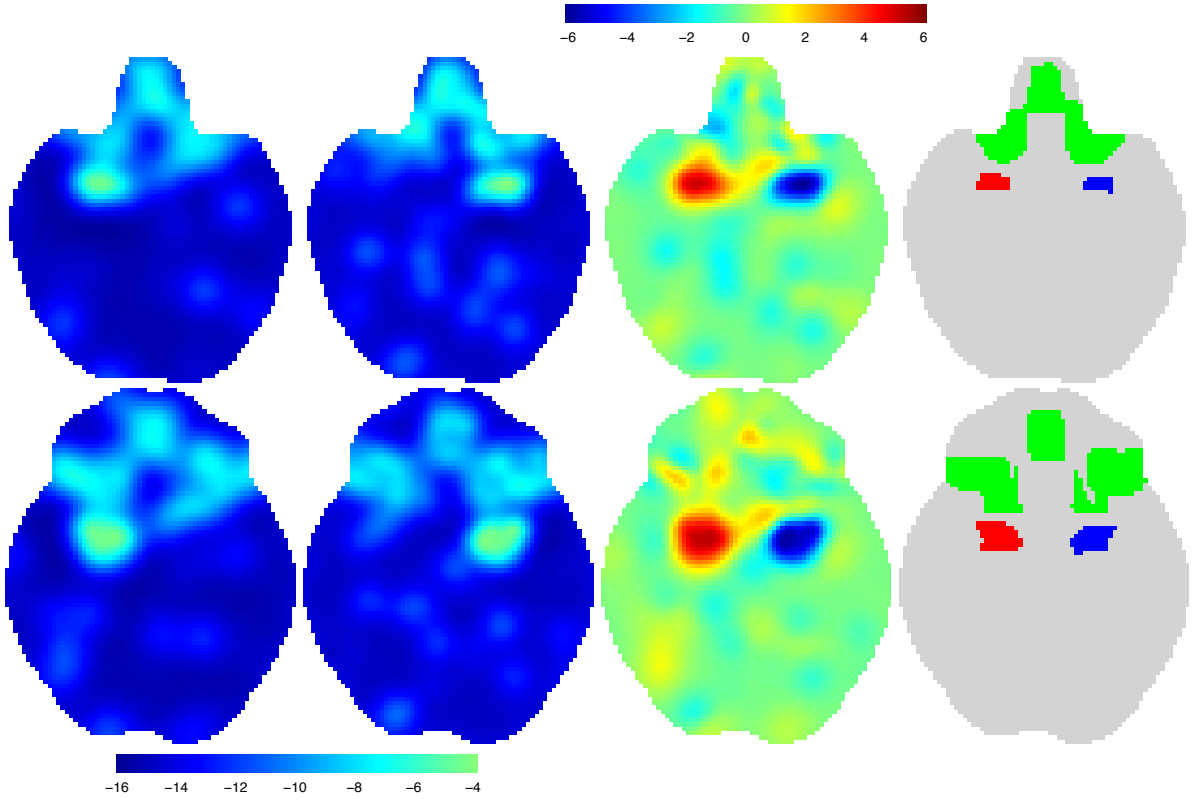

Figure 4: Results for simulation setup 2 of Section 2.2. The top row corresponds to axial slice $z = -24$ whereas the bottom row corresponds to axial slice $z = -16$. Columns 1 and 2 are the estimated log-intensities for type 1 and type 2 respectively. The third column is the standardised mean posterior difference between the two latent Gaussian processes in the corresponding slice; bright colours indicate areas mostly activated by type 1 process. The fourth column shows the regions of the brain systematically activated by the two processes; red for type 1, blue for type 2 and green for both.

3 Analysis of the WM dataset: supplementary plots

This section contains MCMC convergence diagnostics and supplemental material for the real data analysis of Section 5 of the main paper. Posterior traceplots are shown in Figures 5, 6, 7 and 8. The posterior traceplots for the marginal standard deviation parameters σ_k and the correlation decay parameters ρ_k are shown in Figure 5 (top and bottom row, respectively). Figure 6 shows posterior traceplots for the overall mean parameters μ_k (top row), the regression coefficient of the sample size covariate (bottom row, first subplot) and the integrated intensities of verbal and non-verbal studies (bottom row, second and third subplot, respectively). Let v_k^m and v_k^M be the voxels where the k-th latent GP $\boldsymbol{\beta}_k$ has the minimum and maximum mean posterior values, respectively. Posterior traceplots of $(\boldsymbol{\beta}_0)_{v_0^m}$, $(\boldsymbol{\beta}_0)_{v_0^M}$, $(\boldsymbol{\beta}_1)_{v_1^m}$, $(\boldsymbol{\beta}_1)_{v_1^M}$, $(\boldsymbol{\beta}_2)_{v_2^m}$ and $(\boldsymbol{\beta}_2)_{v_2^M}$ are shown in Figure 7. Finally, in Figure 8 we present α_m and α_M, where m and M index the studies with the lowest and largest mean posterior random effects, respectively. All chains are obtained after applying a thinning factor of 15 to the original MCMC chains of length 15,000. Finally, Figure 9 shows the voxelwise posterior variance of $\exp\{\boldsymbol{\beta}_2\}$, the multiplicative age effect on the intensity of both verbal and non-verbal studies.

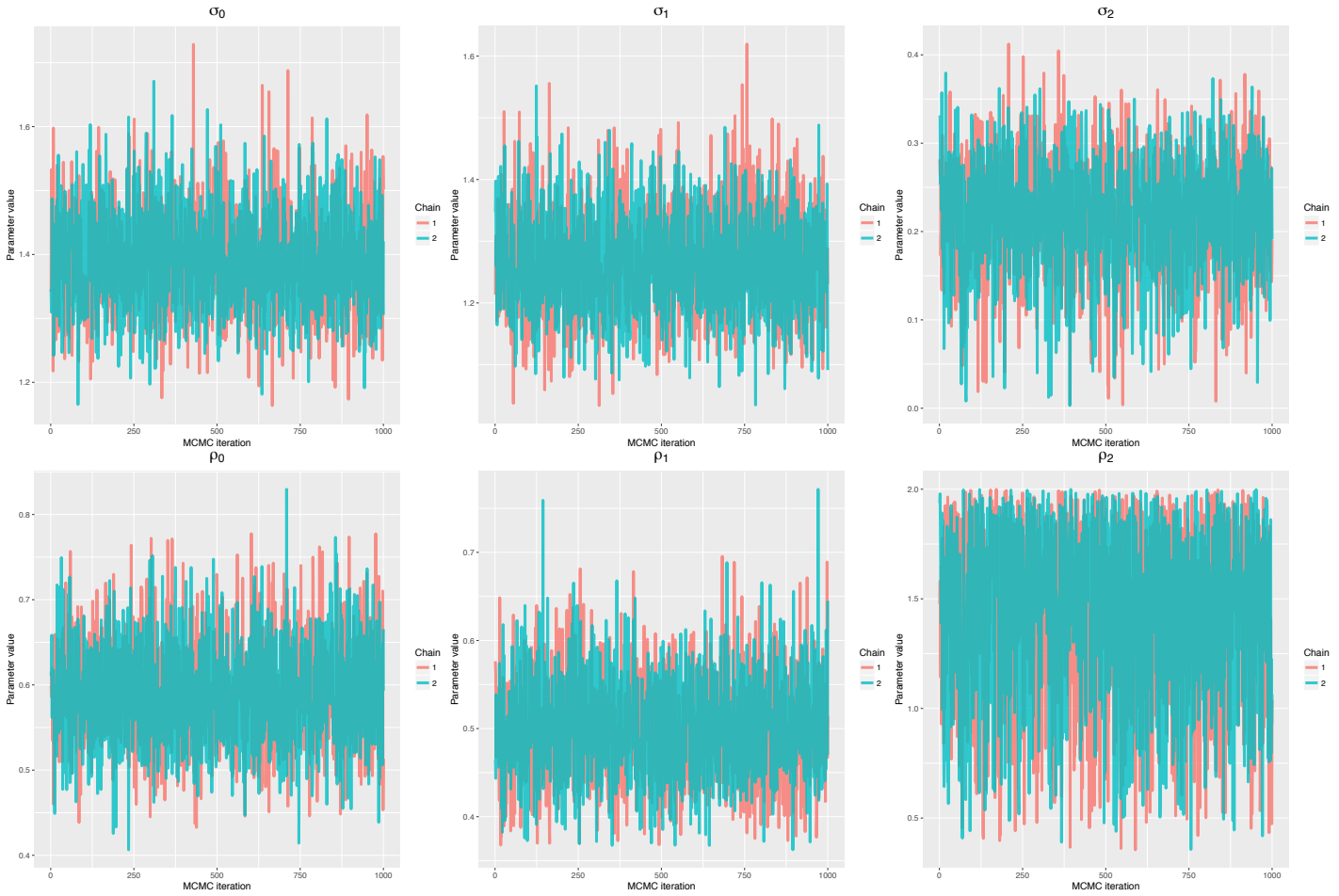

Figure 5: Posterior traceplots for some of the scalar parameters of the LGCP model of Equation 10 of the main paper that we fit to the working memory dataset presented in Section 2 of the main paper. Top row: marginal standard deviations σ_k of the 3 latent GPs. Bottom row: correlation decay parameters ρ_k of the 3 latent GPs. The chains have been obtained after applying a thinning factor of 15 to the original chains of length 15,000.

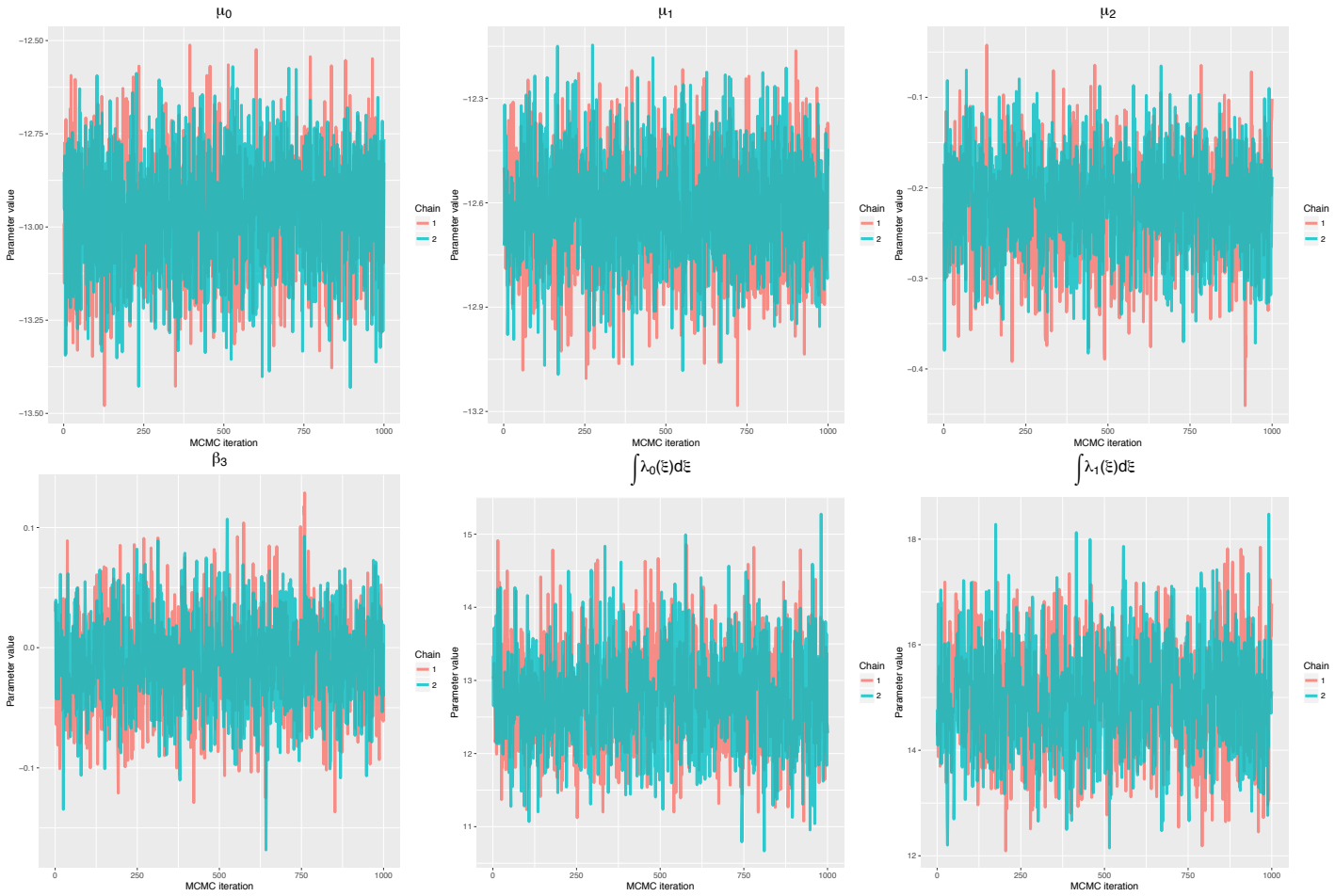

Figure 6: Posterior traceplots for some of the scalar parameters of the LGCP model of Equation 10 of the main paper that we fit to the working memory dataset presented in Section 2 of the main paper. Top row: overall means μ_k of the 3 latent GPs. Bottom row, subplot 1: regression coefficient β_3 for the sample size covariate. Bottom row, subplots 2-3: integrated intensities for verbal and non-verbal studies, respectively. The chains have been obtained after applying a thinning factor of 15 to the original chains of length 15,000.

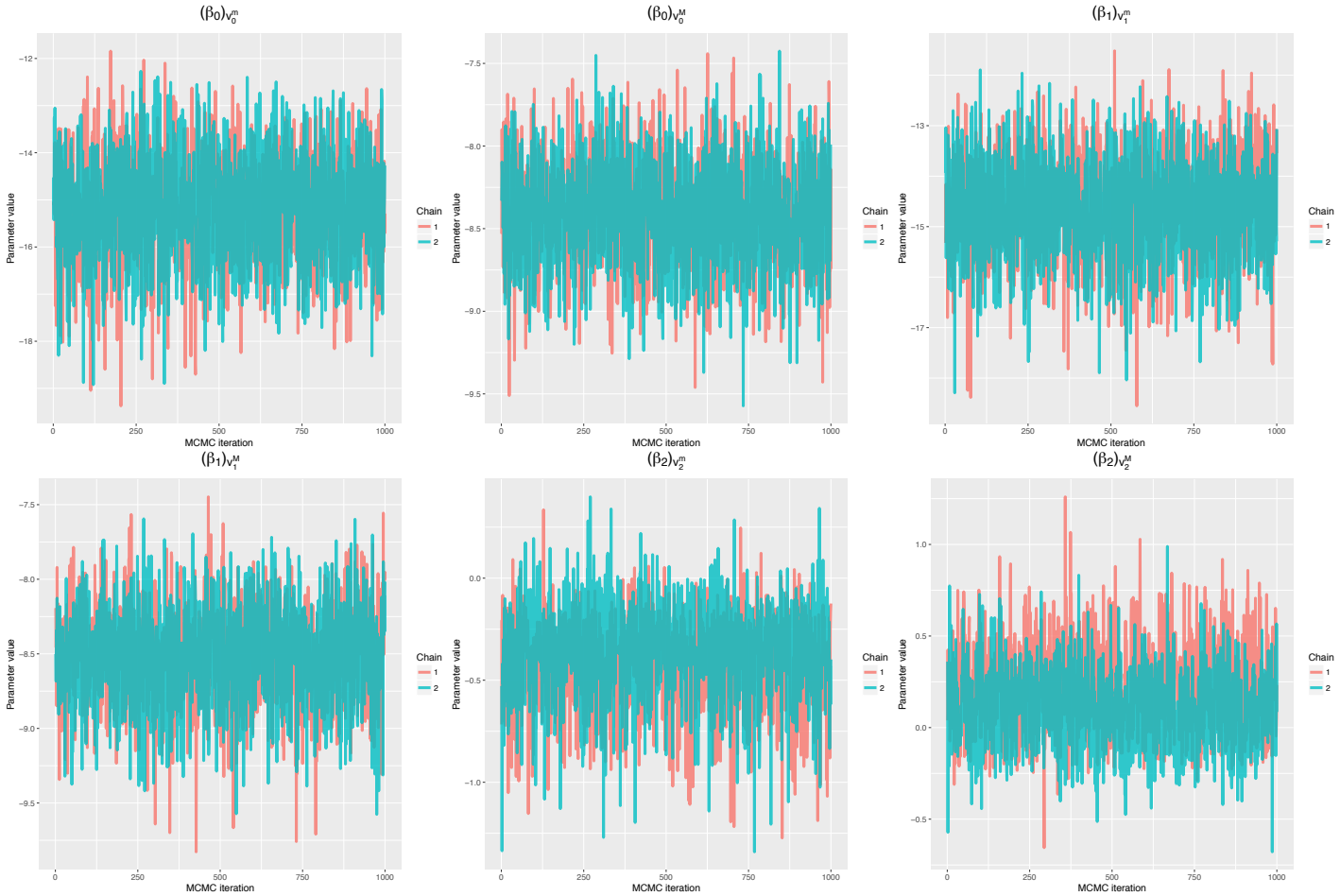

Figure 7: Posterior traceplots for some of the scalar parameters of the LGCP model of Equation 10 of the main paper that we fit to the working memory dataset presented in Section 2 of the main paper. Subplots 1-6 correspond to $(\boldsymbol{\beta}_0)_{v_0^m}$, $(\boldsymbol{\beta}_0)_{v_0^M}$, $(\boldsymbol{\beta}_1)_{v_1^m}$, $(\boldsymbol{\beta}_1)_{v_1^M}$, $(\boldsymbol{\beta}_2)_{v_2^m}$ and $(\boldsymbol{\beta}_2)_{v_2^M}$, where v_k^m and v_k^M represent the voxels with the minimum and maximum mean posterior values of $\boldsymbol{\beta}_k$, respectively. The chains have been obtained after applying a thinning factor of 15 to the original chains of length 15,000.

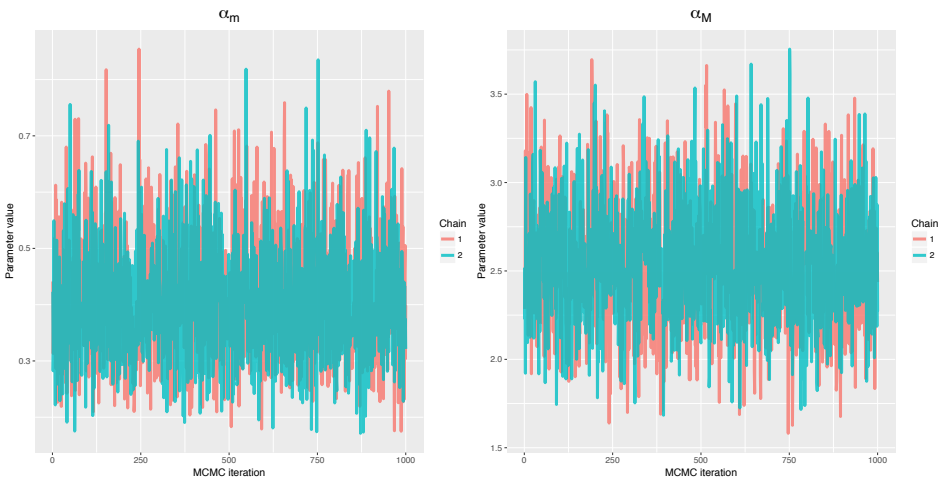

Figure 8: Posterior traceplots for some of the scalar parameters of the LGCP model of Equation 10 of the main paper that we fit to the working memory dataset presented in Section 2 of the main paper. Subplots 1-2 represent α_m and α_M, where m and M index the studies with the lowest and largest mean posterior random effects, respectively. The chains have been obtained after applying a thinning factor of 15 to the original chains of length 15,000.

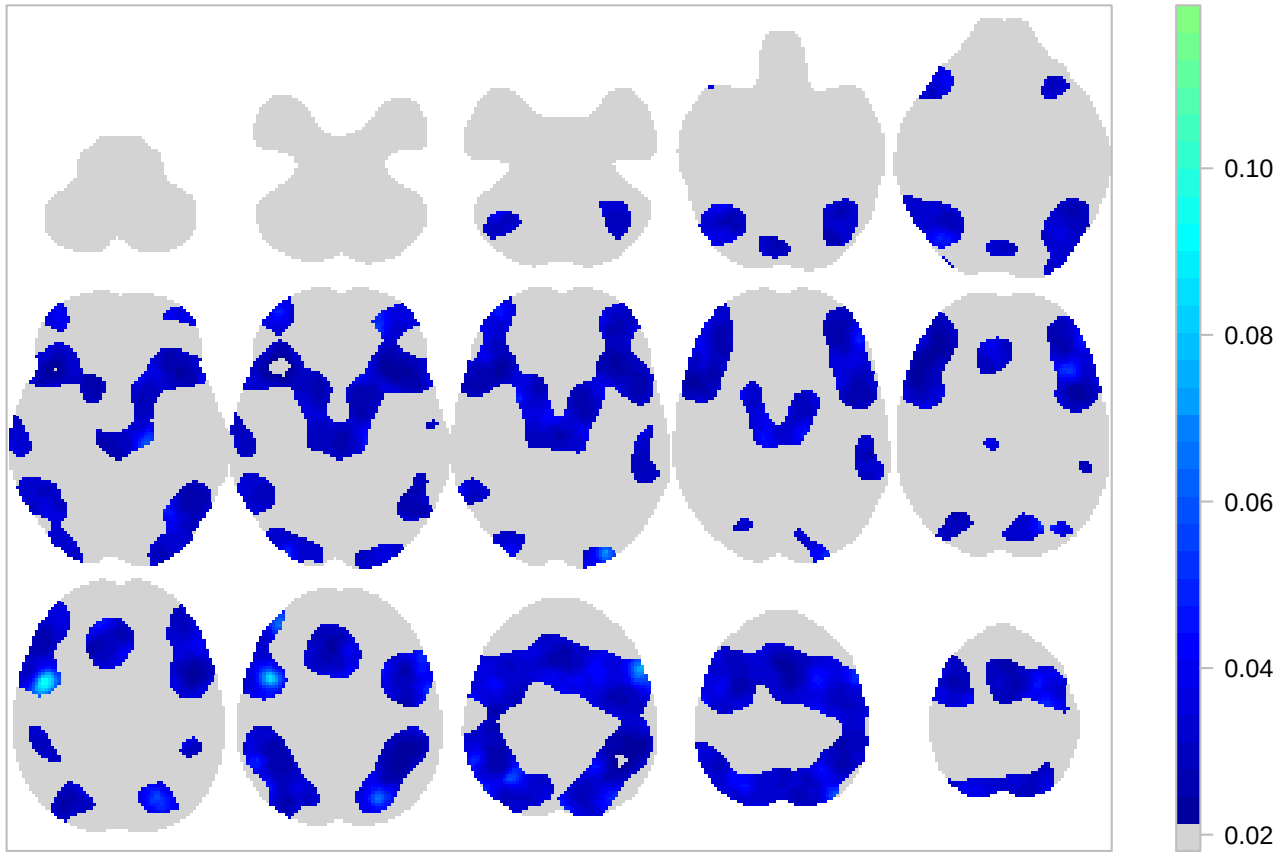

Figure 9: Posterior variance of $\exp\{\boldsymbol{\beta}_2\}$, the multiplicative age effect on the intensity of both verbal and non-verbal studies. Top row shows (from left to right) axial slices $z = -50, -42, -34, -26$ and -18, respectively. Middle row shows axial slices $z = -10, -2, +6, +14$ and $+22$, respectively. Bottom row shows axial slices $z = +30, +38, +46, +54$ and $+62$, respectively. Voxels for which the mean posterior $\boldsymbol{\lambda}$ is low (below the 75% quantile over the brain) have been set to zero.

4 Full brain analysis

In this section we present full brain results for the working memory CBMA conducted in Section 5 of the main paper. Table 2 presents probabilities of activation and expected number of foci for several ROIs, along with their 95% credible intervals. In Table 3 we compare the probability of an activation between working memory studies using verbal and non-verbal stimuli, using the same ROIs. The ROIs have been obtained from the Harvard-Oxford atlas (Desikan et al., 2006). All quantities are based on 1,000 MCMC samples which are obtained after applying a thinning factor of 15 to the original MCMC chains of length 15,000.

Table 2: Working memory CBMA results. Table presents posterior summaries for the % probability of at least one activation in a given ROI (rounded to 2 decimal points), as well as the expected number of foci (rounded to 3 decimal points). All quantities have been calculated based on 1,000 MCMC samples.

		$\mathbb{P}(\mathbf{N_X(B)} \geq \mathbf{1})$			$\int_{\mathbf{B}} \lambda(\xi)\mathbf{d}\xi$		
ROI	**Vol.**	**Mean**	$\mathbf{p_{0.025}}$	$\mathbf{p_{0.975}}$	**Mean**	$\mathbf{p_{0.025}}$	$\mathbf{p_{0.975}}$
Frontal pole	25900	60.24	51.05	66.42	0.926	0.714	1.091
Insular cortex	3613	33.39	26.68	39.36	0.407	0.310	0.500
Superior frontal gyrus	8861	38.27	27.74	43.98	0.484	0.325	0.579
Middle frontal gyrus	8421	60.55	48.23	66.48	0.933	0.658	1.093
Inferior frontal gyrus, pars triangularis	2317	22.54	15.44	27.94	0.256	0.168	0.328
Inferior frontal gyrus, pars opercularis	2335	39.88	31.06	45.96	0.510	0.372	0.615
Precentral gyrus	13967	68.47	59.96	73.72	1.158	0.915	1.336
Temporal pole	8044	9.60	5.41	13.48	0.101	0.056	0.145
Superior temporal gyrus, anterior division	916	1.77	0.80	2.87	0.018	0.008	0.029
Superior temporal gyrus, posterior division	2897	8.89	5.46	11.91	0.093	0.056	0.127
Middle temporal gyrus, anterior division	1425	1.64	0.55	2.88	0.017	0.005	0.029
Middle temporal gyrus, posterior division	4206	9.37	5.18	12.76	0.099	0.053	0.137
Middle temporal gyrus, temporooccipital part	3202	12.84	6.85	17.63	0.138	0.071	0.194
Inferior temporal gyrus, anterior division	1287	1.38	0.38	2.56	0.014	0.004	0.026
Inferior temporal gyrus, posterior division	4138	5.43	2.80	8.12	0.056	0.028	0.085
Inferior temporal gyrus, temporooccipital part	2605	11.49	6.16	16.05	0.122	0.064	0.175
Postcentral gyrus	10638	24.29	17.60	29.24	0.279	0.194	0.346
Superior parietal lobule	4489	36.16	26.16	42.31	0.450	0.303	0.550
Supramarginal gyrus, anterior division	2910	13.19	7.98	17.24	0.142	0.083	0.189
Supramarginal gyrus, posterior division	4071	28.38	19.69	33.72	0.335	0.219	0.411
Angular gyrus	3703	21.69	14.39	26.34	0.245	0.155	0.306

Lateral occipital cortex, superior division	14484	61.52	53.01	67.73	0.959	0.755	1.131
Lateral occipital cortex, inferior division	7490	27.53	18.66	33.57	0.323	0.207	0.409
Intracalcarine cortex	2211	8.36	4.35	11.76	0.087	0.045	0.125
Frontal medial cortex	1539	1.94	0.64	3.62	0.020	0.006	0.037
Juxtapositional lobule cortex	2282	26.46	16.14	32.74	0.308	0.176	0.397
Subcallosal cortex	2176	4.47	1.73	7.16	0.046	0.017	0.074
Paracingulate gyrus	4095	46.22	35.94	52.89	0.622	0.445	0.753
Cingulate gyrus, anterior division	4144	22.61	14.76	28.19	0.257	0.160	0.331
Cingulate gyrus, posterior division	4668	11.07	5.70	15.52	0.118	0.059	0.169
Precuneous cortex	7844	26.46	18.08	32.62	0.308	0.199	0.395
Cuneal cortex	1743	7.75	4.09	10.95	0.081	0.042	0.116
Frontal orbital cortex	5188	36.94	27.27	43.06	0.462	0.318	0.563
Parahippocampal gyrus, anterior division	3313	5.69	2.94	8.43	0.059	0.030	0.088
Parahippocampal gyrus, posterior division	2014	6.01	3.23	8.71	0.062	0.033	0.091
Lingual gyrus	5388	15.62	9.12	20.15	0.170	0.096	0.225
Temporal fusiform cortex, anterior division	1243	1.21	0.39	2.22	0.012	0.004	0.022
Temporal fusiform cortex, posterior division	2951	5.72	3.10	8.42	0.059	0.032	0.088
Temporal occipital fusiform cortex	2458	13.65	8.21	17.69	0.147	0.086	0.195
Occipital fusiform gyrus	3587	23.35	15.64	28.33	0.266	0.170	0.333
Frontal operculum cortex	1062	16.48	9.97	20.60	0.180	0.105	0.231
Central opercular cortex	2578	11.08	6.70	14.67	0.118	0.069	0.159
Parietal operculum cortex	1684	6.23	3.23	8.92	0.064	0.033	0.093
Planum polare	1210	2.64	1.29	3.94	0.027	0.013	0.040
Heschl's gyrus	786	2.51	1.08	3.89	0.025	0.011	0.040
Planum temporale	1442	6.05	3.26	8.45	0.063	0.033	0.088
Supracalcarine cortex	424	1.52	0.66	2.41	0.015	0.007	0.024
Occipital pole	9658	20.42	12.04	25.75	0.229	0.128	0.298
Left cerebral white matter	28034	81.33	75.05	85.13	1.684	1.388	1.906
Left cerebral cortex	82249	99.04	98.05	99.44	4.679	3.938	5.193
Left lateral ventricle	1289	4.10	1.96	6.02	0.042	0.020	0.062

Left thalamus	1591	16.17	9.31	21.64	0.177	0.098	0.244
Left caudate	572	3.60	1.75	5.44	0.037	0.018	0.056
Left putamen	923	7.83	3.91	10.96	0.082	0.040	0.116
Left pallidum	312	5.15	2.58	7.59	0.053	0.026	0.079
Brain stem	8078	13.45	7.80	18.29	0.145	0.081	0.202
Left hippocampus	921	3.38	1.21	5.37	0.034	0.012	0.055
Left amygdala	390	1.41	0.48	2.43	0.014	0.005	0.025
Left accumbens	111	0.88	0.26	1.57	0.009	0.003	0.016
Right cerebral white matter	31216	79.16	72.13	83.13	1.573	1.278	1.780
Right cerebral cortex	86480	98.78	97.36	99.24	4.431	3.636	4.874
Right lateral ventricle	1019	3.36	1.77	4.97	0.034	0.018	0.051
Right thalamus	1398	12.88	7.12	17.43	0.138	0.074	0.192
Right caudate	515	2.68	1.02	4.30	0.027	0.010	0.044
Right putamen	800	6.82	3.39	9.80	0.071	0.034	0.103
Right pallidum	266	3.24	1.08	5.19	0.033	0.011	0.053
Right hippocampus	772	1.78	0.67	2.99	0.018	0.007	0.030
Right amygdala	399	0.86	0.19	1.73	0.009	0.002	0.017
Right accumbens	86	0.64	0.15	1.26	0.006	0.001	0.013

Table 3: Working memory CBMA results. Table presents posterior summaries for the % probability of a reported activation in studies using verbal and non-verbal stimuli. All quantities have been calculated based on 1,000 MCMC samples.

		Verbal			Non-verbal		
ROI	**Vol.**	**Mean**	$\mathbf{p_{0.025}}$	$\mathbf{p_{0.975}}$	**Mean**	$\mathbf{p_{0.025}}$	$\mathbf{p_{0.975}}$
Frontal pole	25900	59.99	46.81	67.86	60.24	41.75	69.92
Insular cortex	3613	32.79	21.14	40.70	33.86	20.43	43.28
Superior frontal gyrus	8861	29.45	18.36	36.70	45.87	29.60	54.29
Middle frontal gyrus	8421	53.05	42.42	60.85	66.66	50.91	74.55
Inferior frontal gyrus, pars triangularis	2317	19.43	10.84	24.95	25.44	14.64	33.90
Inferior frontal gyrus, pars opercularis	2335	43.66	31.68	51.15	35.69	21.49	44.80
Precentral gyrus	13967	64.10	51.56	71.00	72.09	57.82	79.52
Temporal pole	8044	7.54	2.91	11.50	11.58	4.28	18.15
Superior temporal gyrus, anterior division	916	2.31	0.71	4.05	1.23	0.27	2.52
Superior temporal gyrus, posterior division	2897	10.76	5.31	15.14	6.95	2.25	11.32
Middle temporal gyrus, anterior division	1425	2.19	0.48	4.36	1.09	0.15	2.56
Middle temporal gyrus, posterior division	4206	9.58	4.59	14.22	9.12	3.83	14.44
Middle temporal gyrus, temporooccipital part	3202	11.88	5.65	17.75	13.73	5.73	21.06
Inferior temporal gyrus, anterior division	1287	1.76	0.31	3.69	0.99	0.10	2.53
Inferior temporal gyrus, posterior division	4138	4.85	1.59	8.09	5.98	1.47	10.31
Inferior temporal gyrus, temporooccipital part	2605	7.45	3.00	11.83	15.31	6.95	22.66
Postcentral gyrus	10638	20.37	11.48	26.10	27.95	16.64	35.68
Superior parietal lobule	4489	38.81	26.05	46.48	33.24	19.67	42.96
Supramarginal gyrus, anterior division	2910	12.78	5.90	17.83	13.56	6.23	19.83
Supramarginal gyrus, posterior division	4071	29.93	19.40	36.61	26.70	14.58	35.17
Angular gyrus	3703	24.30	15.51	31.09	18.91	9.32	25.61
Lateral occipital cortex, superior division	14484	55.45	42.51	63.87	66.55	52.60	74.58

Lateral occipital cortex, inferior division	7490	23.69	14.54	30.51	31.06	18.71	40.58
Intracalcarine cortex	2211	6.51	2.61	10.12	10.15	4.37	15.19
Frontal medial cortex	1539	1.21	0.24	2.62	2.66	0.60	5.58
Juxtapositional lobule cortex	2282	21.21	12.33	28.23	31.23	18.26	41.74
Subcallosal cortex	2176	4.20	1.08	7.66	4.72	1.03	9.24
Paracingulate gyrus	4095	42.91	29.21	51.04	49.14	33.89	59.57
Cingulate gyrus, anterior division	4144	17.69	10.95	24.25	27.15	14.86	36.16
Cingulate gyrus, posterior division	4668	8.06	3.60	12.42	13.93	5.98	21.66
Precuneous cortex	7844	22.88	12.61	30.03	29.76	18.38	39.96
Cuneal cortex	1743	6.71	1.97	10.43	8.76	3.29	14.54
Frontal orbital cortex	5188	37.26	25.37	45.24	36.48	24.30	45.43
Parahippocampal gyrus, anterior division	3313	5.93	2.24	9.59	5.44	1.65	9.35
Parahippocampal gyrus, posterior division	2014	5.80	2.27	9.47	6.21	2.46	10.81
Lingual gyrus	5388	15.61	7.56	21.30	15.56	7.52	22.72
Temporal fusiform cortex, anterior division	1243	1.52	0.33	3.13	0.90	0.18	2.11
Temporal fusiform cortex, posterior division	2951	5.26	2.06	8.46	6.17	2.13	10.49
Temporal occipital fusiform cortex	2458	8.87	4.19	13.17	18.12	8.94	26.00
Occipital fusiform gyrus	3587	26.68	17.03	33.74	19.78	10.47	27.65
Frontal operculum cortex	1062	15.27	9.08	20.31	17.64	9.10	23.87
Central opercular cortex	2578	10.04	4.95	14.41	12.08	5.64	17.46
Parietal operculum cortex	1684	7.35	2.65	11.29	5.08	1.56	9.05
Planum polare	1210	2.37	0.79	4.02	2.91	0.98	5.02
Heschl's gyrus	786	3.07	0.99	5.47	1.94	0.34	3.79
Planum temporale	1442	7.38	3.50	10.90	4.69	1.55	8.26
Supracalcarine cortex	424	0.95	0.29	1.74	2.07	0.51	3.64
Occipital pole	9658	19.84	10.66	26.91	20.91	11.10	29.20
Left cerebral white matter	28034	76.33	66.31	81.39	85.13	75.35	89.94
Left cerebral cortex	82249	98.88	97.31	99.41	99.14	97.59	99.63
Left lateral ventricle	1289	3.12	1.30	5.16	5.05	1.65	8.64
Left thalamus	1591	14.56	7.44	20.87	17.69	7.35	26.38

Left caudate	572	3.49	1.25	5.71	3.69	1.09	6.53
Left putamen	923	6.97	2.01	11.15	8.67	2.25	13.88
Left pallidum	312	6.26	2.38	10.29	4.01	1.24	7.29
Brain stem	8078	10.42	5.26	15.72	16.33	6.90	25.06
Left hippocampus	921	3.46	1.08	6.28	3.29	0.74	6.39
Left amygdala	390	1.21	0.16	2.59	1.60	0.25	3.37
Left accumbens	111	1.07	0.15	2.29	0.69	0.09	1.49
Right cerebral white matter	31216	73.44	65.11	78.96	83.50	74.00	88.46
Right cerebral cortex	86480	98.22	96.50	98.97	99.12	97.60	99.61
Right lateral ventricle	1019	2.66	0.96	4.35	4.05	1.21	6.74
Right thalamus	1398	15.17	8.00	21.79	10.48	3.36	16.86
Right caudate	515	2.51	0.76	4.45	2.85	0.68	5.34
Right putamen	800	5.91	1.94	9.49	7.69	2.56	12.98
Right pallidum	266	2.74	0.65	4.97	3.73	0.93	6.91
Right hippocampus	772	1.77	0.52	3.42	1.78	0.46	3.55
Right amygdala	399	0.86	0.11	2.03	0.87	0.13	2.17
Right accumbens	86	0.78	0.13	1.84	0.50	0.08	1.30

References

Desikan, R. S., F. Sgonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage 31*(3), 968–980.

Kang, J., T. D. Johnson, T. E. Nichols, and T. D. Wager (2011). Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *Journal of the American Statistical Association 106*(493), 124–134.