ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2019)
68, *Part* 1, *pp.* 217–234

Bayesian log-Gaussian Cox process regression: applications to meta-analysis of neuroimaging working memory studies

Pantelis Samartsidis,
University of Cambridge, UK

Claudia R. Eickhoff and Simon B. Eickhoff,
Heinrich-Heine University Düsseldorf, and Forschungszentrum Jülich, Germany

Tor D. Wager,
University of Colorado at Boulder, USA

Lisa Feldman Barrett,
Northeastern University, Boston, USA

Shir Atzil,
Hebrew University of Jerusalem, Israel

Timothy D. Johnson
University of Michigan, Ann Arbor, USA

and Thomas E. Nichols
University of Oxford, UK

[Received January 2017. Final revision May 2018]

Summary. Working memory (WM) was one of the first cognitive processes studied with functional magnetic resonance imaging. With now over 20 years of studies on WM, each study with tiny sample sizes, there is a need for meta-analysis to identify the brain regions that are consistently activated by WM tasks, and to understand the interstudy variation in those activations. However, current methods in the field cannot fully account for the spatial nature of neuroimaging meta-analysis data or the heterogeneity observed among WM studies. In this work, we propose a fully Bayesian random-effects metaregression model based on log-Gaussian Cox processes, which can be used for meta-analysis of neuroimaging studies. An efficient Markov chain Monte Carlo scheme for posterior simulations is presented which makes use of some recent advances in parallel computing using graphics processing units. Application of the proposed model to a real data set provides valuable insights regarding the function of the WM.

Keywords: Functional magnetic resonance imaging; Metaregression; Random-effects meta-analysis; Working memory

Address for correspondence: Pantelis Samartsidis, Medical Research Council Biostatistics Unit, University of Cambridge, Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK.
E-mail: pantelis@mrc-bsu.cam.ac.uk

1. Introduction

1.1. The working memory

Humans depend on working memory (WM) for many behaviours and cognitive tasks. WM includes both the retention of information (also known as short-term memory), as well as the manipulation of information over a short duration. An example of the former is remembering a phone number until you dial it, whereas an example of the latter is building a 'mental map' while receiving directions. WM is impaired in a number of neurological and psychiatric diseases, most notably in all forms of dementia.

With its central role in everyday behaviour and implication in disease, WM has been frequently studied with functional brain imaging techniques like functional magnetic resonance imaging (MRI). Functional MRI is sensitive to changes in blood flow, volume and oxygenation level in the brain and provides a non-invasive way to identify regions of the brain that are associated with a given task or behaviour. However, each functional MRI study has traditionally had very small samples, rarely exceeding 20. Thus, there is a need for meta-analysis methods to pool information over studies, separating consistent findings from those occurring by chance, as well as metaregression methods (Greenland, 1994) to understand heterogeneity in terms of study-specific characteristics.

1.2. Neuroimaging meta-analyses

In functional MRI there are two broad approaches for meta-analysis. When the full statistical images from each study are available, i.e. effect sizes and associated standard errors for all voxels in the brain, an intensity-based meta-analysis can proceed by means of standard meta-analytic methods (see Hartung *et al.* (2008) for an overview). However, these statistic images (more than 200000 voxels) traditionally have not been shared by researchers. Instead, researchers publish only the x, y, z brain atlas co-ordinates of the local maxima in significant regions of the statistic image. We call these co-ordinates the foci (singular focus). When only foci are available then a co-ordinate-based meta-analysis (CBMA) is conducted. As can be expected, the transition from full images to the lists of foci involves a heavy loss of information (Salimi-Khorshidi *et al.*, 2009). However, since the vast majority of researchers rarely provide the full images, CBMA constitutes the main approach for functional MRI meta-analysis.

Most work in the field is focused on so-called *kernel-based* methods such as activation likelihood estimation (Turkeltaub *et al.*, 2002; Eickhoff *et al.*, 2012), multilevel kernel density analysis (Wager *et al.*, 2004, 2007) and signed differential mapping (Radua and Mataix-Cols, 2009; Radua *et al.*, 2012). Roughly, these methods construct a statistic map as the convolution of the foci (precisely, this is a convolution of a Dirac delta function located at each focus with a given kernel) with three-dimensional spatial kernels, but do not exactly correspond to traditional kernel density estimation. In particular, these methods give special treatment to foci that appear close in one study, decreasing their influence relative to dispersed points. Areas of the map with large values suggest brain regions of consistent activation across studies. For statistical inference, the map is thresholded by reference to a Monte Carlo distribution under the null hypothesis of no consistent activation across studies. Kernel-based methods are not based on an explicit probabilistic model and hence often lack interpretability. Moreover, for some methods it is difficult to obtain standard errors and hence only p-values are reported for each voxel. Some of these approaches cannot accept study level covariates, and thus cannot conduct metaregression, and all are massively univariate in that they have no model of spatial dependence and can make only limited probabilistic statements about sets of voxels.

Recently, some model-based methods have been proposed to address the limitations of kernel-based methods, such as the Bayesian hierarchical independent Cox cluster process model of Kang *et al*. (2011), the Bayesian non-parametric binary regression model of Yue *et al*. (2012), the hierarchical Poisson–gamma random-field model of Kang *et al*. (2014) and the spatial Bayesian latent factor model of Montagna *et al*. (2018). However, most of these methods do not allow for metaregression. Further, current model-based approaches do not account for dependence that is induced when a single publication reports the results of multiple studies using the same cohort of participants. (In this work, we refer to 'study' as the result of one statistical map; typically a publication will report results from several maps.)

1.3. Contribution and outline

The contributions of this work are twofold. The first contribution is methodological. In particular, we propose a Bayesian spatial point process model, an extension of the log-Gaussian Cox process (LGCP) model (Møller *et al*., 1998) that can account for study-specific characteristics as explanatory variables, thus allowing for meta-regression. Compared with the model of Montagna *et al*. (2018), which is the only existing co-ordinate-based metaregression method, our model has two advantages. Firstly, it is less mathematically complex and therefore easier to communicate to practitioners and to elicit prior distributions for its parameters. Secondly, by introducing random-effect terms, our model can capture heterogeneities that cannot be captured by the covariates and also reduce biases that are caused by the assumption that studies in the meta-analysis are independent of each other.

The second contribution of this paper is to conduct a meta-analysis of WM functional MRI studies using the model proposed. Even though previous meta-analyses of WM studies exist (Wager and Smith, 2003; Owen *et al*., 2005; Rottschy *et al*., 2012), none of these studies uses some of the available model-based methods and hence the inferences that they provide are limited. Further, our analysis quantifies the effect of some important covariates and thus provides new insights regarding the function of WM.

The remainder of this paper is structured as follows. In Section 2 we present the data under investigation and state the questions that our meta-analysis wishes to answer. Motivated by the data in Section 2, we introduce our LGCP model in Section 3. The algorithm that is used for posterior inference is presented in Section 4. The results of the real data analysis can be found in Section 5. Finally, Section 6 summarizes our findings and sets some possible directions for future research.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from `https://osf.io/pgn2a/`.

2. Motivating data set

Our investigations are motivated by data from Rottschy *et al*. (2012). The data have been retrieved from 89 publications on WM but some of these publications conduct multiple studies (experiments). The average number of studies per publication is 1.76 (range 1–7). Overall, we include 157 studies in the meta-analysis and the total number of foci is 2107. As well as the foci, for each study we observe the stimulus type (where 102 studies used verbal stimuli and 55 studies used non-verbal stimuli), the sample size (mean 14.94; standard deviation SD 5.64) and the average age of the participants (mean 32 years; SD 10.99 years). Table 1 gives more descriptive statistics, whereas a graphical representation of the data can be found in Fig. 1. Note that the data set that we use is a subset of the data set of Rottschy *et al*. (2012); this is because of missing values for the covariate age.

Table 1. Data summaries

	Minimum	*Median*	*Mean*	*Maximum*
Data set composition				
Studies per publication	1	1	1.76	7
Foci per study	1	11	13.42	55
Participants per study	6	14	14.94	41
Mean participant age (years)	21.25	29.20	32.00	75.11
Verbal				
Foci per study	1	10	11.83	39
Participants per study	7	14	14.91	41
Mean participant age (years)	21.80	30.12	33.80	75.11
Non-verbal				
Foci per study	2	15	16.36	55
Participants per study	6	13	14.98	33
Mean participant age (years)	21.25	28.00	28.64	61

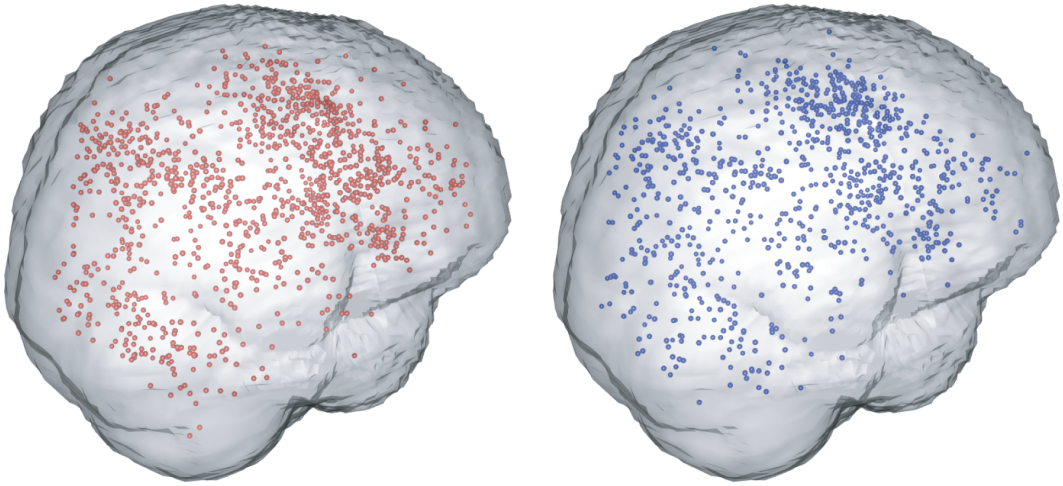

Fig. 1. Graphical representation of the meta-analysis data set: the data consist of 2107 foci from 157 studies on WM; of these, 1207 are obtained from studies using (a) verbal stimuli (●) whereas the remaining 900 are obtained from studies using (b) non-verbal stimuli (●) (the code used to generate this figure is courtesy of Jian Kang)

Our meta-analysis aims to address the following questions related to the function of WM.

(a) What are the regions of the brain that are consistently engaged by WM across studies?
(b) Do these regions differ depending on the type of stimulus presented to the participants?
(c) Is the organization of WM affected by age?
(d) Does sample size affect the total number of activations reported?

To ensure that the answers to these questions are not driven by influential publications conducting multiple studies, our investigations should account for such dependences.

3. A model for co-ordinate-based meta-analysis metaregression

To address the questions that were raised in Section 2, we propose a model for CBMA meta-regression. First, we set the notation. Suppose that there are a total I studies in the meta-analysis and that each study i comes with a point pattern $\mathbf{x}_i$, a set of foci $x_{ij} \in \mathcal{B} \subset \mathbb{R}^3$, where $\mathcal{B}$ is the support of the analysis, which is usually set from a standard atlas of the brain, and $j = 1, \ldots, n_i$, where n_i is the number of foci in a study. Additionally, suppose that for each point pattern there is a set of K study-specific characteristics, $\{z_{ik}\}_{k=1}^{K}$. Henceforth, we shall occasionally refer to these characteristics as covariates.

We assume that each point pattern $\mathbf{x}_i$ is the realization of a Cox point process $\mathbf{X}_i$ defined on $\mathcal{B}$, driven by a random intensity $\lambda_i(\cdot)$. We can then model the intensity function at each point $\xi \in \mathcal{B}$ as

$$\lambda_i(\xi) = \alpha_i \exp\left\{ \sum_{k=0}^{K^*} \beta_k(\xi) z_{ik} + \sum_{k=K^*+1}^{K} \beta_k z_{ik} \right\}, \tag{1}$$

where α_i is the random effect of study i, $\beta_k(\cdot)$ are the regression coefficients for the covariates that have a local effect ($k = 0, \ldots, K^*$), z_{ik} are covariate values where $k = 0$ is for the intercept ($z_{i0} = 1$) and β_k are the regression coefficients for the covariates that have a global (homogeneous) effect ($k = K^* + 1, \ldots, K$).

Equation (1) defines a spatial log-linear model over the brain. Foci are more likely to occur in regions of the brain with high intensity values whereas we expect almost no foci in regions as the intensity approaches 0. The exact rates are given by the properties of a Cox process. In particular, given $\lambda_i(\cdot)$, the expected number of foci in any bounded $B \subseteq \mathcal{B}$ is a Poisson random variable with mean $\int_B \lambda_i(\xi) \mathrm{d}\xi$ (Møller and Waagepetersen, 2004).

The inclusion of the random-effect terms is an important feature of our model. Firstly, by assuming that $\alpha_i = \alpha_j$ for studies i and j retrieved from the same publication, we relax the assumption of independence between their reported activations. This assumption is taken by all existing CBMA approaches but is unlikely to hold for studies from the same publication. For example, a multistudy publication will typically engage the same participants in all of its experiments. By using a common random effect for studies from the same publication, our model prevents publications with several studies from driving the estimates of the regression coefficients. Secondly, the random effects can allow for additional variability in the total number of foci that cannot be captured by the Poisson log-linear model. In a recent study, Samartsidis *et al.* (2017) found that CBMA data do show such overdispersion and thus inclusion of the random-effect terms can potentially improve the fit to the data.

Separation of the covariates into those with a localized and those with a global effect should be done with caution. If we are interested in investigating whether the effect of a covariate varies from one region of the brain to another, such as age in our application, a spatially varying regression coefficient is needed. However, the total number of parameters that are associated with a spatially varying effect is large and therefore assigning a spatially varying coefficient to a covariate with a global effect may substantially increase the uncertainty that is associated with the other model parameters. To determine whether a spatially varying coefficient for a covariate is required, we can fit two models: one that assumes that the covariate has a global effect and one that assumes a local effect. If the more complex model improves the fit to the data substantially (as determined by a goodness-of-fit measure, e.g. posterior predictive checks (Gelman *et al.*, 1996; Leininger and Gelfand, 2017)), then it should be preferred for inference instead of the simple model. Sometimes, it is plausible to assume a global effect solely based on

prior expectation. For instance, a covariate for multiple-testing correction can be assumed to have a global effect; for studies not applying any corrections, we expect false positive results to appear uniformly across the brain.

A Bayesian model is defined with prior distributions on model parameters, which here include the functional parameters $\beta_k(\cdot)$ $(k = 0, \ldots, K^*)$ and scalar parameters β_k $(k = K^* + 1, \ldots, K)$. A natural way to proceed is to assume that $\beta_k(\cdot)$ are realizations of Gaussian processes and that the β_k have normal distributions. That way, when $\alpha_i = 1$, the right-hand side of equation (1) is also a Gaussian process, and each point process is an LGCP (Møller *et al.*, 1998). The LGCP is a flexible model for spatial point data that can account for aggregation (Møller *et al.*, 1998; Møller and Waagepetersen, 2007) or even repulsion between points (Illian *et al.*, 2012a) and has therefore found applications in several fields such as disease mapping (Benes *et al.*, 2002; Liang *et al.*, 2009) and ecology (Møller and Waagepetersen, 2003; Illian *et al.*, 2012b).

By the definition of a Cox process, $\mathbf{X}_i$ is a Poisson point process on $\mathcal{B}$ conditional on $\lambda_i(\cdot)$ (Møller and Waagepetersen, 2004). The density (Radon–Nikodym derivative) of this point process with respect to the unit rate Poisson process is

$$\pi(\mathbf{x}_i|\lambda_i) = \exp\left\{|\mathcal{B}| - \int_{\mathcal{B}} \lambda_i(\xi)\mathrm{d}\xi\right\} \prod_{x_{ij}\in\mathbf{x}_i} \lambda_i(x_{ij}), \tag{2}$$

for $i = 1, \ldots, I$, with $|\mathcal{B}|$ denoting the volume of the brain. We can view $\pi(\mathbf{x}_i|\lambda_i)$ as the density of the sampling distribution of the data. If we further assume independent studies, then the posterior distribution of the model parameters conditional on the foci is given, up to a normalizing constant, by

$$\pi(\{\alpha_i\}_{i=1}^I, \{\beta_k(\cdot)\}_{k=0}^{K^*}, \{\beta_k\}_{k=K^*+1}^K | \{\mathbf{x}_i\}_{i=1}^I) \propto \prod_{i=1}^I \pi(\mathbf{x}_i|\lambda_i) \times \prod_{i=1}^I \pi(\alpha_i) \prod_{k=1}^{K^*} \pi\{\beta_k(\cdot)\} \prod_{k=K^*+1}^K \pi(\beta_k), \tag{3}$$

where $\pi(\alpha_i)$, $\pi\{\beta_k(\cdot)\}$ and $\pi(\beta_k)$ are the priors on the random-effects and functional and scalar parameters respectively, which we discuss as the priors below in Section 3.2.

3.1. Choice of correlation function

We shall assume an isotropic, Gaussian correlation structure, i.e. for points $\xi, \xi' \in \mathcal{B}$ we have

$$\mathrm{corr}\{\beta_k(\xi), \beta_k(\xi')\} = \exp(-\rho_k \|\xi - \xi'\|^{\delta_k}), \tag{4}$$

where $\rho_k > 0$ are the correlation decay parameters and $\delta_k = 2$ for all $k = 1, \ldots, K^*$. For numerical stability with the discrete Fourier transform (see Section 4) we set $\delta = 1.9$ in our implementations. The same correlation structure was used by Møller *et al.* (1998) and Møller and Waagepetersen (2003) in the context of LGCPs.

A Gaussian correlation function is used instead of alternative correlation structures (see for example Rasmussen and Williams (2005)) because it allows us to calculate the gradient of the posterior with respect to the correlation parameters ρ_k, which we use to design an efficient algorithm for posterior simulations (see Section 4 for details). Further, in exploratory work using other correlation structures, our neuroscientist colleagues preferred the appearance of results from Gaussian correlation, perhaps because of the pervasive use of Gaussian kernel smoothing in functional MRI. Finally, it is well known that estimating the correlation parameters for

more flexible correlation structures can be extremely challenging in practice; see for example discussions by Zhang (2004) and Diggle *et al.* (2013) for the Matérn correlation function.

3.2. Posterior approximation

Calculation of the posterior in equation (3) requires the evaluation of the infinite dimensional Gaussian processes $\beta_k(\cdot), k=0, \ldots, K^*$, which we approximate with a finite dimensional distribution. Following Møller *et al.* (1998) and Benes *et al.* (2002), we consider the discretization of the three-dimensional volume with a regular rectangular grid $W \supset \mathcal{B}$. We use V cubic cells (i.e. voxels) in W with volume $A=a^3$, where a is the length of the side. In neuroimaging, analysis with 2 mm^3 cubic voxels is typical, leading to a box shaper grid of about 1 million voxels, of which about 200000 are in the brain or cerebellum. For simplicity, we consider both grey matter and white matter voxels in our implementations. Voxels are indexed $v=1, \ldots, V$, and the co-ordinate of v is the location of the centre $\boldsymbol{\nu}_v \in \mathbb{R}^3$.

For any $k=0, \ldots, K^*$, the Gaussian process $\beta_k(\cdot)$ can be now approximated with a step function which is constant within each voxel v and equal to the value of $\beta_k(\cdot)$ at the location of the centre, i.e. $\beta_k(\boldsymbol{\nu}_v)$. Waagepetersen (2004) showed that the accuracy of this approximation improves as $a \to 0$. By definition, $\boldsymbol{\beta}_k = (\beta_k(\boldsymbol{\nu}_1), \ldots, \beta_k(\boldsymbol{\nu}_V))$ are multivariate Gaussian vectors. We parameterize $\boldsymbol{\beta}_k$ as

$$\boldsymbol{\beta}_k = \mu_k \mathbf{1}_V + \sigma_k \mathbf{R}_k^{1/2} \boldsymbol{\gamma}_k, \tag{5}$$

where μ_k are the overall (scalar) means, $\mathbf{1}_V$ is a V-vector of 1s, σ_k are the marginal standard deviations, $\mathbf{R}_k$ are the $V \times V$ correlation matrices with elements $(\mathbf{R}_k)_{ij} = \exp(-\rho_k \|\boldsymbol{\nu}_i, \boldsymbol{\nu}_j\|^2)$ and $\boldsymbol{\gamma}_k$ are the *a priori* $\mathcal{N}_V(\mathbf{0}, \mathbf{I}_V)$ vectors, $k=0, \ldots, K^*$. The same parameterization was used by Møller *et al.* (1998) and Christensen and Waagepetersen (2002) and was advocated by Christensen *et al.* (2006) because it allows for computationally efficient posterior simulations.

Priors for the V-vectors $\boldsymbol{\gamma}_k$ are induced by the parameterization of equation (5). The priors for the remaining model parameters are set as follows. We assign weakly informative $\mathcal{N}(0, 10^8)$ priors to the scalar parameters μ_k, σ_k and β_k. Further, we assume that $\rho_k \sim \text{Uni}(3.5 \times 10^{-3}, 0.1)$, which we found corresponded to smoothness ranges found in single-study functional MRI statistic maps. Finally, to ensure identifiability, we *a priori* let $\alpha_i \sim \mathcal{G}(\kappa, \kappa)$. In our analyses, we set $\kappa = 10$ since we expect 90% of the multiplicative random effects to be within the interval [0.5,1.5].

Once the latent Gaussian processes have been approximated, we can also approximate λ_i with a step function as before. The intensities at the centre of each voxel are given by

$$\boldsymbol{\lambda}_i = \alpha_i \exp\left\{ \sum_{k=0}^{K^*} (\mu_k \mathbf{1}_V + \sigma_k \mathbf{R}_k^{1/2} \boldsymbol{\gamma}_k) z_{ik} + \sum_{k=K^*+1}^{K} \beta_k z_{ik} \mathbf{1}_V \right\}, \tag{6}$$

where $\boldsymbol{\lambda}_i$ is the V-vector: the discretized intensity. We shall write $\lambda_{iv} = (\boldsymbol{\lambda}_i)_v$ for the v-element of study i's intensity. The approximated posterior is

$$\pi(\boldsymbol{\theta} | \{\mathbf{x}_i\}_{i=1}^I) \propto \prod_{i=1}^{I} \left\{ \exp\left(-\sum_v A_v \lambda_{iv} \right) \prod_{j=1}^{n_i} \lambda_{iv(x_{ij})} \right\} \pi(\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta} = \{\{\alpha_i\}_{i=1}^I, \{\mu_k\}_{k=1}^{K^*}, \{\sigma_k\}_{k=1}^{K^*}, \{\rho_k\}_{k=1}^{K^*}, \{\boldsymbol{\gamma}_k\}_{k=1}^{K^*}, \{\beta_k\}_{k=K^*+1}^{K}\}$, A_v takes the value A when $\boldsymbol{\nu}_v \in \mathcal{B}$ and 0 otherwise, $v(x_{ij})$ is the index of the voxel containing x_{ij} and $\pi(\boldsymbol{\theta})$ is the joint prior distribution of the parameters. The posterior distribution in equation (7) is still an-

alytically intractable because of the presence of an unknown normalizing constant and thus we need to resort to Monte Carlo simulation or approximation techniques to obtain samples from it. The method that we use is described in Section 4.

4. Sampling algorithm details

Bayesian methodology for inference on LGCPs can be broadly divided into two main categories: simulation-based approximations of the posterior, such as Markov chain Monte Carlo (MCMC) sampling (Møller *et al*., 1998) and elliptical slice sampling (Murray *et al*., 2010), and deterministic approximations to the posterior such as integrated nested Laplace approximations (Illian *et al*., 2012a; Simpson *et al*., 2016) and variational Bayes methods (Jaakkola and Jordan, 2000). In a recent study, Taylor and Diggle (2014) compared the Metropolis-adjusted Langevin algorithm (MALA) with the integrated nested Laplace approximation algorithm and found that both methods gave similar results. In our application, we choose to use simulation-based methods because application on our three-dimensional problem is more straightforward.

We propose a hybrid MCMC algorithm to sample from the posterior (7), where parameters are updated in two blocks. The first block includes the random-effect terms $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^{I}$, whereas the second block includes the remaining model parameters $\boldsymbol{\theta}^* = \boldsymbol{\theta} \setminus \boldsymbol{\alpha}$. The gamma prior is conjugate for the elements of $\boldsymbol{\alpha}$; hence, they are simulated from their full conditional distributions given the remaining model parameters; see section 1.5 in the Web-based supplementary materials for details. Even though it is possible, we choose not to update $\boldsymbol{\alpha}$ jointly with $\boldsymbol{\theta}^*$ because that would increase the computation time of our algorithm.

Sampling from the full conditional of $\boldsymbol{\theta}^*$ given $\boldsymbol{\alpha}$ is challenging because of its dimensionality. Girolami and Calderhead (2011) showed that, of all possible strategies, their Riemann manifold Hamiltonian Monte Carlo (HMC) sampler is the computationally most efficient for LGCPs in a two-dimensional setting. Unfortunately, application in this problem (a three-dimensional setting) is prohibitive as it would require the inversion of a huge $V \times V$ tensor matrix. Alternatives to Riemann manifold HMC sampling include the MALA and the standard HMC (Duane *et al*., 1987; Neal, 2011) algorithm. We choose to use HMC sampling because Girolami and Calderhead (2011) found that it is more efficient compared with the MALA in a two-dimensional setting. This finding was confirmed in our preliminary two-dimensional simulation studies with synthetic CBMA data, where the HMC outperformed the MALA in terms of computational efficiency (mixing–running time trade-off).

HMC sampling initially appeared in the physics literature by Duane *et al*. (1987) under the name *hybrid Monte Carlo* sampling, and later emerged in the statistics literature by Neal (2011). HMC sampling emulates the evolution of a particle system which is characterized by its position $\mathbf{q}$ and momentum $\mathbf{p}$ over time. In our case, $\mathbf{q}$ will be the parameter vector of interest $\boldsymbol{\theta}^*$, and $\mathbf{p}$ will be introduced artificially from an $\mathcal{N}_d(0, \mathbf{M})$ distribution, with d being the dimensionality of the problem and $\mathbf{M}$ the mass matrix. The dynamics of the system are described by a set of differential equations, known as Hamilton's equations.

An HMC algorithm alternates between moves for the position vector $\boldsymbol{\theta}^*$ and the momentum vector $\mathbf{p}$ on the basis of Hamilton's equations. If the solutions of the equations can be found analytically then moves will be deterministic; if not, numerical integration is required and an acceptance–rejection step must be performed to account for integration error. Integration is done in fictitious time ϵL, where ϵ is the *step size* and L is the *number of steps*. Typically the *leapfrog integrator* is employed, which for $L = 1$ and starting at time t is performed as (Neal, 2011)

$$\left.\begin{aligned}
\mathbf{p}\left(t+\frac{\epsilon}{2}\right) &= \mathbf{p}(t)+\frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}^*}\log[\pi\{\boldsymbol{\theta}^*(t)\mid\{\mathbf{x}_i\}_{i=1}^I,\boldsymbol{\alpha}\}],\\
\boldsymbol{\theta}^*(t+\epsilon) &= \boldsymbol{\theta}^*(t)+\epsilon\mathbf{M}^{-1}\mathbf{p}\left(t+\frac{\epsilon}{2}\right),\\
\mathbf{p}(t+\epsilon) &= \mathbf{p}\left(t+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}^*}\log[\pi\{\boldsymbol{\theta}^*(t+\epsilon)\mid\{\mathbf{x}_i\}_{i=1}^I,\boldsymbol{\alpha}\}].
\end{aligned}\right\} \tag{8}$$

Overall, if the method is applied correctly, it will produce samples from the desired full conditional distribution $\pi(\boldsymbol{\theta}^*\mid\{\mathbf{x}_i\}_{i=1}^I,\boldsymbol{\alpha})$. Gradient expressions for the elements of $\boldsymbol{\theta}^*$, including correlation parameters ρ_k, can be found in the Web-based supplementary materials' section 1. Since it is well known that grouping of variables can lead to samplers with faster convergence properties (Park and van Dyk, 2009), we choose to update all elements of $\boldsymbol{\theta}^*$ jointly by using the HMC algorithm. The solutions to Hamilton's equations are not available analytically so we need to use the leapfrog integrator and to include an accept–reject step at the end of it.

Our sampler requires the specification of a step size ϵ and a total number of leapfrog steps L for the HMC step. Hoffman and Gelman (2014) showed how tuning can be achieved automatically but when we applied this method to our problem the running time was increased substantially. Therefore we use an alternative approach to tune these parameters. The step size is automatically adjusted during the burn-in phase of the HMC algorithm to give an overall acceptance rate that is close to the 65% that was suggested by Neal (2011). In particular, if ϵ_t is the step size at iteration t and q_{t_1} is the acceptance rate over the past t_1 iterations, then every t_2 iterations we calculate the new step size ϵ_t' as

$$\epsilon_t' = \begin{cases} 0.9\epsilon_t & q_{t_1} < 0.60,\\ \epsilon_t & 0.60 \leqslant q_{t_1} \leqslant 0.70,\\ 1.1\epsilon_t & q_{t_1} > 0.70. \end{cases} \tag{9}$$

Specifically we use $t_1 = 100$ and $t_2 = 10$. A similar approach was employed by Marshall and Roberts (2012) for the MALA. The latter (number of leapfrog steps) is always fixed at $L = 50$. We took this approach because we found that, for our LGCP application, the mixing properties of the algorithm scale linearly with L but also with the total number of HMC iterations. Hence one can use a relatively large L and few iterations or relatively smaller L and more iterations; the total computation time stays relatively constant.

The last tuning parameter in the HMC algorithm is the variance–covariance matrix of the zero-mean normal momentum parameters $\mathbf{M}$. To our knowledge, there is only limited off-the-shelf methodology on how to adjust $\mathbf{M}$. As a starting point we set $\mathbf{M} = \mathbf{I}$. Neal (1996) suggested that if an estimate of the posterior variance $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}^*}$ is available then good practice is to set $\mathbf{M} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}^*}^{-1}$. In principle, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}^*}$ can be estimated during the burn-in phase of the HMC algorithm but in practice this is not possible because of the dimensionality of the problem. In our simulations, we found that the mean posterior variance of the elements of the $\boldsymbol{\gamma}_k$ was higher compared with the scalar parameters, followed by β_k or σ_k and then ρ_k. Especially for the ρ_k the scale is typically much smaller compared with the other parameters in our applications and so we use $100\rho_k$ instead of ρ_k. After the reparameterization we found that setting the mass for parameters of $\boldsymbol{\gamma}_k$, β_k, σ_k and ρ_k equal to 1, 9, 16 and 25 respectively worked well in most of our implementations on simulated and real data. However, users might need to adjust these parameters if mixing of the chains is slow. For example, estimates of the posterior variance of the scalar parameters can be obtained on the basis of preliminary runs of the algorithm for a few iterations. In section 2 of the Web-based supplementary materials, we perform a series of simulations studies which demonstrate

that the proposed HMC algorithm can efficiently sample from the posterior distribution of the high dimensional parameter vector $\boldsymbol{\theta}^*$.

The most computationally demanding part of the algorithm is calculation of the large matrix vector products $\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k$ appearing in the intensity functions of equation (6). Luckily, an elegant solution to this problem was given by Møller *et al*. (1998) based on *circulant embedding* that was first proposed by Dietrich and Newsam (1993) and Wood and Chan (1994). The key to the approach is the linear algebra result that a circulant matrix has the discrete Fourier basis as its eigenvectors. $\mathbf{R}_k$ is not circulant but is block Toeplitz and can be embedded in a $2V \times 2V$ matrix that is circulant. Thus the matrix square root, inversion and multiplication can be accelerated by using (the highly efficient) discrete Fourier transform of the embedded matrix and manipulating Fourier coefficients, followed by the inverse discrete Fourier transform and extracting the appropriate submatrix or subvector. See Rue and Held (2005), section 2.6.2, for more details.

We close this section by stressing that, despite the massive dimensionality of the parameter vector, the problem has a very high degree of parallelization. Intensities can be evaluated in blocks of thousands of voxels simultaneously, making the algorithm suitable for implementation in a *graphics processing unit*. The most computationally intensive part of our model, namely operations with discrete Fourier transforms, is also amenable to parallelization and there are libraries such as Nvidia's cuFFT library that are designed for this specific task. Overall, we believe that implementation of the LGCP model described above will soon become a routine task for any moderately powerful graphics processing unit device.

5. Analysis of the working memory data set

5.1. Model, algorithm details and convergence diagnostics

For $i = 1, \ldots, 157$ we fit the model

$$\boldsymbol{\lambda}_i = \alpha_i \exp\left(\beta_0 d_{i0} + \beta_1 d_{i1} + \beta_2 \text{age}_i + \beta_3 \frac{1}{\sqrt{n_i}} \mathbf{1}_V \right), \tag{10}$$

where d_{i0} and d_{i1} are indicator variables of verbal and non-verbal stimuli respectively and n_i is the total number of participants in study i. Continuous parameters were standardized before implementation.

We ran the MCMC algorithm that was described in Section 4 for 22000 iterations, discarding the first 7000 as a burn-in. The algorithm ran for approximately 30 h on an Nvidia Tesla K20c graphics processing unit card. We then applied a thinning factor of 15 to the chains and therefore end up with 1000 draws from the posterior distribution of the model parameters. The total number of leapfrog steps is set to $L = 50$ and the step size is initialized at $\epsilon = 0.00001$. We use a diagonal mass matrix with units specified in Section 4. A preliminary run of the algorithm revealed that the posterior variance of the scalar parameters ρ_2 and σ_2 of β_2 was higher compared with the corresponding parameters of β_0 and β_1. Therefore, to improve mixing of the algorithm, we set the mass parameters to 1 and 4 for ρ_2 and σ_2 respectively.

Convergence of the MCMC chain is assessed visually by inspection of posterior trace plots for the model parameters. We run a total of two MCMC chains to examine whether they all converge to the same values. Posterior trace plots are shown in the Web-based supplementary materials' section 3. Because of the large number of parameters we mainly focus on the scalar parameters of the model and some summary statistics; see section 3 in the Web-based supplementary materials for more details. Results indicate that our chains have converged to their stationary distribution. This is verified by the fact that posterior values from the two runs overlap one another for all the quantities that we examine.

5.2. Results

Fig. 2 shows the mean posterior of $\boldsymbol{\lambda}$, the average intensity of a WM study, where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{\mathrm{v}} + \boldsymbol{\lambda}_{\mathrm{nv}})/2$, $\boldsymbol{\lambda}_{\mathrm{v}}$ is the intensity for verbal WM studies and $\boldsymbol{\lambda}_{\mathrm{nv}}$ is for non-verbal WM studies (the mean age and number of participants are set equal to the average values in our data set). We can see that WM engages several regions of the brain. The regions that are mostly activated are the frontal orbital cortex (axial slice $z = -10$, left), the insular cortex ($z = -10$, right, and $z = -2$, left and right), the precentral gyrus ($z = 30$, left), Broca's areas ($z = 22$ and $z = 30$, bilateral), the angular gyrus ($z = 46$, left), the superior parietal lobule ($z = 46$, right) and the paracingulate gyrus ($z = 46$, middle).

Our results are qualitatively similar to results that were obtained by Rottschy *et al.* (2012) who used the activation likelihood estimation method. However, our model-based approach enables us to derive several quantities of interest along with credible intervals (CIs) that cannot be obtained by any of the kernel-based methods. For example, we may calculate the probability of observing at least one focus in a set of voxels, e.g. a region of interest (ROI) or the entire brain. Table 2 summarizes the posterior distribution of $\mathbb{P}\{N_{\mathbf{X}}(B) \geqslant 1\}$, the probability of observing at least one focus in B, for several ROIs B. A full brain analysis can be found in the Web-based supplementary materials' section 4. The division of the brain into ROIs is done according to the Harvard–Oxford atlas (Desikan *et al.*, 2006).

We use posterior intensities $\boldsymbol{\lambda}_{\mathrm{v}}$ and $\boldsymbol{\lambda}_{\mathrm{nv}}$ to compare activation between the two types of studies in our sample, namely studies using verbal and studies using non-verbal stimuli. We

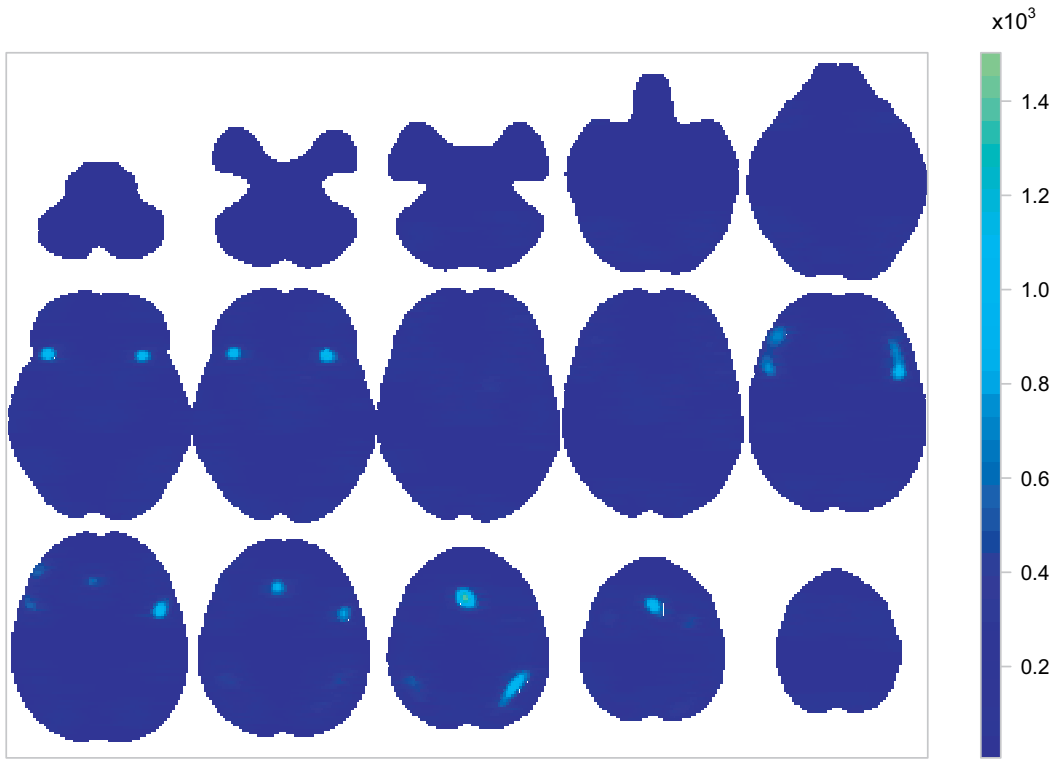

Fig. 2. Voxelwise mean posterior of λ, the average intensity of a WM study: the top row shows (from left to right) axial slices $z = -50, -42, -34, -26, -18$; the middle row shows axial slices $z = -10, -2, 6, 14, 22$; the bottom row shows axial slices $z = 30, 38, 46, 54, 62$

Table 2 Posterior percentage probabilities of observing at least one focus for several ROIs†

ROI	*Mean (%)*	*95% CI (%)*	*Results (%) for verbal stimulus*	*Results (%) for non-verbal stimulus*
Frontal orbital cortex	36.94	[27.27,43.06]	37.26	36.48
Insular cortex	33.39	[26.68,39.36]	32.79	33.86
Precentral gyrus	68.47	[59.96,73.72]	64.10	72.09
Inferior frontal gyrus	39.88	[31.06,45.96]	43.66	35.69
Angular gyrus	21.69	[14.39,26.34]	24.30	18.91
Superior parietal lobule	36.16	[26.16,42.31]	38.81	33.24
Paracingulate gyrus	46.22	[35.94,52.89]	42.91	49.14

†All quantities have been calculated based on 1000 MCMC samples.

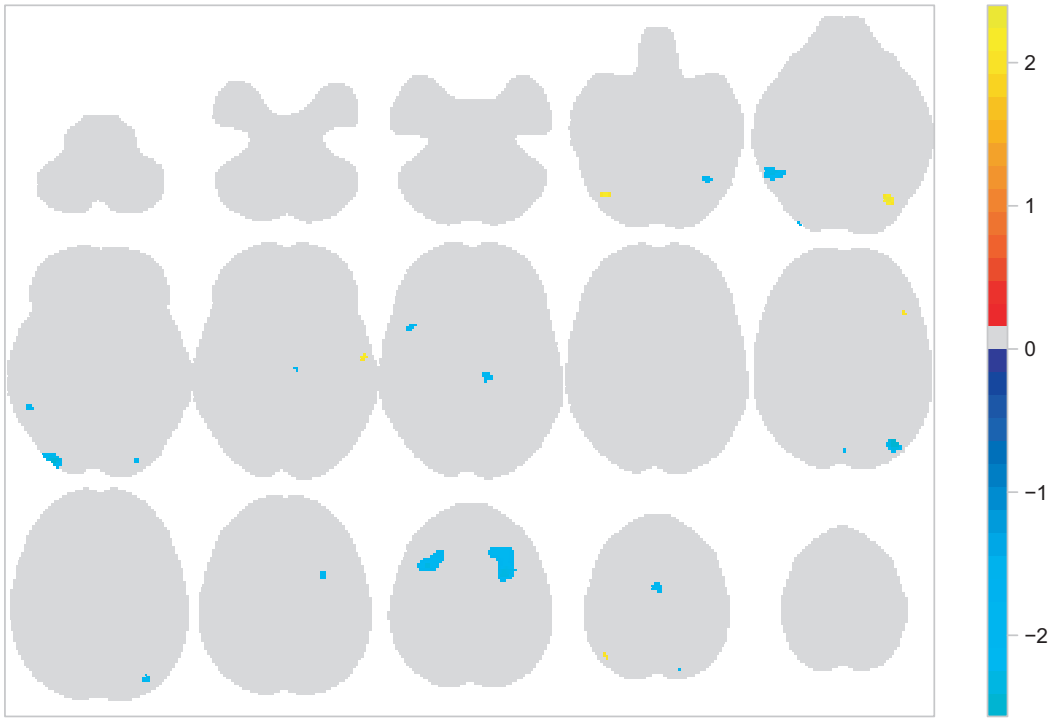

Fig. 3. Voxelwise mean standardized posterior difference between β_1 and β_2, the intensities of studies using verbal and non-verbal stimuli respectively: the top row shows (from left to right) axial slices $z = -50, -42, -34, -26, -18$; the middle row shows axial slices $z = -10, -2, 6, 14, 22$; the bottom row shows axial slices $z = 30, 38, 46, 54, 62$; voxels for which the mean posterior λ is low (below the 75% quantile over the brain) or the absolute mean standardized posterior difference is less than 2 have been set to 0

start with an ROI analysis. In particular, for each type and ROI we calculate the probability of at least one focus observed as explained above. These are shown in Table 2 for a few ROIs, whereas a full brain analysis of the two types can be found in section 4 in the Web-based supplementary materials. We see that, even though the two types show similar patterns of activation, there are several ROIs where the probabilities of at least one focus have CIs with little

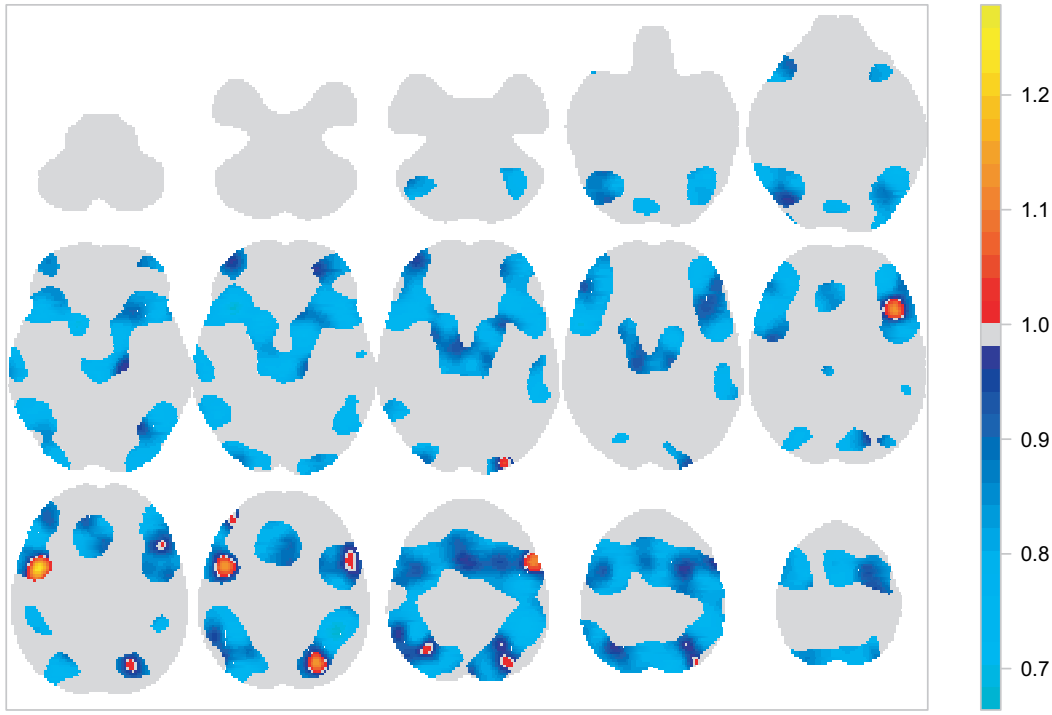

Fig. 4. Mean posterior of $\exp(\beta_2)$, the multiplicative age effect on the intensity of both verbal and non-verbal studies: the top row shows (from left to right) axial slices $z = -50, -42, -34, -26, -18$; the middle row shows axial slices $z = -10, -2, 6, 14, 22$; the bottom row shows axial slices $z = 30, 38, 46, 54, 62$; voxels for which the mean posterior λ is low (below the 75% quantile over the brain) have been set to 1

overlap. The main differences are found in the superior frontal gyrus, the middle frontal gyrus, the lateral occipital cortex, superior division and the inferior frontal gyrus, *pars opercularis*. A voxel-by-voxel comparison is also feasible. To answer this, we use the mean standardized posterior difference $(\beta_{0v} - \beta_{1v})/\text{sd}(\beta_{0v} - \beta_{1v})$. This is shown in Fig. 3. Large positive values indicate regions that are activated by verbal stimuli more than non-verbal stimuli. Such regions appear in the occipital fusiform gyrus ($z = -18$, right). On the basis of the mean standardized posterior difference, regions that are mostly activated in studies using non-verbal stimuli are located in the middle frontal gyrus ($z = 46$).

Our results provide evidence that age has an important effect on the function of WM. The point estimate for the overall age effect μ_2 is -0.22 (95% CI $[-0.337, -0.120]$) thus suggesting that we expect a decrease of 20% in the total number of reported activations per study, each time the average age of the participants increases by 10.99 years. Localized age effects can be identified through the posterior distribution of $\exp(\beta_2)$, the mean of which is shown in Fig. 4. The map represents the multiplicative effect that an increase in the average participant age by 10.99 years has on the intensity of both verbal and non-verbal studies. Large negative age effects can be found near the left putamen ($z = -2$ and $z = -10$, middle), the insular cortex ($z = -2$, left) and near the superior parietal lobule ($z = 38$ and $z = 46$, right). A positive age effect is found near the precentral gyrus ($z = 30$, left). However, because of the limited number of studies, the posterior variance of these estimates is large in some regions of the brain; see Fig. 9 in the Web-based supplementary materials' section 3.

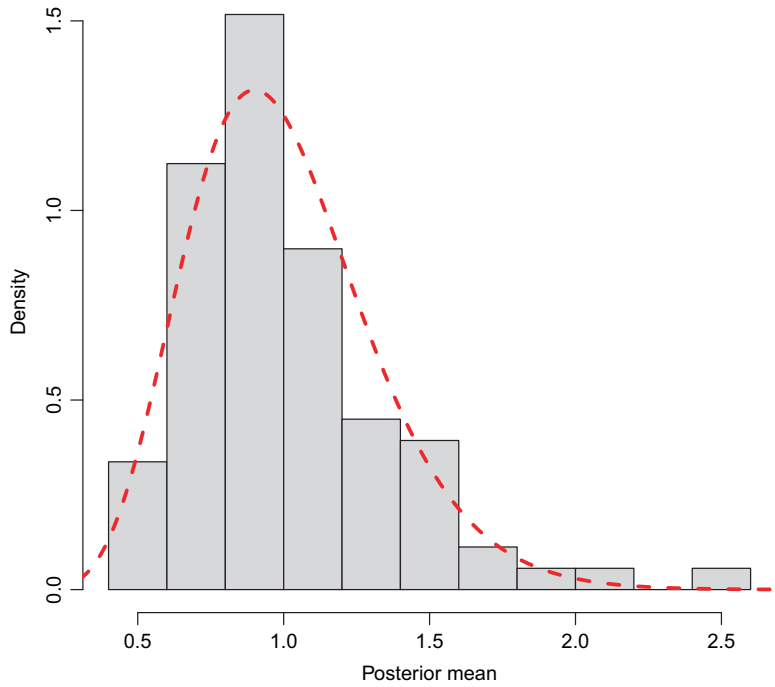

Fig. 5. Histogram of the mean posterior random-effect terms α_i (– – –, density of the gamma prior): we plot only the 89 unique random effects, one for each publication considered in the meta-analysis; means are based on a sample of 1000 MCMC draws from the posterior

The 95% CI for the sample size covariate is $[-0.088, 0.064]$, thus indicating that there is no significant effect on the total number of reported activations. The result is counterintuitive as we would expect that studies with few participants would be underpowered and thus detect fewer activations. Thus, further investigation is required.

Fig. 5 shows the mean posterior of the 89 unique random-effect terms α_i: one for each publication considered. We see that, although most of the mass is near 1, there are publications whose mean posterior random effect is different from 1, thus suggesting that observed variability of the foci counts is larger compared with what can be explained by the Poisson log-linear model. The importance of allowing for this additional variability can be seen by comparing the proposed random-effects model with the standard LGCP model, which we also fitted to the data. We use posterior predictive checks (Gelman *et al.*, 1996) to assess how well the two models fit the data. For each study and MCMC draw we simulate, from the posterior predictive distribution of $N_{\mathbf{X}_i}(\mathcal{B})$, the total number of foci, given the covariates. On the basis of these draws, we calculate the 95% predictive intervals of $N_{\mathbf{X}_i}(\mathcal{B})$ and check whether they contain the observed values. For our model, the coverage of the intervals is 90% compared with 66% obtained by using the standard LGCP model, which implies that our model provides a better fit with the data compared with the standard LGCP. A comparison of the predictive intervals that takes into account the length of these intervals can be based on the mean interval score (Gneiting and Raftery, 2007). This is 22.45 and 76.93 for the random-effects and standard LGCP models respectively, thus suggesting that the inclusion of α_i leads to improved prediction of the study counts.

Some of the estimated effects are affected by inclusion of the random-effect terms. For in-

stance, the expected number of foci for verbal studies is estimated as 12.80 (95% CI [11.57,14.14]) by the random-effects LGCP as opposed to 11.67 (95% CI [10.97,12.36]) by the fixed effects LGCP model. One possible explanation for this is that our model is assigning a low random effect to publications that systematically report only a few foci. Such a behaviour is desired since, for example, this under-reporting could be solely due to researchers' preference. Further, the random-effects model provides CIs that fully account for the uncertainty in the regression coefficients. For example, the 95% CI for the overall age effect μ_2 provided by the fixed effects LGCP is $[-0.309, -0.151]$, shorter than the CI that is provided by our model.

6. Discussion

In this work, we have presented a new CBMA model: an extension of the LGCP model. To our knowledge, this is the first application of the random-effects LGCP with covariates in a three-dimensional problem with multiple realizations. The model has an appealing interpretation being a spatial generalized linear model and several interesting inferences can be obtained based on the properties of the spatial Poisson process that cannot be obtained with the commonly used kernel-based approaches. An advantage of our model compared with most of the existing methods is the inclusion of covariates in the analysis, thus allowing for metaregression. Finally, a novel feature that was introduced in our work is the inclusion of random-effect terms which can account for additional heterogeneity in the total number of activations, compared with the standard Poisson model.

Application of our model on a meta-analysis of WM studies has given valuable insights regarding the data. Although our maps for the overall pattern of WM activations (Fig. 2) and the differential effect of verbal *versus* non-verbal WM tasks (Fig. 3) reflect previous findings by Rottschy *et al.* (2012), our fully Bayesian approach enabled us to make direct inference on the probability of any foci and expected number of foci. Our model found no regions with evidence of different rates of foci between verbal and non-verbal WM tasks (Web-based supplementary materials' section 4, Table 3). Importantly, our model enables a metaregression, and we examined the effect of age and found no strong effects but generally negative effects of age on the number of foci.

There are a few limitations to our work. Firstly, even though we found that the MCMC algorithm proposed performed well in most of the applications that were considered, we believe that there is room for further improvement. For example, one can consider adaptive schemes to adjust the mass matrix $\mathbf{M}$ of the HMC algorithm automatically, which we found is crucial for the mixing properties of the algorithm. Secondly, we are currently not considering the problem of learning the hyperparameter κ that controls the posterior variability of the random-effect terms, but rather we make use of our prior expectations to tune it. However, since we found that results are sensitive to the specification of κ, it is plausible to consider estimating it along with the remaining model parameters.

Our work can be extended in several ways. One possible direction for future research is to perform a comparison of existing methodologies that can be used for posterior inference with the proposed LGCP model in the context of CBMA. However, given the computation time that is required to apply these methods to a three-dimensional problem, such a comparison might be too long. Another potential future direction is to study the conditions, such as sample size or minimum number of foci, under which it is possible to estimate several global or spatially varying effects by using the LGCP. Such work can be of importance for practical implementations since it will provide some guidance regarding the complexity of metaregression models that can be fitted to a given data set.

Another open problem is how to use some additional information about the foci such as p-values or T-scores. These values can be attached as marks to the existing point patterns. Such an approach can enrich the inferences that are obtained from a CBMA by characterizing the magnitude of activation in each region as opposed to the localization of activations, which is the question that current methods address. Finally, it is worth considering a zero-truncated LGCP model. The reason is that several CBMAs use data from databases such as BrainMap (Laird *et al*., 2005), where only studies with at least one focus are registered. For such applications, a model that does not account for the zero-truncation can provide biased intensity estimates, especially when the expected number of foci per study is low. Currently, very few of the existing approaches propose adjustments for this potential problem.

Acknowledgements

This work was largely completed while PS and TEN were at the University of Warwick, Department of Statistics. PS, TDJ and TEN were supported by National Institutes of Health grant 5-R01-NS-075066; TEN was supported by Wellcome Trust fellowship 100309/Z/12/Z and National Institutes of Health grant R01 2R01EB015611-04. The work that is presented in this paper represents the views of the authors and not necessarily those of the National Institutes of Health or the Wellcome Trust Foundation.

References

Benes, V., Bodlák, K., Møller, J. and Waagepetersen, R. P. (2002) Bayesian analysis of log Gaussian Cox processes for disease mapping. *Technical Report*. Department of Mathematical Sciences, Aalborg University, Aalborg.

Christensen, O. F., Roberts, G. O. and Sköld, M. (2006) Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *J. Computnl Graph. Statist.*, **15**, 1–17.

Christensen, O. F. and Waagepetersen, R. P. (2002) Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics*, **58**, 280–286.

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

Dietrich, C. R. and Newsam, G. N. (1993) A fast and exact method for multidimensional Gaussian stochastic simulations. *Wat. Resour. Res.*, **29**, 2861–2869.

Diggle, P. J., Moraga, P., Rowlingson, B. and Taylor, B. M. (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Statist. Sci.*, **28**, 542–563.

Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987) Hybrid Monte Carlo. *Phys. Lett.* B, **195**, 216–222.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F. and Fox, P. T. (2012) Activation likelihood estimation meta-analysis revisited. *NeuroImage*, **59**, 2349–2361.

Gelman, A., Meng, X.-L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sin.*, **6**, 733–807.

Girolami, M. and Calderhead, B. (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B, **73**, 123–214.

Gneiting, T. and Raftery, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *J. Am. Statist. Ass.*, **102**, 359–378.

Greenland, S. (1994) Invited commentary: a critical look at some popular metaanalytic methods. *Am. J. Epidem.*, **140**, 290–296.

Hartung, J., Knapp, G. and Sinha, B. K. (2008) *Statistical Meta-analysis with Applications*. Hoboken: Wiley.

Hoffman, M. and Gelman, A. (2014) The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, **15**, 1593–1623.

Illian, J. B., Sørbye, S. H. and Rue, H. (2012a) A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Statist.*, **6**, 1499–1530.

Illian, J. B., Sørbye, S. H., Rue, H. and Hendrichsen, D. K. (2012b) Using INLA to fit a complex point process model with temporally varying effects—a case study. *J. Environ. Statist.*, **3**, 1–25.

Jaakkola, T. and Jordan, M. (2000) Bayesian parameter estimation via variational methods. *Statist. Comput.*, **10**, 25–37.

Kang, J., Johnson, T. D., Nichols, T. E. and Wager, T. D. (2011) Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *J. Am. Statist. Ass.*, **106**, 124–134.
Kang, J., Nichols, T. E., Wager, T. D. and Johnson, T. D. (2014) A Bayesian hierarchical spatial point process model for multi-type neuroimaging metaanalysis. *Ann. Appl. Statist.*, **8**, 1561–1582.
Laird, A. R., Lancaster, J. J. and Fox, P. T. (2005) Brainmap: the social evolution of a human brain mapping database. *Neuroinformatics*, **3**, 65–77.
Leininger, T. J. and Gelfand, A. E. (2017) Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Baysn Anal.*, **12**, 1–30.
Liang, S., Carlin, B. P. and Gelfand, A. E. (2009) Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *Ann. Appl. Statist.*, **3**, 943–962.
Marshall, T. and Roberts, G. (2012) An adaptive approach to Langevin MCMC. *Statist. Comput.*, **22**, 1041–1057.
Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.
Møller, J. and Waagepetersen, R. P. (2003) An introduction to simulation-based inference for spatial point processes. In *Spatial Statistics and Computational Methods* (ed. J. Møller), ch. 4, pp. 143–198. Berlin: Springer.
Møller, J. and Waagepetersen, R. P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall–CRC.
Møller, J. and Waagepetersen, R. P. (2007) Modern statistics for spatial point processes. *Scand. J. Statist.*, **34**, 643–684.
Montagna, S., Wager, T., Feldman Barrett, L., Johnson, T. D. and Nichols, T. E. (2018) Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data. *Biometrics*, **74**, 342–353.
Murray, I., Adams, R. P. and MacKay, D. J. (2010) Elliptical slice sampling. *J. Mach. Learn. Res. Wrkshp Conf. Proc.*, **9**, 541–548.
Neal, R. M. (1996) *Bayesian Learning for Neural Networks*. New York: Springer.
Neal, R. M. (2011) MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (eds S. Brooks, A. Gelman, G. L. Jones and X. Meng), ch. 5, pp. 113–162. Boca Raton: Chapman and Hall–CRC.
Owen, A. M., McMillan, K. M., Laird, A. R. and Bullmore, E. (2005) N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mappng*, **25**, 46–59.
Park, T. and van Dyk, D. A. (2009) Partially collapsed Gibbs samplers: illustrations and applications. *J. Computnl Graph. Statist.*, **18**, 283–305.
Radua, J. and Mataix-Cols, D. (2009) Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *Br. J. Psychiatr.*, **195**, 393–402.
Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N. and Surguladze, S. (2012) A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur. Psychiatr.*, **27**, 605–611.
Rasmussen, C. E. and Williams, C. K. I. (2005) *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.
Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T. and Eickhoff, S. B. (2012) Modelling neural correlates of working memory: a coordinate-based meta-analysis. *NeuroImage*, **60**, 830–846.
Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman and Hall–CRC.
Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D. and Nichols, T. E. (2009) Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, **45**, 810–823.
Samartsidis, P., Montagna, S., Laird, A. R., Fox, P. T., Johnson, T. D. and Nichols, T. E. (2017) Estimating the number of missing experiments in a neuroimaging meta-analysis. *Preprint bioRxiv 225425*.
Simpson, D., Illian, J., Lindgren, F., Sørbye, S. and Rue, H. (2016) Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, **103**, 49–70.
Taylor, B. M. and Diggle, P. J. (2014) INLA or MCMC?: A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *J. Statist. Computn Simuln*, **84**, 2266–2284.
Turkeltaub, P. E., Eden, G. F., Jones, K. M. and Zeffiro, T. A. (2002) Metaanalysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, **16**, 765–780.
Waagepetersen, R. P. (2004) Convergence of posteriors for discretized log Gaussian Cox processes. *Statist. Probab. Lett.*, **66**, 229–235.
Wager, T. D., Jonides, J. and Reading, S. (2004) Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage*, **22**, 1679–1693.
Wager, T. D., Lindquist, M. and Kaplan, L. (2007) Meta-analysis of functional neuroimaging data: current and future directions. *Socl Cogn. Affect. Neursci.*, **2**, 150–158.
Wager, T. D. and Smith, E. E. (2003) Neuroimaging studies of working memory. *Cogn. Affect. Behav. Neursci.*, **3**, 255–274.
Wood, A. T. A. and Chan, G. (1994) Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Computnl Graph. Statist.*, **3**, 409–432.

Yue, Y. R., Lindquist, M. A. and Loh, J. M. (2012) Meta-analysis of functional neuroimaging data using Bayesian nonparametric binary regression. *Ann. Appl. Statist.*, **6**, 697–718.
Zhang, H. (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Statist. Ass.*, **99**, 250–261.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Web-based supplementary materials for "Bayesian log-Gaussian Cox process regression: with applications to meta-analysis of neuroimaging working memory studies"'.