*i*Map4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling

Junpeng Lao[1] · **Sébastien Miellet**[1,2] · **Cyril Pernet**[3] · **Nayla Sokhn**[1] · **Roberto Caldara**[1]

Abstract A major challenge in modern eye movement research is to statistically map *where* observers are looking, by isolating the significant differences between groups and conditions. As compared to the signals from contemporary neuroscience measures, such as magneto/electroencephalography and functional magnetic resonance imaging, eye movement data are sparser, with much larger variations in space across trials and participants. As a result, the implementation of a conventional linear modeling approach on two-dimensional fixation distributions often returns unstable estimations and underpowered results, leaving this statistical problem unresolved (Liversedge, Gilchrist, & Everling, 2011). Here, we present a new version of the *i*Map toolbox (Caldara & Miellet, 2011) that tackles this issue by implementing a statistical framework comparable to those developed in state-of-the-art neuroimaging data-processing toolboxes. *i*Map4 uses univariate, pixel-wise linear mixed models on smoothed fixation data, with the flexibility of coding for multiple between- and within-subjects comparisons and performing all possible linear contrasts for the fixed effects (main effects, interactions, etc.). Importantly, we also introduced novel nonparametric tests based on resampling, to assess statistical significance.

Finally, we validated this approach by using both experimental and Monte Carlo simulation data. *i*Map4 is a freely available MATLAB open source toolbox for the statistical fixation mapping of eye movement data, with a user-friendly interface providing straightforward, easy-to-interpret statistical graphical outputs. *i*Map4 matches the standards of robust statistical neuroimaging methods and represents an important step in the data-driven processing of eye movement fixation data, an important field of vision sciences.

Human beings constantly move the eyes to sample visual information of interest from the environment. Eye fixations deliver inputs with the highest resolution to the human visual cortex from the fovea, as well as blurry, low-spatial-frequency information from peripheral vision (Rayner, 1998). Thus, isolating statistically *where* and *how long* fixations are deployed to process visual information is of particular interest to behavioral researchers, psychologists, and neuroscientists. Moreover, fixation mapping has a wide range of practical applications in determining marketing strategies and the understanding of consumer behaviour (Duchowski, 2002).

Conventional eye movement data analyses rely on the estimation of probabilities of occurrence of fixations and saccades (or their characteristics, such as duration or length) within predefined regions of interest (ROIs), which are at best defined a priori—but often also defined a posteriori, on the basis of data exploration, which inflates the Type I error rate. Another issue with ROIs is of course that other important information not included in the ROI is discarded. In a continuous effort to circumvent the limitations of the ROI approach

✉ Junpeng Lao
  junpeng.lao@unifr.ch

✉ Roberto Caldara
  roberto.caldara@unifr.ch

[1] Department of Psychology, University of Fribourg, Faucigny 2, 1700 Fribourg, Switzerland

[2] Department of Psychology, Bournemouth University, Bournemouth, UK

[3] Centre for Clinical Brain Sciences and Neuroimaging Sciences, University of Edinburgh, Edinburgh, UK

(for a detailed discussion on this point, see Caldara & Miellet, 2011), we previously developed an unbiased, data-driven approach to compute statistical fixation maps of eye movements: the *i*Map toolbox (Caldara & Miellet, 2011). From the very first version, the toolbox was developed as a MATLAB open source toolbox freely available for download online. The previous versions (1 and 2) made use of Gaussian smoothing and the random field theory as a statistical engine (Caldara & Miellet, 2011), which is one of the standard methods applied in statistical analyses for functional magnetic resonance imaging (fMRI) data (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). Version 3 introduced pixel-wise *t* test and bootstrap clustering in order to generate self-contained statistical maps (Miellet, Lao, & Caldara, 2014). However, all of the previous versions of *i*Map still suffered from a major limitation: They could only contrast two conditions at a time.

A major revision of the toolbox was necessary to enable the analysis of more complex experimental designs routinely used in the field. One of the most suitable and obvious statistical solutions to overcome this problem would be to implement a general linear model, a widespread approach in both behavioral and neural-imaging data analyses. In fact, many modern procedures for hypothesis testing, such as the *t* test, analysis of variance (ANOVA), regression, and so forth, belong to the family of general linear models. However, eye movement data are a sparse production of visual perceptual sampling. Unlike neuroimaging data, eye movement data contain many empty cells with little to no data points across the tested space (e.g., all of the pixels in an image). This caveat engenders a statistical problem when the same statistical inference procedure is applied on each pixel, regardless or whether or not its data are missing. To account for the sparseness and the high variation of spatial eye movement data, we developed a specific novel approach for smoothed fixation maps, which was inspired by the statistical framework implemented in diverse state-of-the-art neuroimaging data-processing toolboxes: statistical parametric mapping (SPM; Penny et al., 2011), Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), and LIMO EEG (Pernet, Chauveau, Gaspar, & Rousselet, 2011). In the simplest case, users can apply a massive univariate, pixel-wise linear mixed model (LMM) on the smoothed fixation data with the subject considered as a random effect, which offers the flexibility to code for multiple between- and within-subjects comparisons. Our approach allows users to perform all possible linear contrasts for the fixed effects (main effects, interactions, etc.) from the resulting model coefficients and the estimated covariance. Importantly, we also introduced a novel nonparametric statistical test based on resampling (permutation and bootstrap spatial clustering) to assess the statistical significance of the linear contrasts (Pernet, Latinus, Nichols, & Rousselet, 2015; Winkler, Ridgway, Webster, Smith, & Nichols, 2014).

In the next section, we briefly describe the key concepts of the LMM approach. We then introduce the *novel* nonparametric statistical approach on the fixed effects that we implemented in *i*Map4, which uses a resampling procedure and spatial clustering. We also report a validation of the proposed resampling procedures, and illustrate how *i*Map4 can be used, with both a subset of data from a previous study and computer-simulated data. Finally, we give an overview of future development and discuss technical insights on eye fixation mapping.

Linear mixed models

In this part, we outline the key elements and concepts of LMMs in comparison with general linear models (GLM) and hierarchical linear models (HLM). Mixed models represent a complex topic, and the discussion of many underlying mathematical details goes beyond the scope of this article. For general, thoughtful introductions to mixed models, users of the toolbox should refer to Raudenbush and Bryk (2002) and McCulloch, Searle, and Neuhaus (2011). Users may also wish to consult the documentation and help files of the *LinearMixedModel* class in the MATLAB Statistics Toolbox for details about parameter estimation and the available methods (www.mathworks.com/help/stats/linearmixedmodel-class.html).

Statistical hypothesis testing methods that make use of the analysis of variance (regression, *t* test, ANOVA, analysis of covariance, etc.) are the most popular methods of data analysis in many fields of research. Commonly used in psychology and neuroimaging studies, these methods could all be written as particular cases of GLM:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_t x_{ti} + \varepsilon_i,$$
$$\varepsilon_i \sim N(0, \sigma^2), \tag{1}$$

where y_i is the *i*th experiment measure and $\beta_1, \beta_2, \ldots, \beta_t$ are the model coefficients. The error term ε_i is normally distributed with mean 0 and variance σ^2. Alternatively, the GLMs (Eq. 1) could be expressed in matrix form:

$$Y = X\beta + \varepsilon,$$
$$\varepsilon \sim N(0, \sigma^2 I), \tag{2}$$

where matrix $X = [x_1, x_2, \ldots, x_t]$ is the design matrix, and I is an *n*-by-*n* identity matrix (*n* being the total number of observation). Usually, one of the columns in X is 1, so that the model includes a constant or intercept coefficient that represents the overall mean. It is worth noting that the design matrix could be parameterized in a different way. In conventional psychology or behavioral researches, a sigma-restricted parameterization is often applied. In a sigma-restricted design matrix, X is full rank and invertible, and the degrees of freedom are equal to the number of columns. In comparison, many types of neuroimaging analysis software prefer a cell-

mean model or an overparameterized design matrix at the single-subject level (Penny et al., 2011; Pernet et al., 2011). Such software uses an overparameterized design matrix, and its solution to Eq. 2 is given by projecting the response vector **Y** to the pseudo-inverse of the design matrix **X**. The form of the design matrix is important, since it codes different experiment designs and the intended statistical testing. In *iMap4*, the design matrix of the fixed effect can be a cell-mean model, sigma-restricted model (for the Type III ANOVA), or the offset from a reference model (for Type I ANOVA).

The coefficient estimations $\hat{\boldsymbol{\beta}}$ could be found easily by ordinary least squares or other, more robust methods. Finally, statistical inferences on the model estimations could be expressed in different forms, depending on the types of design matrix. In the case of sigma-restricted parameterization, we can separate the design matrix $\mathbf{X^{sr}}$ and the vector of parameters β^{sr} into two parts (Kherad-Pajouh & Renaud, 2010):

$$\mathbf{X}^{sr} = [\mathrm{X}_1 \ \mathrm{X}_2] \, , \beta^{sr} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

where $\mathbf{X_1}$ and β_1 are the components of interest, with the corresponding hypotheses:

$$\mathrm{H}_0 \ : \ \beta_1 \ = \ 0 \ \text{versus} \ \mathrm{H}_1 \ : \ \beta_1 \neq 0. \tag{3}$$

Given the Gaussian distribution of the error ε and the existence of the inverse or general inverse of design matrix $\mathbf{X^{sr}}$, we can get the statistics for the F test by means of ANOVA with the following equations (for simplicity, in Eqs. 4–8 we denote $\mathbf{X} = \mathbf{X^{sr}}$):

$$\mathrm{H} = \mathrm{X}\left(\mathrm{X}^\mathrm{T}\mathrm{X}\right)^{-}\mathrm{X}^\mathrm{T} \tag{4}$$

$$\mathrm{X}_{\mathrm{resid}} = \left(I\text{-}\mathrm{X}_2\left(\mathrm{X}_2^\mathrm{T}\mathrm{X}_2\right)^{-}\mathrm{X}_2^\mathrm{T}\right)\mathrm{X}_1 \tag{5}$$

$$\mathrm{H}_{\mathrm{resid}} = \mathrm{X}_{\mathrm{resid}}\left(\mathrm{X}_{\mathrm{resid}}^\mathrm{T}\mathrm{X}_{\mathrm{resid}}\right)^{-}\mathrm{X}_{\mathrm{resid}}^\mathrm{T} \tag{6}$$

$$\mathrm{dfe} = \text{Number of observations}-\mathrm{rank}(\mathbf{X}) \tag{7}$$

$$F = \frac{\mathbf{Y}^\mathbf{T}\,\mathbf{H_{resid}}\,\mathbf{Y}/\mathrm{rank}(\mathbf{X_1})}{\mathbf{Y}^\mathbf{T}\left(\boldsymbol{I} - \mathbf{H}\right)\mathbf{Y}/\left(\mathrm{dfe}\right)} \, , \tag{8}$$

where **H** represents the hat matrix of the linear model in Eq. 2; it projects the response vector **Y** onto the column space of **X**. $\mathbf{H_{resid}}$ is the hat matrix of the hypothesis in Eq. 3, and dfe is the model's degrees of freedom. F has a Fisher–Snedecor distribution $\mathcal{F}\left(\mathrm{rank}\left(\mathrm{X}_1\right), \mathrm{dfe}\right)$.

As a comparison, in an overparameterized design matrix or cell-mean model design matrix with design matrix $\mathbf{X^{cm}}$ and the

vector of parameters $\beta^{\mathbf{cm}}$, the statistics of various effects are performed by linear combinations of the coefficient $\beta^{\mathbf{cm}}$. For example, the equivalent Hypothesis (3) could be expressed as:

$$\mathrm{H}_0 \ : \ \mathbf{c}*\beta^{cm} \ = \ 0 \ \text{versus} \ \mathrm{H}_1 \ : \ \mathbf{c}*\beta^{cm} \neq 0, \tag{9}$$

where rank(**c**) = rank($\mathbf{X_1}$) and $\mathbf{c}*\beta^{cm} = \beta_1$ in the sigma-restricted parameterization model. The related F test is then given by the quartic form of the linear contrast matrix **c** and the inverse of the covariant matrix of $\beta^{\mathbf{cm}}$ (for simplicity, in Eqs. 10 and 11 we denote $\mathbf{X} = \mathbf{X^{cm}}$):

$$\mathrm{MSE} = \mathbf{Y}^\mathbf{T}\left(\boldsymbol{I} - \mathbf{H}\right)\mathbf{Y}/(\mathrm{dfe}) \tag{10}$$

$$F = \frac{\left(\mathbf{c}*\beta^{cm}\right)^\mathbf{T}\left(\mathrm{MSE}*\mathbf{c}\left(\mathbf{X^T X}\right)^{-}\mathbf{c^T}\right)^{-}\left(\mathbf{c}*\beta^{cm}\right)}{\mathrm{rank(c)}} , \tag{11}$$

where **H** and dfe are computed using Eqs. 4 and 7, respectively. Moreover, it could be proved that Eqs. 8 and 11 are equivalent. The related details and mathematical proofs could be found in many textbooks (e.g., Christensen, 2011).

The GLM $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ could be easily extended into a generalized form with $\varepsilon \sim \mathbf{N}(0, \sigma^2 \mathbf{V})$ where $\mathbf{V}$ is some known positive definite matrix. Moreover, if a more specific structure of the error ε is available, the GLM (Eq. 2), which has one random-effect term (the error ε), could be further extended into a mixed model. Mixed models include additional random-effect terms that can represent the clusters or classes. In a typical neuroimaging study, this could be the subjects or groups. In the following example, we consider a simplified case in which only the subject is considered as the additional random effect. This type of model is one of the most widely used models in both fMRI and electroencephalography (EEG).

As a demonstration, here we consider a random intercept and slope model, with both the intercept (i.e., the overall mean of each subject) and the slope (i.e., the differences among conditions within each subject) varying independently. This type of HLM, or so-called two-level linear model, takes the form of an expansion of Eq. 1 into:

$$\begin{aligned} y_{ij} &= \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{tj}x_{tij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{N}\left(0, \sigma^2\right) \\ \beta_{1j} &= \beta_{10} + b_{1j}\,, \quad b_{1j} \sim \mathrm{N}\left(0, \sigma_1^2\right) \\ &\vdots \\ \beta_{tj} &= \beta_{t0} + b_{tj}\,, \quad b_{tj} \sim \mathrm{N}\left(0, \sigma_t^2\right), \end{aligned}$$
$$\tag{12}$$

where j stands for the jth subject. After substituting the subject-level parameters in the first-level model, Eq. 12 becomes

$$y_{ij} = \left(\beta_{10} + b_{1j}\right)x_{1ij} + \left(\beta_{20} + b_{2j}\right)x_{2ij} + \cdots + \left(\beta_{t0} + b_{tj}\right)x_{tij} + \varepsilon_{ij}$$

$$= \underbrace{\beta_{10}x_{1ij} + \beta_{20}x_{2ij} + \cdots + \beta_{t0}x_{tij}}_{fixed\ effects} + \underbrace{b_{1j}x_{1ij} + b_{2j}x_{2ij} + \cdots + b_{tj}x_{tij}}_{random\ effects} + \ \varepsilon_{ij}$$

If we express the subject-level predictor x_{tij} in the random effects by the term z_{tij}, we get the LMM

$$y_{ij} = \underbrace{\beta_{10}x_{1ij} + \beta_{20}x_{2ij} + \cdots + \beta_{t0}x_{tij}}_{fixed\ effects} + \underbrace{b_{1j}z_{1ij} + b_{2j}z_{2ij} + \cdots + b_{tj}z_{tij}}_{random\ effects} + \varepsilon_{ij}$$

which corresponds to the standard form of LMMs:

$$Y = X\beta + Zb + \varepsilon \tag{13}$$

$b \sim N(0, \sigma^2 D)$, $\varepsilon \sim N(0, \sigma^2 I)$, b and ε are independent from each other, where $\sigma^2 D$ is the covariance matrix for the random effects. In the example here, D would be a j-by-j identity matrix. An alternative form of Eq. 12, as applied in LIMO EEG or SPM, can be found in Friston, Stephan, Lund, Morcom, and Kiebel (2005, Eq. 1).

HLMs are specific cases of LMMs. In a mixed model, factors are not necessarily hierarchical. Moreover, crossed factors between fixed effects and random effects are much easier to model in mixed than in hierarchical models. In additions, the fixed effects and random effects are estimated simultaneously in mixed models, which is not always the case in hierarchical models.

Parameter estimation in mixed models is much more complicated than in GLM or HLM. Assuming that the model in Eq. 13 has the error covariance matrix R: var $(Y|b) = R$, this model is equivalent to $y \sim N(X\beta, V)$, $V = ZDZ^T + R$. The estimation of the fixed effects β requires prior knowledge of V, which is usually unavailable. In practice, the variance component V is commonly replaced by an estimation $\hat{V}$ based on one of several approaches, such as ANOVA, maximum likelihood (ML) estimation, restricted maximum likelihood (ReML) estimation, or Monte Carlo approximation (McCulloch, Searle, & Neuhaus, 2011; Pinheiro & Bates, 2000). In general, the model-fitting procedure of LMM is implemented in major statistical packages (e.g., R and Stata) by solving Henderson's mixed model equation. *i*Map4 calls the MATLAB class *LinearMixedModel* from Statistics Toolbox (versions R2013b or above) to estimate the coefficients (fixed effect β and random effect b) and the covariance matrix V with various options (key concepts with regard to parameter estimations can be found in the MATLAB documentation: www.mathworks.com/help/stats/estimating-parameters-in-linear-mixed-effects-models.html). In brief, model coefficients are estimated by ML or ReML, and the pattern of the covariance matrix of the random effects (D) could take the form of a full covariance matrix, a diagonal covariance matrix, or other symmetry structure.

Statistical inferences in LMM are also much more complex than in a GLM. In a balanced design, or with the variance component V known, hypothesis testing of the fixed effect

follows Eq. 8 or 11 as an exact test. However, in an unbalanced design with random effects, no exact F statistics are available, since biases in the estimation usually result in an unknown distribution of F (Kherad-Pajouh & Renaud, 2015). Although F and t values are available as approximate tests in most statistical packages, Baayen, Davidson, and Bates (2008) discouraged the usage of t or F statistics, and especially report of the p value, in mixed models. Other approaches have also been proposed. For example, likelihood ratio tests could be performed to test composite hypotheses by comparing the desired model with the reduced model. However, there are many constraints on the application of likelihood ratio tests (e.g., the method of model fitting and selection of the reduced model). Moreover, running multiple copies of similar LMMs is computationally expensive, especially in the context of pixel-wise testing, such as in *i*Map4.

Besides the practical problem of statistical inferences with LMM, another main challenge in the application of LMM to spatial eye movement data is the Type I error from multiple comparisons. To resolve these issues, we adopted resampling techniques for null-hypothesis statistical testing, as is suggested in neuroimaging analysis with GLM or HLM (Pernet et al., 2015; Winkler et al., 2014). Nonparametric statistics using Monte Carlo simulation are ideal for both parameter estimation and hypothesis testing (Baayen et al., 2008; Kherad-Pajouh & Renaud, 2015). In *i*Map4, we adapted a simplified version of the permutation test suggested by Winkler et al. (2014) and a bootstrap clustering method similar to the one applied in LIMO EEG (Pernet et al., 2011). Details of the proposed algorithm and preliminary validation result are described in the following section.

Pixel-wise modeling and spatial clustering

Although the generation mechanism of eye movement data is still largely under debate, recent theories and applications suggest that a spatial model is the most appropriate to consider the statistical analysis of fixations, especially their location distribution. For example, Barthelmé, Trukenbrod, Engbert, and Wichmann (2013) recommended using the point process framework to infer how fixations are distributed in space. Although we endorse this fruitful approach and its Bayesian nature, here we aimed to resolve this problem from the opposite perspective. Instead of inferring from the spatial

distribution of the fixations, we inferred on each location in the search space (i.e., each pixel within the eyetracker's recordable range or each pixel in the visible stimuli). In other words, we addressed the question: "How long is this pixel being fixated (or what is the probability of this pixel being fixated) in the function of the experimental conditions?", by formally applying mixed models independently on each pixel, we have

$$\mathbf{Y}(s) = \mathbf{X}\boldsymbol{\beta}(s) + \mathbf{Z}\boldsymbol{b}(s) + \boldsymbol{\varepsilon}(s) \qquad (14)$$

For $s \in \boldsymbol{D}$ of the search space.

The complete procedure as implemented in *i*Map4 is explained in Fig. 1. The eye movement data for each participant are concatenated into one input data matrix. *i*Map4 first partitions the data matrix into a fixation characteristic matrix (red box) and an experiment condition information matrix (green box). The fixation characteristic matrix contains a fixation's spatial location (x and y), the fixation duration, and an order index of each fixation. The experiment condition matrix contains an index of each subject, an index of each trial/item, and the different levels of each experimental condition. Fixation durations are then projected into the two-dimensional space according to their x- and y-coordinates at the single-trial level. *i*Map4 then smooths the fixation duration map by convoluting it with a two-dimension Gaussian kernel function:

$$\text{Kernel} \sim \text{N}\left(0, \sigma^2 I\right),$$

where I is a two-by-two identity matrix and the full width at half maximum (FWHM) of the kernel is *1° of visual angle* as the default setting.

This step is essential to account for the spatial uncertainty of eye movement recordings (both mechanical and
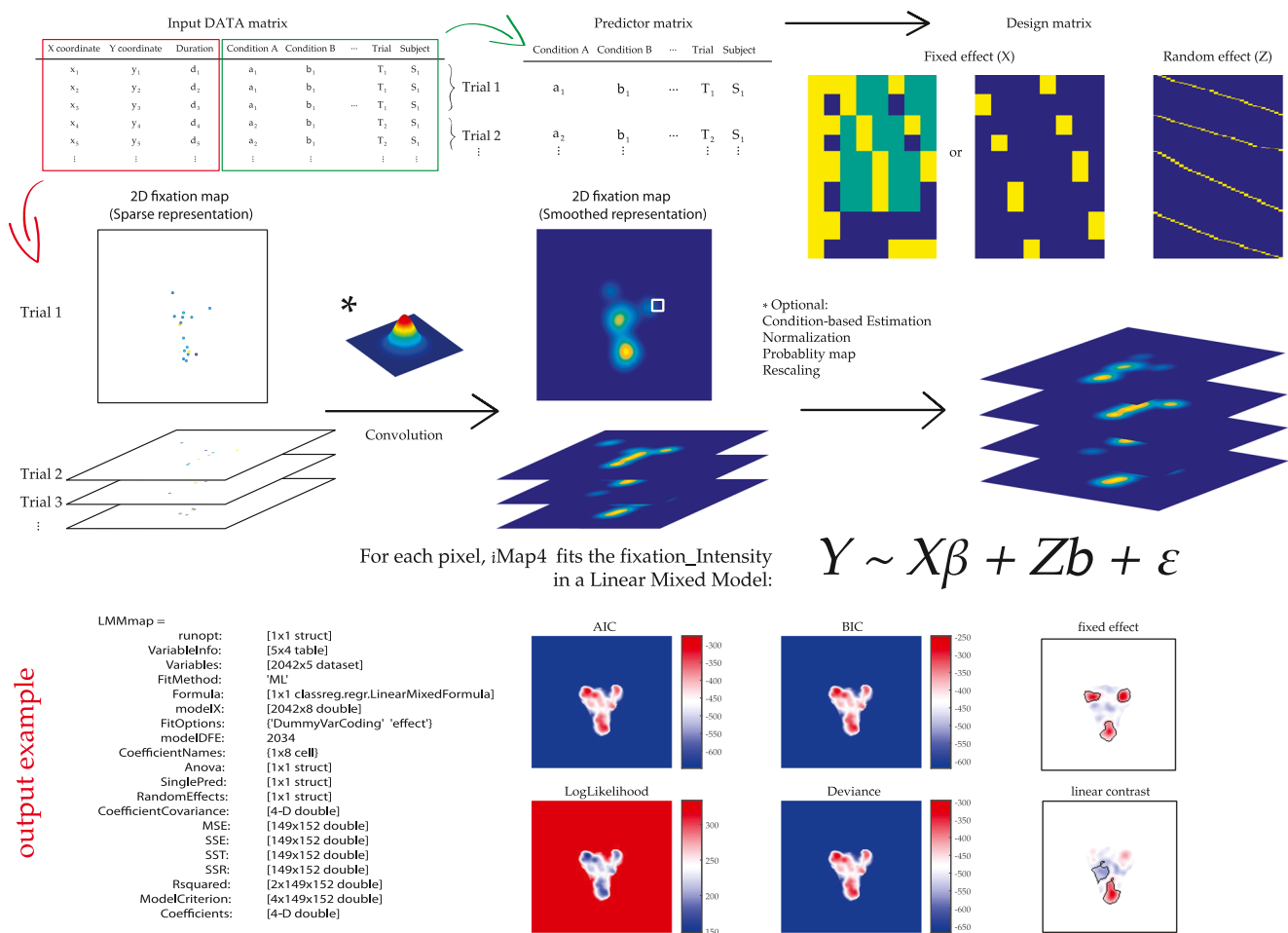

Fig. 1 Illustration of the basic processing steps implemented in *i*Map4. The input data matrix is partitioned into an eye movement matrix and predictor matrix. Fixation durations are projected into the two-dimensional space according to their x- and y-coordinates at the single-trial level for each participant. The experimental information of each trial is also summarized in a predictor table. Subsequently, the sparse representation of the fixation duration map is smoothed by convoluting it with a two-dimensional Gaussian kernel function, kernel~N $(0, \sigma^2 \boldsymbol{I})$.

After estimating the fixation bias of each condition independently for all observers (by taking the expected values across trials within the same condition), *i*Map4 models the 3-D smoothed fixation map (*item**xSize*ySize) independently for each pixel using an LMM. The result is saved as a MATLAB structure in *LMMmap*. *i*Map4 offers many parametric and nonparametric methods for hypothesis testing and multiple-comparison correction

physiological) and the sparseness of the fixation locations. The Gaussian kernel could also be replaced by other 2-D spatial filters to best suit the research question.

The resulting smoothed fixation map is a 3-D matrix. The last two dimensions of the fixation matrix are the sizes of the stimuli or search space. The information of each entry in the first dimension is stored in a predictor table, which is generated from the experiment condition matrix. Each experiment condition can be coded at the single-trial level in the predictor table, or as one entry by taking the average map across trials.

In addition, iMap4 provides a robust estimation option by applying Winsorization in order to limit extreme values in the smoothed fixation matrix. The goal here is to reduce the effect of any potential outliers. Additional options include: spatial normalization (z-scored map or probability map), spatial down-sampling (linear transformation using *imresize* in MATLAB) to optimize computing speed, and mask creation to exclude irrelevant pixels.

The resulting 3-D fixation matrix is then modeled in a LMM as the response variable. The results are saved as a MATLAB structure (*LMMmap*, as in the examples below). The fields of *LMMmap* are nearly identical to the output from the *LinearMixedModel* class. For each modeled pixel, iMap4 saves the model criterion, variances explained, error sum of squares, coefficient estimates, and their covariance matrix for both fixed and random effects, and the ANOVA results for the LMM. Additional modeling specifications, as well as other model parameters, including the LMM's formula, design matrix for fixed and random effect, and residual degrees of freedom, are also saved in *LMMmap*. Linear contrasts and other analyses based on variance or covariance can be performed afterward from the model-fitting information. Any other computation on the *LinearMixedModel* output can also be replicated with *LMMmap*.

One of the crucial assumptions of pixel-wise modeling is that all pixels are independent and identically distributed. Of course, this assumption is never satisfied, neither before nor after smoothing. To ensure valid inferences on activity patterns in a large 2-D pixel space, we applied nonparametric statistics to resolve the biases in parameter estimation and problems arising from multiple comparisons. We developed two resampling-based statistical hypothesis-testing methods for the fixed-effect coefficients: a universal permutation test and a universal bootstrap clustering test.

The resampling tests on the model coefficient for fixed effects β operate on the fixed-effect-related variances. To do so, we simply removed the variance associated with the random effects from the response matrix:

$$\mathbf{Y}_{\text{fixed}}(s) = \mathbf{X}\beta(s) + \varepsilon(s) = \mathbf{Y}(s) - \mathbf{Z}b(s), \quad (15)$$

For $s \in \mathbf{D}$ of the search space.

For any permutation test, iMap4 performs the following algorithms on $\mathbf{Y}_{\text{fixed}}$ for each pixel.

Algorithm 1

For a given hypothesis or linear contrast c (as in Eq. 9), iMap4

– Performs a linear transformation on the design matrix $\mathbf{X}$ to get a new design matrix $\mathbf{M}$ so that the partitioning of $\mathbf{M} = [\mathbf{M}1, \mathbf{M}2]$. Then iMap4 computes the new coefficients by projecting $\mathbf{Y}_{\text{fixed}}$ to the pseudo-inverse of $\mathbf{M}$. The design matrix $\mathbf{M}$ is created so that the original hypothesis testing is equivalent to the hypothesis regarding the $\mathbf{M}1$ coefficients. The matrix transformation and partition are the same as the algorithm described in Winkler et al. (2014, Appx. A).

– Computes the residuals related to the hypothesis by subtracting the variance accounted for by $\mathbf{M}2$ from $\mathbf{Y}_{\text{fixed}}$, to get $\mathbf{Y}_{rr}$.

– Fits $\mathbf{Y}_{rr}$ to $\mathbf{M}$ by solving $\mathbf{Y}_{rr} = \mathbf{M}\beta_{\mathbf{m}} + \varepsilon$, and gets the statistical value $\mathbf{F}_{rr}$ of $\mathbf{M}1$ according to Eqs. 10 and 11. Note that to replicate the original hypothesis testing on the fixed effect, the new contrast c' is just used to partition $\mathbf{M}$ into $\mathbf{M}1$ and $\mathbf{M}2$.

– Permutes the rows of the design matrix $\mathbf{M}$ to obtain the new design matrix $\mathbf{M}^*$.

– Fits $\mathbf{Y}_{rr}$ to $\mathbf{M}^*$ and gets the $\mathbf{F}_{rr}^*$ of $\mathbf{M}1^*$.

– Repeats the previous two steps a large number of times (k resamplings/repetitions), and the p value is then defined as in Eq. 16. Importantly, the family-wise error rate (FWER) corrected p value is computed by comparing the largest $\mathbf{F}_{rr}^*$ across all tested pixels in one resampling with the original $\mathbf{F}_{rr}$:

$$p = \frac{\left(\# \ \mathbf{F}_{rr}^* \geq \mathbf{F}_{rr} \right)}{k}. \quad (16)$$

Algorithm 1 is a simplified version of Winkler et al. (2014, Algorithm 1): The resampling table includes permutation but not sign-flipping, which assumes the errors to be independent and symmetric. Thus, the underlying assumptions are stronger than with classical permutations, which require only exchangeable errors (Winkler et al., 2014).

Importantly, this test is exact only under a balanced design with no missing values and only subjects as a random effect. As was previously shown in Kherad-Pajouh and Renaud (2015), a general and exact permutation approach for mixed-model designs should be performed on modified residuals that have up to second-moment exchangeability. This is done to satisfy the important assumptions for repeated measures ANOVA: normality and the sphericity of errors. However, there are strict requirements to achieve this goal: careful transformation and partition of both the fixed- and random-effects design matrices, and removal of the random effects related to $\mathbf{M}2$ (Kherad-Pajouh & Renaud, 2015). In iMap4, we perform an approximation version by removing all random effects to increase the efficiency and speed of the huge amount of

resampling computation in our pixel-wise modeling algorithm. Validation and simulation data set indeed showed that the sensitivity and the false alarm rate of the proposed algorithm were not compromised.

Algorithm 2

iMap4 performs the following algorithm on $\mathbf{Y_{fixed}}$ for each pixel as the bootstrap clustering approach.

- For each unique categorical variable, iMap4 removes the conditional expectations from $\mathbf{Y_{fixed}}$ for each pixel. A random shuffling is then performed on the centered data to acquire $\mathbf{Y_c}$, so that any potential covariance is also disrupted. This is done to construct the true empirical null-hypothesis distribution in which all elements and their linear combinations in $\mathbf{Y_c}$ have expected values equal to 0.
- Randomly draws with replacement from $\{\mathbf{X}, \mathbf{Z}, \mathbf{Y_c}\}$ equal numbers of subjects $\{\mathbf{X^*}, \mathbf{Z^*}, \mathbf{Y_c^*}\}$.
- Fits $\mathbf{Y_c}^*$ to $\mathbf{X}^*$ by solving $\mathbf{Y_c}^* = \mathbf{X}^*\beta^* + \varepsilon$. For a given hypothesis or linear contrast c (as in Eq. 9), iMap4 computes the statistics value $\mathbf{F}^*$ according to Eqs. 10 and 11, and their parametric p value under the GLM framework.
- Thresholds the statistical maps $\mathbf{F}^*$ at $p^* \leq .05$ and records the desired maximum cluster characteristics across all significant clusters. The cluster characteristics considered are cluster mass (summed F value within a cluster), cluster extent (size of the cluster), and cluster density (mean F value).
- The previous three steps are repeated a large number of times, to get the cluster characteristic distribution under the null hypothesis.
- Thresholds the original statistical map $\mathbf{F}$ at $p \leq .05$ and compares the selected cluster characteristic with the value of the null distribution corresponding to the 95th percentile. Any cluster with the chosen characteristic larger than this threshold is considered significant.

The bootstrap clustering approach is identical to the bootstrap procedure described by Pernet et al. (2011; Pernet et al., 2015) if only a subject intercept is considered as the random effect. In addition, Algorithm 2 extents the philosophy and approach presented by Pernet et al. (2011; Pernet et al., 2015) to nonhierarchical mixed-effect models.

It is worth noting that we implemented in iMap4 a high-performance algorithm to minimize the computational demands of the large amount of resampling. The model fitting in both resampling approaches makes use of ordinary least squares. The inversion of the covariance matrices (required for Eq. 11) is computed on the upper triangular factor of the Cholesky decomposition. Calculation of the quartic form (as in Eq. 11) for all pixels is optimized by constructing a sparse

matrix of the inverse of the covariance matrix. More details of these algrebraic simplifications can be found in the *imapLMMresample* function in iMap4.

Other multiple-comparison correction methods, such as Bonferroni correction, false discovery rate, or random field theory (RFT), could also be applied. A threshold-free cluster enhancement algorithm could also be applied on the statistical (*F*-value) maps as an option after the permutation and bootstrap clustering procedures (Smith & Nichols, 2009).

We performed a validation study to assess the Type I error rate when applying the permutation and bootstrap clustering approach for hypothesis testing. We used a balanced repeated measures ANOVA design with a two-level between-group factor and a three-level within-group factor. A total population of 134 observers (67 in each group) was drawn from previous face-viewing eye movement studies. We centered the cell means for the whole dataset to obtain the validation dataset under the null hypothesis (similar to Step 1 in Algorithm 2). Thus, we used real data to warrant realistic distributions and centered them to ensure that H0 was confirmed. Any significant output from iMap4 performed on this dataset would be considered as a false alarm (Type I error).

The validation procedure followed the steps below: We first randomly sampled without replacement a balanced number of subjects from both groups. We then ran iMap4 under the default settings and performed hypothesis testing on the two main effects and the interaction. To estimate the FWER, we computed the frequency of significant output under different statistics and MCC settings. Preliminary results based on 1, 000 randomizations with a sample size of $n \in [8, 16, 32, 64]$ showed that with an alpha of .05, the FWERs were indeed all under. 05 using nonparametric statistics (see Fig. 2b for the permutation test, and Fig. 2c and d for the bootstrap clustering test). More simulations considering a wider range of scenarios will be required to understand fully the behavior of the proposed approaches, although the cluster stats are likely to behave as in Pernet et al. (2015).

Graphical user interface (GUI) and command line handling

iMap4 runs on MATLAB 2013b and above, since it requires some essential functions and classes from the Image Processing Toolbox and Statistics Toolbox in these versions. iMap4 will execute in parallel on multicores or distributed workers, when available.

We recommend that users install iMap4 as a MATLAB application. The users can call iMap4 directly in the MATLAB command window after installation. A general GUI will open upon >>iMAP, called in the command window or when launching the app (Fig. 3a). The users can then import the fixation data, load a preprocessed data matrix for LMM, or
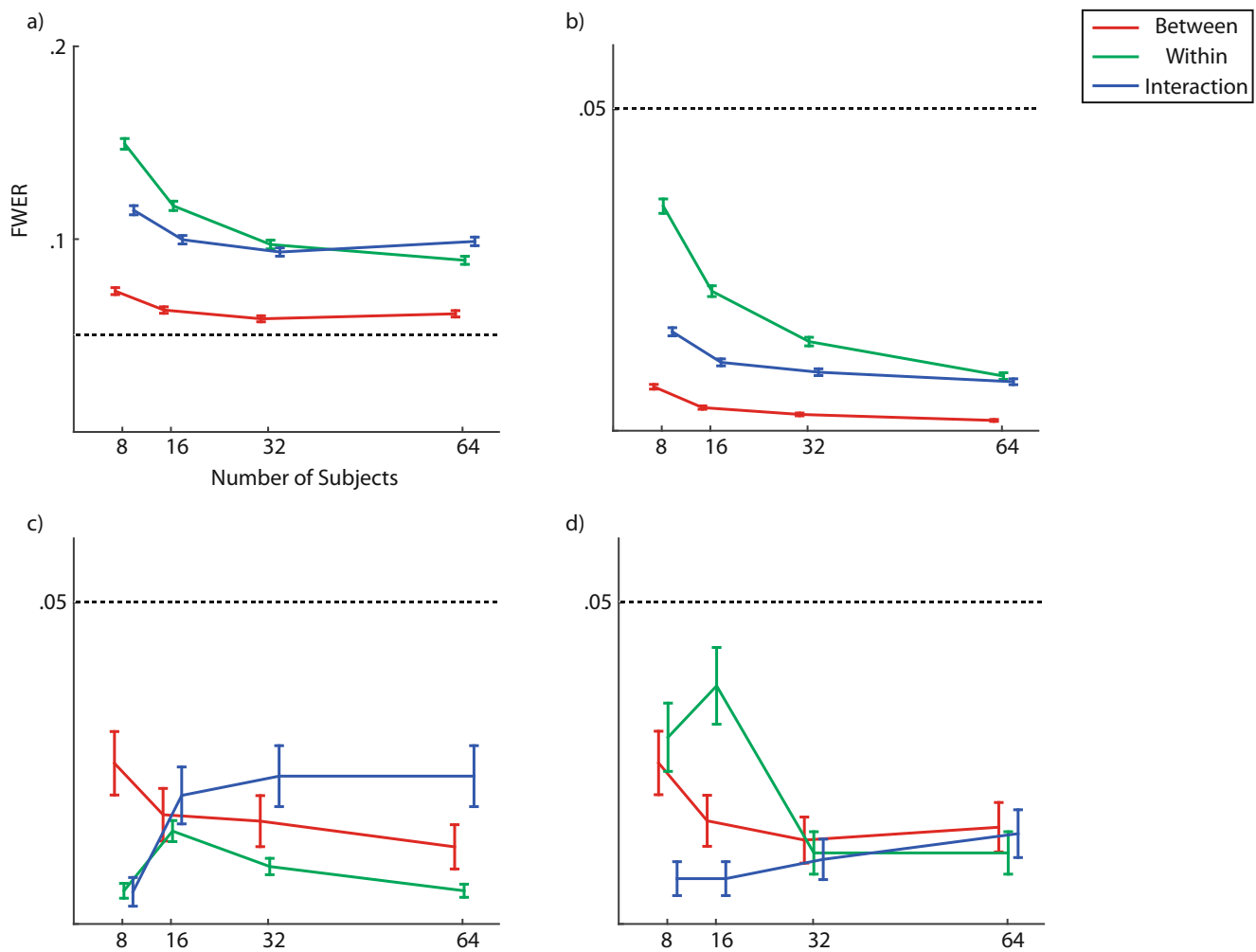
Fig. 2 Validation results of the proposed resampling procedure for statistical inference. (**a**) Family-wise error rates (FWERs) using the uncorrected parametric *p* values. All FWERs are significantly above .05. (**b**) FWERs using the permutation approach (Algorithm 1). (**c**) FWERs using the proposed bootstrap-clustering approach (Algorithm 2) thresholds on cluster mass. (**d**) FWERs using the proposed bootstrap-clustering approach (Algorithm 2) thresholds on cluster extent. Notice that the FWERs of panels **a** and **b** are computed at the pixel level (i.e., the proportions of false-positive pixels across simulations), and the FWERs of panels c and d are calculated at the test level (i.e., the percentages of any false positives per test for the 1,000 simulations). Error bar shows the standard errors

display the modeling results and perform statistical hypothesis testing. These main steps have their own independent GUIs: Create Fixation Matrix (Fig. 3b), Linear Mixed Model (Fig. 3c), and Display Results (Fig. 3d). Although most features of *i*Map4 could be obtained via these GUIs, we encourage advanced users to use command lines, especially for the additional options specification of the *LinearMixedModel* class. A short example of the command-line handling of the main functions is shown in Fig. 3e. A user guidebook containing the instructions for each step can be accessed via the Help button. We have also provided datasets with tutorial files to explain practically how to use *i*Map4. As a demonstration, two examples based on real and simulation data are given in the next section. MATLAB scripts of the examples are part of the *i*Map4 installation package.

Applications to real and simulation data

In the following examples, we illustrate *i*Map4's flexibility and power with two real data sets and a computer simulation. All material and codes presented here are available in the *i*Map4 installation package.

Example 1

We consider first a subset of participants from Bovet, Lao, Bartholomée, Caldara, & Raymond, (2016), as a demonstration

Fig. 3 The main graphical user interfaces of *i*Map4 (**a**–**d**) and example ▶ command lines handling the core functions (**e**). For more details, please refer to the online guidebook and demonstration codes

a)

b)

c)

d)
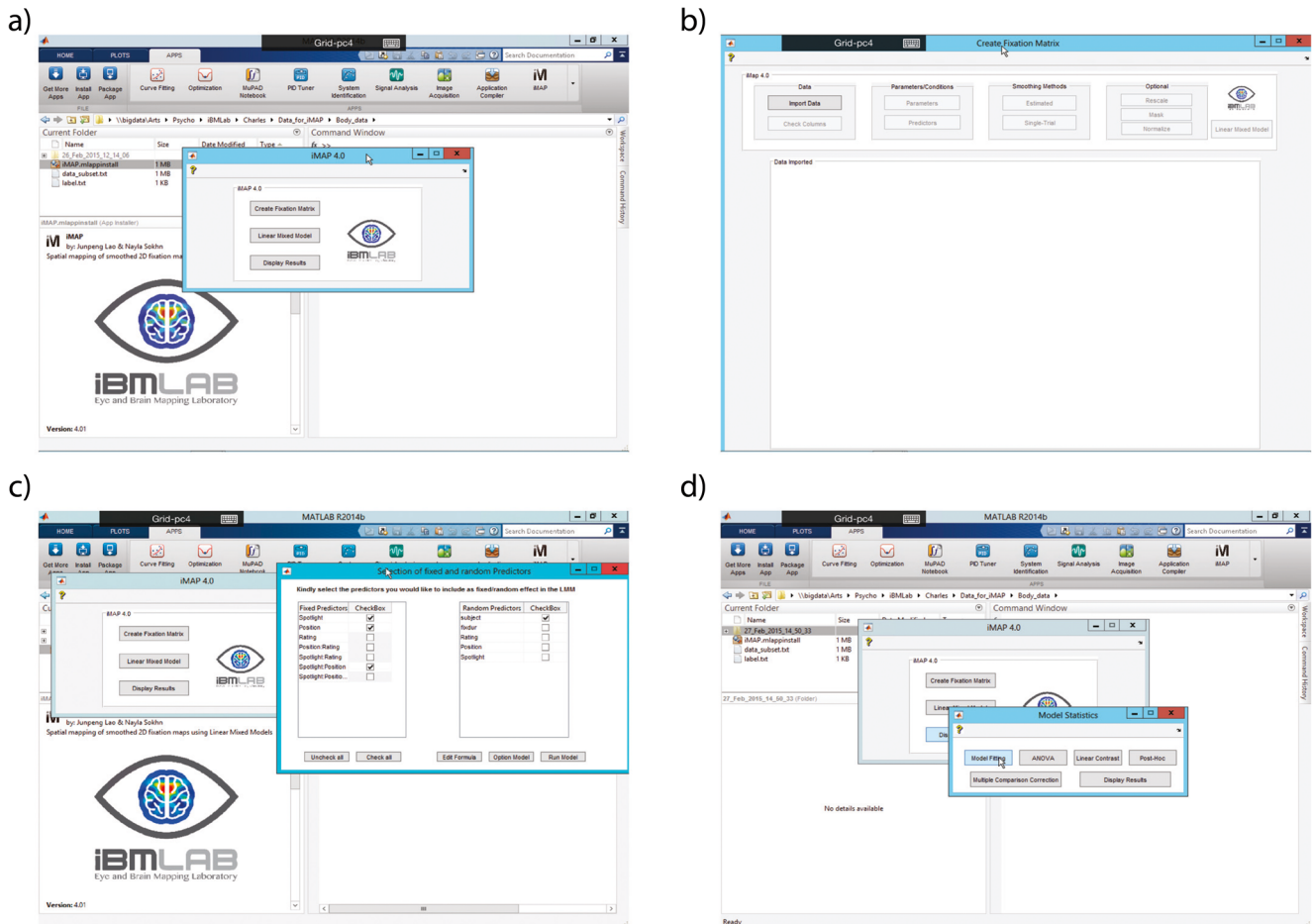
e)

```
>>  opt.singlepredi = 1; opt.parallelname = 'grid2';
>>  [LMMmap,lmexample] = imapLMM(FixMap, PredictorM, Mask, opt, ...
        'PixelIntensity ~ BetweenSbj + WithinSbj + BetweenSbj:WithinSbj + (fixDur|Subject)', ...
        'DummyVarCoding', 'effect', 'FitMethod', 'REML');   % model fitting
>>
>>  opt1.type = 'model';   % display model fitting
>>  [StatMap] = imapLMMcontrast(LMMmap, opt1);
>>  imapLMMdisplay(StatMap, 0, 'Backgroundimage.jpg');   % output figure
>>
>>  opt2.type = 'fixed'; opt2.alpha = .05;   % display ANOVA result for the fixed effect
>>  [StatMap] = imapLMMcontrast(LMMmap, opt2);
>>  imapLMMdisplay(StatMap, 0, 'Backgroundimage.jpg');   % output figure
>>  mccopt.methods = 'bootstrap'; mccopt.grouping = 'BetweenSbj'; mccopt.nboot = 1000;
        mccopt.bootopt = 2; % multiple comparison correction on cluster using bootstrap
>>  [StatMap_mcc] = imapLMMmcc(StatMap, LMMmap, mccopt, FixMap);
>>  imapLMMdisplay(StatMap_mcc, 0, 'Backgroundimage.jpg');   % output figure
>>
>>  [PostHoc]=imapLMMposthoc(StatMap_mcc, RawMap, LMMmap);    % perform post-hoc
```

of the analysis procedure in *i*Map4. A step-by-step demonstration is available in the user guidebook and example code.

In short, the dataset consists of eye movement data from 20 male observers during a gaze-contingent study. Observers viewed computer-rendered female bodies in different conditions and performed a behavioral task (i.e., subjective ratings of bodily attractiveness). This was a within-subjects design with two experimental manipulations: the viewing condition (three levels: *2° spotlight*, *4° spotlight*, or *natural viewing*) and body orientation (two levels: *front view* or *back view*). The aim of the study was to evaluate the use of visual information for bodily attractiveness evaluation in the male observers. Other details of the experiment can be found in the article.

Fixation durations were projected into the two-dimensional space according to their coordinates at the single-trial level. The fixation duration maps were first smoothed at 1° of visual angle. We used the "estimated" option by taking the expected values across trials within the same condition independently for each observer. To reduce the computation time, we downsampled the fixation map to 256*205 pixels and applied a mask to only model the pixels with average durations longer than half of the minimum fixation duration input.

Before proceeding to the modeling step, we visualized the preprocessed fixation maps and the descriptive statistics to get a sense of the data. For each of the categorical conditions, *i*Map4 outputs the mean fixation map for each level. Descriptive statistics for the following eye movement measures are saved in a matrix and will be plotted in a histogram or boxplot: number of fixations, sum of fixation durations

(total viewing time), mean fixation duration, total path length (total eye movement distance in pixels), and mean path length. See Fig. 4 for an example of the descriptive-results output.

We applied a full model on the fixation duration map without any spatial normalization:

$$
\begin{aligned}
\text{Pixel}_{\text{Intensity}(x, y)} \sim\ & \text{Viewing condition} \ + \ \text{Body orientation} \\
& + \ \text{Viewing condition} * \text{Body orientation} \\
& + \ (\text{fixation duration} \,|\, \text{subject}), \quad x, y \in \text{fixation map resolution}.
\end{aligned}
$$

Notice that the mean fixation duration for each condition and subject were treated as random effects to control for the variation across individuals. The parameters were fitted with restricted maximum likelihood estimation (ReML).

We encourage users to interpret the result from *i*Map4 in the following way. First, check the model fitting by displaying the model criteria. For example, Fig. 5a shows the R^2 values or multiple-correlation coefficients, which represent the proportions of variability in the fixation matrix explained by the fitted model. Interpretation of the result should be drawn with caution if the R^2 values are too low. The users can then proceed to test their hypotheses, such as through ANOVA or linear contrast, and perform multiple-comparisons corrections (Fig. 5b and c). A post-hoc analysis is applicable if any interaction is presented, or if any condition contains multiple levels. The user can select one or more significant area(s) as data-driven ROI(s) for the post-hoc analysis. *i*Map4 performs *t* tests between any pairs of categorical conditions within this ROI by using the raw input values from the nonsmoothed
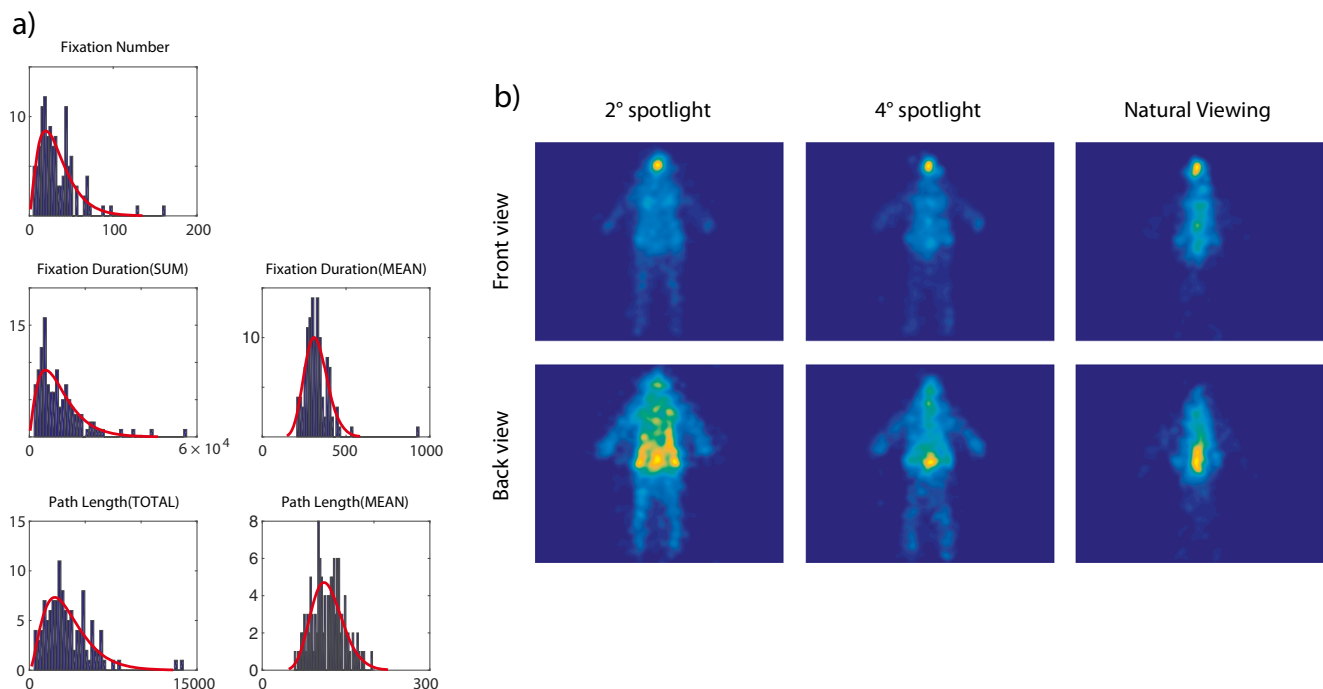

Fig. 4 Descriptive results from *i*Map4 on the real dataset. (**a**) Five eye movement measures plotted in histograms. In this case, fixation durations are in milliseconds and path lengths are in pixels. (**b**) Mean fixation maps of all levels of the categorical conditions
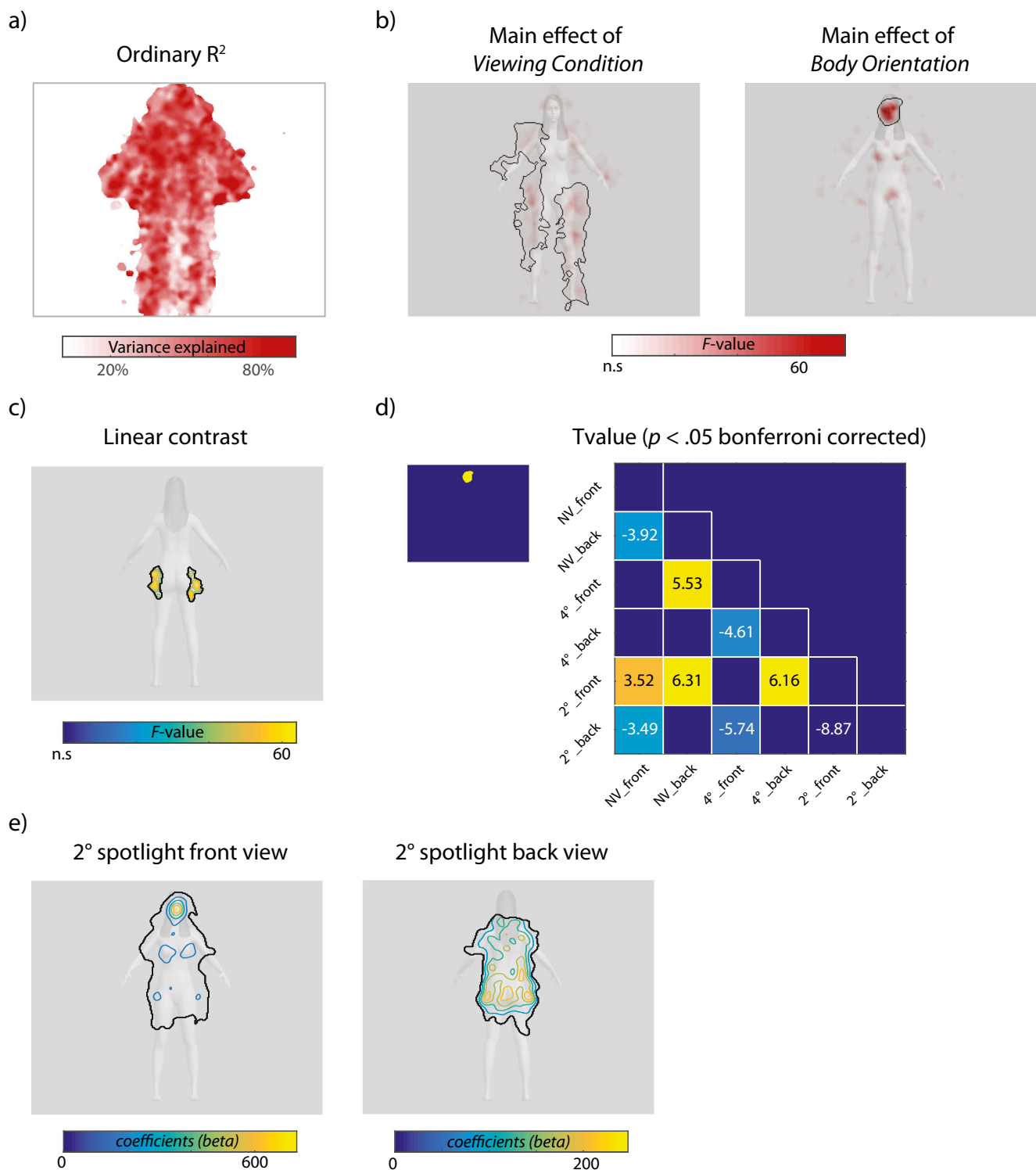
Fig. 5 *i*Map4 results for Bovet et al. (2016) with different output styles. (**a**) Ordinary R^2 values for the fitted model. (**b**) ANOVA results of the main effects and interaction. Here the intensity represents the F values. *i*Map4 only displays significant maps. (**c**) Statistical results of the linear contrast [*2° spotlight–natural viewing*] in the *back view* condition. Here the F value is represented on a contour map. (**d**) Post-hoc analysis in the selected mask. The mask is generated from the significant region of the *body orientation* effect (left panel). The *t* test results are shown in the matrix in the right panel (labeled conditions on the *y*-axis minus those labeled on the *x*-axis). Only significant results are shown ($p < .05$, Bonferroni corrected). (**e**) One-tailed *t* tests against the average over all fixation intensities for the *2° spotlight front view* and *2° spotlight back view* conditions. The solid black lines contain the significant regions for all of the panels above

fixation matrix (Fig. 5d). In addition, users can compute the above-average or above-chance fixation intensity for each categorical predictor (Fig. 5e).

Example 2

As a second demonstration, we reanalyzed the full dataset from one of our previous studies, Miellet, He, Zhou, Lao, and Caldara (2012).

Previous studies testing Western Caucasian (WC) and East Asian (EA) observers had shown that people deploy different eye movement strategies during free viewing of faces. WC observers fixate systematically toward the eyes and mouth, following a triangular pattern, whereas EA observers predominantly fixated at the center of the face (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Caldara, Zhou, & Miellet, 2010). Moreover, human observers can flexibly adjust their eye movement strategies to adapt to environmental constraints, as has been shown using different gaze-contingent paradigms (Caldara, Zhou, & Miellet, 2010; Miellet et al., 2012). In our 2012 study, we tested two groups of observers in a face task in which their foveal vision was restricted by a blind spot. This was a mixed design with the culture of the observers as the between-subjects factor (WCs or EAs) and the blind spot size as the within-subjects factor (four level: *natural viewing*, *2° blindspot*, *5° blindspot*, or *8° blindspot*). For more details of the experiment, please refer to Miellet et al. (2012).

Using *i*Map4, we created the single-trial 2-D fixation duration map and smoothed at 1° of visual angle. Importantly, to keep in line with Miellet et al. (2012), spatial normalization was performed by *z*-scoring the fixation map across all pixels independently for each trial (the results are identical without spatial normalization in this example). We also applied a mask generated with the default option. No down-sampling was performed. We then applied a full model on the single-trial fixation duration map made used of the "single-trial" option in *i*Map4:

$$\text{Pixel}_{\text{Intensity}(x,\,y)} \sim \text{ Observer culture} + \text{Blindspot size}$$
$$+ \text{ Observer culture*Blindspot size}$$
$$+ \text{ (1|subject)}, \quad x, y \in \text{fixation map resolution.}$$

Only the subject predictor was treated as a random effect, and the model was fitted using ML.

After model fitting, we performed an ANOVA to test the two main effects and their interactions. We applied a bootstrap clustering test using a cluster density of 1,000 as the criterion. We found a significant interaction and a main effect of *blind spot* size, but not a main effect of culture (see Fig. 6a). This result replicates the findings in Miellet et al. (2012). Moreover, by performing a linear contrast of the model coefficients, we reproduced Fig. 2 from Miellet et al. (2012). The results using *i*Map4 are shown in Fig. 6b.

Example 3

We also used simulated data to illustrate the use of *i*Map4 with continuous predictor. We created a dataset and manually introduced an effect between the numbers of fixations and the subjective rating on a single-trial level. Moreover, to maximize the simulation's efficiency, different linear relationships were introduced simultaneously. For each subject, we generated a data matrix through the two following steps:

- In a 4*4 grid, we introduced a different linear relationship in each cell between fixation number and subjective rating. Figure 7a shows the linear relationships we introduced for one subject. We varied the slope and the strength of the linear association. The correlation was strongest on the top row ($r = .9$), and there was no correlation on the bottom row ($r = 0$). The slope varied among [1, 0.4, –0.2, –0.8] across the columns. Note that each dot on a scatterplot represents one trial, and the dots with the same rating (value on the *x*-axis) across subplots belong to the same trial. The resulting matrices after this step were a one-dimensional array *Rating* and a two-dimensional matrix **P** (matrix size: *16 * number of trials*)
- The spatial locations of fixations were generated using linear Gaussian random fields. For each trial, we created a Gaussian mixture model *gm* using the *gmdistribution* class in MATLAB. The Gaussian mixture model *gm* contains 16 (4*4) 2-D Gaussian distribution components. The center of each component aligned with the center of each grid, and the covariance was an identity matrix with 1° of visual angle on the diagonal. Crucially, the mixing proportion of each component was decided by the column of the specific trial in **P**. A number of random fixations were then generated from this Gaussian mixture model *gm*. See Fig. 7b for a realization of one random trial for one subject.

The dataset contained 20 subjects performing 100 trials, each with an average fixation number of 58.02. Figure 7c shows the average map for fixation numbers. We fitted a simple model with ReML:

$$\text{Pixel\_Intensity}_{(x,\,y)} \sim 1 + \text{Rating}$$
$$+ \text{ (1|subject)}, \quad x, y \in \text{screen resolution.}$$

The significant regression coefficients of *Rating* are shown in Fig. 7d. *i*Map4 accurately rejected the null hypothesis for most conditions when there was a significant relationship. For the most robust effect ($r = .9$), *i*Map4 accurately estimated the coefficients. It also correctly reported a null result for $r = 0$. Moreover, *i*Map4 did not report any significant effect for the weakest relationship (slope = –0.2, $r = .3$), due to the lack of power. Indeed, further
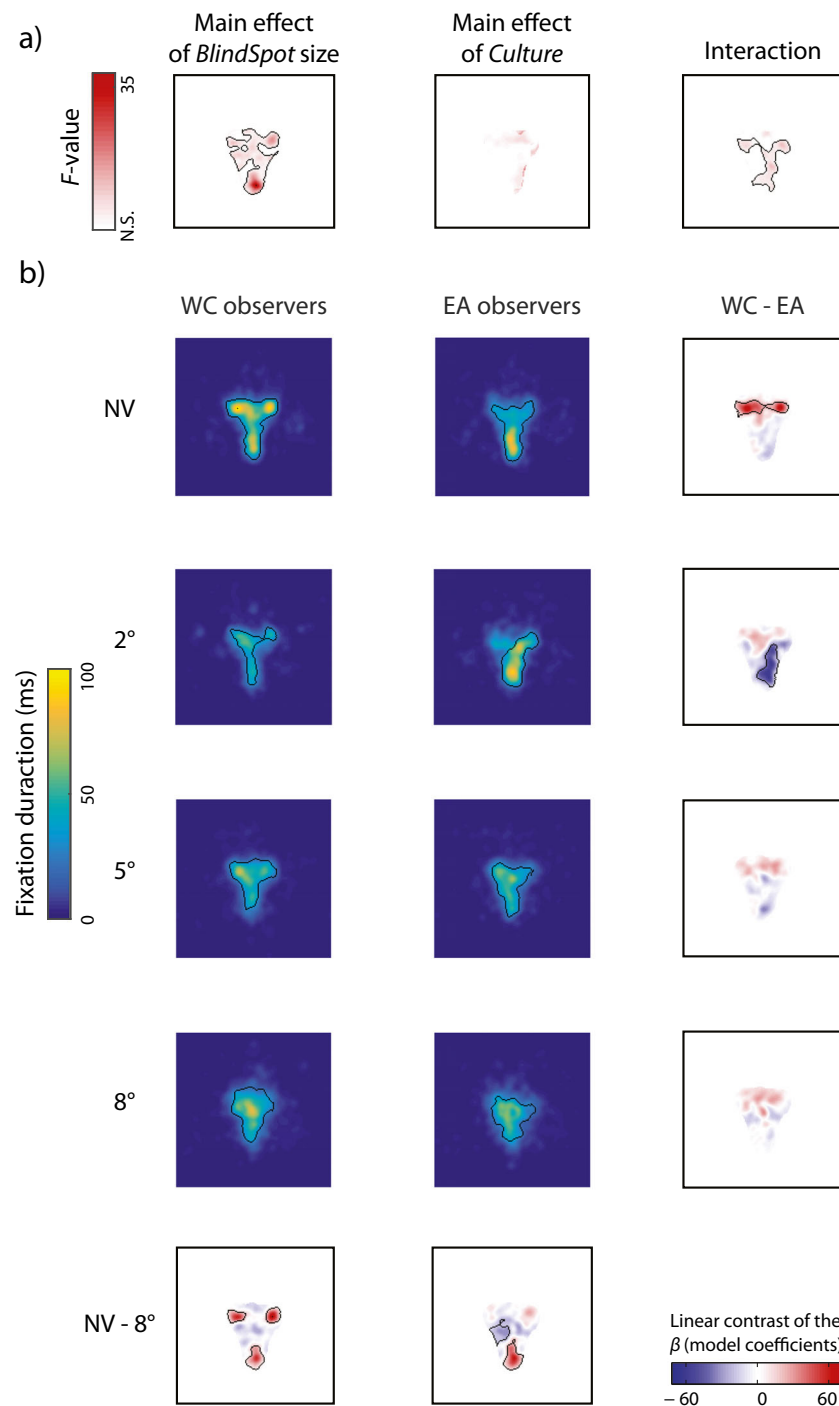
Fig. 6 *i*Map4 results for Miellet et al. (2012). (**a**) ANOVA results of the linear mixed model. (**b**) Replication of the Fig. 2 results for Miellet et al. (2012), using linear contrasts of the model coefficients. The solid black lines contain the significant regions for all of the panels above

simulations showed that increasing the numbers of fixations, trials, or subjects would lead to significance.

Discussion and future developments

In the present article, we have reported a major update of *i*Map, a toolbox for statistical fixation mapping of eye movement data.

While keeping unchanged the general data-driven philosophy of *i*Map, we significantly improved the underlying statistical engine, by incorporating pixel-wise LMMs and a variety of robust nonparametric statistics. Crucially, the new analysis pipeline allows for the testing of complex designs while controlling for a wide range of random factors. We also implemented a full GUI to make this approach more accessible to MATLAB beginners. Examples from empirical and

computer-simulated datasets showed that this approach has a slightly conservative FWER under H0, while remaining highly sensitive to actual effects (e.g., Fig. 6d). The present method represents a significant advance in eye movement data analysis, particularly for analyzing experimental designs using normalized visual stimuli. In fact, *i*Map4 uses a statistical inference method similar to those in fMRI and magnetoencephalography/EEG analysis. The interpretation of the statistical maps is simply done by looking at which stimulus features/pixels relate to the significant areas (after multiple-comparison correction). This procedure is similar to the interpretation of fMRI results: after a significant region is revealed, we can use its spatial coordinates to check in which part of the cortex the region activated above chance level is located.

As a powerful statistical tool, LMMs are gaining popularity in psychological research and have previously been applied in eye movement studies (e.g., Kliegl, Masson, & Richter, 2010). Similarly, particular cases of LMM, such as HLM or two-level models, are now standard data-processing approaches in neuroimaging studies. As a general version of HLMs, LMMs are much more flexible and powerful than other multilevel models. Most importantly, an exact same LMM could be applied to behavior, eye movement, and neuroimaging data, bridging these different measures to allow drawing more direct and complete conclusions.

However, there are both theoretical and practical challenges in using LMM for the statistical spatial mapping of fixation data. First, the fixation locations are too sparse to directly apply pixel-wise modeling. Similarly to previous versions of *i*Map, we used spatial smoothing of the fixation locations, a preprocessing step necessary to account for the measurement error of eyetrackers and the imprecision of the physiological system (i.e., the human eye). The second issue is selecting the appropriate hypothesis testing for LMM and the multiple-comparison problems caused by modeling massive number of pixels in nonbalanced designs. We addressed this issue by applying nonparametric statistics based on resampling and spatial clustering. Another important challenge is the constraint of computational resources. Parameter estimations using LMM, the pixel-wise modeling approach, and resampling techniques are very computationally demanding and time-consuming. To produce a useful but also usable tool, we adapted many advanced and novel algorithms, such as parallel computing. Preprocessing options such as downsampling and applying a mask also significantly decrease the computational time of *i*Map4.

The comparison among ROIs/areas of interest, iMap 2.0, and the current version

In classical eye movement data analyses, particularly those considering fixation locations, the main challenge for statsially identifying the regions that have been fixated above chance level lies in the fact that we are facing a high-dimensional data space. Mathematically, each pixel represents one dimension that could be potentially important. However, it is trivial to say that many of these dimensions are redundant and could be reduced to a particular set of representations or features. In other words, eye fixation data points are embedded in a high-dimensional pixel space, but they actually occupy only a subspace with much lower dimensionality (Belkin & Niyogi, 2003). Indeed, in similar high-dimensional datasets, a low-dimensional structure is often assumed and is naturally the main focus for investigation. Thus, by arbitrarily choosing one or multiple ROIs, one can represent the high-dimensional dataset as a low-dimensional manifold. The fixation map thus projects into this manifold, and all the pixels within the same ROI are then considered as being in the same dimension. In this case, each ROI represents one feature. Such a method is comparable to early neural network and many other linear dimension reduction methods in the machine-learning literature with hand-coded features (LeCun, Haffner, Bottou, & Bengio, 1999; Sorzano, Vargas, & Montano, 2014).

The early versions of *i*Map (1 and 2) adopted a similar logic, but relied on RFT to isolate data-driven features. Therefore, the fixation bias in each pixel was projected into a lower-dimensional subspace, resulting in fixation clusters. Second-level statistics were then computed at the cluster level instead of the pixel level to perform statistical inference (Miellet, Lao, & Caldara, 2014).

From *i*Map 3 onward, we took a very different approach. We used spatial clustering and multiple-comparison correction to avoid the use of second-level statistics to perform statistical inference. In *i*Map4, the fixation bias is similarly modeled on each pixel using a flexible yet powerful statistical model: the LMM. The LMM, in combination with nonparametric statistics and a spatial clustering algorithm, directly isolates the significant pixels. As a result, the *i*Map4 outputs can be interpreted intuitively and straightforwardly at the map level (i.e., by visualizing the areas reaching significance from the tested hypothesis).

Parameter settings and statistical choices

Our aim was and still is the development of a data-driven and fully automatized analysis tool. However, even in *i*Map4

Fig. 7 *i*Map4 results on the simulation dataset. (**a**) Linear relationships ▶ being introduced into the 4*4 grid. The *x*-axis shows the *z*-scored rating, and the *y*-axis shows the expected number of fixations. The slopes between *y* and *x* are the same within each column ([1, 0.4, –0.2, –0.8], respectively), whereas the correlation rho is the same within each row ([0.9, 0.6, 0.3, 0]), respectively). (**b**) One realization of a random trial for one subject. The left panel shows the raw fixation locations; the right panel shows the smoothed fixation number map. (**c**) The average fixation map across all trials for the 20 subjects. (**d**) Estimated relationships between rating and fixation number (regression coefficients). The black circles indicate statistical significance
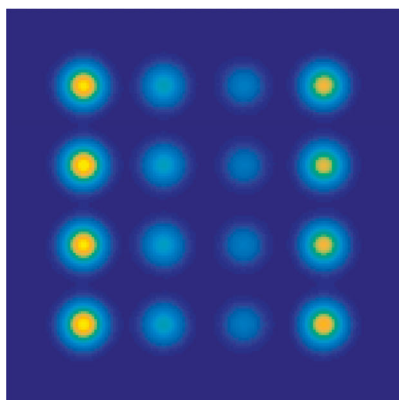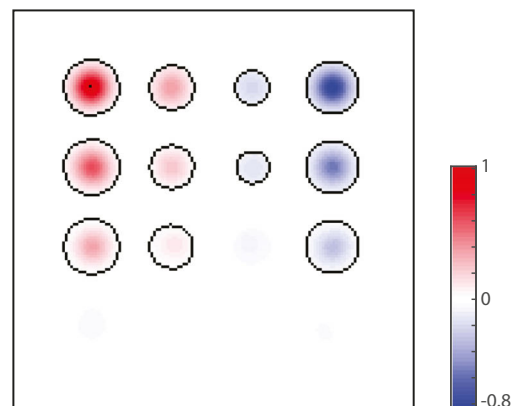
a)

$y = x + c$ $y = .4 * x + c$ $y = -.2 * x + c$ $y = -.8 * x + c$

$r = 0.9$

$r = 0.6$

$r = 0.3$

$r = 0$

Response

Rating
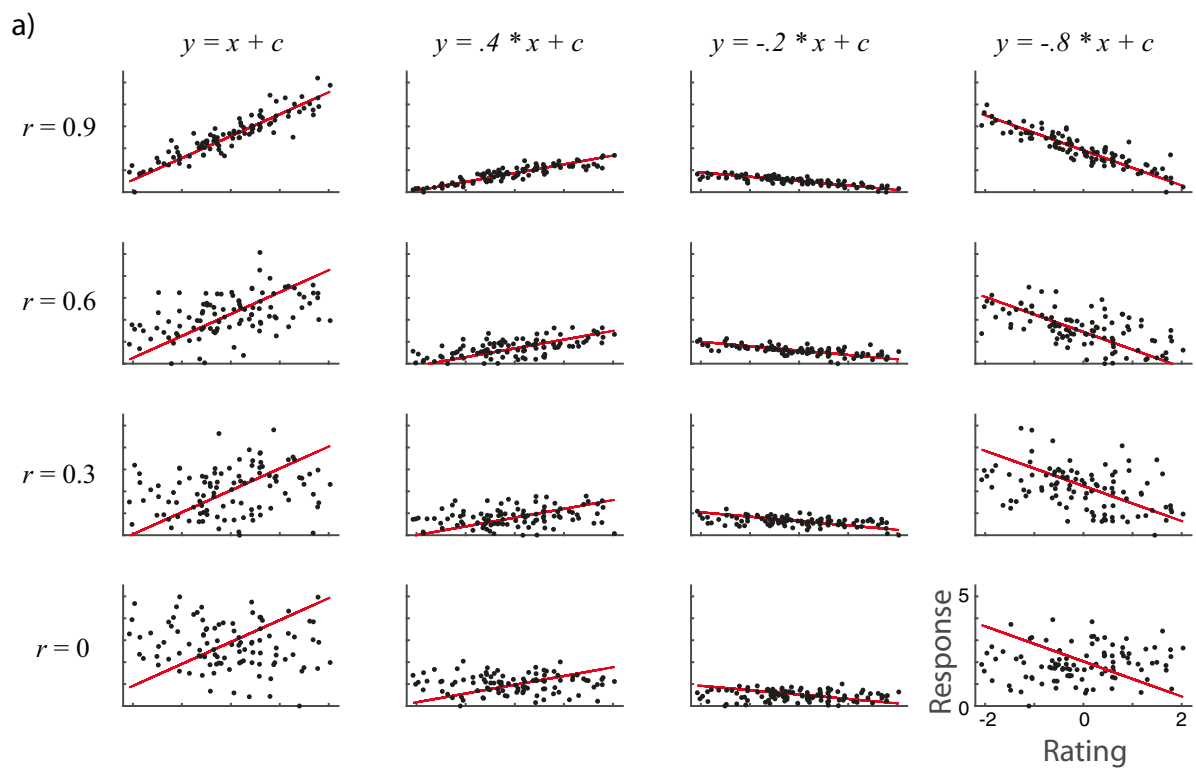
b)
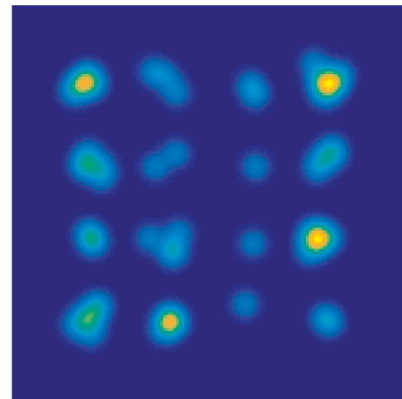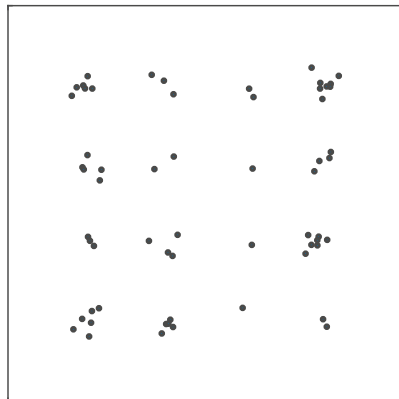
c) Fixation number map (mean)

d) Coefficient estimation *(beta)*

some of the parameters in the analysis rely on a user's expertise and subjective choices, which thus should be considered carefully before use. These parameters include the kernel size for the smoothing procedure, the spatial down-sampling and masking, the spatial normalization, and the choice of statistics.

The rationale for the determining the kernel size for the smoothing procedure has been previously discussed (Caldara & Miellet, 2011), and the majority of the arguments we put forward in this previous article still hold true. Here, we would remind users that the spatial smoothing procedure mainly resolves the sparseness of fixation data. It also partially accounts for the spatial covariance, which is ignored in univariate pixel-wise modeling. Finally, it accounts for the recording errors from eyetrackers, such as drift during the calibration, pupil-size variations, and so forth.

We also recommend that users perform down-sampling and apply a mask before modeling their data. This step is important to reducing computational demands (time, memory, etc.). In general, we recommend that the down-sampling factor not be bigger than half of the smoothing kernel size. In other words, if the FWHM of the Gaussian kernel is 10 pixels, the rescale factor should be less than 5. We are currently running further simulations and validations to investigate the best parameters under different settings, and hopefully will provide a statistical data-driven solution for this choice in future updates.

Spatial normalization (via a z-scored or probability map) is available as an option in iMap4. Spatial normalization used to be a standard preprocessing procedure in previous versions of iMap. However, the hypotheses tested on raw fixation duration/number maps are fundamentally different from their spatially normalized versions. Importantly, after spatial normalization, the interpretation of results should be drawn on a spatially *relative* bias instead of on the *absolute* differences. Of course, if the viewing duration in each trial is constant within an experiment, spatial normalization will not make any difference.

For iMap4 we developed two main, nonparametric statistics based on resampling techniques. It is worth noting that different applicability comes with the choice of permutation tests versus bootstrap spatial-clustering tests. In our own experience during empirical and simulation studies, permutation tests are more sensitive for studies with small sample sizes; the bootstrap-clustering approach usually gives more homogeneous results but is biased toward bigger clusters. We suggest that users adopt a "wisdom of crowds" approach and look at the agreement among different approaches before concluding on the data analysis (Marbach et al., 2012). Nonconvergent results should be interpreted carefully.

An alternative to pixel-wise approaches

In recent years, other frameworks have been also developed to model eyetracking data (Boccignone, 2015). One

such approach is the aforementioned Poisson point process model (Barthelmé et al., 2013). It is a well-established statistical model when the point (fixation) occurrence is the main concern. Under some transformation, the Poisson point processes model of fixation occurrences could be expressed and modeled as a logistic regression, making it straightforward to apply using conventional statistical software (Barthelmé & Chopin, 2015). For example, Nuthmann and Einhäuser (2015) made use of logistic mixed models to determine the influence of low- and high- visual properties in scene images on eye movements. Moreover, smooth effect and spatial covariants could be captured by applying regression splines in a generalized additive model, as demonstrated in Barthelmé and Chopin (2015).

Importantly, the point process model addresses different questions than does iMap. It is most appropriate when the effect of spatial location is considered irrelevant, a nuisance effect, or a fixed intercept (see, e.g., Barthelmé & Chopin, 2015; Nuthmann & Einhäuser, 2015). As a comparison, in iMap the parameters of interest are location specific, varying from pixel to pixel. In other words, the differences or effects among different conditions are location-specific, forming a complex pattern in two dimensions. These high dimension effects are more natural and easy to model using a pixel-wise model, as in iMap4.

Conclusion and future development

In conclusion, we have presented an advanced eye movement analysis approach using LMMs and nonparametric statistics: iMap4. This method is implemented in MATLAB with a user-friendly interface. We aimed to provide a framework for analyzing spatial eye movement data with the most sophisticated statistical modeling to date. The procedure described in the present article currently represents our best attempt to conform with the conventional null-hypothesis testing, while providing options for robust statistics. We currently are still working on many improvements, including functions to compare different fitted models, statistics on the random-effect coefficients, and replacing LMMs with generalized LMMs for modeling fixation numbers (Bolker, Brooks, Clark, Geange, Poulsen, Stevens, & White, 2009). In the future, we will also switch our focus to Bayesian statistics and the generative model (such as the Gaussian process) in an effort to develop a unified model of statistical inference for eye movement data (Jaynes & Bretthorst, 2003).

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Barthelmé, S., & Chopin, N. (2015). The Poisson transform for unnormalised statistical models. *Statistics and Computing, 25,* 767–780.

Barthelmé, S., Trukenbrod, H., Engbert, R., & Wichmann, F. (2013). Modeling fixation locations using spatial point processes. *Journal of Vision, 13*(12), 1. doi:10.1167/13.12.17

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation, 15,* 1373–1396.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3,* e3022.

Boccignone, G. (2015). *Advanced statistical methods for eye movement analysis and modeling: A gentle introduction.* arXiv preprint. arXiv:1506.07194.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution, 24,* 127–135. doi:10.1016/j.tree.2008.10.008

Bovet, J., Lao, J., Bartholomée, O., Caldara, R., & Raymond, M. (2016). Mapping females' bodily features of attractiveness. *Scientific Reports, 6,* 18551. doi:10.1038/srep18551

Caldara, R., & Miellet, S. (2011). *i*Map: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods, 43,* 864–878. doi:10.3758/s13428-011-0092-x

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "*spotlight*" reveals universal information use for face recognition. *PLoS ONE, 5,* e9708. doi:10.1371/journal.pone.0009708

Christensen, R. (2011). *Plane answers to complex questions: The theory of linear models.* Berlin: Springer.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34,* 455–470.

Friston, K. J., Stephan, K. E., Lund, T. E., Morcom, A., & Kiebel, S. (2005). Mixed-effects and fMRI studies. *NeuroImage, 24,* 244–252.

Jaynes, E. T., & Bretthorst, G. L. (2003). *Probability theory: The logic of science.* Cambridge: Cambridge University Press.

Kherad-Pajouh, S., & Renaud, O. (2010). An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. *Computational Statistics and Data Analysis, 54,* 1881–1893.

Kherad-Pajouh, S., & Renaud, O. (2015). A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs. *Statistical Papers, 56,* 947–967. doi:10.1007/s00362-014-0617-3

Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18,* 655–681.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). *Object recognition with gradient-based learning Shape, contour and grouping in computer vision* (pp. 319–345). New York: Springer.

Liversedge, S., Gilchrist, I., & Everling, S. (2011). *The Oxford handbook of eye movements.* Oxford: Oxford University Press.

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., . . . Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods, 9,* 796–804. doi:10.1038/nmeth.2016

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2011). *Generalized, linear, and mixed models.* New York: Wiley.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When East meets West: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research, 5*(5), 1–12.

Miellet, S., Lao, J., & Caldara, R. (2014). An appropriate use of iMap produces correct statistical results: A reply to McManus (2013) "iMAP and iMAP2 produce erroneous statistical maps of eye-movement differences.". *Perception, 43,* 451–457.

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences, 1339,* 82–96.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 9.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: The analysis of functional brain images: The analysis of functional brain images.* San Diego: Academic Press.

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: A toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational Intelligence and Neuroscience, 2011,* 11.

Pernet, C. R., Latinus, M., Nichols, T. E., & Rousselet, G. A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods, 250,* 85–93. doi:10.1016/j.jneumeth.2014.08.003

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Berlin: Springer.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks: Sage.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44,* 83–98.

Sorzano, C. O. S., Vargas, J., & Montano, A. P. (2014). *A survey of dimensionality reduction techniques.* arXiv preprint. arXiv:1403.2877.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage, 92,* 381–397.

Supplemental material

The full *i*Map4 toolbox and the data samples shown in this article are freely available for download (http://ibmlab.github.io/iMap4/). A full user guide can be found separately from https://github.com/iBMLab/iMap4/wiki