Estimating Cortical Feature Maps with Dependent Gaussian Processes

Nicholas J. Hughes and Geoffrey J. Goodhill

Abstract—A striking example of brain organisation is the stereotyped arrangement of cell preferences in the visual cortex for edges of particular orientations in the visual image. These "orientation preference maps" appear to have remarkably consistent statistical properties across many species. However fine scale analysis of these properties requires the accurate reconstruction of maps from imaging data which is highly noisy. A new approach for solving this reconstruction problem is to use Bayesian Gaussian process methods, which produce more accurate results than classical techniques. However, so far this work has not considered the fact that maps for several other features of visual input coexist with the orientation preference map and that these maps have mutually dependent spatial arrangements. Here we extend the Gaussian process framework to the multiple output case, so that we can consider multiple maps simultaneously. We demonstrate that this improves reconstruction of multiple maps compared to both classical techniques and the single output approach, can encode the empirically observed relationships, and is easily extendible. This provides the first principled approach for studying the spatial relationships between feature maps in visual cortex.

Index Terms—Gaussian processes, multitask learning, neuroimaging, visual cortical maps

1 Introduction

The detailed statistical analysis of structures in the brain relies critically upon accurate reconstructions of those structures from noisy and incomplete neuroimaging data. Robust methods for performing such estimation are therefore of vital importance in neuroscience. A prime example of this problem is in the study of topographic maps in the primary visual cortex (V1) of many mammals. Here neurons are selective for the orientation of edges and contours in the visual field [1] and are arranged spatially such that cells with similar orientation preferences are located nearby. The overall structure of these orientation preference (OP) maps is quite stereotypical [2], [3] and common features of the statistical structure of these maps have been discovered (e.g., [4]). Intrinsic signal optical imaging is the standard technique for wide-area imaging these maps [5], [6], however it produces very noisy data.

Recently an important new method was introduced for improving the estimates of OP maps from such data, using a Bayesian approach based on Gaussian processes (GP) [7]. This utilises a prior distribution over maps which captures the important lower-order statistics of map structure. The technique has been extended to analyse maps from animals reared in abnormal visual environments, and is a powerful framework for investigating the mathematical nature of map structure and its plasticity in the face of changed visual environments [8]. However, so far this work has ignored the fact that OP maps do not exist in isolation, but rather coexist in V1 with maps for several other features of the visual input. These include the eye of origin (ocular dominance, OD) [9], [10], and the spacing of patterns and textures (spatial frequency, SF) [11], [12], [13]. These maps are dependent on one another through complex spatial relationships [11], [14], [15], [16] (Fig. 1). While these relationships have been described empirically, and have been reproduced by models of map development [17], [18], a formal approach to describe the spatial dependencies between maps is currently lacking.

Here we propose a rigorous mathematical framework for both the joint estimation of these maps from imaging data, and for studying the nature and plasticity of the spatial relationships between maps, by extending the Gaussian process approach to consider multiple maps simultaneously. Critically, this involves formulating a joint prior for OD and OP maps, and thus constitutes the first formal description of the underlying relationship between these two maps. This allows for the correlations in map structure across different features to be taken into account, to both improve map estimates and to study their relationships. To test this method we apply it to maps sampled from our joint prior and then corrupted with noise, and demonstrate its superiority both to the classical vector averaging approach to map estimation, and to the model used by [7] which does not take into account the relationships between maps. Finally we demonstrate the extensibility of this model by adding a map for spatial frequency, and use it to investigate the nature of the relationships between all three maps. This work demonstrates the importance and utility of applying multitask learning techniques to the analysis of noisy imaging data.

All source-code for the methods described here are available at github.com/nickjhughes/dgpmaps.

• *The authors are with the Queensland Brain Institute and School of Mathematics and Physics, University of Queensland, St Lucia QLD 4072, Australia.*
E-mail: nicholas.hughes1@uqconnect.edu.au, g.goodhill@uq.edu.au.

Manuscript received 17 July 2015; revised 17 May 2016; accepted 23 Oct. 2016. Date of publication 1 Nov. 2016; date of current version 12 Sept. 2017. Recommended for acceptance by A. J. Storkey.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2016.2624295

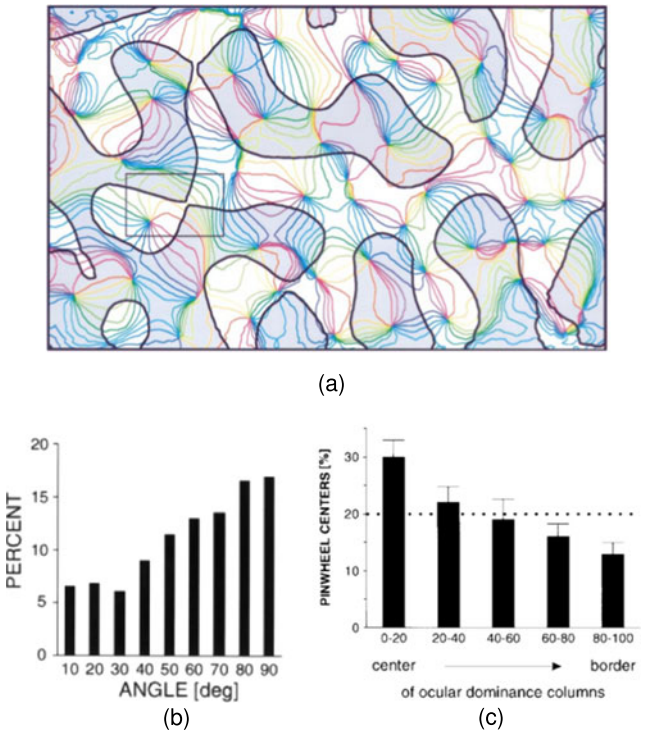


Fig. 1. Topographic features maps in primary visual cortex have stereotypical structures and spatial relationships. (a) Contours of an orientation preference (coloured lines) and an ocular dominance (black lines) map, demonstrating the tendency for their intersection at angles close to perpendicular, and for OP pinwheels to lie far from the OD contours. (b) The distribution of crossing angles of the two maps' contours for a particular pair of maps, quantifying the trend observed in panel a. (c) The relative location of OP pinwheels in the OD map, averaged over maps, quantifying the trend observed in panel a. Figures reproduced from [11].

2 Gaussian Process Regression Map Estimation

Gaussian process regression is a non-parametric regression technique [19] which has been used to estimate orientation preference maps from noisy imaging data [7], [8]. Here we briefly describe the method in general.

We represent an orientation preference map as a two-dimensional complex field $\vec{m}_{\text{op}} : \mathbb{R}^2 \to \mathbb{C}$, mathematically equivalent to a two-dimensional vector field, where the preferred orientation and the strength of that preference at each point are given by half the argument and the magnitude, respectively, of the complex field. The argument is halved as orientation is periodic in π rather than 2π radians. We represent an ocular dominance map as a two-dimensional scalar field $m_{\text{od}} : \mathbb{R}^2 \to \mathbb{R}$, where zero represents an equal preference for the two eyes. We also represent a spatial frequency map as a two-dimensional scalar field $m_{\text{sf}} : \mathbb{R}^2 \to \mathbb{R}$.

Let $m : \mathbb{R}^2 \to \mathbb{R}$ be a two-dimensional scalar field representing a feature map, such as an OD map or one of the vector components of an OP map, as above. Let $r_i : X \to \mathbb{R}$ be one of N noisy observations of m on a rectangular grid of points $X \subset \mathbb{R}^2$, for example the recorded response to a single presentation of a stimulus in a functional imaging experiment. We assume that we can write $r_i(x) = m(x) + \epsilon_i(x)$, $x \in X$, i.e., we observe the true underlying map corrupted by some possibly correlated Gaussian noise.

A GP prior is a distribution over maps, defined by a mean function, which we assume to be zero here for simplicity, and a covariance function K which specifies the auto-covariance between points in the map. For example this could be a difference-of-Gaussians function for the auto-covariance of an OP map [7], [8].

To perform GP regression, the prior distribution defined at the observed points X is conditioned on the data to give a posterior distribution, the mean of which is the Bayes optimal estimate of the underlying map m. The prior on the set of observed points is $p(m) = \mathcal{N}(m; 0, \Sigma_{\text{prior}})$, $\Sigma_{\text{prior}} = K(X, X)$, where $K(X, X)$ denotes the matrix formed by evaluating the covariance function K on all pairs of points in X. This prior will have hyperparameters, namely the parameters of K, which can be fit in a number of ways (see below).

According to Bayes' rule, the posterior distribution is proportional to the product of the prior and the likelihood,

$$p(m|r_1, \ldots, r_N) \propto p(m)p(r_1, \ldots, r_N|m).$$

The likelihood of a single trial is $p(r_i|m) = \mathcal{N}(r_i; m, \Sigma_\epsilon)$, where Σ_ϵ is the noise covariance matrix, which can be inferred from the data in a variety of ways (see below). The likelihood of all trials is the product of the individual likelihoods, assuming each is an independent observation of the true map

$$\begin{aligned} p(r_1, \ldots, r_N|m) &= \prod_{i=1}^{N} \mathcal{N}(r_i; m, \Sigma_\epsilon) \\ &\propto \mathcal{N}\left(m; \frac{1}{N}\sum_{i=1}^{N} r_i, \frac{1}{N}\Sigma_\epsilon\right). \end{aligned}$$

As the likelihood and prior are both Gaussian, so is the posterior,

$$\begin{aligned} p(m|r_1, \ldots, r_N) &\propto \mathcal{N}(m; \mu_{\text{post}}, \Sigma_{\text{post}}), \\ \Sigma_{\text{post}} &= (N\Sigma_\epsilon^{-1} + \Sigma_{\text{prior}}^{-1})^{-1}, \\ \mu_{\text{post}} &= \Sigma_{\text{post}}\Sigma_\epsilon^{-1}\sum_{i=1}^{N} r_i. \end{aligned}$$

The mean of the posterior μ_{post} is the best estimate of the underlying map m, and the covariance Σ_{post} gives an indication of error in the fit, with its diagonal giving the variance in the estimate at each point.

We used the same method as [7] to infer the noise covariance matrix Σ_ϵ when estimating maps. Briefly, the covariance between repeated responses to each stimulus was calculated and then averaged across stimuli to construct an initial estimate of the noise covariance. An iterative process of deriving the posterior mean and fitting the noise covariance to the residuals of the posterior mean and the data was then performed, with the fitting performed with a factor analysis expectation maximisation algorithm at each step. As there are more pixels in the maps than trials, the full noise covariance cannot be inferred from the data, and so the noise covariance is assumed to be low rank and of the form $\Sigma_\epsilon = G_\epsilon G_\epsilon^\top + D_\epsilon$, where G_ϵ is of size $n \times q$, where n is the number of pixels in the map and q is the rank of the noise, and D_ϵ is diagonal. We used $q = 2$ throughout, and assumed the noise covariance was the same for each feature map in the joint model.

Sampling maps from a GP prior is performed by multiplying the Cholesky factor of the covariance matrix by a

vector of random numbers sampled from the standard normal distribution [19].

3 Multiple Output Gaussian Processes

When applied to orientation preference maps by [7] and [8], the GP map estimation method considered two outputs, namely the two vector components of the OP map. However, they were treated independently as the two components are uncorrelated in all the cases considered. In contrast, in a case where there are multiple correlated outputs, for example the spatially dependent OP and OD feature maps in primary visual cortex, a full GP which considers all outputs simultaneously must be used. Here we construct such a full correlated GP model for multiple feature maps.

In general, let $m_i : \mathbb{R}^2 \to \mathbb{R}$, $1 \leq i \leq N$, be a collection of N two-dimensional scalar fields representing feature maps, with corresponding auto-covariance functions $K_{i,i}$ and cross-covariance functions $K_{i,j}$. For simplicity we assume the fields have zero mean. Three of these maps could represent, for example, an OP map $\vec{m}_{\text{op}} = m_1 + im_2$ and an OD map $m_{\text{od}} = m_3$. A joint GP prior considering this entire collection of maps, which are defined on the two-dimensional grid of points $X \subset \mathbb{R}^2$, is defined by the block covariance matrix

$$\Sigma_{\text{prior}} = \begin{pmatrix} K_{1,1}(X,X) & \dots & K_{1,N}(X,X) \\ \vdots & \ddots & \vdots \\ K_{N,1}(X,X) & \dots & K_{N,N}(X,X) \end{pmatrix}.$$

With this covariance function defined, GP regression is performed as for a single output. However, Σ_{prior} is a valid covariance matrix if and only if it is positive semidefinite, that is, if and only if $z^\top \Sigma_{\text{prior}} z \geq 0$ for all column vectors z of the appropriate size [19]. In general, a block matrix formed from smaller covariance matrices is not itself positive semidefinite, but must have greater restrictions placed on its blocks. For example, a real matrix $X = \begin{pmatrix} A & B \\ B^\top & D \end{pmatrix}$ is positive semidefinite if and only if both A and $D - B^\top A^{-1} B$ are positive semidefinite [20]. This means that even if the auto-covariance and cross-covariance matrices for a number of maps are valid, the block matrix formed by them is not guaranteed to be a valid covariance matrix. There are several alternate characterisations of positive semidefiniteness (e.g., that all of the eigenvalues of the matrix are non-negative), but it is difficult to define any general restrictions on the cross-covariance functions such that a valid prior is formed [19].

However, rather than define the auto-covariance and cross-covariance functions between outputs in a GP directly, another method is to define them as the result of convolution integrals of filtering kernels which describe the structure of the outputs based on filtering Gaussian noise [21]. This solves the issue of how to define valid cross-covariance functions, and is a convenient and tractable way to extend the GP framework to multiple outputs. Here we briefly describe this method.

Considering a GP with a single output, samples can be taken from it by filtering standard Gaussian noise (i.e., distributed as $\mathcal{N}(0,1)$) with the filtering kernel defined by the auto-covariance function of the GP. Viewed this way, the auto-covariance function of a single output GP can be defined as the result of convolving a filtering kernel $h : \mathbb{R}^2 \to \mathbb{R}$ with itself, $K_{1,1} = h * h$. This can be extended to multiple correlated outputs by defining each output as the sum of the results of filtering different noise sources with different kernels. For example, consider two outputs of a GP defined by $Y_i = U_i + V_i$, $1 \leq i \leq 2$, where the V_i result from filtering independent noise sources with kernels k_i, while the U_i come from filtering a noise source shared by the two outputs with kernels h_i. These two outputs will be correlated through their component which comes from the shared noise source. Defined as convolutions, the auto-covariance and cross-covariance functions for the two outputs are

$$\begin{aligned} K_{1,1} &= k_1 * k_1 + h_1 * h_1, \\ K_{2,2} &= k_2 * k_2 + h_2 * h_2, \\ K_{1,2} &= k_1 * k_2, \\ K_{2,1} &= k_2 * k_1. \end{aligned}$$

The covariance matrix for this multiple output GP is then just the block matrix described above constructed using these covariance functions. A single output GP using this framework is effectively equivalent to the standard single output GP case (e.g., as presented in [19]), but when applied to multiple output GPs solves the issue of how to define cross-covariance functions by guaranteeing that the resulting overall covariance matrix will be valid [21]. The only requirement placed on the kernels is that they be absolutely integrable, which is satisfied by most useful functions.

The filtering kernel method for constructing joint Gaussian process priors requires the evaluation of the convolution integrals which define the covariance functions. This can be done analytically, assuming the integrals result in a closed form [21], however to avoid the tedious task of evaluating several integrals analytically we chose to evaluate the convolutions numerically. That is, all of the kernels were evaluated on a sufficiently large grid for the particular problem being considered and convolved numerically, which was a computationally efficient and simple approach.

The process of deriving the posterior distribution involves the inversion of the prior covariance matrix (see above), which will contain $n^2 = N|X|^2$ entries for a model over N maps and $|X|$ pixels. As naive matrix inversion has a computational complexity of $O(n^3)$, deriving the posterior becomes computationally intractable on modern hardware for even moderately sized maps of more than 100 pixels square. For simplicity we only used maps small enough to avoid this issue, but there are techniques which make the application of our methods to larger maps computational tractable. For example, [7] used a low-rank approximation of the prior covariance matrix to deal with this problem, generated via an incomplete Cholesky decomposition, and the same approach can be applied to the multiple output case. Briefly, the Cholesky factor G is constructed one row at a time until a sufficient level of accuracy or a set rank has been achieved. This results in a triangular matrix G of size $n \times q$, with $q \ll n$, where $GG^\top$ provides an accurate low-rank approximation of the prior covariance matrix if the filtering kernels used drop off as a function of distance sufficiently quickly, which most useful kernels do [22].

TABLE 1
Joint Prior Model Hyperparameters

Kernel(s)	Parameter	Value(s)	Units	Description
G (doG)	α	1	–	Scaling
	σ	2	pixels	Wavelength
G' (doG)	α'	$[0, 2]$	–	Scaling
	σ'	2	pixels	Wavelength
H, V (Gabors)	β	0.2	–	Scaling
	γ	0.75	–	Eccentricity
	ρ	2	pixels	Wavelength
	λ	1.2	pixels	Period
G_{sf} (doG)	α_{sf}	0.25	–	Scaling
	σ_{sf}	4 (2σ)	pixels	Wavelength

The parameters of the filtering kernels which define the covariance functions of the joint prior, the range of values we consider, the units where appropriate, and the purpose of each. Also see Fig. 2. doG: Difference-of-Gaussians (Equation (1)). H, V: horizontal and vertical Gabors (Equations (2) and (3)).

4 A Joint Prior for Multiple Feature Maps

To construct a Gaussian process prior which considers multiple feature maps jointly using this method, we need filtering kernels which describe their structure and can also be used to model their interdependence.

Rojer and Schwartz [23] introduced a purely descriptive model of OP maps, consisting of a pair of two-dimensional fields of Gaussian noise with zero mean and unit variance which are filtered with a band-pass filter and combined to form a two-dimensional vector field. The result, when interpreted as an OP map, matches the lower-order statistics of empirical maps found in several different species. The same model, when applied to a single field of noise, recreates the structure of OD maps well. For the purpose of comparison with other models of OP and OD maps, [17] extended the model of [23] to treat the OD map as a precursor to the two components of the OP map, thereby introducing dependencies between the maps which match the spatial relationships seen in empirical maps. We use these ideas in the derivation of our GP model.

Rather than convert a band-pass filter to the space domain, for simplicity we directly used a difference-of-Gaussians (doG) filter, the same as used by [7] and [8],

$$G(\vec{x}, \vec{x}') = \alpha^2 \left[\exp\left(\frac{-|\vec{x} - \vec{x}'|^2}{4\sigma^2} \right) + \frac{1}{4} \exp\left(\frac{-|\vec{x} - \vec{x}'|^2}{16\sigma^2} \right) - \frac{4}{5} \exp\left(\frac{-|\vec{x} - \vec{x}'|^2}{10\sigma^2} \right) \right], \tag{1}$$

where $\vec{x}$ and $\vec{x}'$ are two points in $\mathbb{R}^2$, and α and σ are parameters controlling the magnitude and wavelength of the filter, respectively (Table 1). A doG filter is a space-domain representation of a band-pass filter, as the two Gaussians are low-pass filters which suppress high-frequency spatial information, and thus the subtraction of one from the other results in only spatial frequencies lying between the two cut off wavelengths. [17] note that any band-pass filter is sufficient to replicate their results.

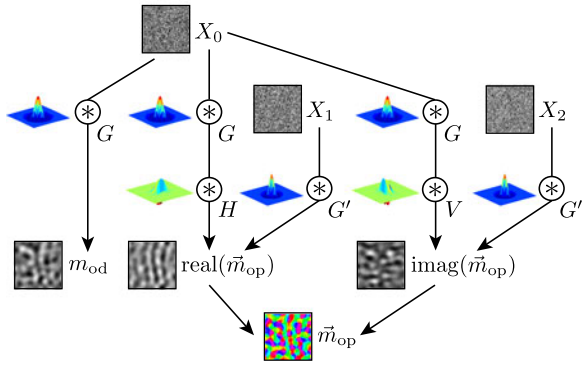


Fig. 2. Multiple output Gaussian process prior for multiple feature maps. Schematic of our multiple output model using the filtering kernel method to construct covariance functions. The independent Gaussian noise sources X_0, X_1, and X_2 are convolved ($*$) with difference-of-Gaussians filtering kernels G and G', and horizontal and vertical Gabor filtering kernels H and V, to generate an ocular dominance map, m_{od}, and the two components, $\text{real}(\vec{m}_{\text{op}})$ and $\text{imag}(\vec{m}_{\text{op}})$, of an orientation preference map $\vec{m}_{\text{op}}$. The relative scaling of the two difference-of-Gaussians filtering kernels controls the dependence of the maps through the proportion of the shared noise source X_0 which influences the orientation preference map.

Erwin et al. [17] define the OP map as a result of vectorially combining the horizontal and vertical directional gradients of the OD map generated as above. The gradient calculation cannot be used directly in the filtering kernel approach, but can be well approximated using a pair of Gabor filters, with a horizontal,

$$H(\vec{x}, \vec{x}') = \beta \exp\left(-\frac{(x_2 - x_2')^2 + \gamma^2 (x_1 - x_1')^2}{2\rho^2} \right) \sin\left(\frac{2\pi (x_2 - x_2')}{\lambda} \right), \tag{2}$$

and vertical

$$V(\vec{x}, \vec{x}') = \beta \exp\left(-\frac{(x_1 - x_1')^2 + \gamma^2 (x_2 - x_2')^2}{2\rho^2} \right) \sin\left(\frac{2\pi (x_1 - x_1')}{\lambda} \right), \tag{3}$$

filter for the derivatives in the x and y directions respectively. Here $\vec{x} = (x_1, x_2)$ and $\vec{x}' = (x_1', x_2')$ are two points in $\mathbb{R}^2$, and β, γ, ρ, λ are parameters controlling the scaling, eccentricity, wavelength, and period of the filters, respectively Table 1. These filters, along with the doG filter given above Equation 1, are absolutely integrable, and so satisfy the requirements of the filtering kernel method. Using these three filters we can replicate the model of [17] in the filtering convolution framework (Fig. 2, X_0 components only), although embedding this in a GP model allows for its use in map estimation and analysis, rather than just construction, and therefore goes significantly beyond the original model.

In principle, there are many ways this model could be extended to allow for a modulation of the dependence between the maps. Here we chose to add a separate component to each of the components of the OP map, so that they are the sum of the Gabor-filtered OD map as above as well as a separate noise source convolved with an independent

doG filter G' (Equation (1) Fig. 2). This allows a continuum of dependence to be modelled, from completely independent, where the shared source of noise X_0 does not affect the OP map, to one of full dependence, where the independent sources X_1 and X_2 do not affect the OP map, equivalent to the simple model described above. The covariance functions of this prior are then the following:

$$\begin{aligned}
K_{1,1} &= G * G \\
K_{1,2} &= G * (G * H) \\
K_{1,3} &= G * (G * V) \\
K_{2,1} &= K_{1,2} \\
K_{2,2} &= (G * H) * (G * H) + G' * G' \\
K_{2,3} &= (G * H) * (G * V) \\
K_{3,1} &= K_{1,3} \\
K_{3,2} &= K_{2,3} \\
K_{3,3} &= (G * V) * (G * V) + G' * G'.
\end{aligned}$$

In general G and G' have different sets of parameters, which we label (α, σ) and (α', σ'), respectively (Table 1). The level of dependence between the two maps is parametrised by the ratio of the two magnitude parameters α/α'. For values larger than one, the maps are more correlated, for values smaller the maps are less correlated, and for a value of zero the maps are completely independent. A summary of the parameters is given in Table 1. This full model, which extends the basic ideas of [17] and [23] to allow for the quantification and control of dependence between the OP and OD maps, therefore allows not only the construction of maps, but also their improved estimation from imaging data by taking advantage of the known correlations between the structure of the maps.

5 Spatial Map Relationships

To confirm that the empirical spatial relationships between the maps are captured by this model, we quantified the relationships with two standard metrics: the distribution of crossing angles of the intersections of the contours of two maps, and the location of the OP map pinwheels relative to the OD map.

5.1 Metrics

To calculate the crossing angle distributions, we calculated the zero-level contour of the OD map, or the 0, 22.5, ... , 157.5 degree contours of the OP map. The angle between the contours of two maps was then calculated at all points of intersection, defined as the difference between the orientations of the tangents of the two contours at the intersection point. The distributions of orientations were then constructed by binning each orientation difference into one of nine bins centred on 5, 15, ... , 85 degree, and the resulting counts normalised to have a sum of one. For a baseline comparison, we also calculated the expected distribution for unrelated maps, by shuffling the pairs of maps sampled from priors and comparing unrelated maps, and making the same number of these comparisons as actual pairs sampled. This distribution should follow the first quarter of a sine curve [24].

To measure the relationship between the location of pinwheels and the OD domains we used the metric described in [11] with a slight modification. [11] binned each OD map into 10 equally sized regions, corresponding to values ranging from the centres of contralateral eye regions to the centres of ipsilateral eye regions, and calculated the bin corresponding to each pinwheel. The bins corresponding to the same distance from each centre were then pooled, leaving five bins corresponding to values from the centre to the border of the OD columns. We used a slightly different approach, by binning the positive and negatives values in each OD map into five equally sized regions, and then performed the same pooling of pinwheel locations. This modification considers the regions on either side of zero separately, rather than binning all the values into 10 equal bins, and therefore respects OD borders and takes into account contralateral bias in the OD map. Pinwheel locations were defined in orientation maps using the contour integration method [4], [25]: an integral of orientation preference was calculated along a closed path of radius one pixel around each pixel, and the pixel was defined as a pinwheel if the integral was $\pm\pi$ radians. For connected clusters of such pixels, the cluster's centre of mass was defined as the pinwheel location.

5.2 Measured Relationships

Figs. 3a and 3b show a pair of maps sampled from this joint prior ($\alpha' = 1$, other parameters as in Table 1). We calculated the crossing angles of the contours of pairs of maps (Fig. 3c, $\alpha' = 1$), and the relationships between the pinwheels and the OD map domains (Fig. 3d, $\alpha' = 1$). The stereotypical tendency for the map contours to cross at higher angles than expected by chance, and the tendency for OP map pinwheels to lie in the centre of OD map domains, are both successfully captured by our model.

5.3 Parametrising Map Interdependence

The design of the OP map in our model, as both resulting from independent sources and the gradient of the OD map (Fig. 2), allows the level of dependence between the maps to be varied. This is controlled by the relative magnitude of the parameters α and α' of the two difference-of-Gaussians filters G and G'.

We sampled maps from the prior with varying values of α', while keeping $\alpha = 1$ constant, and analysed the map relationships as above (Figs. 3c and 3d). A value of $\alpha' = 0$ indicates full dependence, and as α' increases the maps become effectively independent and the spatial relationships disappear. For values above $\alpha' = 2$, the map relationships disappear and become indistinguishable from those between unrelated maps (crossing angles: $p = 0.06$, $\chi^2 = 16.3$; pinwheel locations: $p = 0.43$, $\chi^2 = 4.89$; chi-square tests). These two parameters therefore allow for the fine quantitative control of the level of dependence between the two maps.

6 Estimating Maps

The above results demonstrate that our joint prior captures the empirical spatial relationships between OP and OD maps, and that the level of dependence can be varied. We now apply the joint prior to estimate maps from constructed noisy data to demonstrate the method's superiority to classical map estimation techniques.

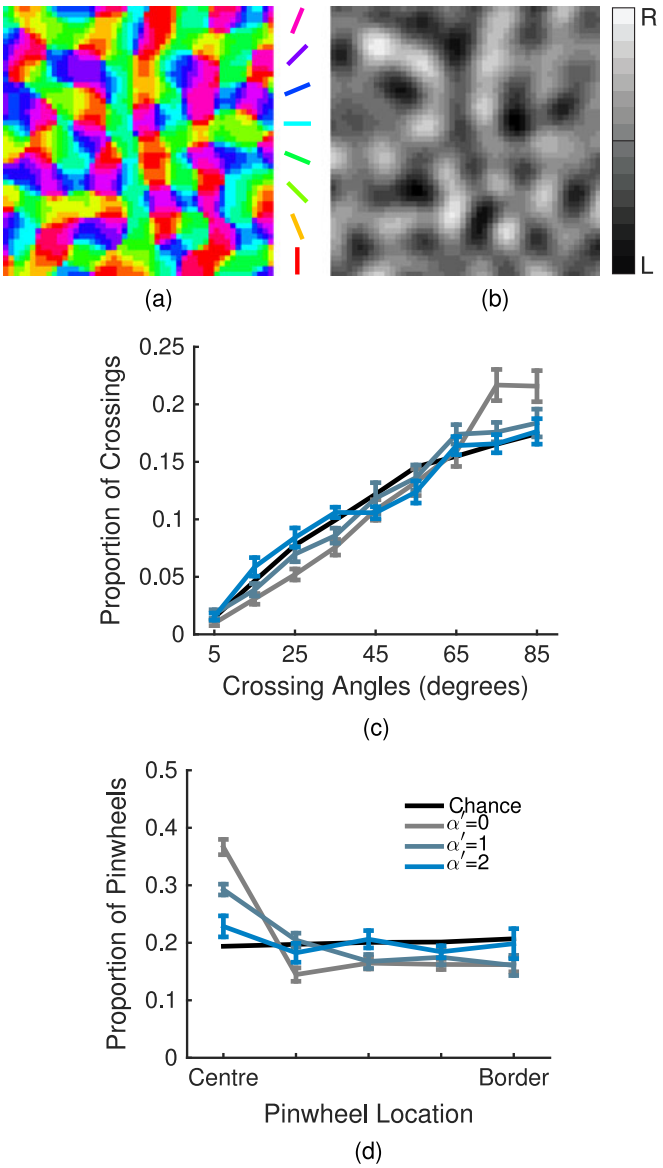


Fig. 3. Maps sampled from the prior have spatial relationships matching empirical maps, and their level of dependence is parametrised. (a–b) An example pair of maps, orientation preference and ocular dominance, sampled from the prior defined by the model given in Fig. 2. Parameters as in Table 1, $\alpha' = 1$. (c) Contour intersection angle distributions as a function of α' (coloured as per legend), showing a preference for angles near perpendicular for small values of the dependence parameter, and a shift towards the chance distribution (black line) for increasing values of the parameter. The $\alpha' = 0$ and 1 ($p = 0.027$, $\chi^2 = 18.8$), and $\alpha' = 0$ and 2 ($p < 5 \times 10^{-7}$, $\chi^2 = 46.8$), distributions are significantly different (chi-square tests). (d) Relative pinwheel location distributions as a function of α' (coloured as per legend), showing a preference for locations near the centre of ocular dominance domains for small values of the dependence parameter, and a shift towards the chance distribution (black line) for increasing values of the parameter. The $\alpha' = 0$ and 1 ($p = 0.016$, $\chi^2 = 14.0$), and $\alpha' = 0$ and 2 ($p < 3 \times 10^{-5}$, $\chi^2 = 28.5$), distributions are significantly different (chi-square tests). Mean and standard error of the mean shown for ten sampled pairs of maps.

6.1 Constructing Noisy Data

We constructed noisy data as a model of experimental data but with known true maps for comparison. To do this we sampled a pair of maps, m_{od} and $\vec{m}_{\text{op}}$, from an instance of the joint prior with a given set of parameters. We then modelled the responses of each pixel in the map to the presentation of an orientation $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ to one of the eyes $e \in \{-1, 1\}$ as

$$r_{\theta,e} = e m_{\text{od}} + \cos(2\theta)\text{real}(\vec{m}_{\text{op}}) + \sin(2\theta)\text{imag}(\vec{m}_{\text{op}}).$$

This was repeated for all combinations of orientations and eyes and each response was replicated $n_{\text{trials}} = 10$ times to simulate multiple trials of the same stimulus. Random, correlated, Gaussian noise was then added to all the trials using the same factor analysis method as [7]. The correlated component of the noise was half the magnitude of the independent component variance σ^2_{noise} in all cases.

6.2 Vector Averaged Map Estimates

Maps were estimated from imaging data by a weighted averaging process, usually referred to as vector averaging for OP maps. Giving the collection of responses $r_{\theta,e,j}$ as defined above, the vector averaged OD map was defined by

$$m_{\text{od}} = \frac{1}{n_{\text{trials}}} \sum_{j=1}^{n_{\text{trials}}} \frac{1}{4} \sum_{\theta} r_{\theta,2,j} - \frac{1}{n_{\text{trials}}} \sum_{j=1}^{n_{\text{trials}}} \frac{1}{4} \sum_{\theta} r_{\theta,1,j},$$

and the OP map by

$$\vec{m}_{\text{op}} = \sum_{\theta} \frac{1}{n_{\text{trials}}} \sum_{j=1}^{n_{\text{trials}}} \frac{1}{2} \sum_{e} r_{\theta,e,j} \cos(2\theta) + i r_{\theta,e,j} \sin(2\theta).$$

To provide a fair comparison with the GP posterior estimates, these maps were then both low-pass and high-pass Gaussian filtered, with wavelengths for each chosen to maximise the correlation between the vector averaged maps and the corresponding true maps.

6.3 Inferring Hyperparameters

To perform GP regression on a dataset the parameters of the prior (Table 1) must be inferred from the data. One method to do this is via maximising the marginal likelihood [8], which is a quantity which specifies how likely the given dataset is to come from a particular prior. Maximising over the hyperparameters then gives the prior most suitable for a particular dataset. However, optimising the marginal likelihood of this joint model is both algorithmically complicated and computationally demanding. We therefore chose to use the simpler approach of directly matching the covariance functions of the prior to the empirical covariance functions of the data [7]. Specifically, we estimated the hyperparameters by calculating the full normalised auto-covariance matrices of the OD map and both components of the OP map, and then fit the values of the parameters by optimising the fit between these empirical auto-covariances and the auto-covariances resulting from the prior. The quality of the fit was measured by the correlation between the auto-covariance matrices, summed over all three map components.

6.4 Prior Sampled Maps

We sampled a pair of maps from the joint prior (Figs. 4a and 4e; parameters as in Table 1, $\alpha' = 1$), and added noise to simulate noisy experimental data as described above (Figs. 4b and 4f; $\sigma^2_{\text{noise}} = 6$). We then estimated the maps by both simple vector averaging and spatial filtering (Figs. 4c and 4g), and by fitting the parameters of the prior to this data as described above and deriving the posterior using this estimated joint prior and the data (Figs. 4d and 4h).

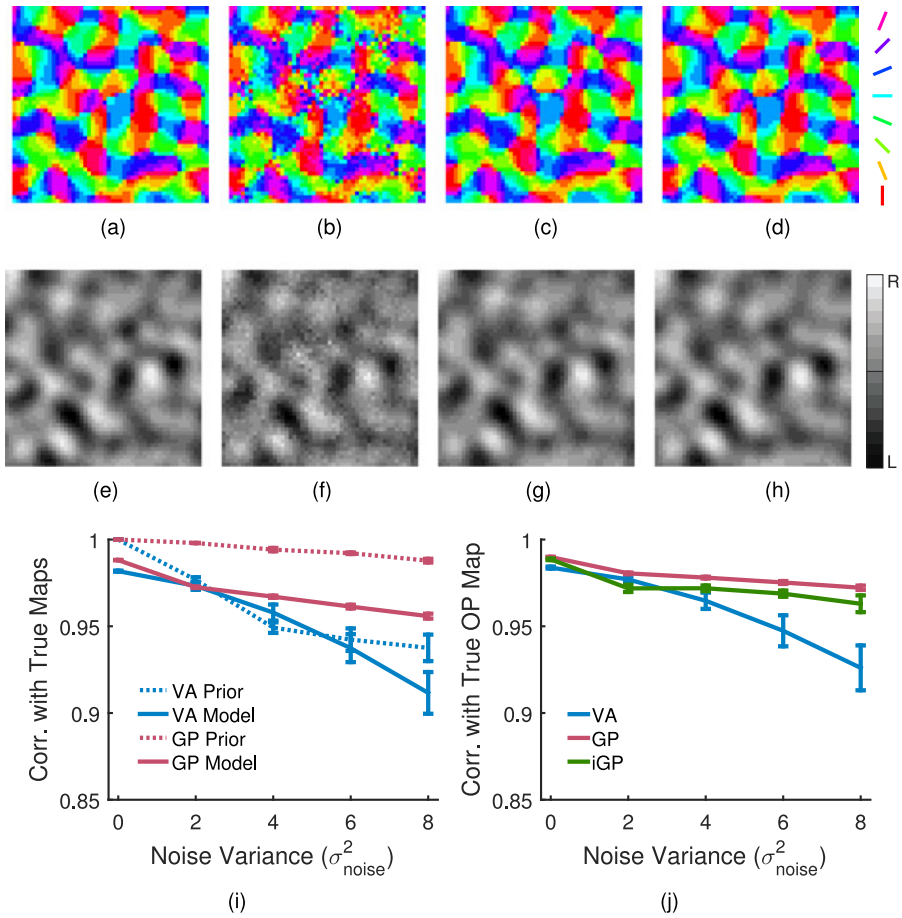


Fig. 4. Estimating maps from noisy data. (a,e) A pair of maps sampled from the joint prior with parameters as in Table 1 and $\alpha' = 1$. (b,f) Noisy observations of the same pair of maps, corrupted with correlated Gaussian noise ($\sigma^2_{\text{noise}} = 6$). (c,g) Optimally filtered vector averaged maps constructed from the noisy observations. Correlation with true maps: 0.95. (d,h) The posterior mean estimates using the joint prior, constructed from the same noisy data as the vector averaged maps. Correlation with true maps: 0.99. (i) The quality of maps estimated with optimally filtered vector averaging (VA, blue) and our joint prior Gaussian process model (GP, red), measured as the correlation with the true maps, as a function of the level of added correlated noise σ^2_{noise}. Results for both maps sampled from the joint prior (dotted) and maps generated with the elastic net model (solid) are shown. The GP estimate is significantly superior in all cases except the noiseless case on maps sampled from the prior (two sample t-tests, $p < 0.001$), and for all noise levels greater than $\sigma^2_{\text{noise}} = 4$ on maps from the model (two sample t-tests, $p < 0.01$). (j) The quality of OP maps generated with the elastic net model, estimated with vector averaging (VA), our joint prior model (GP), and an independent GP model that does not take into account the OD map (iGP), as a function of the level of noise added. The GP estimate is significantly better than the iGP estimate for $\sigma^2_{\text{noise}} = 0, 2, 4, 6$ (two sample t-tests, $p < 0.05$). Mean and standard error of the mean shown for ten different pairs of maps for each level of noise.

The maps estimated via vector averaging were less accurate (correlation with true maps: 0.95) than the maps estimated with the joint GP prior (correlation: 0.99).

To assess the effect of the level of added noise on these results, we performed the above for a range of noise levels ($\sigma^2_{\text{noise}} = 0$ to 8), and calculated the correlation between the vector averaged and posterior estimated maps, and the true maps (Fig. 4i, dotted lines). In each case the joint prior map estimate is significantly better than the optimally filtered vector averaged map, except in the noiseless case $\sigma^2_{\text{noise}} = 0$ as expected (two sample t-tests, $p < 0.001$). The joint prior estimates also show less variance across different map samples with the same level of noise added, demonstrating the technique's robustness.

6.5 Computational Model Maps

By estimating maps sampled from our joint prior, the GP method has an advantage not present when analysing experimental data, in that the form of the prior is known to be a perfect fit to the maps. To confirm that this did not unfairly affect the above results, we performed the same analysis on maps generated by the elastic net model as described in [26], using the same method of adding correlated noise and generating simulated imaging data. The elastic net model is a dimension reduction algorithm for determining how preferences for features of the input space, such as visual field position and edge orientation, are arranged in the cortex [25]. It folds a two-dimensional net in a higher-dimensional feature space in an attempt to minimise a trade-off between coverage of the input space and continuity of the cortical representation. We only included visual field position, orientation preference, and ocular dominance dimensions in our simulations. The parameters and training regime used were as described in [26], except that we generated maps of size 50×50 pixels, and only generated maps with normal visual input. We generated 10 different maps from which simulated noisy experimental data was constructed, as described above. The simulations were performed using the same custom MATLAB software as used by [26], a more general version of which is available online at http://faculty.ucmerced.edu/

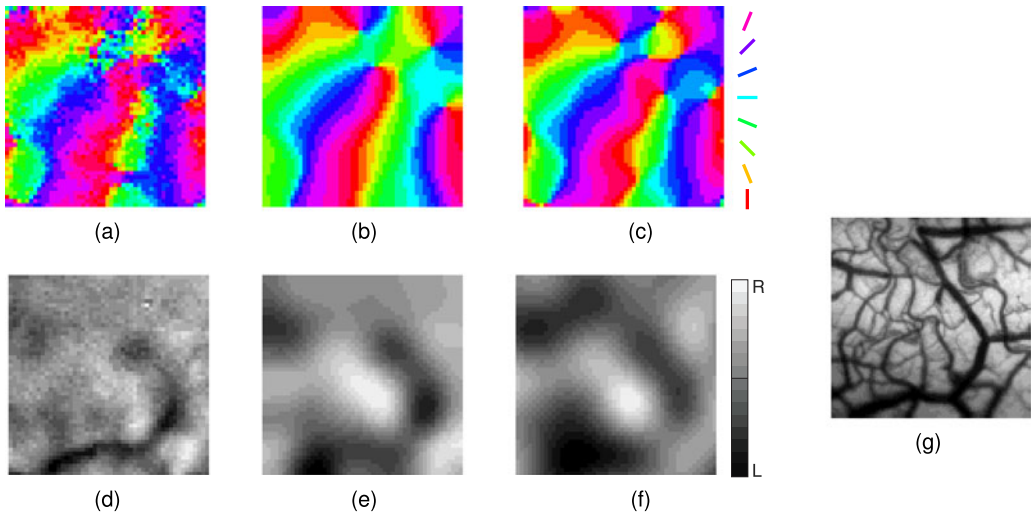


Fig. 5. Estimating maps from experimental data. (a,d) A pair of maps estimated with vector averaging from previously published experimental data [27]. (b,e) Optimally filtered vector averaged maps, filtered to maximise the correlation with the joint prior estimate. (c,f) The posterior mean estimates using the joint prior. (g) The same region of cortex imaged with a different wavelength light to highlight the vasculature patterns present. The vector averaged maps are more strongly affected by artefacts induced by these patterns than the joint prior estimate.

mcarreira-perpinan/research/EN.html. The analysis of these maps show the same trend as for the prior sampled maps above (Fig. 4i, solid lines), with the joint prior map estimation being of a significantly higher quality for all noise levels greater than $\sigma^2_{\text{noise}} = 4$ (two sample t-tests, $p < 0.01$).

6.6 Independent Map Estimates

To assess the improvement in map estimation provided by considering multiple maps simultaneously in our joint prior, as compared to considering each map separately, we also estimated OP maps using an independent model. This model used only the independent difference-of-Gaussians components of the OP map from our joint model (i.e., only the X_1 and X_2 components in Fig. 2), thereby being effectively the same as the model used by [7]. The two parameters of this model, α' and σ', were estimated as described above, with only the OP map being taken into account. We estimated maps from noisy data constructed from maps generated with the elastic net model as a function of added noise (Fig. 4j), and the maps estimated with the joint model were of a significantly higher quality for intermediate levels of noise (two sample t-tests, $p < 0.05$, $\sigma^2_{\text{noise}} = 0, 2, 4, 6$), which are approximately the levels of noise seen in optical imaging data. The difference in quality declines as noise is increased, indicating that when high levels of correlated noise are present in the data, accurately modelling those correlations is more important for quality estimates than taking into account the correlations between the different features maps.

6.7 Experimental Data

To confirm that the joint prior method can be successfully applied to experimental data, we estimated maps from previously published experimental data consisting of intrinsic signal optical images of the primary visual cortex of a single, normally-reared cat [27]. See the original publication for full experimental details. The data consisted of the cortical response to 70 presentations of four different oriented gratings (0, 45, 90, 135 degree) to either of the two eyes. Orientation preference and ocular dominance maps were estimated from this data in the same way as described above.

The unfiltered vector averaged maps are of a low quality (Figs. 5a and 5d), and are therefore usually filtered with arbitrarily chosen parameters (Figs. 5b and 5e). Here we optimally filtered these maps by maximising their correlation with the joint prior estimate. In contrast, the joint prior estimate requires no arbitrary choices, with its parameters chosen optimally as described above (Figs. 5c and 5f). Additionally the estimate is improved by the shared information between the estimation of the OP and OD maps provided by the joint prior. The patterns of vasculature present in the same region of cortex as these maps (Fig. 5g) can be a source of noise artefacts in the estimated maps. Due to the correlated noise model used by the Gaussian process model, the joint prior estimate is less affected by these artefacts that the vector averaged maps.

7 Extension to More Maps

In this framework, it is simple to add more feature maps to the prior, and in theory any number of maps could be represented [28]. While the relationships between the OD and OP maps have been extensively explored experimentally, the relationships between these and other maps in V1 are less clear. To demonstrate the extensibility of this model, we added a third feature map that is present in V1, that for spatial frequency. We did this by defining the SF map as a further filtered version of the OD map, therefore coupling the map with the OD map and by extension also the OP map (Fig. 6a). We used another doG filter G_{sf} for this (Equation (1)), with double the wavelength of the OD map (Table 1), matching the auto-covariance of the SF map and the map relationships proposed by [16]. This adds the following covariance functions to the above model, where the fourth output is the SF map

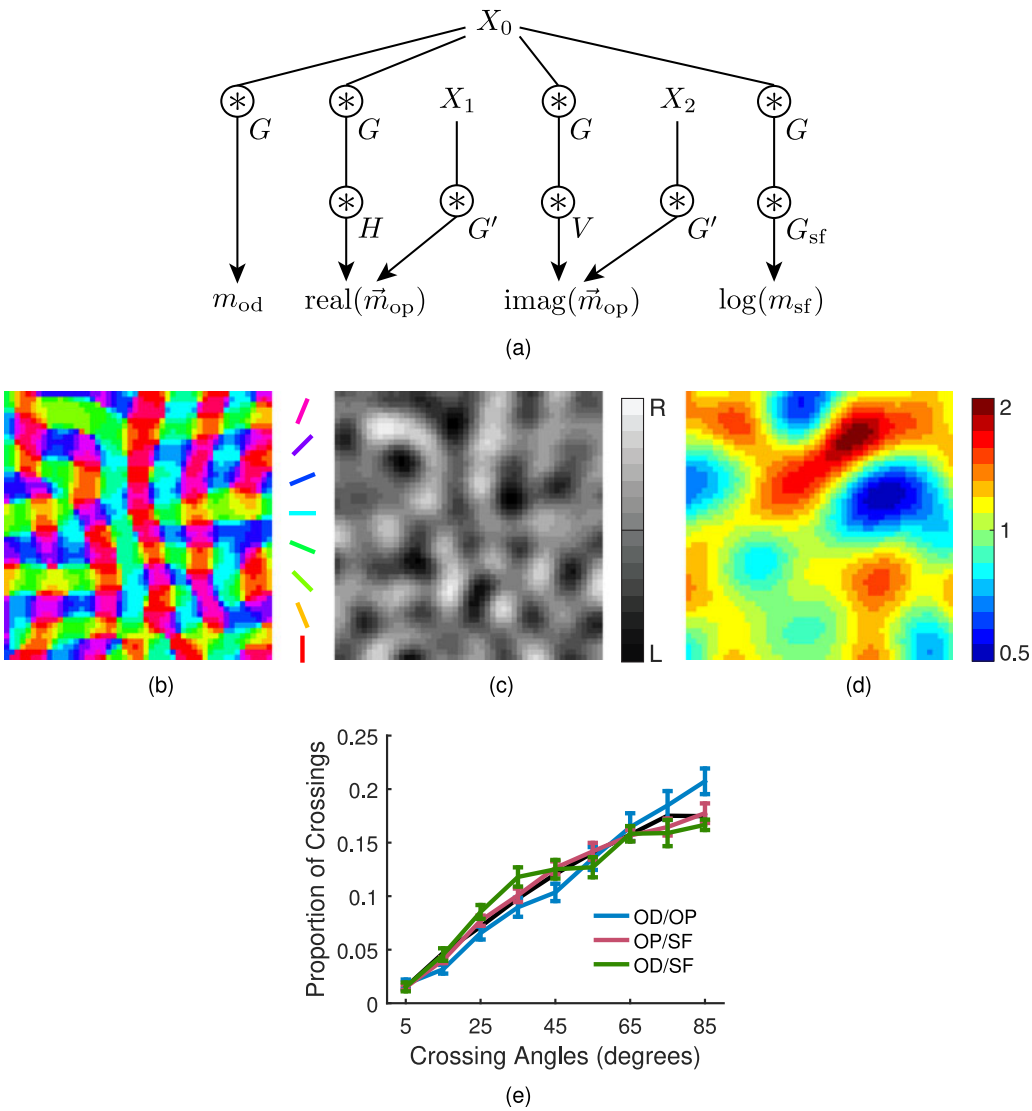


Fig. 6. A joint prior for three visual cortical maps. (a) Schematic of the model, in the same format as Fig. 2. A spatial frequency component has been added, defined as the ocular dominance map filtered with a difference-of-Gaussians kernel G_{sf} with twice the wavelength of G. (b–d) A triplet of maps, orientation preference, ocular dominance, and spatial frequency, sampled from this prior (parameters as in Table 1, $\alpha' = 0.25$). The values and units of the SF map are arbitrary, but follow a log-normal distribution, and when interpreted as cycles per degree match well with experimental data [12]. (e) Crossing angles between all three pairs of maps. The OD and OP maps (blue) intersect at angles larger than expected by chance (black), while the OD and SF maps intersect at more acute angles (green). Mean and standard error of the mean shown for ten sampled triples of maps.

$$
\begin{aligned}
K_{1,4} &= G * (G * G_{\mathrm{sf}}) \\
K_{2,4} &= (G * H) * (G * G_{\mathrm{sf}}) \\
K_{3,4} &= (G * V) * (G * G_{\mathrm{sf}}) \\
K_{4,1} &= K_{1,4} \\
K_{4,2} &= K_{2,4} \\
K_{4,3} &= K_{3,4} \\
K_{4,4} &= (G * G_{\mathrm{sf}}) * (G * G_{\mathrm{sf}}).
\end{aligned}
$$

However, spatial frequency is a non-negative quantity and a regular Gaussian process, with normally distributed values, is therefore an inappropriate model. SF preference appears to be roughly log-normally distributed [12], [29], and we therefore treat the fourth output of this model as the log of the SF map, $\log(m_{\mathrm{sf}})$, rather than as the SF map itself (Fig. 6a). Applying the exponential operator to this output then provides the spatial frequency map, which will have a log-normal distribution as its log is a Gaussian process. The application of non-linear monotonic functions to the outputs of GPs is known as warping, and it is simple to use such a warped GP model for regression and therefore map estimation [30].

A triplet of maps sampled from this prior is shown in Figs. 6b, 6c, and 6d. We quantified the spatial relationships between these maps with the crossing angles between all three pairs of maps, as described above and using the contours of four logarithmically-spaced values of the SF map from 0.5 to 2 (Fig. 6e). The OD and OP maps have the same relationships as in the model without the SF map, as expected. In contrast, the OD and SF map contours tend to intersect at more acute angles than random maps, indicating that their contours are more parallel than orthogonal, mirroring the prediction of [16]. The OP and SF maps have intersection angles matching the distribution for unrelated maps, demonstrating that this model suggests there is no spatial relationship, at least by this metric, between these maps.

The tendency for the OD and SF maps to cross at more acute angles than random maps in this model is revealed by the crossing angle distributions (Fig. 6e) rather than by the parameter α' which was used to measure correlations between the OD and OP maps. In this particular model we only parametrised the perpendicularity of maps or lack thereof, with α', and we do not have a parameter to measure such a parallel relationship. However a model could be constructed which would provide such a parametrisation, for example by the use of different Gabor filters in a similar arrangement to the current model (Fig. 6a), which demonstrates the flexibility and extensibility of this framework for investigating map relationships.

8 Discussion

Here we have constructed a Gaussian process framework for the study of multiple visual feature maps together, including an explicit model of their spatial relationships. This not only allows for improved estimates of empirical maps from noisy imaging data compared to classical techniques, but also the explicit specification of the spatial correlations between maps. The use of a Bayesian prior to encode the known features of maps and their dependencies is integral to this improved estimation. This is the first principled, quantitative approach to describing and studying the spatial relationships between topographic maps, which have previously been explored experimentally in detail but have so far lacked mathematical analysis.

While the specific model we constructed here for OD and OP maps was inspired by the earlier work of [17] and [23], the multiple output Gaussian process framework allows these original models to be taken much further. The specific control of the components which constitute each map, allowing dependencies to be finely quantitatively controlled, the addition of a noise model, and the reconstruction of maps from imaging data, all go significantly beyond the original work we have built upon.

The single map version of this GP framework was used by [8] to conduct a finer quantitative analysis of the striped-reared OP maps from [31] than previously available. By using the values of the inferred parameters of the GP prior, changes in several aspects of map structure caused by the abnormal visual input were found. Our expanded joint prior model allows the same analysis to be performed on several maps together, for example on OP and OD maps in tandem, allowing for analysis of the plasticity of the spatial relationships between the maps.

While the spatial relationships between the OD and OP maps have been experimentally studied in detail, the dependences between these maps and the map for spatial frequency preference are less clear. Several studies have suggested varying levels of dependence between the SF map and others [11], [12], [16], including independence [32]. To demonstrate the extensibility of our model we added an SF map to our joint prior for OD and OP maps, designed based on the map relationships suggested by [16]. However this framework is not restricted to this particular design, but can support any relationships between maps, and is therefore a useful tool for the theoretical exploration of map relationships which are so far experimentally unclear. Whether all three of these maps have a mutual tendency to intersect at angles close to perpendicular, or if one pair of maps have contours which are closer to parallel, is one particular pair of hypothesis which our framework allows the exploration of.

More generally Gaussian processes have found use not only in regression, as we use them here for map estimation, but increasingly in a variety of areas in machine learning including latent variable models [33] and deep learning [34]. This raises the exciting prospect of using our work here as a solid foundation for the exploration of map structure and relationships with these machine learning techniques, allowing for the learned extraction of interesting features and representations of maps.

References

[1] D. H. Hubel and T. N. Wiesel, "Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc. London B*, vol. 198, pp. 1–59, 1977.

[2] G. Blasdel and G. Salama, "Voltage-senstitive dyes reveal a modular organization in monkey striate cortex," *Nature*, vol. 321, pp. 579–585, 1986.

[3] T. Bonhoeffer and A. Grinvald, "Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns," *Nature*, vol. 353, pp. 429–431, 1991.

[4] M. Kaschube, M. Schnabel, S. Löwel, D. M. Coppola, L. E. White, and F. Wolf, "Universality in the evolution of orientation columns in the visual cortex," *Science*, vol. 330, no. 6007, pp. 1113–1116, 2010.

[5] N. Pouratian and A. W. Toga, "Optical imaging based on intrinsic signals," in *Brain Mapping: The Methods*, 2nd ed., A. Toga and J. Mazziotta, Eds. London, U.K.: Academic, 2002, pp. 97–140.

[6] A. Zepeda, C. Arias, and F. Sengpiel, "Optical imaging of intrinsic signals: Recent developments in the methodology and its applications," *J Neuroscience Methods*, vol. 136, pp. 1–21, 2004.

[7] J. H. Macke, S. Gerwinn, L. E. White, M. Kaschube, and M. Bethge, "Gaussian process methods for estimating cortical maps," *NeuroImage*, vol. 56, no. 2, pp. 570–581, 2011.

[8] N. J. Hughes, J. J. Hunt, S. L. Cloherty, M. R. Ibbotson, F. Sengpiel, and G. J. Goodhill, "Stripe-rearing changes multiple aspects of the structure of primary visual cortex," *NeuroImage*, vol. 95, pp. 305–319, 2014.

[9] S. LeVay, M. P. Stryker, and C. J. Shatz, "Ocular dominance columns and their development in layer IV of the cat's visual cortex: A quantitative study," *J. Comparative Neurology*, vol. 179, pp. 223–244, 1978.

[10] P. A. Anderson, J. Olavarria, and R. C. van Sluyters, "The overall pattern of ocular dominance bands in cat visual cortex," *J. Neuroscience*, vol. 8, no. 6, pp. 2183–2200, 1988.

[11] M. Hübener, D. Shoham, A. Grinvald, and T. Bonhoeffer, "Spatial relationships among three columnar systems in cat area 17," *J. Neuroscience*, vol. 17, no. 23, pp. 9270–9284, 1997.

[12] N. P. Issa, C. Trepel, and M. P. Stryker, "Spatial frequency maps in cat visual cortex," *J. Neuroscience*, vol. 20, no. 22, pp. 8504–8514, 2000.

[13] L. Sirovich and R. Uglesich, "The organization of orientation and spatial frequency in primary visual cortex," *Proc. Nat. Academy Sci. United States America*, vol. 101, no. 48, pp. 16941–16946, 2004.

[14] E. Bartfeld and A. Grinvald, "Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex," *Proc. Nat. Academy Sci. United States America*, vol. 89, no. 24, pp. 11905–11909, 1992.

[15] K. Obermayer and G. Blasdel, "Geometry of orientation and ocular dominance columns in monkey striate cortex," *J. Neuroscience*, vol. 13, pp. 4114–4129, 1993.

[16] I. Nauhaus, K. J. Nielsen, A. A. Disney, and E. M. Callaway, "Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex," *Nature Neuroscience*, vol. 15, no. 12, pp. 1683–1690, 2012.

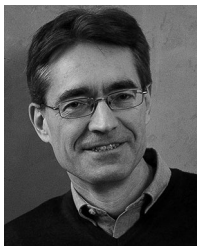
[17] E. Erwin, K. Obermayer, and K. Schulten, "Models of orientation and ocular dominance columns in the visual cortex: A critical comparison," *Neural Comput.*, vol. 7, no. 3, pp. 425–468, 1995.

[18] N. V. Swindale, "The development of topography in the visual cortex: A review of models," *Network*, vol. 7, no. 2, pp. 161–247, 1996.

[19] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[20] F. Zhang, *The Schur Complement and Its Applications*. Berlin, Germany: Springer, 2005.
[21] P. Boyle and M. Frean, "Dependent Gaussian processes," in *Advances in Neural Information Processing Systems 17*. Cambridge, MA, USA: MIT Press, 2005, pp. 217–224.
[22] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, 2002.
[23] A. S. Rojer and E. L. Schwartz, "Cat and monkey cortical columnar patterns modeled by bandpass-filtered 2D white noise," *Biol. Cybern.*, vol. 62, pp. 381–391, 1990.
[24] R. R. A. Morton, "The expected number and angle of intersections between random curves in a plane," *J. Appl. Probability*, vol. 3, no. 2, pp. 559–562, 1966.
[25] M. A. Carreira-Perpiñán, R. J. Lister, and G. J. Goodhill, "A computational model for the development of multiple maps in primary visual cortex," *Cerebral Cortex*, vol. 15, no. 8, pp. 1222–1233, 2005.
[26] C. E. Giacomantonio, M. R. Ibbotson, and G. J. Goodhill, "The influence of restricted orientation rearing on map structure in primary visual cortex," *NeuroImage*, vol. 52, no. 3, pp. 875–883, 2010.
[27] S. L. Cloherty, N. J. Hughes, M. A. Hietanen, P. S. Bhagavatula, G. J. Goodhill, and M. R. Ibbotson, "Sensory experience modifies feature map relationships in visual cortex," *eLife*, vol. 5, 2016, Art. no. e13911.
[28] N. V. Swindale, "How many maps are there in visual cortex?" *Cerebral Cortex*, vol. 10, no. 7, pp. 633–643, 2000.
[29] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, "Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex," *J. Physiology*, vol. 283, pp. 101–120, 1978.
[30] E. Snelson, C. E. Rasmussen, and Z. Ghahramani, "Warped Gaussian processes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 337–344.
[31] F. Sengpiel, P. Stawinski, and T. Bonhoeffer, "Influence of experience on orientation maps in cat visual cortex," *Nature Neuroscience*, vol. 2, no. 8, pp. 727–732, 1999.
[32] D.-S. Kim, Y. Matsuda, K. Ohki, A. Ajima, and S. Tanaka, "Geometrical and topological relationships between multiple functional maps in cat primary visual cortex," *NeuroReport*, vol. 10, no. 12, pp. 2515–2522, 1999.
[33] M. K. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, vol. 9, pp. 844–851.
[34] A. C. Damianou and N. D. Lawrence, "Deep Gaussian processes," in *Proc. 16th Int. Conf. Artif. Intell. Statist.*, 2013, vol. 31, pp. 207–215.

Nicholas J. Hughes received the BSc (Hons.) degree in mathematics and computational science from the University of Queensland, and the PhD degree in computational neuroscience jointly from the School of Mathematics and Physics, Queensland Brain Institute, University of Queensland.



Geoffrey J. Goodhill received the BSc (Joint Hons) degree in mathematics and physics from Bristol University, the MSc degree in artificial intelligence from Edinburgh University, and the PhD degree in cognitive science from Sussex University. He then did postdoctoral work in computational neuroscience with, Edinburgh University, Baylor College of Medicine, and the Salk Institute. He started his own lab with Georgetown University Medical Center, where he was awarded tenure in 2001. In 2005 he moved to the University of Queensland. From 2005-2010 he was Editor-in-Chief of *Network: Computation in Neural Systems*, and he is currently an associate editor of *Neural Computation* and *Brain Informatics*, and on the editorial board of *Nature Scientific Reports*.