NeuroImage xxx (2017) 1–15

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Bayesian population receptive field modelling

Peter Zeidman [a,*], Edward Harry Silson [b], Dietrich Samuel Schwarzkopf [c,d], Chris Ian Baker [b], Will Penny [a]

[a] The Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, UK
[b] Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, 20892-1366, USA
[c] Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, UK
[d] UCL Institute of Cognitive Neuroscience, 17-19 Queen Square, London, WC1N 3AR, UK

ARTICLE INFO

Keywords:
pRF
Bayesian
Retinotopy
Mapping
Population receptive field
Modelling

ABSTRACT

We introduce a probabilistic (Bayesian) framework and associated software toolbox for mapping population receptive fields (pRFs) based on fMRI data. This generic approach is intended to work with stimuli of any dimension and is demonstrated and validated in the context of 2D retinotopic mapping. The framework enables the experimenter to specify generative (encoding) models of fMRI timeseries, in which experimental stimuli enter a pRF model of neural activity, which in turns drives a nonlinear model of neurovascular coupling and Blood Oxygenation Level Dependent (BOLD) response. The neuronal and haemodynamic parameters are estimated together on a voxel-by-voxel or region-of-interest basis using a Bayesian estimation algorithm (variational Laplace). This offers several novel contributions to receptive field modelling. The variance/covariance of parameters are estimated, enabling receptive fields to be plotted while properly representing uncertainty about pRF size and location. Variability in the haemodynamic response across the brain is accounted for. Furthermore, the framework introduces formal hypothesis testing to pRF analysis, enabling competing models to be evaluated based on their log model evidence (approximated by the variational free energy), which represents the optimal tradeoff between accuracy and complexity. Using simulations and empirical data, we found that parameters typically used to represent pRF size and neuronal scaling are strongly correlated, which is taken into account by the Bayesian methods we describe when making inferences. We used the framework to compare the evidence for six variants of pRF model using 7 T functional MRI data and we found a circular Difference of Gaussians (DoG) model to be the best explanation for our data overall. We hope this framework will prove useful for mapping stimulus spaces with any number of dimensions onto the anatomy of the brain.

1. Introduction

There are many examples of neuronal populations which represent stimulus spaces. In the auditory cortex, the 1-dimensional space of sound frequencies is mapped onto the surface of the brain (Merzenich and Brugge, 1973; Moerel et al., 2012). In the visual system, retinotopic mapping has revealed that the 2-dimensional plane of the retina is mapped multiple times onto the surface of visual cortex (e.g. Holmes, 1945). Place cells in the bat hippocampus respond maximally to a specific location in 3-dimensional space (Palacci et al., 2013) and conceptual knowledge may be represented neuronally in spaces of two dimensions or more (Constantinescu et al., 2016). Populations of neurons can be characterised by their receptive fields – the area(s) of N-dimensional space to which they maximally respond. In this paper, we introduce a generic framework for mapping stimulus spaces onto the brain and for performing hypothesis testing. We illustrate this approach in the context of visual population receptive field (pRF) mapping.

To enable pRF mapping, model parameters are required which capture the response of neuronal populations to experimental stimuli. The spatial distribution of these parameters across the brain can reveal large-scale topographic features, such as the presence of retinotopic maps. pRF mapping depends upon building generative models of imaging timeseries - we seek to understand how stimuli cause a change in spatially extended patterns of neuronal activity, which in turns cause the timeseries we measure using medical imaging devices. For functional MRI (fMRI), this involves modelling neuro-vascular coupling and the BOLD response (Kumar and Penny, 2014), which is an inherently nonlinear mapping. For instance, the BOLD response has a refractory period which depends on

* Corresponding author.
E-mail address: peter.zeidman@ucl.ac.uk (P. Zeidman).

http://dx.doi.org/10.1016/j.neuroimage.2017.09.008
Received 6 January 2017; Received in revised form 28 June 2017; Accepted 5 September 2017
Available online xxxx

the interstimulus interval (Friston et al., 1998). Furthermore, brain regions differ in the extent of their vascularization, giving rise to regional differences in BOLD response. Typically, pRF mapping experiments use a canonical haemodynamic response function, which may be determined on a per-subject basis. Here, to obtain the best possible estimators of neural activity for constructing pRF maps, we specified and estimated a non-linear model for each voxel's fMRI timeseries which included a biologically motivated differential equation model of neurovascular coupling and the BOLD response (Buxton et al., 2004; Stephan et al., 2007).

The objective of modelling (and of science more generally) is to test hypotheses. In the context of pRF mapping, hypotheses may be specified explicitly or implicitly. For instance, He et al. (2015) tested the explicit hypothesis that pRF position is modulated by perceived 3D space. Other pRF studies have been exploratory, for instance examining the reorganisation of visual field maps after lesions or disease (e.g. Levin et al., 2010). In studies such as this, there is an implicit hypothesis that pRF parameters will deviate from their normal range in specific areas of cortex. Despite the popularity of pRF mapping, a framework for formal hypothesis testing is currently lacking.

Here we introduce a set of tools for probabilistic (Bayesian) model fitting and inference in the context of pRF mapping, which could offer several benefits to experimenters. The optimal method for testing hypotheses is to compare the likelihood of the data under one model (or hypothesis) against the likelihood of the data under another model (Neyman and Pearson, 1933). For instance, an experimenter may wish to test whether certain brain regions have receptive fields with an excitatory centre and inhibitory surround, as identified by Hubel and Wiesel (1959) with single unit recordings. Such receptive fields may be modelled using a Difference of Gaussians (DoG) function (Rodieck, 1965), which can also capture the neuronal response at the level of voxels in fMRI data (Zuiderbaan et al., 2012). Alternatively, if the evidence for an inhibitory surround is lacking, a simple excitatory receptive field may be the better model (as applied to fMRI data by Dumoulin and Wandell, 2008). This kind of question, regarding which of several models is the best explanation of the available imaging data, may be addressed by comparing the evidence for the fMRI data under competing models at each point in the brain.

Models, including pRF models, cannot simply be compared based on the percentage variance they explain (their accuracy). Such a comparison ignores complexity – any model with more (independent) parameters will explain more of the variance, with the added risk of overfitting the noise and failing to generalise. One solution is to use cross-validation across datasets (e.g. Zuiderbaan et al., 2012) to approximate the model evidence, which offers control for over-fitting (i.e. assesses generalisability). However, this does not facilitate model comparison within a single dataset. In the framework proposed here, an approximation of the log model evidence is used known as the negative variational free energy (Friston et al., 2007; Penny, 2012). This quantity, estimated for each pRF model, is the accuracy of the model minus its complexity. By comparing models based on their free energy, the experimenter can select the simplest model that explains the most variance. Furthermore, by taking into account the covariance between parameters, the free energy is a more sensitive approximation to the log model evidence than other approximations such as the AIC or BIC.

As well as enabling competing models to be compared, the framework we propose has advantages for parameter-level inference, which may be of particular relevance for exploratory pRF studies. Here, parameters such as the pRF's size are each represented as a (normal) probability distribution, with both an expected value and variance/covariance (uncertainty). Thus, the uncertainty of parameter estimates may be expressed when visualising the pRF and when making comparisons within and between subjects. Uncertainty about the parameter estimates may arise from multiple sources – observation noise, subjects' movement, as well as any covariance among parameters. Also, it may not always be possible to confidently assign variance in the measured signal to either neuronal or haemodynamic causes. By estimating the full covariance among neuronal and haemodynamic parameters, we ensure that any uncertainty induced by ambiguity between these parameters is accounted for when visualising the pRF or testing hypotheses.

Here, we generalise an approach previously introduced in the context of tonotopic mapping (Kumar and Penny, 2014), making several novel contributions. We extend the method to stimuli of any dimension, and demonstrate its application in the context of visual (retinotopic) pRF mapping (Section 3.1, 3.2). We evaluate the face validity and robustness to noise of the method using simulated data (Section 3.3), and evaluate test-retest reliability across scanning runs using empirical data (Section 3.4). Finally, we demonstrate the use of this method for hypothesis testing (Section 3.5), by comparing the evidence for two established forms of pRF model: a Gaussian response function (Dumoulin and Wandell, 2008) and a Difference of Gaussians (centre-surround) response function (Zuiderbaan et al., 2012). Within each category of model we also compared the evidence for circular, elliptical and angled (rotated) receptive fields. We do not suggest drawing any firm conclusions about these models from the results we present here, which only uses data from a single subject. Instead, our aim is to demonstrate the statistical apparatus for comparing models, which we hope will prove useful for larger empirical studies. All of the methods described and evaluated here are made available to experimenters via a freely available software toolbox (Appendix A).

2. Material and methods

2.1. Participants

Empirical data were acquired as part of a previously reported study (Silson et al., 2015). Data from one participant is included here. All participants in the previous study had normal or corrected-to-normal vision and gave written informed consent. The National Institutes of Health Institutional Review Board approved the consent and protocol (93-*M*-0170, NCT00001360).

2.2. Data acquisition

Data were acquired using a Siemens 7 T Magnetom scanner in the Clinical Research Centre on the National Institutes of Health campus (Bethesda, MD). Partial EPI volumes of the occipital and temporal cortices were acquired using a 32-channel head coil (42 slices; $1.2 \times 1.2 \times 1.2$ mm; 10% interslice gap; TR, 2 s; TE, 27 ms; matrix size, 170×170; FOV, 192 mm). Anatomical T1 weighted volumes were acquired before the experimental runs. Standard MPRAGE (Magnetization-Prepared Rapid-Acquisition Gradient Echo) and corresponding GE-PD (Gradient Echo–Proton Density) images were collected and the MPRAGE images were then normalized by the GE-PD images, for use as high-resolution anatomical data for the fMRI data analysis.

2.3. Task and procedure

Naturalistic scene images were presented through a bar aperture that gradually traversed the visual field (Fig. 1). During each 36 s sweep, the aperture took 18 evenly spaced steps (each 2s or 1TR) to traverse the entire screen (Dumoulin and Wandell, 2008). Eight of these sweeps formed one run, in the following order: L-R, BR-TL, T-B, BL-TR, R-L, TL-BR, B-T, and TR-BL. There were 16 identical runs per participant. The scene stimuli, which covered a circular area (21° diameter) changed every 400 ms (5 per aperture position). During runs, participants performed a colour-detection task at fixation, indicating via button press when the white fixation dot changed to red. Colour fixation changes occurred semi-randomly, with ~2 colour changes per sweep.

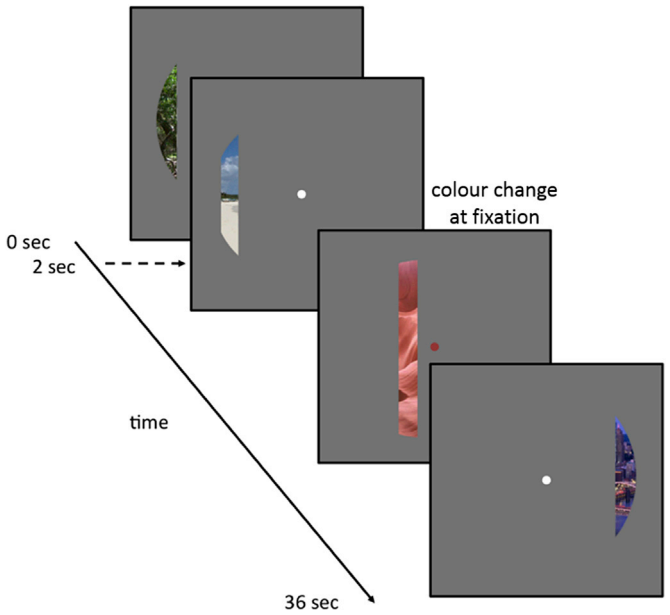

Fig. 1. Example stimulus frames used for pRF mapping. Scene images covering a circular (21° diameter) area of the display were presented through a bar aperture that moved gradually through the visual field. A single sweep across the visual field took 36 s and consisted of 18 equally timed (2 s) steps of equal width. In each run, the aperture completed eight sweeps (2 orientations, 4 directions). Participants were required to maintain fixation and indicate the detection of a colour change at fixation, via a button press. Adapted from Silson et al., (2015) *(license CC BY 4.0)*.

2.4. Pre-processing

As reported in Silson et al. (2015), functional data were initially motion-corrected using the Analysis of Functional NeuroImages (ANFI) software package (http://afni.nimh.nih.gov/afni), following removal of eight "dummy" volumes to allow stabilisation of the magnetic field. Subsequent analyses were conducted using SPM (Statistical Parametric Mapping) version 12 (http://www.fil.ion.ucl.ac.uk/spm/).

2.5. Timeseries extraction

pRF modelling was performed on a per-voxel basis. To enable this, we first specified a General Linear Model (GLM) using SPM, which included 6 motion regressors per session, in addition to regressors modelling the mean of each session. The software toolbox we introduce here (BayesPrf) expects a binary image mask (i.e. a three-dimensional image in NIFTI format) to define the voxels to include in the analysis. For each hemisphere, we created a mask which identified voxels in occipital cortex which were also on the cortical surface (half way through the cortical depth). To select the occipital cortex, we created a binary mask of all voxels in the SPM analysis, from the back of the brain towards the corpus callosum (6mm posterior to the back of the splenium). Next, we created masks of the voxels on the cortical surface. We used the standard processing pipeline in the Freesurfer toolbox (version 5.3, https://surfer.nmr.mgh.harvard.edu/) to extract the cortical surface, and we implemented software functions in Matlab to import the surfaces as Matlab structures and NIFTI images. We then took the intersection of the posterior brain mask and the cortical surface images to provide masks for the left and right hemispheres, confined to the back of the brain.

All voxels from within the combined masks (the intersection of occipital cortex and cortical surface) were extracted. They were high-pass filtered (with cut-off 128 s) to remove low frequency scanner drift, whitened to remove auto-correlation over time, and the motion confounds and session means were regressed out using the GLM described above. This procedure was repeated for each scanning run and the resulting timeseries were averaged over runs. Data from the 8 odd-numbered runs were used for the main analysis and the 8 even-numbered runs were kept aside to evaluate test-retest reliability (Section 3.4).

2.6. Model specification

We specified a generative model which predicted BOLD timeseries given visual stimuli. The model (based on Kumar and Penny, 2014) consisted of two parts – a neuronal model describing the brain's response to the stimuli, and an observation model describing the change in neurovascular coupling and BOLD response caused by the neuronal activity.

2.6.1. Stimulus specification

At each time step t, a set of points U_t was illuminated on the screen, where each point was defined by an x- and y-coordinate:

$$U_t = \{p_1, \ldots, p_n\}$$

$$p_i = (x \in \mathbb{R}, y \in \mathbb{R}) \tag{1}$$

The on-screen coordinates were discretised to give a reduced stimulus resolution of 41 × 41 pixels and expressed in units of degrees of visual angle.

2.6.2. Neuronal model

The neuronal activity $z(t)$ was modelled as a multivariate normal probability density function N:

$$z(t) = \beta \sum_{p_i \in U_t} N(p_i|\mu, \Sigma)$$

$$N(p_i|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(p_i-\mu)^T\Sigma^{-1}(p_i-\mu)} \tag{2}$$

This function is commonly used in pRF mapping (since Dumoulin and Wandell, 2008), although we used a normalized rather than the unnormalised density N. The response was summed over each illuminated pixel on the screen p_i and scaled by parameter β. The model was parameterised by k-dimensional vector μ, which specified the position of the pRF in k-dimensional stimulus space, and k-dimensional covariance matrix Σ which specified the width and rotation of the receptive field. Here we used stimuli of dimension $k = 2$, specified by their Cartesian coordinates (x, y) in visual space.

The neuronal parameters (μ, β and the elements of matrix Σ) needed certain constraints to be placed on their values. For instance, β needed to be positive whereas μ was constrained within the stimulated area of the visual field. However, the optimisation algorithm we employed (variational Laplace) expected normally distributed prior distributions for each parameter, constrained only by their variance. To overcome this, we substituted each parameter for a latent variable, which could be freely adjusted by the optimisation algorithm. Within the pRF model, these latent variables were transformed to be within the desired range of values, before being used to calculate the neuronal response in Equation (2). Next, we detail these transformations for each parameter.

pRF centre parameters. We constrained the centre of the pRF to fall within the circular area of the screen that was illuminated during scanning. While the pRF approach can model the influence of stimulation from beyond the stimulated area, estimating the parameters for such a model is a particularly ill-posed problem (any number of parameter combinations could give rise to the same observations). To constrain the pRF location we introduced latent variables l_ρ and l_θ, which controlled the distance and angle (polar coordinates) of the pRF location relative to the centre of the visual field. These variables were freely adjusted by the optimisation algorithm during model fitting, and were transformed to polar coordinates (ρ, θ) within the model as follows:

$$\rho = r^{*}NCDF\left(l_{\rho}, 0, 1\right)$$

$$\theta = 2\pi^{*}NCDF(l_{\theta}, 0, 1) - \pi \quad (3)$$

where NCDF is the cumulative density function for the univariate normal distribution. These transforms ensured that the distance ρ of the pRF from the centre of the visual field was constrained between 0 and $r = 10.5^{\circ}$ and the polar angle θ was constrained between $-\pi$ and π radians. A second transform then converted from polar to Cartesian coordinates:

$$\mu_{x} = \rho^{*}\cos(\theta)$$

$$\mu_{y} = \rho^{*}\sin(\theta)$$

$$\mu = (\mu_{x}, \mu_{y}) \quad (4)$$

Which then entered the model as μ in Equation (2). We could therefore guarantee that whatever values were selected for latent variables l_{ρ} and l_{θ} by the optimisation algorithm, the pRF location (μ_{x}, μ_{y}) would fall within the circular stimulated area.

pRF width parameter. The width of the pRF was parameterised by its standard deviation σ. To constrain the width to be in the range $[r_0\ r]$, we introduced a latent variable l_{σ}. This latent variable could be freely adjusted by the optimisation algorithm and was transformed within the model to give the pRF width:

$$\sigma = (r - r_0)^{*}NCDF(l_{\sigma}, 0, 1) + r_0 \quad (5)$$

We set r_0, the smallest allowed pRF width, to 0.5°. The pRF model which forms the first part of the results (Section 3.1–3.4) had a receptive field which was circular in shape (isotropic). This was modelled by setting the covariance matrix Σ in Equation (2) as follows:

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad (6)$$

Scaling parameter. Parameter β scaled the neuronal response (see Equation (2)). It was constrained to be positive by substituting it for latent variable l_{β} :

$$\beta = \exp(l_{\beta}) \quad (7)$$

Priors. To complete the specification of the neuronal model, each latent variable $(l_{\rho}, l_{\theta}, l_{\beta})$ was assigned a prior (normal) distribution, representing our beliefs before model fitting. The latent variables representing the pRF location l_{ρ} and l_{θ} had priors $N(0, 1)$. The transformation to constrain these (Equation (3)) converted these normal prior distributions to bounded flat distributions over ρ and θ, with all values in the allowed ranges equally likely (this is known as the probability integral transform). The second transformation, from polar to Cartesian coordinates (Equation (4)), changed this to a peaked distribution around the origin, with a 95% confidence interval of [-8.6° 8.6°], a 99% confidence interval of [-9.8° 9.8°] and 100% confidence interval of the maximum pRF size, [-10.5° 10.5°]. We considered this a reasonable prior given the over-representation of the fovea in cortex (e.g. Azzopardi and Cowey, 1993).

We set the prior on the latent variable representing the pRF width l_{σ} to $N(0, 1)$. Because of the CDF transformation (Equation (5)), this gave an equal prior probability of all widths in the allowed range. We set l_{β} to have prior distribution $N(-2, 5)$, which translated to an expected β parameter of 0.14 with 95% confidence interval [0.003 5.35]. This large variance reflected our uncertainty about this parameter. We found that having a small positive prior value for this parameter provided a reasonable starting point for the optimisation algorithm.

In summary, we implemented priors that would not impose strong hypotheses on the parameters, but would keep them within reasonable ranges. The prior on the pRF size parameter was a (bounded) flat prior and so acted as a simple constraint. The priors on the pRF centre coordinates were sufficiently permissive to enable any location within the stimulated field to be represented, with an expectation that pRFs are less likely to be centered in the periphery (>8.5°) than towards the fovea. This will be an important matter for construct validation (the comparison of this method against extant tools which don't make distributional assumptions), which we return to in the discussion. Another direction for future development may be to replace these fixed priors with empirical priors, either based on published studies or based on the group of subjects (a class of approaches referred to as Parametric Empirical Bayes).

2.6.3. Haemodynamic model

The predicted neuronal activity from the pRF model was fed into the extended Balloon model (Buxton et al., 2004; Stephan et al., 2007), which is a series of differential equations describing how neuronal activity causes a change in blood flow, and how blood flow causes the BOLD response. Most parameters of this model are fixed, based on previous empirical measurements, however three parameters are estimated on a voxel-wise basis: the transit time τ, the rate of signal decay κ, and the ratio of intra-to extra-vascular signal ε. Certain parameters in this model are field-strength specific, which we adjusted for our 7 T data according to recommendations by Heinzle et al. (2016). We set r_0, the intravascular relaxation rate, to 340, the frequency offset at the outer surface of magnetized vessels to 197.86 Hz and the prior distribution on $\log\varepsilon$ to the normal distribution with mean −3.99 and variance 0.83. The 3 free neurovascular/haemodynamic parameters were concatenated with the parameters of the neuronal model and estimated simultaneously.

The model was completed by setting prior expectations on the observation noise, which was assumed to be I.I.D. and normally distributed with mean zero. We set the prior log precision of the noise to 6 with variance (uncertainty) 1/55, which are the default values used with Dynamic Causal Modelling (Friston et al., 2003), a modelling framework within SPM which uses similar model specification and estimation procedures as are applied here.

2.7. Model estimation

For each voxel separately, the model was fitted to the data using the variational Laplace algorithm (Friston et al., 2007), implemented in the Matlab function spm_nlsi_gn.m, which is included with the SPM software package. This Bayesian estimation procedure provides two important estimates. The first is an approximation of the log model evidence, which is the log probability of observing the data y given the model m, $\log p(y|m)$. The approximation, known as the negative variational free energy (henceforth 'free energy'), may be expressed as follows:

$$\log p(y|m) = F(m) + KL[q(\theta|m)\ ||\ p(\theta|y, m)] \quad (8)$$

here, F is the free energy approximation of the model evidence, which we estimate. The second term is a distance measure (the Kullback-Leibler divergence, KL) between two distributions: the estimated parameters $q(\theta|m)$ and the true parameters $p(\theta|y, m)$. We cannot compute this, as we do not know the true parameters. However, because the log model evidence $\log p(y|m)$ is fixed, we know that by maximising the free energy $F(m)$, we minimise the second term and get the closest approximation of the model evidence. The free energy is defined in full elsewhere (see Appendix B of Penny, 2012), however it is important to recapitulate that it may be decomposed into two parts (Beal and Ghahramani, 2003):

$$F(m) = accuracy(m) - complexity(m) \quad (9)$$

The accuracy is the fit of the model to the data. The complexity is the difference (the KL divergence) between the prior distribution of the parameters and the estimated (posterior) distribution of the parameters. This definition of complexity gives the free energy advantages over other

approximations such as the AIC and BIC (Penny, 2012). Rather than simply counting the number of parameters in the model, the KL divergence takes into account the full covariance among parameters, meaning that parameters which are estimated to be independent contribute more to the complexity term than those which co-vary. Using the free energy, we compared pRF models to find the most accurate and least complex explanation for our data.

The algorithm also provides the estimated parameters $q(\theta|m)$ which maximise the free energy. These are computed under the Laplace assumption, which means that all parameters – prior and posterior – are normally distributed. Thus for each parameter, whether neuronal or haemodynamic, we have an expected value (mean) and uncertainty (variance), as well as the covariance among parameters. We report these parameters and their uncertainty in various forms throughout the paper.

We performed model estimation separately for timeseries from each voxel. To improve performance, we initialized the VL model fitting algorithm with pRF parameters identified using a grid search (Dumoulin and Wandell, 2008) with a canonical haemodynamic response function. To further reduce the total estimation time, we developed software functions to divide the voxels across a parallel computer cluster and combine the results following estimation, which are included in the software implementation (Appendix A). Here we divided the estimation across a 192-core cluster computer (running CentOS 6.3 with a clock speed ranging from 2.1 Ghz to 2.5 Ghz). On each core of this cluster, estimation took an average of 116 s per voxel to complete. In total there were 21,230 voxels, taking around 3 h 34 min for the whole analysis.

2.8. Inference

2.8.1. Parameter map generation

Having estimated the pRF models in each voxel, we generated maps of the estimated pRF parameters across the brain, projected onto the cortical surface (Section 3.1). We thresholded these parametric maps to only include voxels where the pRF parameters had moved away from their priors (or had become more precise than the priors). To perform this thresholding, we compared the evidence (approximated by the free energy) for each estimated pRF model against a nested model, in which the pRF parameters were fixed at their prior means. If the evidence for these two models was similar, it meant that our knowledge about the pRF parameters did not improve as a result of seeing the data. On the other hand, if the evidence for the full model was stronger than the evidence for the nested model, it would mean that we can be more confident about these parameters after seeing the data.

To compute this model comparison, for each voxel's pRF model we specified a nested model m_N, in which the prior variance of the pRF location and width parameters (μ_x, μ_y, σ) were fixed at zero. We then computed the ratio of the evidence of the full model m_F against this nested model m_N. This ratio is called the Bayes factor, and the log of this ratio is approximately equal to the difference in the free energy of each model:

$$\ln BF = \ln \frac{p(y|m_N)}{p(y|m_F)} \cong F_N - F_F \tag{10}$$

To compute the free energy of the nested models, F_N, we did not need to separately estimate them using the estimation algorithm. Instead, we derived the nested models' evidence and parameters analytically using Bayesian Model Reduction (Friston and Penny, 2011), which is a generalisation of the Savage-Dickey ratio used in classical statistics. This calculation was performed on the order of milliseconds for each voxel. Having computed the log Bayes factor for each voxel, we then computed the posterior probability of the full model m_F:

$$p(m_N|y) = \frac{1}{1 + \exp(-\log BF)}$$

$$p(m_F|y) = 1 - p(m_N|y) \tag{11}$$

We thresholded the parametric maps to only include voxels where the posterior probability of the full model was at least 0.95. These surviving voxels were used to generate the maps in Fig. 2.

2.8.2. Plotting haemodynamic response

To illustrate the estimated variability in BOLD response across the brain, we computed the impulse response function of each pRF model (those with at least 0.95 probability, above) and averaged these responses across voxels (Section 3.1). The response functions were computed using Volterra kernels, as described by Friston et al. (1998). We also identified voxels where having freely varying haemodynamic parameters increased the free energy, relative to having fixed HRF parameters (i.e. a canonical HRF). We used the same Bayesian Model Reduction procedure above to compare each full pRF model against the nested model with the prior variance on the haemodynamic parameters set to zero. We plot the results as a posterior probability map, thresholded at $p > 0.95$ (Fig. 3, right).

2.8.3. Plotting pRFs

We plotted the receptive field of an example pRF model before and after fitting it to the data (Fig. 4). The prior receptive field was computed by taking 1000 samples from model's prior multivariate distribution over the parameters, and for each sample, computing the response of the neuronal function (Equation (2)) at evenly spaced locations in the visual field (Fig. 4A). These responses were then averaged. This is referred to as the prior predictive density (PD) q for each point in space (x, y):

$$q(x, y) = \int g(x, y, \theta) p(\theta) d\theta \tag{12}$$

where $g(x, y, \theta)$ is the response of the neuronal (pRF) function to a single point in space with parameters θ. By using this approach, we were able to plot the pRF while accounting for uncertainty over its location and size. The posterior response was similarly computed using the estimated parameters following model fitting.

2.8.4. Bayesian Parameter Averaging

In order to summarise pRF models' expected values and covariance across voxels, we computed the Bayesian Parameter Average (BPA). The BPA is the average of the parameters across voxels, weighted by the precision of each estimate. The data entering BPA was the vector of estimated pRF parameters μ_v for each voxel v, together with the precision matrix Λ_v (inverse of the covariance matrix). The BPA provides a probability distribution over the parameters, with vector of means M and covariance matrix C:

$$\begin{aligned} \Lambda &= \sum_{v=1}^{N} \Lambda_v \\ C &= \Lambda^{-1} \\ \mathrm{M} &= \Lambda^{-1} \sum_{v=1}^{N} \mu_v \Lambda_v \\ p(\mathrm{M_i} > 0) &= 1 - NCDF(0; |\mathrm{M_i}|, C_{i,i}) \end{aligned} \tag{13}$$

The final line of equation (13) is the probability that any given parameter i, averaged across voxels, is different to zero. $NCDF(x; a, b)$ is the normal cumulative distribution function with mean a and variance b. In this formulation, we made the simplifying assumption that each voxel's data were independent.

2.8.5. Computing correlations and entropy maps

To compute the correlation between parameters we converted the covariance matrix C from the BPA (described above) to a correlation matrix in the normal way. To visualise the spatial distribution of parameter uncertainty, we created maps of the negative entropy, defined

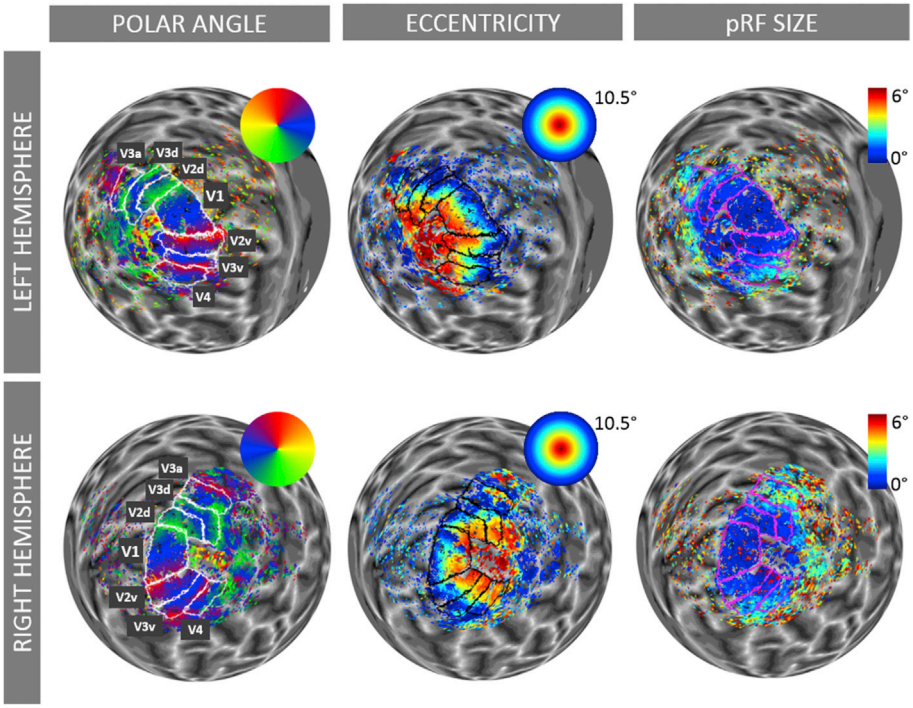

Fig. 2. Parameter maps after estimating the basic pRF model. Left: Polar angle of the pRF, with colours corresponding to the colour wheel (inset). Middle: Distance of the pRF centre from the fovea (eccentricity), with colours indicating degrees of visual angle. Right: Width parameter (standard deviation), with the colour bar indicating degrees of visual angle - smaller pRFs are in blue, larger in red. Maps are thresholded at model posterior probability >0.95 (see methods Section 2.8.1) and projected onto a spherical projection of the cortical surface.

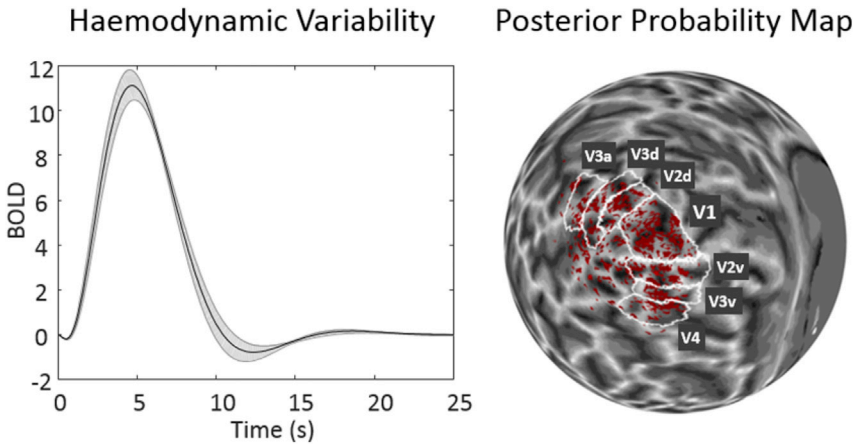

Fig. 3. Analysis of haemodynamic parameters in the pRF model. **Left**: The variability of the estimated BOLD impulse response function across voxels, in response to a stimulus at time 0 of duration 0. The solid line is the mean across voxels and the shaded area is mean ±1 standard deviation across voxels. **Right**: The probability of pRF models with free haemodynamic parameters, relative to models with a fixed (canonical) HRF. Shown for left hemisphere only. Thresholded at P > 0.95.

for each voxel v as:

$$h_v = -\ln|\Sigma_v| \tag{14}$$

where $|\ldots|$ is the determinant and Σ_v is the estimated covariance matrix of the model for voxel v. The negative entropy h_v of a multivariate Gaussian has units of nats, where more positive values indicate greater certainty in the parameters.

2.8.6. Simulation

To evaluate the face validity of the approach, we performed simulations. First, we estimated the empirical signal-to-noise ratio (SNR) of our data. We did this by randomly selecting an estimated pRF model, which had been fitted to the timeseries from a voxel in left hemisphere near the midline (within V1, the delineation of which is described in section 2.8.8). We integrated this model over time to give the predicted BOLD timeseries y and computed the residual timeseries r. We calculated the SNR based on the standard deviation:

$$SNR = \frac{\sigma_y}{\sigma_r} = 1.06 \tag{15}$$

Next we chose a random 1000 voxels from our empirical data (which had survived the test for significance described in section 2.8.1), and we used the estimated pRFs from these voxels to generate 1000 simulated BOLD timeseries. We then added five levels of observation noise to each timeseries: from a level noisier than our example voxel (SNR 0.5) to less noisy (SNR 1.5). We fitted pRF models to these simulated timeseries and correlated the estimated parameters against the 'true' parameters that generated the data. Note that for all validation analyses, we converted

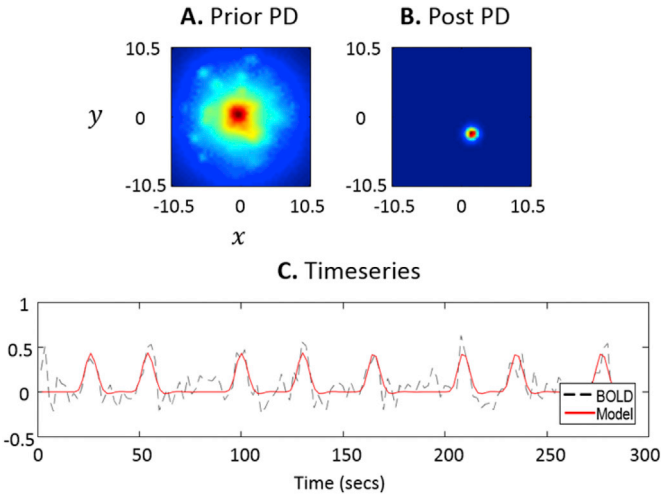

Fig. 4. Estimated pRF and fitted timeseries for an example voxel in left V1 (voxel 2334). **A.** The response of the model to each position in visual space prior to Bayesian model fitting. This is the prior Predictive Density (PD), which uses sampling to take into account uncertainty over the parameters. The prior model parameters were sampled 1000 times and the predicted responses generated and averaged. The units of the axes are degrees of visual angle. **B.** The posterior PD, which is the estimated pRF response after Bayesian estimation. **C.** The modelled timeseries (red) and observed timeseries (grey).

from latent variables (e.g. l_ρ and l_θ) to the underlying parameters which entered the model (e.g. ρ and θ), as these are the quantities of interest to experimenters.

2.8.7. Alternative models

We compared the evidence for different pRF models to address two questions. The first was whether a single Gaussian or a Difference-of-Gaussians (DoG) response function would be a better explanation for the data. The DoG function has an excitatory centre and inhibitory surround. Our second question was what shape of receptive field would best explain the data (circular, elliptical or elliptical with rotation). Thus, we formed a 2×3 factorial model space to address these questions. Here we describe the modifications made to the basic circular (isotropic) model to form these alternative models.

Elliptical models. We specified models which replaced the pRF width (standard deviation) parameter σ with separate parameters representing horizontal width σ_x and vertical width σ_y. This enabled the pRF to take on an elliptical shape. This was implemented by changing the covariance matrix Σ (see Equation (2)) to include two separate parameters:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \tag{16}$$

Latent variables were introduced for each of the two width parameters, as for the single width parameter in the basic model, with the same priors.

Elliptical models with rotation. Enabling the pRF to rotate required the introduction of a parameter P representing the correlation coefficient:

$$\sigma_P = \mathrm{P}*\sigma_x*\sigma_y$$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_P \\ \sigma_P & \sigma_y^2 \end{bmatrix} \tag{17}$$

Parameter P was constrained within the range -1 to 1 by substituting it for a latent variable l_P . The latent variable was transformed within the model as follows:

$$\mathrm{P} = 2*NCDF(l_\mathrm{P}, 0, 1) - 1 \tag{18}$$

Difference of Gaussians. We created versions of the circular, elliptical and elliptical + rotation models which used a DoG response function (Rodieck, 1965; Zuiderbaan et al., 2012). The Gaussian distribution was complemented by a second distribution, which had certain constraints to ensure that it would act as an inhibitory surround. For the basic case of a circular receptive field, the modified response function was as follows:

$$z(t) = \beta_c \sum_{p_i \in U_t} N(p_i|\mu, \Sigma_\mathrm{c}) - \beta_s \sum_{p_i \in U_t} N(p_i|\mu, \Sigma_\mathrm{s})$$

$$\Sigma_c = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

$$\Sigma_\mathrm{s} = \begin{bmatrix} (\sigma^2 + \sigma_d^2) & 0 \\ 0 & (\sigma^2 + \sigma_d^2) \end{bmatrix}$$

$$\beta_s = max(\beta_c - \beta_d, 0)$$

$$(\beta_\mathrm{c}, \beta_\mathrm{d}, \sigma_\mathrm{d}) > 0 \tag{19}$$

here, the neuronal response z was the difference of two normal distributions. The first represented the excitatory response of the centre of the pRF. It was parameterised by scaling parameter β_c, centre location μ and covariance matrix Σ_c. The second normal distribution represented the inhibitory surround of the pRF. It has the same location as the excitatory centre μ, but different scaling β_s and covariance Σ_s. The difference in scaling between the two pRFs was controlled by parameter β_d and the difference in the pRFs' widths was controlled by the parameter σ_d. Positivity constraints were enforced on the β scaling parameters and σ_d. This formulation, which introduced 2 new free parameters compared to the previous model (β_d, σ_d), ensured that the response of the centre was positive and the response of the surround was negative.

Using a latent variable, we constrained parameter σ_d to fall into the range $[0\ r]$ with a flat prior. The difference in scaling of the surround relative to the centre of the pRF, β_s, was constrained to be positive using a log parameter. This translated to a prior expectation for β_s of 0.05 with 95% confidence interval [0.001 1.97].

2.8.8. Bayesian model comparison

We compared the evidence for our data under 6 different pRF models, using the free energy, described above. For each voxel we had 6 numbers - the free energy of each competing model - which we collated into a [voxels x 6] matrix. We submitted this matrix for analysis using a random effects (RFX) model (Stephan et al., 2009). Typically, this approach is used in the context of studying groups of subjects. One assumes that each subject's data were generated by one of the models in the comparison, and the RFX model estimates the probability that any randomly selected subject's data were generated by each model. These are known as the expected probabilities of each model. Additionally, it gives the probability that any one model is better than all the others in the comparison – these are the protected exceedance probabilities. (The word 'protected' refers to a statistical correction to account for the possibility that all models are equally likely (Rigoux et al., 2014).) Here, we simply used multiple voxels instead of multiple subjects, and the RFX model identified which pRF best explained the data in spatially extended regions of the brain.

After performing model comparison at the level of all included voxels, we next performed comparisons on a region of interest (ROI) basis. We manually delineated the visual fields V1, V2d, V2v, V3d, V3v, V3a and V4 (where d = dorsal, v = ventral). This was performed by exporting the parameter maps described above to the Delineation Tool in the SamSrf Toolbox (https://doi.org/10.6084/m9.figshare.1344765.v23). The manually defined ROIs were then imported from Freesurfer format (as used by the Delineation Tool) back into NIFTI format. Simple software

tools for this export/import procedure are supplied with the toolbox described here, and we made further use of the Delineation Tool to overlay our parameter maps with the ROI boundaries (Figs. 2, 3, 6 and 8).

Strictly, the RFX procedure described above assumes that the sources of data (the voxels) are independent. This assumption was violated here due to the spatial smoothness of the fMRI data, but nonetheless this method allowed us to provide summary measures for each ROI (see section 3.5.2). For completeness, we repeated the pRF analyses and model comparison while removing the issue of spatial smoothness assumptions, by using the mean timeseries for each ROI rather than each individual voxel. The results of this analysis are illustrated in Supplementary Fig. 4 and are summarised in Section 3.5.2 of the results.

3. Results

3.1. Parameter estimates

After estimating a pRF model for each voxel, we produced maps of the parameters (Fig. 2). These confirmed that the parameters were in keeping with the established retinotopic arrangement of occipital cortex, with clearly defined boundaries between the visual fields (Fig. 2, left).

The remaining parameters controlled the scaling of the neuronal response (β), the transit time (τ), the rate of decay (κ) and ratio of intra-to extra-vascular signal (ε), all of which were estimated on a voxel-wise basis together with the pRF parameters. To investigate the variability in neurovascular/haemodynamic response across voxels, we calculated the estimated BOLD impulse response function for each voxel and summarised these responses by their mean and standard deviation (Fig. 3, left). The BOLD responses varied across voxels, in terms of the height of the response and the undershoot which follows, but there was limited variability in the latency of the peak. The consistency of BOLD responses across voxels may have been increased by selecting voxels from a consistent depth from the cortical surface – a consideration which may be interesting to revisit when modelling laminar-specific fMRI responses (Heinzle et al., 2016).

We then quantitatively investigated whether there was benefit in estimating the haemodynamic parameters in a voxel-wise manner. To address this, we first identified the voxels where the neural parameters had moved away from their priors (posterior model probability > 0.95). Then for each of these voxels we performed a Bayesian model comparison, of the full pRF model against a reduced model where the haemodynamic parameters were fixed at their prior expectations (i.e. their prior variance was set to zero). This effectively produced alternative models with a canonical or fixed HRF. Out of 9762 voxels, there were 1791 voxels (18%) with strong evidence (p > 0.95) for the model with free haemodynamic parameters over the fixed model. There were no voxels with strong evidence for the fixed model. The voxels which benefited from having free haemodynamic parameters were distributed across occipital cortex, rather than being localised to any particular region of interest (Fig. 3, right). Therefore, we may conclude that the added complexity cost of having three free HRF parameters per voxel is outweighed by the increased accuracy which they provide, across occipital cortex.

The prior pRF response, posterior pRF response and fitted timeseries for an example voxel are shown in Fig. 4. In Fig. 4A, the response of the pRF model is shown for each point in retinotopic space, with the model's parameters set to their prior means and variances. To account for uncertainty, we sampled from the parameters according to their prior probability distributions, then generated the pRF response based on the sampled parameters. Repeating this 1000 times and averaging the responses gave the prior Predictive Density (PD) shown. Fig. 4B shows the response of the model with parameters estimated from the data (the posterior parameters), demonstrating a punctate pRF estimate in the right visual field. This plot again similarly takes into account uncertainty over the parameters through sampling (the posterior PD). Strikingly, the uncertainty visible as the diffuse pattern in the prior PD has been explained away after seeing the data. This close fit to the data can be also be seen qualitatively in the modelled timeseries (Fig. 4C). It is also noticeable that the post-stimulus undershoot in the BOLD response was not well captured, which is a known limitation of the haemodynamic model we are applying (Havlicek et al., 2015). We return to this issue in the discussion.

3.2. Correlation and entropy

A key advantage of the Bayesian approach is that the uncertainty of each parameter (the variance) is estimated, as well as the covariance among parameters. The covariance is important because parameters which strongly co-vary cannot be precisely estimated, and this uncertainty should be taken into account when making inferences. To investigate the covariance among parameters, we averaged models over voxels using Bayesian Parameter Averaging (see methods Section 2.8.4) and then transformed the averaged covariance matrix to give a correlation matrix. This is shown in Fig. 5, where the upper left quadrant shows the correlations among neuronal (pRF) parameters, the bottom right quadrant shows correlations among haemodynamic parameters and the top right quadrant shows the correlation between neuronal and haemodynamic parameters.

Reassuringly, the parameters representing the location of the pRF in visual space (l_ρ, l_θ) had only weak correlation with other parameters. However, the pRF width parameter σ was strongly correlated with the scaling parameter l_β (Pearson's correlation 0.63). In practice, this means that one could increase the pRF width or increase the scaling parameter and, to some extent, get the same change in the predicted timeseries. This phenomenon is not specific to the Bayesian approach proposed here, and may be of concern for any pRF estimation method, despite this correlation not generally being quantified. (The same issue also pertains to the Difference of Gaussian model – for completeness we provide correlation matrices for all models we examined, in Supplementary Fig. 1.) The correlation matrix also revealed statistical dependencies among the haemodynamic parameters, as well as between the scaling parameter l_β and the haemodynamic parameters. In practice, correlations among the haemodynamic parameters are unlikely to be a problem, as experimenters rarely need to make inferences about these parameters individually.

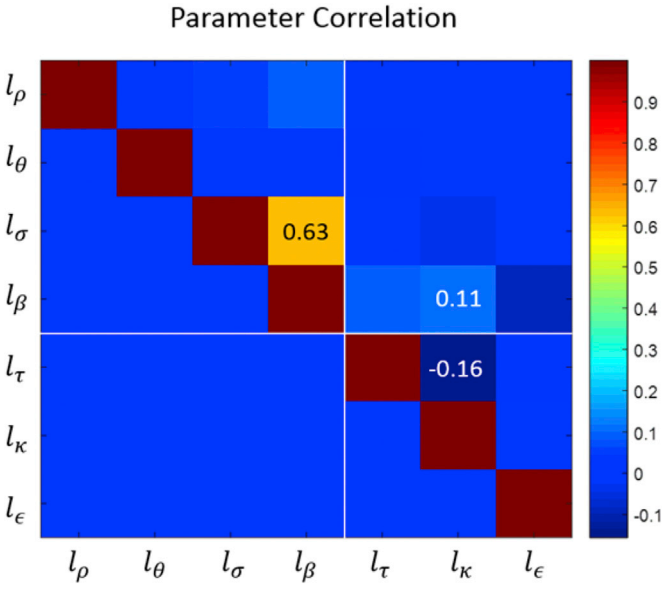

Fig. 5. Correlation among model parameters (latent variables), averaged over voxels. There were 4 neuronal (pRF) parameters $(\rho, \theta, \sigma, \beta)$ *and 3 neuro-vascular/haemodynamic parameters* $(\tau, \kappa, \varepsilon)$. For clarity, only values above the diagonal are shown and entries with an absolute value greater than 0.10 are labelled.

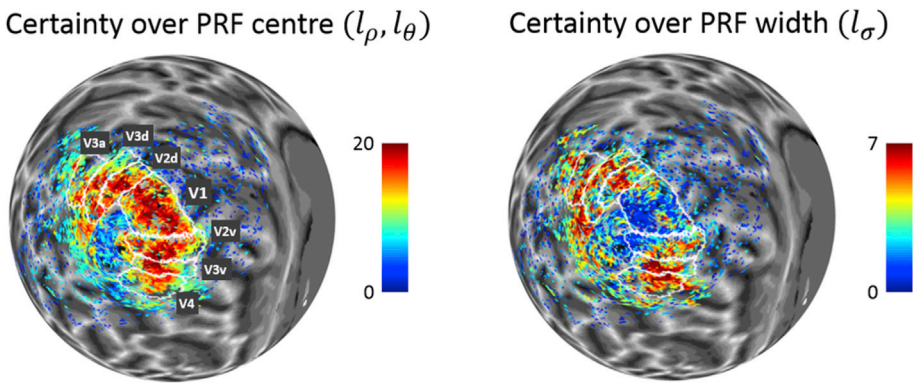

Fig. 6. Negative entropy maps showing certainty over the pRF location (left) and certainty over the pRF width (right) after estimation of the models. More positive values (red) indicate greater certainty. Colours indicate negative entropy in units of nats. Images are thresholded to only display voxels at p > 0.95 (see section 2.8.1).

Having covariance among parameters induces uncertainty in the parameter estimates, in addition to any uncertainty induced by observation noise. Next, we mapped this uncertainty by computing the negative entropy for each pRF model, where more positive values indicate greater certainty about the parameter estimates. We computed separate negative entropy maps for the pRF location (l_ρ, l_θ) and width (l_σ) parameters. The map for pRF location (Fig. 6, left) showed that we could be confident about the position of the pRF in the majority of voxels in the early visual fields (V1, V2), but fewer voxels show high certainty in the higher visual fields (V3a, V4). There was less confidence overall for pRF width than for pRF location – the maximum negative entropy was 6.70 nats for pRF width compared to 20.40 nats for pRF location (higher is more confident). Surprisingly, in early visual cortex - where we would expected to have been most certain about pRF width - we were in fact least certain (Fig. 6, right). This could be because smaller pRFs, which typify early visual cortex, require higher resolution stimuli in order to gain confident estimates about their size. Additionally, uncertainty over pRF width may have been induced by covariance with the scaling parameter l_β, identified above.

3.3. Face validity

We next performed two forms of validation: we used simulations to establish the face validity of the approach and we used empirical data to evaluate test-retest reliability across runs.

We fitted pRF models to 1000 simulated timeseries where the 'true' parameters which generated the data were known, under varying levels of observation noise. The expected values of the pRF position closely matched the true parameters (Fig. 7, left), while the accuracy of the width parameter was slightly lower and dropped more markedly with SNR. While all correlations were highly significant (all p < 1.3e-11), the neurovascular/haemodynamic parameters $(\tau, \kappa, \varepsilon)$ were estimated less accurately and were more sensitive to the SNR than the pRF parameters. This may have been due to the covariance among the haemodynamic parameters, reducing their identifiability. The accuracy of the θ polar angle parameter was slightly lower than we had expected at lower SNRs (correlation 0.94 at SNR 1 and correlation 0.83 at SNR 0.5). We found this was an artefact of the use of polar coordinates. There were pRFs with parameter θ close to $-\pi$ radians in the test set and close to $+\pi$ radians in

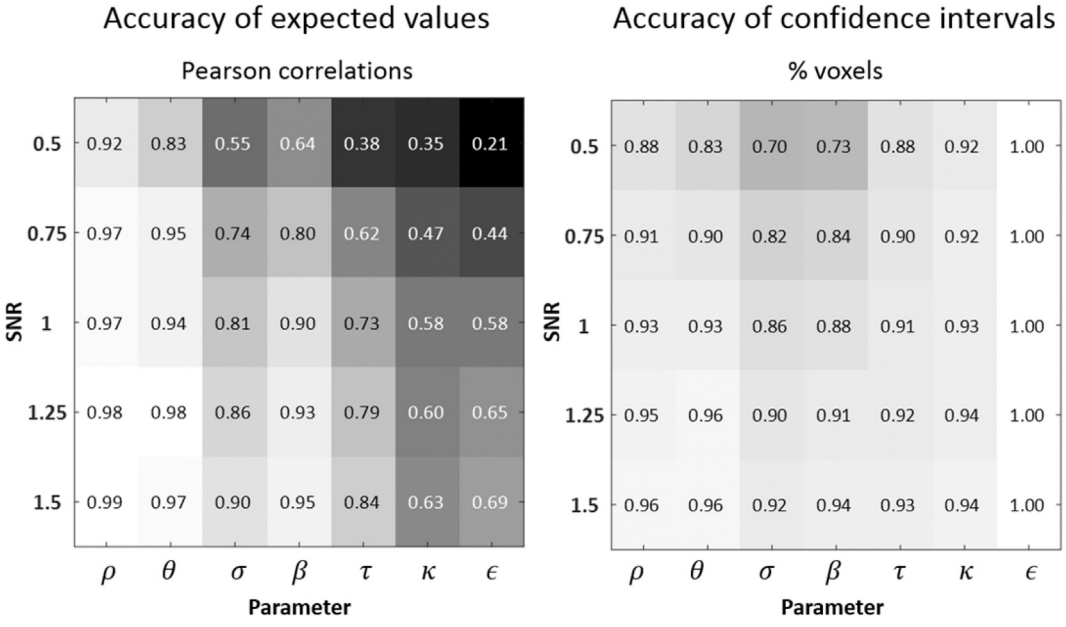

Fig. 7. Accuracy of parameter estimates assessed by simulation. ***Left:*** Pearson's correlations between the parameters used to simulate data and parameters estimated from those simulated data. The vertical axis lists 5 levels of observation noise, quantified as the signal-to-noise ratio (SNR). Darker colours indicate lower correlations. Parameters are as for Fig. 5. ***Right:*** The percentage of voxels where the estimated 95% (posterior) confidence interval contained the parameter which generated the data.

the validation set, or vice versa. Despite being adjacent in visual space, the difference in sign reduced the test-reliability. When we corrected for this by converting from polar to Cartesian coordinates, accuracy of the θ parameter was near ceiling (correlation > 0.99) at SNR 1.5, SNR 1.25 and SNR 1, reducing to correlation 0.98 at SNR 0.75 and correlation 0.96 at SNR 0.5.

We also investigated the accuracy of the estimated uncertainty of the parameters (Bayesian confidence intervals). We anticipated that the posterior 95% confidence interval of each parameter would include the "true" parameter which generated the simulated data, approximately 95% of the time. For each level of SNR and for each parameter, we computed the fraction of the 1000 voxels which included the 'true' parameter in their estimated 95% confidence interval (Fig. 7, right). Near the SNR of our example voxel (SNR 1.06), the confidence intervals behaved as expected for the pRF location, with slightly lower performance than expected for the pRF width. With reducing SNR, the accuracy of the confidence intervals decayed slowly (in particular the width parameter), perhaps due to the appearance of local optima in the parameter space, coupled with the correlation between the width and beta parameters. However, overall, these results demonstrate that our estimates of uncertainty were robust to noise, even for the haemodynamic parameters which had the least accurate expected values.

3.4. Test-retest reliability

Next we tested the consistency of the parameter estimates across scanning runs on empirical data. The main analysis, above, used only the 8 odd-numbered runs. Here we used the 8 even-numbered runs as a validation set, and the resulting parameter map of polar angle (which is generally the most informative map for delineating visual fields) is shown in Fig. 8, with the remaining parameter maps shown in Supplementary Fig. 2. Here, the white lines are the borders of the visual fields from the main dataset, and by overlaying these on the validation dataset (Fig. 8, right), it can be seen that the maps are highly consistent. To quantify this we correlated the expected values of the parameters from each half of the data on a voxel-wise basis. We limited this to within the area of the visual field regions of interest and excluded voxels where the pRF parameters failed to deviate from their priors ($p > 0.95$). Note that this correlation analysis made a strong assumption about the success of coregistration – it assumed that the voxels in all runs were perfectly aligned. Nevertheless, the correlation between pRF location parameters (transformed to Cartesian coordinates) was high (Table 1). The width parameter was much less consistent, which speaks to the difficulty of recovering this parameter in pRF models. That said, the parameter map of width still showed striking consistency across datasets (Supplementary Fig. 2).

Table 1
Correlations between parameters estimated from alternating runs.

	Pearson's correlation
μ_x	0.89
μ_y	0.91
σ	0.50
β	0.70
τ	0.83
κ	0.77
ε	0.75

3.5. Model comparison

Having validated the modelling framework, we next used it to demonstrate testing hypotheses. We asked two questions. First, whether the neuronal population within each voxel had only an excitatory response to stimuli within their receptive fields, or whether they also exhibited an inhibitory response at the periphery of their receptive fields. We represented these two alternatives using models with a single Gaussian response function, and models with a Difference of Gaussians (DoG) response function with an excitatory centre and inhibitory surround, as has previously been introduced (Zuiderbaan et al., 2012). We also asked which of three shapes of receptive field would offer the best explanation for our data – circular (Model 1), elliptical (Model 2), or elliptical with rotation (Model 3). We formally addressed these questions by comparing the evidence for the data under competing models (Bayesian model comparison). The model that wins in such a comparison is the model that strikes the optimal balance between accuracy and complexity.

We formed a model space akin to a 2 × 3 factorial design. The first factor was the response function: models either had a single Gaussian or a DoG response function. The second factor was the shape of the receptive field (x3). We specified each of these 6 models and estimated the free energy (log model evidence) within each voxel. Thus, we had six free energies per voxel, representing the relative evidence for each model, which we pooled across voxels using a random effects analysis (see methods section 2.8.8 for details).

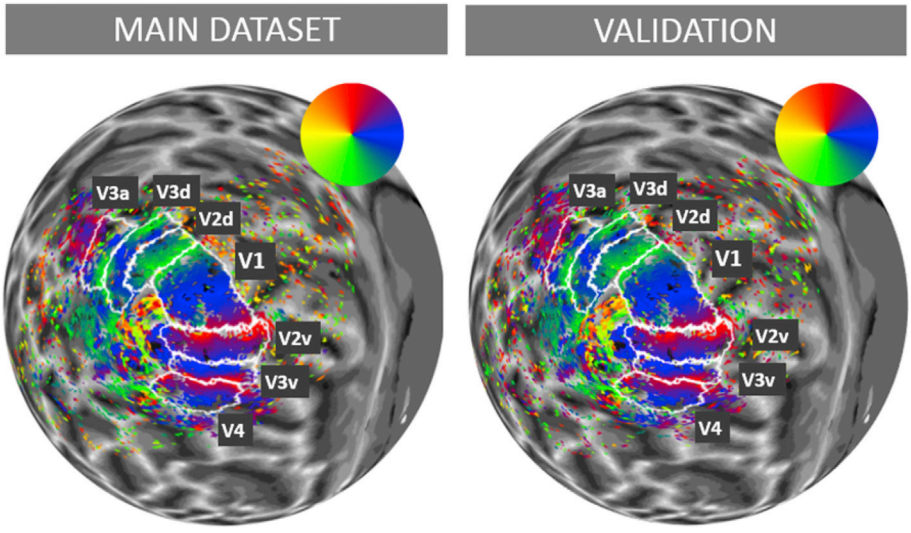

Fig. 8. Test-retest reliability. Polar angle maps are shown for left hemisphere from the main dataset (left) and the validation dataset (right). White lines show delineation of the visual fields on the main dataset in both images.

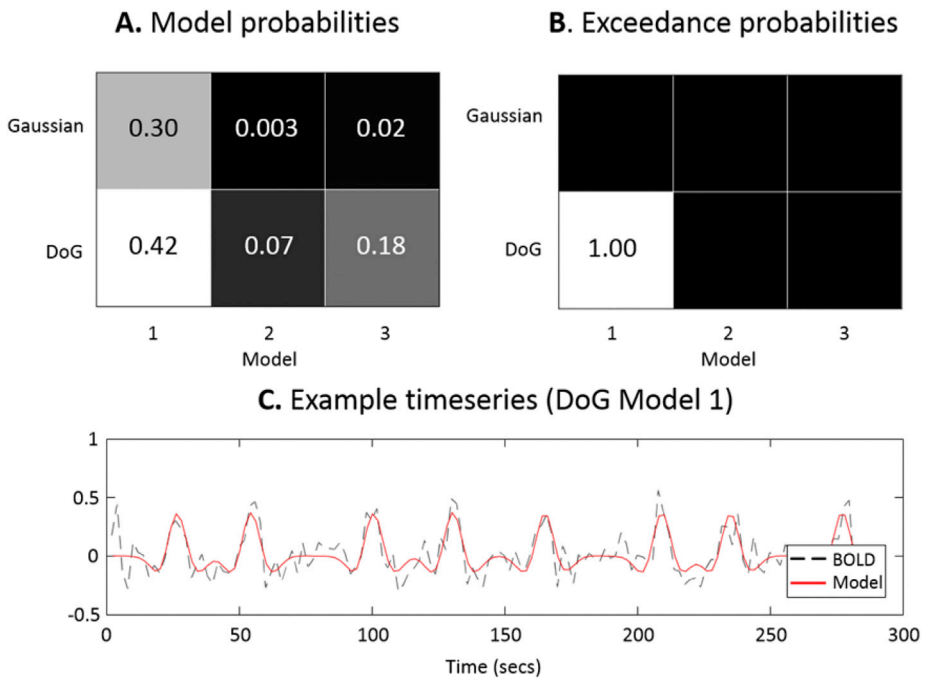

Fig. 9. Comparison of 6 pRF models across all included voxels. Models either had a single Gaussian or Difference of Gaussians (DoG) receptive field (rows) and were circular (Model 1), elliptical (Model 2) or elliptical with rotation (Model 3), shown in the columns. **A**. The expected probability of each model. **B**. The protected exceedance probability (PXP) of each model, which is the probability that each model is better than the all the others in the comparison. **C**. An example timeseries for the winning model (DoG Model 1), fitted to data from the same voxel as shown in Fig. 4.

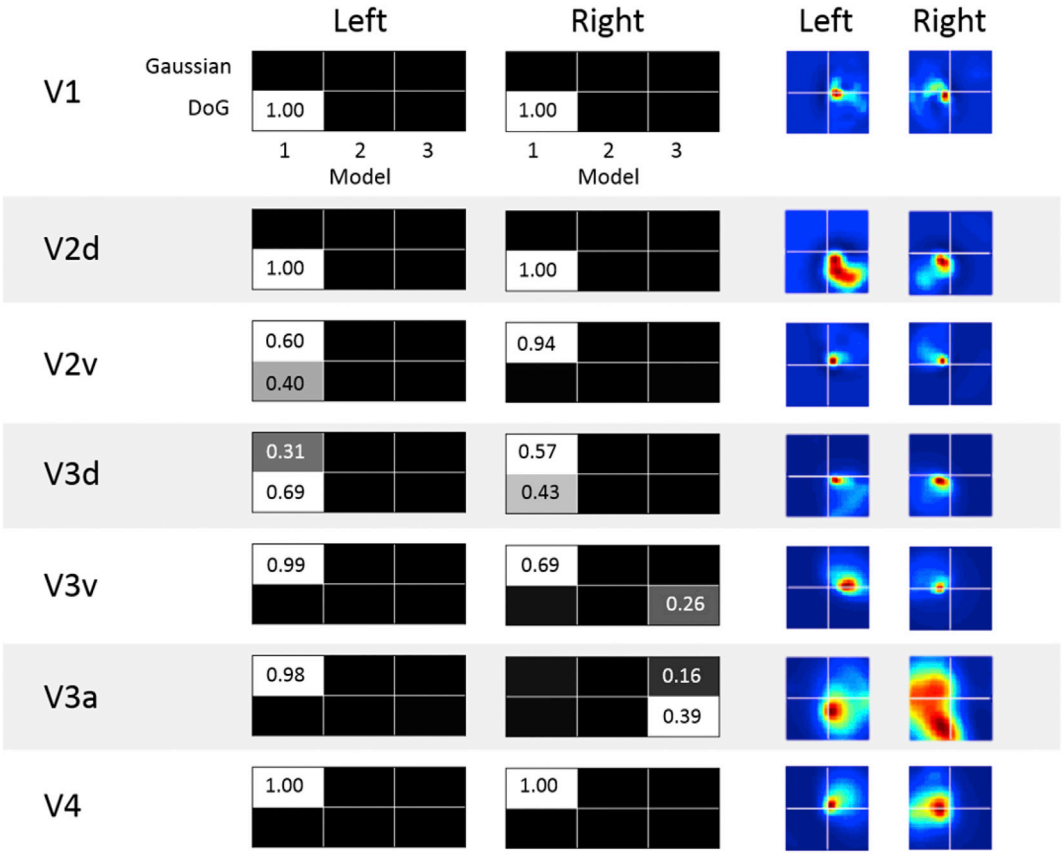

Fig. 10. Results of comparing 6 different pRF models (black grids in the left 2 columns) in 7 regions of interest. **Left columns**: Each grid shows a 2 × 3 factorial model space. The models either employed a 2D Gaussian distribution or a Difference of Gaussians (DoG). For each type there were three variants: Circular (Model 1), Elliptical (Model 2) or Elliptical with Rotation (Model 3). The shades of grey/white and labels indicate the Protected Exceedance Probability (PXP), which is the probability that a model is better than all the others being tested. **Right columns:** The summed pRFs of the winning model within each region (see text for details). Colours are scaled within each plot individually, representing arbitrary units of neuronal activity. Region identities are detailed in Section 2.8.8 and illustrated in Fig. 2.

3.5.1. Whole volume analysis

Fig. 9A shows the expected probability of each model. If a voxel were picked at random, these are the probabilities of each model having generated that data. The model with the strongest probability was the DoG Model 1 – that is, a model with circular receptive field and excitatory/inhibitory dynamics ($p = 0.42$). This means that, if choosing a voxel at random, there would be a 42% chance that DoG model 1 would be the best model. The next best model was the single Gaussian Model 1 with probability 0.30 and the third best was the elliptical and rotated DoG Model 3 with probability 0.18. The remaining models all had much weaker probability. Fig. 9B shows the Protected Exceedance Probability (PXP), which is the probability that any one model is better than all other models in the comparison. With probability approaching certainty, DoG Model 1 was the winning model, and therefore the best explanation for any randomly chosen voxel's data. An example timeseries fitted by DoG Model 1 is shown in Fig. 8C. We also took the opportunity to test whether the single Gaussian model and DoG model gave similar estimates of the pRF location. We correlated the expected values of the Gaussian and DoG models' location parameters, converted to Cartesian coordinates. They were highly consistent, with a Pearson's correlation of 0.96 in the x-coordinate and 0.97 for the y-coordinate (Supplementary Fig. 3). In summary, this comparison demonstrates that the circular DoG model provides a better explanation for the data than the single Gaussian model, and that having elliptical and angled pRFs does not, at the level of the whole visual cortex, contribute enough to the model evidence to outweigh the added complexity they induce.

3.5.2. Regions of interest analysis

We next investigated whether the 6 competing pRF models had different explanatory power in different brain regions. We repeated the model comparison within each of 7 regions of interest, using manually delineated visual field maps.

Our first question was whether there was evidence for an excitatory-inhibitory receptive field, captured by the DoG models. While DoG models were the best explanation for the data overall (see above) we found variability among the visual fields (Fig. 10, black grids). In V1 and V2d, a circular DoG model best explained the data, whereas single Gaussian models were preferred in right V2v, left V3v, left V3a and bilateral V4. The results in left V2v, bilateral V3d, right V3v and right V3a were not conclusive with no clear winning model, which may indicate heterogeneous populations of responses in these regions. We conclude from this analysis that the DoG models had the most consistent evidence at the lowest level of the visual hierarchy (V1), whereas the evidence was mixed in V2, and higher visual regions (V3,V4) generally preferred a simpler Gaussian model.

Our second question regarded the shape of the receptive field. In most regions (V1, V2d, V2v, V3d, left V3a, V4), the model with a simple circular receptive field (Model 1) was the clear winner. Therefore, modelling an elliptical shape and rotation did not improve the model fit sufficiently to outweigh the added complexity. The clear exception was right V3a, where the two strongest models were angled DoG model ($p = 0.39$) and the angled Gaussian model ($p = 0.16$). In right V3v, the evidence was divided between the simple Gaussian model and the angled DoG model ($p = 0.26$).

Having identified the best model in each region, we next visualised the area of the visual field covered by these models' receptive fields. For each ROI, we summed the (posterior) receptive field of the overall winning model across voxels, to form coverage maps (Fig. 10, right columns). This confirmed that each pRF responded primarily to the contralateral side of retinotopic space, with dorsal regions responding to the lower visual field and ventral regions responding to the upper visual field. Regions V3a and V4 spanned the contralateral upper and lower visual fields. To quantify the angle estimates in right V3a, we averaged the parameters across voxels from the best model (DoG with rotation) using Bayesian Parameter Averaging. This gave a probability distribution representing the average correlation coefficient (rotation parameter), with an expected value of −0.10 and probability of being non-zero $p = 1.00$. This represents a slight rotation towards the fovea.

Finally, we performed an analysis to investigate whether similar results would be found by fitting pRF models to the mean timeseries of each region, rather than fitting a pRF model to every voxel independently (Supplementary Fig. 4). We did not have strong hypotheses about the model comparison result, as the model which best explains the average response for an extended region is not necessarily the best model for explaining individual voxels' timeseries. On the one hand, we may expect a simpler model to win, as less data is involved and any nuances of the individual timeseries are lost through averaging. On the one hand, the signal should have a much higher SNR, with observation noise all but eliminated through central limit theorem. Therefore, we might expect the data to support more complex models. We found that every region, with the exception of left V3d, favoured a DoG response function rather than a simple Gaussian. Whereas individual voxels' receptive fields were best described with a circular model, the averaged timeseries in most regions were best explained with an angled and rotated receptive field (DoG Model 3). Despite these differences, the pRF response functions in most regions were remarkably similar to the summed voxel-level response functions, although generally covering a slightly larger area (compare Fig. 10 and Supplementary Fig. 4).

4. Discussion

In this paper we introduced a Bayesian framework for specifying, estimating and comparing pRF models. The framework is generic and so can be applied to stimulus spaces of any dimension; here we evaluated it for mapping 2D visual receptive fields. Having validated the method using simulations and evaluated its test-retest reliability, we demonstrated how it may be used to test hypotheses about neuronal response functions, at the level of individual voxels and within predefined regions of interest.

A key advantage of this approach is that the uncertainty of each parameter is quantified by computing the full parameter covariance matrix. This gave rise to a striking observation – there was a strong correlation between the neuronal scaling parameter l_β and the pRF width parameter l_σ. In other words, increasing either parameter would, to a certain extent, give the same change in the modelled BOLD response, which limits the precision with which either parameter can be estimated. An appropriate way of dealing with this is to take account of the uncertainty caused by these correlations when making inferences. Here, we did this by plotting the pRF's posterior predictive density (PD) to visualise the receptive field, which accounts for the uncertainty in the pRF position and width. This made clear that with the relatively high SNR data we used, which had been scanned at 7 T and averaged over 8 runs, the receptive field was estimated with high precision despite the covariance between parameters. It will be interesting to see how the precision of the parameters vary when acquiring empirical data at lower field strengths. As suggested by our simulations, with reduced SNR, taking account of uncertainty becomes increasingly important.

In the generative model implemented here, we included an established model of neurovascular coupling and BOLD response. Whereas most pRF studies use a canonical haemodynamic response function, which may be tailored on a subject-by-subject basis, we estimated parameters of the haemodynamic model simultaneously with the pRF parameters, on a voxel-wise basis. (Readers familiar with Dynamic Causal Modelling will note that we were effectively estimating a single region DCM for every voxel.) This comes with a computational cost – estimation takes longer than standard approaches to pRF estimation. However, our results demonstrate the utility of this voxel-wise model. There was variability in the peak and the undershoot of the BOLD response across voxels (Fig. 3), which would not be modelled using a fixed BOLD response across voxels. A model comparison demonstrated that the accuracy gained by estimating the haemodynamic parameters outweighed their complexity cost. Reassuringly, the parameters of the

haemodynamic model were only weakly correlated with the parameters representing the pRF location and width (Fig. 5), meaning that neuronal and haemodynamic contributions to the BOLD timeseries could be separated with high confidence.

We demonstrated testing hypotheses about pRF response function or shape by embodying each hypothesis as a model and comparing the evidence for each. It has previously been stated that "pRF models … are not centered on statistical hypothesis testing … Rather, the strategy follows the constructivist philosophy of creating models to account for an increasingly large range of stimuli in many visual areas." (Wandell and Winawer, 2015, p. 350). Developing pRF models that are as general as possible requires identifying those which offer the best trade-off between accuracy and complexity. A penalty for complexity (also called an Occam factor) is required when comparing models, otherwise the most complex model will always be considered the best (explain the most variance), at the risk of overfitting the noise and failing to generalise. Finding the best of several models in terms of balancing accuracy and complexity is achieved by identifying the model with the highest model evidence, a procedure referred to as Bayesian model comparison. Here we used an approximation of the log model evidence, the free energy, to score models. Unlike other approximations such as the AIC and BIC, the free energy takes into account the full covariance among parameters (Penny, 2012), meaning that independent parameters add more to the model's complexity than parameters which are correlated.

By comparing the free energy of different models (Bayesian model comparison), we asked whether the response in each voxel was best described as a single multivariate normal distribution, or as a Difference of Gaussians (DoG) model with excitatory centre and inhibitory surround (Rodieck, 1965; Zuiderbaan et al., 2012). The DoG model had previously been found to explain more variance in V1, V2 and V3, but offered little or no improvement in higher level regions V3a, hV4 and LO1 (Zuiderbaan et al., 2012). Similarly, we found DoG models to be better than single Gaussian models in V1 and V2. In several higher level regions (V3v, left V3a, V4), we found strong evidence for the simpler single Gaussian model. The fact that Zuiderbaan et al. (2012) found little or no difference in explained variance in these higher regions speaks to the notion that the model evidence will favour the simpler model unless there are sufficient gains in accuracy to outweigh the added complexity.

We found that the single Gaussian pRF model, coupled with the Balloon model of BOLD response, failed to capture the majority of the post-stimulus undershoot in the BOLD response (Fig. 4C). These parts of the timeseries appear to have fitted slightly better using a DoG model, and may coincide with the inhibitory activity in centre-surround dynamics (Fig. 9C), although a noticeable post-stimulus undershoot remains unmodelled suggesting this is not a sufficient explanation for the data. A recently proposed upgrade to the BOLD response model we used here offers a more physiologically plausible representation of the post-stimulus undershoot (Havlicek et al., 2015), which could be included in our framework. Model comparisons could then be performed to ensure that there is still strong evidence for a DoG neuronal response function, in the context of a haemodynamic model which better explains the BOLD undershoot.

Our second question regarded the shape of the receptive field. We compared three receptive field shapes – circular, elliptical and elliptical with rotation – and tested the evidence for models with these specifications in sub-regions of the brain. In most regions, a simple circular model was the winner - meaning the data could not support a more complex explanation involving an elliptical shape or rotation. We would not, however, wish to draw strong conclusions from these region-of-interest analyses, as we were only using data from a single subject and we only selected these particular models to demonstrate the model comparison framework. With the analyses presented here we hope to have demonstrated that pRF models can be formerly compared based on their model evidence, to address interesting hypotheses. Of the models we tested, we found the most parsimonious (simplest) model that explained the most data was a Difference of Gaussians model with a circular (isotropic) receptive field (Fig. 9).

A potential drawback of the model fitting approach used here is that it involves approximations. With variational Bayes, the true multivariate probability density over the parameters is approximated using simpler distributions – here, the multivariate normal distribution. This is referred to as variational Bayes under the Laplace approximation. (The probability density of the parameters should not be confused with the spatial distribution of the parameters' expected values across the brain – we used a mass-univariate approach which makes no assumptions about spatial distribution). Sampling methods have been used to validate the Laplace approximation in the context of linear models (Friston et al., 2007) as well as with Dynamic Causal Modelling (DCM) for fMRI (Chumbley et al., 2007), which is a connectivity modelling framework that shares the same basic structure, haemodynamic Balloon model and model estimation methods as we have used here. Based on this previous validation work, we had some confidence that the Laplace approximation was appropriate here, further evidenced by the pRF parameter maps recapitulating the expected features of the visual fields (Fig. 2). Nonetheless, a quantitative evaluation of the Laplace assumption in the context of pRF mapping would be ideal, which would usually be addressed through the use of sampling methods.

Our validation plan follows the typical development of methods in SPM. With this first paper, we have established face validity (using simulations) and predictive validity (using delineation of known visual fields with empirical data). The next step will be to establish construct validity, by comparing our results to extant pRF modelling tools. A recent study (Quax et al., 2016) demonstrated successful estimation of pRF parameters in a Bayesian setting using sampling methods (slice sampling MCMC), and applying both approaches to the same data would provide construct validation of the variational approximations. Furthermore, recent developments in gradient-based sampling, intended for use with models which have large numbers of parameters, may offer improved performance and stability over established sampling approaches (Sengupta et al., 2016, 2014). These sampling methods could also be used to validate the assumptions of the framework described here.

A more general limitation of pRF models is that they are phenomenological. They are very effective at describing the summed receptive field of a voxel or a brain region, but they give no insight into how this receptive field arises from the underlying neuronal circuitry. Developments in modelling neuronal circuitry using functional imaging data, Dynamic Causal Modelling (Friston, 2003), have come about by iteratively developing models which afford greater model evidence and biological plausibility than their predecessors. By providing a framework for evaluating models, we hope to facilitate the development of biologically plausible generative models for pRFs. These are likely to be spatiotemporal models which explain how lateral connections between neuronal populations give rise to the distributed pattern of activity across the cortical sheet. Neural field models (Pinotsis et al., 2012) have been successfully used to explain contrast dependent gain control in visual cortex (Pinotsis et al., 2014) and could in principle be extended from electrophysiological modelling to fMRI.

Future work will be in several directions. Estimation speed of the pRF models was generally over 100 s per voxel (see Section 2.7). As such, we expect this framework to be primarily used in a voxel-wise manner for small regions of interest, or alternatively on summary timeseries (the mean or first principal component) from regions of interest, although the latter is not guaranteed to give identical results (Supplementary Fig. 1). However, there may be opportunities for software optimisation, for instancing removing the overhead of converting between polar and Cartesian coordinates. In the meantime, we have addressed the performance issue by providing software tools for distributed model estimation over parallel processors (Appendix A), and in future this could be extended to take advantage of massively parallel GPU computing, which has already been demonstrated in the context of Dynamic Causal Modelling (Aponte et al., 2016). In terms of applications, the main focus of our ongoing efforts will be to model higher dimensional spaces. The

generality of the approach makes it ideally suited to mapping high dimensional or abstract stimulus spaces onto the brain, with any number of model parameters. We hope that the tools we have implemented (Appendix A) will be useful to researchers in a variety of fields.

Funding sources

The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust 091593/Z/10/Z.

Appendix A. Software implementation

We have developed a suite of software tools for specifying, estimating and comparing pRF models. These are Matlab functions which depend on the open-source SPM software package (http://www.fil.ion.ucl.ac.uk/spm/), and use the same underlying algorithms as Dynamic Causal Modelling (Friston, 2003). This toolbox forms the reference implementation of the Bayesian framework we have described, although we hope to have given sufficient detail in the manuscript that it could be re-implemented in other software packages. The BayespRF toolbox is available from https://github.com/pzeidman/BayespRF.

The toolbox is structured around a series of operators which act upon pRF models. A pRF model, together with one or more timeseries, is represented as a Matlab structure and is stored in a file with the format PRF_<name>.mat. A typical workflow proceeds as follows. One creates a 3D image mask indicating which voxels to include. These could be all voxels in a brain structure, or all voxels on the cortical surface, or be based on the results of an initial analysis using a general linear model (GLM). Timeseries are then extracted using the VOI extraction function in SPM, which also performs all the necessary pre-processing. The resulting file (VOI_<name>.mat) contains both a summary timeseries of all included voxels, and individual timeseries for each voxel. Next, one specifies a pRF model using **spm_prf_analyse('specify',…)**, which provides options for whether to use a summary timeseries or individual voxels' timeseries. If averaging over runs is required, this will be performed at this stage. The model evidence and parameters may then be estimated using **spm_prf_analyse('estimate',…)**. The results of the estimation are reviewed and projected onto 3D orthogonal projections of the brain using the GUI **spm_prf_review(…)**, which will create parameter maps (as shown in Fig. 3) if they do not already exist.

If there are a large number of voxels to be estimated, it may be more tractable to use parallel computing and estimate pRF models for multiple voxels simultaneously. The toolbox provides two methods for this. The estimation function has an option 'useparfor', which will take advantage of the Mathworks Parallel Computing Toolbox, if available. Alternatively, a function is provided to take a multi-voxel pRF file and split it into multiple pRF files: **spm_prf_analyse('split',…)**. These individual pRF files can be estimated on separate machines, for instance on separate nodes of a cluster computer, and then merged once estimation is complete using **spm_prf_analyse('merge',…)**. The results of the merged pRF file may then be reviewed.

Within the pRF model file are stored the priors, posteriors, timeseries data and model configuration. A summary of the most important fields is given in Table A1. Users wishing to design new pRF models, or modify the current ones, may wish to take advantage of the **spm_prf_editor** tool, which provides a GUI for manipulating the parameters of a model.

Table A1
Key fields within the Matlab structure representing a pRF model

Field	Description
M.pE	Prior mean (expectation) of each parameter
M.pC	Prior covariance of the parameters
M.IS	Name of the pRF model function to use
Y.y	Matrix of timeseries data against which to estimate the model
xY.XYZmm	Coordinates (mm) of each timeseries in the brain
U	[1 x T] structure representing the T stimuli displayed during scanning, including onsets, durations and coordinates of stimulated pixels
Ep	Estimated parameters for each voxel
Cp	Estimated covariance of the parameters for each voxel
Eh	Estimated log precision of the noise
F	Variational free energy (approximate evidence) of the model for each voxel

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2017.09.008.

References

Aponte, E.A., Raman, S., Sengupta, B., Penny, W.D., Stephan, K.E., Heinzle, J., 2016. Mpdcm: a toolbox for massively parallel dynamic causal modeling. J. Neurosci. Methods 257, 7–16. http://dx.doi.org/10.1016/j.jneumeth.2015.09.009.

Azzopardi, P., Cowey, A., 1993. Preferential representation of the fovea in the primary visual cortex. Nature 361, 719–721. http://dx.doi.org/10.1038/361719a0.

Beal, M.J., Ghahramani, Z., 2003. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. Bayesian Stat. 7 http://dx.doi.org/10.1152/jn.00664.2010.

Buxton, R.B., Uludag, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. Neuroimage 23 (Suppl. 1), S220–S233. http://dx.doi.org/10.1016/j.neuroimage.2004.07.013.

Chumbley, J.R., Friston, K.J., Fearn, T., Kiebel, S.J., 2007. A Metropolis-Hastings algorithm for dynamic causal models. Neuroimage 38, 478–487. http://dx.doi.org/10.1016/j.neuroimage.2007.07.028.

Constantinescu, A.O., O'Reilly, J.X., Behrens, T.E.J., 2016. Organizing conceptual knowledge in humans with a gridlike code. Sci. (80-.) 352, 1464–1468. http://dx.doi.org/10.1126/science.aaf0941.

Dumoulin, S.O., Wandell, B.A., 2008. Population receptive field estimates in human visual cortex. Neuroimage 39, 647–660. http://dx.doi.org/10.1016/j.neuroimage.2007.09.034.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. Neuroimage 34, 220–234. http://dx.doi.org/10.1016/j.neuroimage.2006.08.035.

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. Neuroimage 56, 2089–2099. http://dx.doi.org/10.1016/j.neuroimage.2011.03.062.

Friston, K.J., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302. http://dx.doi.org/10.1016/S1053-8119(03)00202-7.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.

Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. Magn. Reson. Med. 39, 41–52. http://dx.doi.org/10.1002/mrm.1910390109.

Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., Uludag, K., 2015. Physiologically informed dynamic causal modeling of fMRI data. Neuroimage. http://dx.doi.org/10.1016/j.neuroimage.2015.07.078.

He, D., Mo, C., Wang, Y., Fang, F., 2015. Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. Exp. Brain Res. 233, 3535–3541. http://dx.doi.org/10.1007/s00221-015-4425-3.

Heinzle, J., Koopmans, P.J., den Ouden, H.E.M., Raman, S., Stephan, K.E., 2016. A hemodynamic model for layered BOLD signals. Neuroimage 125, 556–570. http://dx.doi.org/10.1016/j.neuroimage.2015.10.025.

Holmes, G., 1945. Ferrier Lecture: the organization of the visual cortex in man. Proc. R. Soc. Lond. Ser. B 361, 2239–2259. http://dx.doi.org/10.1098/rstb.2006.1832.

Hubel, D.H., Wiesel, T.N., 1959. Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148, 574–591. http://dx.doi.org/10.1113/jphysiol.2009.174151.

Kumar, S., Penny, W., 2014. Estimating neural response functions from fMRI. Front. Neuroinform. 8, 48. http://dx.doi.org/10.3389/fninf.2014.00048.

Levin, N., Dumoulin, S.O., Winawer, J., Dougherty, R.F., Wandell, B.A., 2010. Cortical maps and white matter tracts following long period of visual deprivation and retinal image restoration. Neuron 65, 21–31. http://dx.doi.org/10.1016/j.neuron.2009.12.006.

Merzenich, M.M., Brugge, J.F., 1973. Representation of the cochlear partition on the superior temporal plane of the macaque monkey. Brain Res. 50, 275–296. http://dx.doi.org/10.1016/0006-8993(73)90731-2.

Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. 32, 14205–14216. http://dx.doi.org/10.1523/JNEUROSCI.1388-12.2012.

Neyman, J., Pearson, E.S., 1933. On the problem of the most efficient tests of statistical hypotheses. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 231, 289–337. http://dx.doi.org/10.1098/rsta.1933.0009.

Palacci, J., Sacanna, S., Steinberg, A.P., Pine, D.J., Chaikin, P.M., Yartsev, M.M., Ulanovsky, N., 2013. Representation of three-dimensional space in the Hippocampus of flying bats. Sci. (80-.) 340, 367–372. http://dx.doi.org/10.1126/science.1235338.

Penny, W.D., 2012. Comparing dynamic causal models using AIC, BIC and free energy. Neuroimage 59, 319–330. http://dx.doi.org/10.1016/j.neuroimage.2011.07.039.

Pinotsis, D.A., Brunet, N., Bastos, A., Bosman, C.A., Litvak, V., Fries, P., Friston, K.J., 2014. Contrast gain control and horizontal interactions in V1: a DCM study. Neuroimage 92, 143–155. http://dx.doi.org/10.1016/j.neuroimage.2014.01.047.

Pinotsis, D.A., Moran, R.J., Friston, K.J., 2012. Dynamic causal modeling with neural fields. Neuroimage 59, 1261–1274. http://dx.doi.org/10.1016/j.neuroimage.2011.08.020.

Quax, S.C., van Koppen, T.C., Jylänki, P., Dumoulin, S.O., van Gerven, M.A.J., 2016. Slice-sampled Bayesian PRF Mapping bioRxiv).

Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J., 2014. Bayesian model selection for group studies - Revisited. Neuroimage. http://dx.doi.org/10.1016/j.neuroimage.2013.08.065.

Rodieck, R.W., 1965. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vis. Res. 5, 583–601. http://dx.doi.org/10.1016/0042-6989(65)90033-7.

Sengupta, B., Friston, K.J., Penny, W.D., 2016. Gradient-based MCMC samplers for dynamic causal modelling. Neuroimage 125, 1107–1118. http://dx.doi.org/10.1016/j.neuroimage.2015.07.043.

Sengupta, B., Friston, K.J., Penny, W.D., 2014. Efficient gradient computation for dynamical models. Neuroimage 98, 521–527. http://dx.doi.org/10.1016/j.neuroimage.2014.04.040.

Silson, E.H., Chan, A.W.-Y., Reynolds, R.C., Kravitz, D.J., Baker, C.I., 2015. A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. J. Neurosci. 35, 11921–11935. http://dx.doi.org/10.1523/JNEUROSCI.0137-15.2015.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., Friston, K.J., 2009. Bayesian model selection for group studies. Neuroimage 46, 1004–1017. http://dx.doi.org/10.1016/j.neuroimage.2009.03.025.

Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007. Comparing Hemodynamic Models with DCM. http://dx.doi.org/10.1016/j.neuroimage.2007.07.040.

Wandell, B.A., Winawer, J., 2015. Computational neuroimaging and population receptive fields. Trends Cogn. Sci. http://dx.doi.org/10.1016/j.tics.2015.03.009.

Zuiderbaan, W., Harvey, B.M., Dumoulin, S.O., 2012. Modeling center – surround configurations in population receptive fields using fMRI. J. Vis. 12, 1–15. http://dx.doi.org/10.1167/12.3.10.Introduction.