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.