*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], and Roberto Caldara[1*]

[1]*Department of Psychology, University of Fribourg, Fribourg, Switzerland*

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

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

* Corresponding authors:

Roberto Caldara or Junpeng Lao

Department of Psychology, University of Fribourg

Faucigny 2, 1700 Fribourg, Switzerland

Email: junpeng.lao@unifr.ch or roberto.caldara@unifr.ch

Tel: +41 (26) 300 76 36 / Fax: +41 (26) 300 97 12

Number of pages: 22

Number of figures: 6

Number of words:

Abstract: 239

Introduction: 720

Discussion: 1438

Keywords: Eye Movement Analysis, Statistical mapping, Linear Mixed Model

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. Compared to signals of contemporary neuroscience measures, such as M/EEG and fMRI, 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 and Miellet, 2011) which tackles this issue by implementing a statistical framework comparable to those developped in state-of-the-art neuroimaging data processing toolboxes. *i*Map4 uses univariate, pixel-wise Linear Mixed Models (LMM) on the smoothed fixation data, with the flexibility of coding for multiple between- and within- subject comparisons and performing all the possible linear contrasts for the fixed effects (main effects, interactions, etc.). Importantly, we also introduced novel non-parametric 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.

Introduction

Human beings constantly move the eyes to redirect their gaze in order to sample visual information of interest from the environment. Eye fixations deliver visual inputs with the highest precision from the location falling in the fovea to the human visual cortex, 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 interested to behavioral researchers, psychologists and neuroscientists. Moreover, fixation mapping has a wide range of practical applications in fields such as commercial and consumer industry (Duchowski, 2002).

Conventional eye movement data analyses rely on the probability of occurrence of fixations and saccades (or their characteristics such as duration or length) within predefined Regions of Interest (ROIs), at best defined a priori but often also a posteriori based on data exploration, which inflates the type one error rate. Another issue with ROI is of course that other important information not included in the ROI are discarded. In a continuous effort to circumvent limitations in the use of ROI approach (for a detailed discussion on this point, see Caldara & Miellet, 2011), we previously developped 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 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 were suffering from a major limitation: they could only contrast two conditions at a time.

A major revision of the toolbox was necessary in order to allow the analysis of the more complex experimental designs routinely used in the field. One of the most suitable and obvious statistical solutions to overcome this problem is to implement a General linear model, a wide spread approach both in behavioral and neural imaging data analyses. In fact, many modern hypotheses testing, such as the t-test, ANOVA, regression, etc. belong to the family of general linear models. However, eye movement data are a sparse production of visual perceptual sampling. Contrary to neuroimaging data, there are many empty cells with little to no data points across the tested space (e.g., all the pixels in the image). This caveat engenders a statistical problem when the same statistical inference procedure is applied on each pixel regardless or whether there are missing data or not. To account for the sparseness and the high

variation of spatial eye movement data, we develop 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, the users can apply a massive univariate, pixel-wise Linear Mixed Model on the smoothed fixation data with the subject considered as a random effect, which offers the flexibility to code for multiple between- and within- subject comparisons. Our approach allows the users to perform all the 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 non-parametric statistical test based on resampling (permutation and bootstrap spatial clustering) to assess the statistical significance of the linear contrasts (Pernet et al., 2014; Winkler et al., 2014).

In the next section, we briefly describe the key concepts of the Linear Mixed Model (LMM) approach. We then introduce the *novel* non-parametric statistical approach on the fixed effects we implemented in *i*Map4, which uses a resampling procedure and spatial clustering. We also report a validation of the proposed resampling procedures. We then illustrate how *i*Map4 can be used, with 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 will outline the key elements and concepts of Linear Mixed Models in comparison with General Linear Models (GLM) and Hierarchical Linear Models (HLM). Mixed models are a complex subject and many underlying details are beyond the scope of this paper. For a general thoughtful introduction to mixed models, users of the toolbox should refer to Raudenbush & Bryk (2002) and McCulloch, Searle & Neuhaus (2011). The users may also wish to consult the documentation and the help files of the *LinearMixedModel* class in Matlab Statistics Toolbox™ for details about parameter estimation and available methods (http://www.mathworks.com/help/stats/linearmixedmodel-class.html).

Statistical hypothesis testings that make use of the analysis of variance (regression, *t*-test, ANOVA, ANCOVA, 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 the GLM:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_t x_{ti} + \varepsilon_i \qquad (1)$$

$\varepsilon_i \sim \mathrm{N}\,(0, \sigma^2)$

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 of zero and variance of σ^2. Alternatively, the general linear models (Eq. 1) could be expressed in matrix form:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2)$$

$$\boldsymbol{\varepsilon} \sim \mathbf{N}\,(\mathbf{0}, \boldsymbol{\sigma^2 I})$$

where matrix $\mathbf{X} = [x_1, x_2, \ldots, x_t]$ is the design matrix, and $\boldsymbol{I}$ is an n-by-n identity matrix (n being the total number of observation). Usually, one of the columns in $\mathbf{X}$ is 1 so that the model including a constant or intercept coefficient that represents the overall mean. It is worth noting that the design matrix could be parameterized in different way. In conventional psychology or behavioral researches, a sigma-restricted parameterization is often applied. In a sigma-restricted design matrix, $\mathbf{X}$ is full rank and invertible, and the degrees of freedom is equal to the number of columns. In comparison, many neuroimaging analysis software prefer a cell-mean model or an over-parameterized design matrix in single subject level (Penny et al., 2011; Pernet et al., 2011). In these software, they use an over-parameterized design matrix and its solution to Eq. (2) is given by projecting the response vector $\mathbf{Y}$ to the pseudoinverse of the design matrix $\mathbf{X}$. The form of design matrix is important as it codes different experiment design and the intended statistics testing. In *i*Map4, the design matrix of the fixed effect can be cell-mean model, sigma-restricted model (for the Type III ANOVA) or the offset from reference model (for the Type I ANOVA).

The coefficient estimations $\widehat{\boldsymbol{\beta}}$ 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 $\boldsymbol{\beta^{sr}}$ into two parts (Kherad-Pajouh & Renaud, 2010):

$$\mathbf{X^{sr}} = [\mathbf{X_1} \quad \mathbf{X_2}]\,, \boldsymbol{\beta^{sr}} = \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix}$$

where $\mathbf{X_1}$ and $\boldsymbol{\beta_1}$ are the components of interest with the corresponding hypotheses:

$$\mathrm{H}_0 : \boldsymbol{\beta_1} = \mathbf{0} \text{ vs. } \mathrm{H}_1 : \boldsymbol{\beta_1} \neq \mathbf{0} \qquad (3)$$

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

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-}\mathbf{X}^{\mathbf{T}} \qquad (4)$$

$$\mathbf{X_{resid}} = (\boldsymbol{I} - \mathbf{X_2}\left(\mathbf{X_2^T X_2}\right)^{-}\mathbf{X_2^T})\,\mathbf{X_1} \qquad (5)$$

$$\mathbf{H_{resid}} = \mathbf{X_{resid}}\left(\mathbf{X_{resid}^T X_{resid}}\right)^{-}\mathbf{X_{resid}^T} \qquad (6)$$

$$\text{dfe} = \text{Number of observation} - \text{rank}(\mathbf{X}) \qquad (7)$$

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

where **H** represents the hat matrix of the linear model (2); it projects the response vector **Y** onto the column space of **X**. $\mathbf{H_{resid}}$ is the hat matrix of the hypothesis (3) and dfe is the model degrees of freedom. *F* has a Fisher–Snedecor distribution $\mathcal{F}$ (rank(X_1) , dfe).

As comparison, in an over-parameterized design matrix or cell-mean model design matrix with design matrix $\mathbf{X^{cm}}$ and the vector of parameters $\boldsymbol{\beta}^{\mathbf{cm}}$, the statistics of various effects is performed by linear combinations of the coefficient $\boldsymbol{\beta}^{\mathbf{cm}}$. For example, the equivalent hypothesis of (3) could be expressed as:

$$\mathrm{H}_0 : \mathbf{c} * \boldsymbol{\beta}^{\boldsymbol{cm}} = \mathbf{0} \text{ vs. } \mathrm{H}_1 : \mathbf{c} * \boldsymbol{\beta}^{\boldsymbol{cm}} \neq \mathbf{0} \qquad (9)$$

where rank(**c**) = rank($\mathbf{X_1}$) and $\mathbf{c}*\boldsymbol{\beta}^{\mathbf{cm}} = \boldsymbol{\beta}_{\mathbf{1}}$ 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 covariant matrix of $\boldsymbol{\beta}^{\mathbf{cm}}$ (for simplicity, in Eq. 10 and 11 we denote $\mathbf{X} = \mathbf{X^{cm}}$):

$$\text{MSE} = \mathbf{Y^T}\,(\boldsymbol{I} - \mathbf{H})\,\mathbf{Y} \,/\, (\text{dfe}) \qquad (10)$$

$$F = \frac{(\mathbf{c}*\boldsymbol{\beta}^{\boldsymbol{cm}})^{\mathbf{T}}(\text{MSE}*\mathbf{c}(\mathbf{X^T X})^{-}\mathbf{c^T})^{-}(\mathbf{c}*\boldsymbol{\beta}^{\boldsymbol{cm}})}{\text{rank(c)}} \qquad (11)$$

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

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

considered as the additional random effect. This type of model is one of the most widely used models in fMRI and EEG.

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

$$y_{ij} = \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{tj}x_{tij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{N}\,(0, \sigma^2)$$

$$\beta_{1j} = \beta_{10} + b_{1j}\,, \quad b_{1j} \sim \mathrm{N}\,(0, \sigma_1^2)$$

$$\vdots$$

$$\beta_{tj} = \beta_{t0} + b_{tj}\,, \quad b_{tj} \sim \mathrm{N}\,(0, \sigma_t^2) \qquad (12)$$

where *j* stand for the *j*th subject. After substituting the subject-level parameters in the first-level model, Eq. (12) becomes

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

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

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 Linear Mixed Models:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \qquad (13)$$

$$\boldsymbol{b} \sim \mathbf{N}\left(\mathbf{0}, \boldsymbol{\sigma^2 D}\right), \boldsymbol{\varepsilon} \sim \mathbf{N}\left(\mathbf{0}, \boldsymbol{\sigma^2 I}\right), \boldsymbol{b}\ and\ \boldsymbol{\varepsilon}\ are\ independent\ from\ each\ other$$

where $\boldsymbol{\sigma^2 D}$ is the covariance matrix for the random effects. In the example here $\boldsymbol{D}$ would be a *j*-by-*j* identity matrix. Alternative form of the Eq. (12) as applied in LIMO EEG or SPM could be found in Friston, Stephan, Lund, Morcom, & Kiebel (2005, Eq. 1).

HLM are specific cases of LMM. 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 models compare to hierarchical models. In additions, the fixed effects and the 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. Assumed that model (13) has the error covariance matrix $\boldsymbol{R}$: $var(\boldsymbol{Y}|\boldsymbol{b}) = \boldsymbol{R}$, model (13) is equivalent to $y \sim \mathbf{N}(\boldsymbol{X\beta},\ \boldsymbol{V}), \boldsymbol{V} = \boldsymbol{ZDZ^T} + \boldsymbol{R}$. The estimation of the fixed effects β requires prior knowledge of $\boldsymbol{V}$, which is usually unavailable. In practice, the variance component $\boldsymbol{V}$ is commonly replaced by an estimation $\widehat{\boldsymbol{V}}$ from 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, model fitting procedure of LMM is implemented in major statistical packages (e.g., R and Stata) by solving the Henderson's mixed model equation. *i*Map4 called Matlab class *LinearMixedModel* from Statistics Toolbox ™ (R2013b or above) to estimate the coefficients (fixed effect $\boldsymbol{\beta}$ and random effect $\boldsymbol{b}$) and the covariance matrix $\boldsymbol{V}$ with various options (key concepts in regarding to parameter estimations could be found in Matlab documentation: http://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 ($\boldsymbol{D}$) could take the form of full covariance matrix, diagonal covariance matrix, or other symmetry structure.

Statistical inferences in LMM are also much more complex compared to GLM. In a balanced design, or with the variance component $\boldsymbol{V}$ known, hypothesis testing of the fixed effect follow Eq. 8 or 11 as an exact test. However, in an unbalanced design with random effect, no exact *F*-statistics is available, as there are biases in the estimation that usually result in an unknown distribution of *F* (Kherad-Pajouh & Renaud, 2014). Although *F*- and *t*- values are available as approximate test in most statistical package, Baayen et al (2008) discouraged the usage of *t*- or *F*- statistics and especially the report of *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 of the application of likelihood ratio test (e.g., method of model fitting, selection of the reduced model). Moreover, running multiple copies of similar LMM is computationally expensive, especially in the context of pixel-wise testing such as in *i*Map4.

Besides the practical problem in statistical inferences of 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 technique for null hypothesis of statistical test (NHTS) as suggested in the neuroimaging analysis for GLM or HLM (Pernet, et al., 2014; Winkler,

et al., 2014). Nonparametric statistics using Monte Carlo simulation are ideal both for parameter estimation and hypothesis testing (Baayen et al, 2008; Kherad-Pajouh & Renaud, 2014). 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 fixation especially its location distribution. For example, Barthelmé, Trukenbrod, Engbert, & Wichmann, (2013) recommend using the point process framework to inference how fixations are distributed in space. While we endorse this fruitful approach and its Bayesian nature, here we aim to resolve this problem from an opposite perspective. Instead of inferring from the spatial distribution of the fixation, we infer on each location in the search space (i.e., each pixel of within the eye tracker recordable range or each pixel in the visible stimuli). In other words, we try to answer the question: *"How long is this pixel being fixated (or what is the probability of this pixel to be fixated) in the function of the experimental conditions?"*. Formally, by applying mixed models independently on each pixel, we have:

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

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

The complete procedure as implemented in *i*Map4 is explained in Figure 1. Eye movement data for each participant is concatenated into one input data matrix. *i*Map4 first partition the data matrix into a fixation characteristic matrix (*red box*) and an experiment condition information matrix (*green box*). The fixation characteristic matrix contains fixation spatial location (*x* and *y*), fixation duration, and order index of each fixation. The experiment condition matrix contains index of each subject, index of each trial/item, and 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:

$Kernel \sim N\ (0\ , \sigma^2 I)$, where *I* is a two by two identity matrix and the full width at half maximum (FWHM) of the Kernel is *1° visual angle* as the default setting.

This step is essential to account for the spatial uncertainty of eye movement recording (both mechanical and physiological) and the sparseness of the fixation locations. The Gaussian kernel could also be replaced by other 2D spatial filters to best suit the research question.

The resulting smoothed fixation map is a 3D matrix. The last two dimensions of the fixation matrix are the size of the stimuli/search space. 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, *i*Map4 provides 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 potentially 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 3D 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 *LinearMixedModel* class. For each modeled pixel, *i*Map4 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 on the LMM. Additional modeling specifications, as well as other model parameters including LMM 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 on *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 large 2D pixel space, we applied non-parametric 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 $\boldsymbol{\beta}$ are operated on the fixed effect related variances. To do so, we simply removed the variance associated with the random effects from the response matrix:

$$\boldsymbol{Y_{fixed}}(s) = \boldsymbol{X\beta}(s) + \boldsymbol{\varepsilon}(s) = \boldsymbol{Y}(s) - \boldsymbol{Zb}(s) \qquad (15)$$

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

For any permutation test, *i*Map4 performs the following algorithms on $\boldsymbol{Y_{fixed}}$ for each pixel:

Algorithm 1:

For a given hypothesis or linear contrast $\boldsymbol{c}$ (as in Eq. 9), *i*Map4

- Performs a linear transformation on the design matrix $\boldsymbol{X}$ to get a new design matrix $\boldsymbol{M}$ so that the partitioning of $\boldsymbol{M} = [\boldsymbol{M1}, \boldsymbol{M2}]$**.** Then *i*Map4 computes the new coefficients by projecting $\boldsymbol{Y_{fixed}}$ to the pseudoinverse of $\boldsymbol{M}$**.** The design matrix $\boldsymbol{M}$ is created so that the original hypothesis testing is equivalent to the hypothesis regarding $\boldsymbol{M1}$ coefficients**.** The matrix transformation and partition are the same as the algorithm described in Winkler et al (2014, appendix A)
- Computes the residuals related to the hypothesis by subtracting the variance accounted by $\boldsymbol{M2}$ from $\boldsymbol{Y_{fixed}}$ to get $\boldsymbol{Y_{rr}}$
- Fits $\boldsymbol{Y_{rr}}$ to $\boldsymbol{M}$ by solving $\boldsymbol{Y_{rr}} = \boldsymbol{M\beta_m} + \boldsymbol{\varepsilon}$, and get the statistics value $\boldsymbol{F_{rr}}$ of $\boldsymbol{M1}$ according to Eq. (10) and (11). Note that to replicate the original hypothesis testing on the fixed effect, the new contrast $\boldsymbol{c'}$ is just to partition $\boldsymbol{M}$ into $\boldsymbol{M1}$ and $\boldsymbol{M2}$
- Permutes the rows of the design matrix $\boldsymbol{M}$ to obtain the new design matrix $\boldsymbol{M}^*$
- Fits $\boldsymbol{Y_{rr}}$ to $\boldsymbol{M}^*$ and gets the $\boldsymbol{F_{rr}}^*$ of $\boldsymbol{M1}^*$
- Repeats the previous two steps for a large number of times (*k resamplings/repetitions)*, the *p*-value is then defined as Eq. (16). Importantly, the FWER corrected *p*-value is computed by comparing the largest $\boldsymbol{F_{rr}}^*$ across all tested pixels in one resampling with the original $\boldsymbol{F_{rr}}$.

$$p = \frac{(\#\ \boldsymbol{F_{rr}}^* \geq\ \boldsymbol{F_{rr}})}{k} \qquad (16)$$

Algorithm 1 is a simplified version of Winkler et al (2014, Algorithm 1): the resampling table includes permutation but not sign-flipping. Sign-flipping 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 balance design with no missing value and only subject as the random effect. As previously shown in Kherad-Pajouh and Renaud (2014), 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 sphericity of the error. However, there are strict requirements in order to achieve this goal: careful transformation and partition of both fixed and random effects design matrices, and removal of the random effects related to $\boldsymbol{M2}$ (Kherad-Pajouh and Renaud, 2014). In *i*Map4, we perform an approximation version by removing all random effects to increase the efficiency and the 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 purposed algorithm are not compromised.

*i*Map4 performs the following algorithm on $\boldsymbol{Y_{fixed}}$ for each pixel as the bootstrap clustering approach:

Algorithm 2:

- For each unique categorical variable, *i*Map4 removes the conditional expectations from $\boldsymbol{Y_{fixed}}$ for each pixel. A random shuffling is then performed on the centered data to acquire $\boldsymbol{Y_c}$, 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 $\boldsymbol{Y_c}$ have expected values equal to zero.
- Randomly draws with replacement from $\{\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y_c}\}$ an equal number of subjects $\{\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{Y_c}^*\}$
- Fits $\boldsymbol{Y_c}^*$ to $\boldsymbol{X}^*$ by solving $\boldsymbol{Y_c}^* = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$. For a given hypothesis or linear contrast $\boldsymbol{c}$ (as in Eq. 9), *i*Map4 computes the statistics value $\boldsymbol{F}^*$ according to Eq. (10) and (11) and their parametric *p*-value under the GLM framework.
- Thresholds statistical maps $\boldsymbol{F}^*$ at $p^* \leq .05$ and records the desired maximum cluster characteristics across all significant clusters. Cluster characteristics considered are: cluster mass (summed *F* value within a cluster), cluster extent (size of the cluster), or 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 statistics map $\boldsymbol{F}$ at $p \leq .05$ and compares the selected cluster characteristic with the value of the null distribution corresponding to the 95^{th} 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; 2014) if only subject intercept is considered as the random effect. In addition,

Algorithm 2 extents the philosophy and approach presented by Pernet et al. (2011; 2014) to non-hierarchical mixed-effect models.

It is worth noting that we implemented in *i*Map4 a high-performance algorithm to minimize the computational demands for the large amount of resampling. Model fitting in both resampling approaches makes use of Ordinary Least Squares. The inversion of the covariance matrixes (as 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 inversion of the covariance matrix. More details of these algribra simplifications could be found in the *imapLMMresample* function in *i*Map4.

Other multiple comparison correction methods such as Bonferroni correction, False Discovery Rate (FDR), or Random Field Theory (RFT) could also be applied. A Threshold-Free Cluster Enhancement (TFCE) algorithm could also be applied on the statistical (*F*-value) maps as an option after the permutation and bootstrap clustering procedure (Smith & Nichols, 2009).

We performed a validation study to access the type I error rate when applying the permutation and bootstrap clustering approach for hypothesis testing. We used a balanced repeated measurement ANOVA design with a two-level between-group factor and a three-level within-group factor. A total population of 134 observers (67 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 the 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 *i*Map4 performed on this dataset is considered as false alarm (Type I error).

The validation procedure follows the steps below: we first randomly sampled without replacement a balanced number of subjects from both groups. We then ran *i*Map4 under the default setting and perform hypothesis testing on the two main effects and the interaction. To estimate the Family-wise error rate (FWER), we computed the frequency of significant output under different statistics and MCC setting. Preliminary results based on 1000 randomizations on a sample size of $n \in [8, 16, 32, 64]$ showed that with an alpha of .05, the family-wise error rates are indeed all under .05 using non-parametric statistics (see Figure 2b for permutation test, 2c & 2d for bootstrap clustering test). More simulations considering a wider range of scenarios will be required to understand fully the behavior of the proposed approaches, although cluster stats are likely to behave as in Pernet et al. (2014).

Graphical User Interface and Command Line Handling

*i*Map4 runs on Matlab 2013b and above, as it requires some essential function and class in these versions of Image Processing Toolbox™ and Statistics Toolbox™. *i*Map4 will execute in parallel on multi-cores or distributed workers when available.

We recommend users to install *i*Map4 as a Matlab Application. The users can call *i*Map4 directly in the Matlab command window after installation. A general graphical user interface (GUI) will open upon >>*iMAP* called in the command window or launching the app (Figure 3a). The users can then import the fixation data, load preprocessed data matrix for LMM, or display the modeling results and perform statistical hypothesis testing. These main steps have their own independent GUIs: Prepare Fixation Map (Figure 3b), Linear Mixed Model (Figure 3c), and Display Results (Figure 3d). Although most features of *i*Map4 could be obtained via GUI, we encourage the advanced users to use command lines especially for the additional options specification of the *LinearMixedModel* class. A short example of the command lines handling of the main functions is shown in Figure 3e. A user guidebook containing the instructions for each step can be accessed via the help button. We 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.

Application to Real and Simulation Data

In the following, we illustrate *i*Map4 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,J., Lao, J., Bartholomée, O., Caldara, R., and Raymond, M., (2016) Mapping females' bodily features of attractiveness* as a demonstration 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 twenty male observers during a gaze-contingent study. Observers viewed computer rendered female bodies in different conditions and performed a behavioral task (i.e., subjective rating of bodily attractiveness). This is a within-subject design with two experimental manipulations: the viewing condition (three level: *2° spotlight, 4° spotlight,* or *natural viewing*) and body orientation (two level: *front view* or *back view*). The aim of the study is to evaluate the visual information use for bodily attractiveness evaluation in the male observers. Other details of the experiment can be found in the paper.

Fixation durations were projected into the two-dimensional space according to their coordinates at the single-trial level. Fixation duration maps were first smoothed at 1° of visual angle. We used the "estimated" option by taking the expected values across trial within the same condition independently for each observer. To reduce the computational time, we down-sampled the fixation map to 256*205 pixels, and applied a mask to only model the pixels with average duration larger 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 plot in a histogram or boxplot: number of fixations, sum of fixation durations (total viewing time), mean fixation duration, total path length (total eye-movements distance in pixel) and mean path length. See Figure 4 as an example of the descriptive results output.

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

$$Pixel_{Intensity_{(x,y)}} \sim Viewing\ condition + Body\ orientation + Viewing\ condition * Body\ orientation + (fixation\ duration \mid subject)$$

$$x, y \in fixation\ map\ resolution$$

Notices that 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 the users to interpret the result from *i*Map4 as the following. First, check the model fitting by displaying the model criterions. For example, Figure 5a shows the R-squared values or multiple correlation coefficients, which represents the proportion of variability in the fixation matrix explained by the fitted model. Interpretation of the result should be drawn with caution if the R-squared values are too low. The users can then proceed to test their hypotheses, such as ANOVA or linear contrast, and perform multiple comparisons correction (Figure 5b and 5c). A post-hoc analysis is applicable if any interaction presented, or any condition contains multiple levels. The user can select one or more significant area(s) as data-driven ROI(s) for the post-hoc. *i*Map4 performs *t*-tests between any pairs of categorical conditions within this ROI using the raw input values from the non-smoothed fixation matrix (Figure 5d). In addition, the users can compute the above average or above chance fixation intensity for each categorical predictor (Figure 5e).

Example 2

As a second demonstration, we reanalysed the full dataset from one of our previous paper *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*.

Previous studies testing Western Caucasian (WC) and East Asian (EA) observers showed that people deploy different eye movement strategy during free-viewing of faces. Western Caucasian observers fixate systematicly towards the eyes and mouth, following a triangular pattern, whereas East Asian observers perfominatly 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 movment strategy to adapt to the environmental constraints, as shown using different gaze-contingent paradigm (Caldara, Zhou, & Miellet, 2010; Miellet, He, Zhou, Lao, & Caldara, 2012). In our 2012 study, we tested two groups of observers in a face task where their foveal vision were restricted by a blindspot. This is a mixed design with the culture of the observers as the between-subject factor (WCs or EAs) and the blindspot size as the within-subject factor (four level: *natural viewing, 2° blindspot, 5° blindspot,* or *8° blindspot*). For more details of the experiment, please find Miellet, et al (2012).

Using *i*Map4, we created the single-trial 2D 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 result is 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:

$$Pixel_{Intensity_{(x,y)}} \sim Observer\ culture + Blindspot\ size + Observer\ culture * Blindspot\ size + (1|\ subject) \quad x, y \in fixation\ map\ resolution$$

Only the predictor of subject was treated as random effects and the model was fitted with maximum likelihood estimation (ML).

After model fitting, we perform ANOVA to test the two main effects and their interactions. We applied a bootstrap clustering test using cluster dense as criteria with 1000. We found a significant interaction and the main effect of *Blindspot* size, but not the main effect of culture (see Figure 6a). This result replicates the finding in Miellet, et al (2012). Moreover, by performing linear contrast of the model coefficients, we reproduced the figure 2 as in Miellet, et al (2012). The result using *i*Map4 is shown in Figure 6b.

Example 3

We also used simulated data to illustrate *i*Map4 usage in the single-trial approaches. 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 efficiency, different linear relationships were introduced simultaneously. For each subject, we generated a data matrix following 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 the strongest on the top row (r = 0.9) 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 scatter plot represents one trial, and the dots with the same rating (value on the x-axis) across subplots belong to the same trial. The resulting matrix after this step was a one dimension array *Rating* and a two dimension 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 sixteen (4*4) 2d Gaussian distribution components. The center of each component aligned with the center of each grid, while 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 figure 7b for a realization of one random trial for one subject.

The data set contained 20 subjects performing 100 trials each with an average fixation number of 58.02. Figure 7c shows the average map for fixation number. We fitted a simple model with restricted maximum likelihood estimation (ReML):

$$Pixel\_Intensity_{(x,y)} \sim 1 + Rating + (1 \mid subject) \qquad x, y \in screen\ resolution$$

The significant regression coefficients of *Rating* are shown in Figure 7d. *i*Map4 accurately rejected the null hypothesis for most conditions when there was a significant relationship. For the most robust effect (r= 0.9), *i*Map4 accurately estimated the coefficients. It also correctly reported null result for r = 0. Moreover, *i*Map4 did not report any significant effect for the weakest relationship (slope = -0.2, r = 0.3) due to the lack of power. Indeed, further simulations

showed that increasing either the number of fixations, trials, or subjects would lead to significance.

Discussion and Future developments

In the present paper we reported a major update of *i*Map, a toolbox for statistical fixation mapping of eye movement data. While keeping unchanged the general philosophy of *i*Map, we significantly improved the underlying statistical engine, by incorporating pixel-wise Linear Mixed Models and a variety of robust non-parametric statistics. Crucially, the new analysis pipeline allows the testing of complex designs while controling for a wide range of random factors. We also implemented a full graphical user interface to make this approach more accessible to Matlab beginners. Examples from empirical and computer simulated datasets showed that this approach has a slightly conservatively controlled family-wise error rate under H0, while remaining highly sensitive to actual effects (e.g., Figure 6d). The current method represents a significant advance in eye movement data analysis, particularly for experimental designs using normalized visual stimuli. In fact, *i*Map4 uses a similar statistical inference as in fMRI and M/EEG analysis. The interpretation of the statistical maps is simply done by looking at which stimuli 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 coordinate to check which part of the cortex the region activated above chance level is located.

As a powerful statistical tool, LMM 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 a standard data processing approach in neuroimaging studies. As a general version of HLMs, LMMs are much more flexible and powerful than other multi-level models. Most importantly, an exact same LMM could be applied to behavior, eye movement, and neuroimaging data, bridging these different measures together to draw more direct and complete conclusions.

However, there are both theoretical and practical challenges in using LMM for the statistical spatial mapping of fixation data. Firstly, the fixation locations are too sparse to apply pixel-wise modeling directly. 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 the eye-trackers and the imprecision of the physiological system (i.e., the human eye). The second issue is the appropriate hypothesis testing for LMM and the multiple comparison problems caused by modeling massive number of pixels in non-balanced designs. We addressed this issue by applying non-parametric 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 details like down-sampling and applying a mask also significantly help to decrease the computational time of *i*Map4.

The comparison among ROI/AOI, iMap 2.0, and the current version

In classical eye movement data analyses, particularly those considering fixation locations, the main challenge lies in the fact that we are facing a high-dimensional data space. Mathematically, each pixel represents one dimension that could potentially be 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 high-dimensional pixel space, but they actually only occupy a subspace with much lower dimensionality (Belkin & Niyogi, 2003). Indeed, in similar high dimensional datasets a low dimensional structure is often assumed and 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 in the same dimension. In this case, each ROI is representing one feature. Such a method is comparable to early neural network and many other linear dimension reduction methods in 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 the Random Field Theory to isolate data-driven features. Therefore, the fixation bias in each pixel was projected into a lower-dimensional subspace, resulting in fixation clusters. The 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 in order to perform statistical inference. In *i*Map4, the fixation bias is similarly modeled on each pixel using a flexible yet powerful statistical model: the Linear mixed model. The LMM, in combination with non-parametric statistics and a spatial clustering algorithm, directly isolate 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 there are still some parameters in the analysis that rely on 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 in 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. Here, we would like to remind users that the spatial smoothing procedure mainly resolves the sparseness of fixation data. It also partially accounts for the spatial covariance that is ignored in univariate pixel-wise modeling. Finally, it accounts for the recording errors from the eye trackers, such as drift during the calibration, pupil-size variations, etc.

We also recommend users to perform down-sampling and apply a mask before modeling their data. This step is important to reduce the computational demands (time, memory, etc.). In general, we recommend that the down-sampling factor is not bigger than half of the smoothing kernel size. In another word, 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 provide a statistical data-driven solution for this choice in future updates.

Spatial normalization (Z-scored map or probability map) is available as an option in *i*Map4. Spatial normalization used to be a standard preprocessing procedure in previous versions of *i*Map. However, the hypotheses testing on raw fixation duration/number maps are fundamentally different compared to their spatially normalized version. Importantly, after spatial normalization, the results interpretation should be drawn on the spatially *relative* bias instead of the *absolute* differences. Of course, if the viewing duration in each trial is constant within an experiment, spatial normalization will not make any difference.

*i*Map4 developed two main non-parametric statistics based on resampling techniques. It is worth noting that different applicability comes with the choice of permutation tests vs. 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 homogenous results, but is biased toward bigger clusters. We suggest the users to adopt a "wisdom of the crowd" approach and look at the

agreement among different approaches before concluding on the data analysis (Marbach et al., 2012). Non-convergent results should be interpreted carefully.

Alternative to pixel-wise approaches

In recent years, other frameworks have been also developed to model eye tracking data (Boccignone, 2015). One of 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 occurrence could be expressed and modeled as a logistic regression, making it straightforward to apply using conventional statistical software (Barthelmé & Chopin, 2015). For example, Nuthmann & Einhauser (2015) made use of logistic mixed model to show the influence of low and high visual features in scene image on fixation selection. Moreover, smooth effect and spatial covariant could be captured by applying regression splines in a generalized additive model, as demonstrated in Barthelmé & Chopin (2015).

Importantly, the point process model address different questions compare to *i*Map. It is the most appropriate when the effect of spatial location is considered as irrelevant, nuisance effect, or a fixed intercept (e.g., see Barthelmé & Chopin, 2015; Nuthmann & Einhauser, 2015). As comparison, in *i*Map the parameters of interested are conditioned on some pixels/regions of the image. In another word, the differences or effects among different conditions are location specific, forming a complex pattern in 2D. These high dimension effects are more natural and easy to model using a pixel-wise model as in *i*Map4. **Conclusion and future development**

In conclusion, we have presented an advanced eye-movement analysis approach using LMM and non-parametric statistics: *i*Map4. This method is implemented in Matlab with a user-friendly interface. We aim to provide a framework for analyzing spatial eye-movement data with the most sophisticated statistical modeling to date. The procedure described in the current paper is our best current attempt to keep in line with the conventional null-hypothesis testing, while providing options for robust statistics. There are still many improvements we are currently working on, including functions to compare different fitted models, statistics on the random effect coefficients, and replacing LMM with a Generalized Linear Mixed Model (GLMM) for modeling fixation numbers. 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).

Acknowledgments

The authors declare no competing financial interests. This study was supported by the Swiss National Science Foundation (no 100014_138627) awarded to RC.

Supplemental material

The full *i*Map4 toolbox and the data sample as shown in this paper are freely available for download (http://ibmlab.github.io/iMap4/). A full user guidebook could be downloaded separately from https://github.com/iBMLab/iMap4/blob/master/iMap4%20Guidebook.pdf

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*(4), 390-412.

Barthelmé, S., Trukenbrod, H., Engbert, R., & Wichmann, F. (2013). Modeling fixation locations using spatial point processes. *Journal of vision, 13*(12), 1.

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

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

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation, 15*(6), 1373-1396.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3(8), e3022.*

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

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

Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior research methods, 43*(3), 864-878.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting Culture Under the *'Spotlight'*Reveals Universal Information Use for Face Recognition. *PLoS ONE, 5*(3), e9708.

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*: Springer.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455-470.

Friston, K. J., Stephan, K. E., Lund, T. E., Morcom, A., & Kiebel, S. (2005). Mixed-effects and fMRI studies. *Neuroimage, 24*(1), 244-252.

Jaynes, E. T., & Bretthorst, G. L. (2003). *Probability Theory: The Logic of Science*: 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 & Data Analysis, 54*(7), 1881-1893.

Kherad-Pajouh, S., & Renaud, O. (2014). A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs. *Statistical Papers*, 1-21.

Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18*(5), 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): Springer.

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

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., et al. (2012). Wisdom of crowds for robust gene network inference. *Nature methods, 9*(8), 796-804.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2011). *Generalized, Linear, and Mixed Models*: 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(2):5,* 1-12

Miellet, S., Lao, J., & Caldara, R. (2014). An appropriate use of i Map produces correct statistical results: a reply to McManus (2013)"i MAP and i MAP2 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*(1), 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*: 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. (2014). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods*(0).

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*: SAGE Publications.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 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*(1), 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.

Figure Legend

Figure 1, Illustration of the procedure in *i*Map4. The input data matrix is partitioned into 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 dimensions Gaussian Kernel function $Kernel \sim N\,(0\,, \sigma^2 \boldsymbol{I})$. After estimating the fixation bias of each condition independently for all the observers (by taking the expected values across trial within the same condition), *i*Map4 models the 3D smoothed fixation map (*item*xSize*ySize*) independently for each pixel using a LMM. The result is saved as a Matlab structure in *LMMmap*. *i*Map4 offers many parametric and non-parametric methods for hypothesis testing and multiple comparison correction.

Figure 2, Validation result of the proposed resampling procedure as statistical inference. a) The family-wise error rate using the uncorrected parametric *p*-value. All FWER are significantly above .05. b) The family-wise error rate using the permutation approach (Algorithm 1). c) The

family-wise error rate using the proposed bootstrap clustering approach (Algorithm 2) thresholds on cluster mass. d) The family-wise error rate using the proposed bootstrap clustering approach (Algorithm 2) thresholds on cluster extent. Notice that the FWER of a) and b) are computed at pixel level (i.e., the proportion of false positive pixels across simulations), while the FWER of c) and d) are calculated at test level (i.e., the percentage of any false positive per test for the 1000 simulation). Error bar shows the standard error.

Figure 3, The main GUI of *i*Map4 (a - d) and example command lines handling of the core functions (e). For more details please refer to the online guidebook and the demonstration codes.

Figure 4, Descriptive results from *i*Map4 on the real data set. a) Five eye movement measures plot in a histogram. In this case, fixation duration is in millisecond and path length is in pixel. b) Mean fixation map of all levels of the categorical conditions

Figure 5, *i*Map4 results of Bovet et al. (2015) with different output styles. a) Ordinary R-squared value 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) The statistical results of the linear contrast [*2° spotlight - natural viewing*] in *back view* condition. Here the *F*-value is represented on a contour map. d) The post-hoc analysis in the selected mask. The mask is generated from the significant region of *Body Orientation* effect (left panel). The *t*-test results are shown in the matrix in the right panel (labeled condition in 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 condition *2° spotlight front view* and *2° spotlight back view*. The solid black line circles the significant region for all the above figures.

Figure 6, *i*Map4 results of Miellet et al. (2012). a) ANOVA result of the linear mixed model. b) Replication of figure 2 in Miellet et al. (2012) using linear contrast of the model coefficients. The solid black line circles the significant region for all the above figures.

Figure 7, *i*Map4 results on the simulation data set. a) The 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 slope between *y* and *x* are the same within each column ([1, 0.4, -0.2, -0.8] respectively), while 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 location; right panel shows the smoothed fixation number map. c) The average fixation map across all trial for the 20 subjects. d) Estimated relationship between rating and fixation number (regression coefficient). The black circles indicate statistical significance.

Figure 1

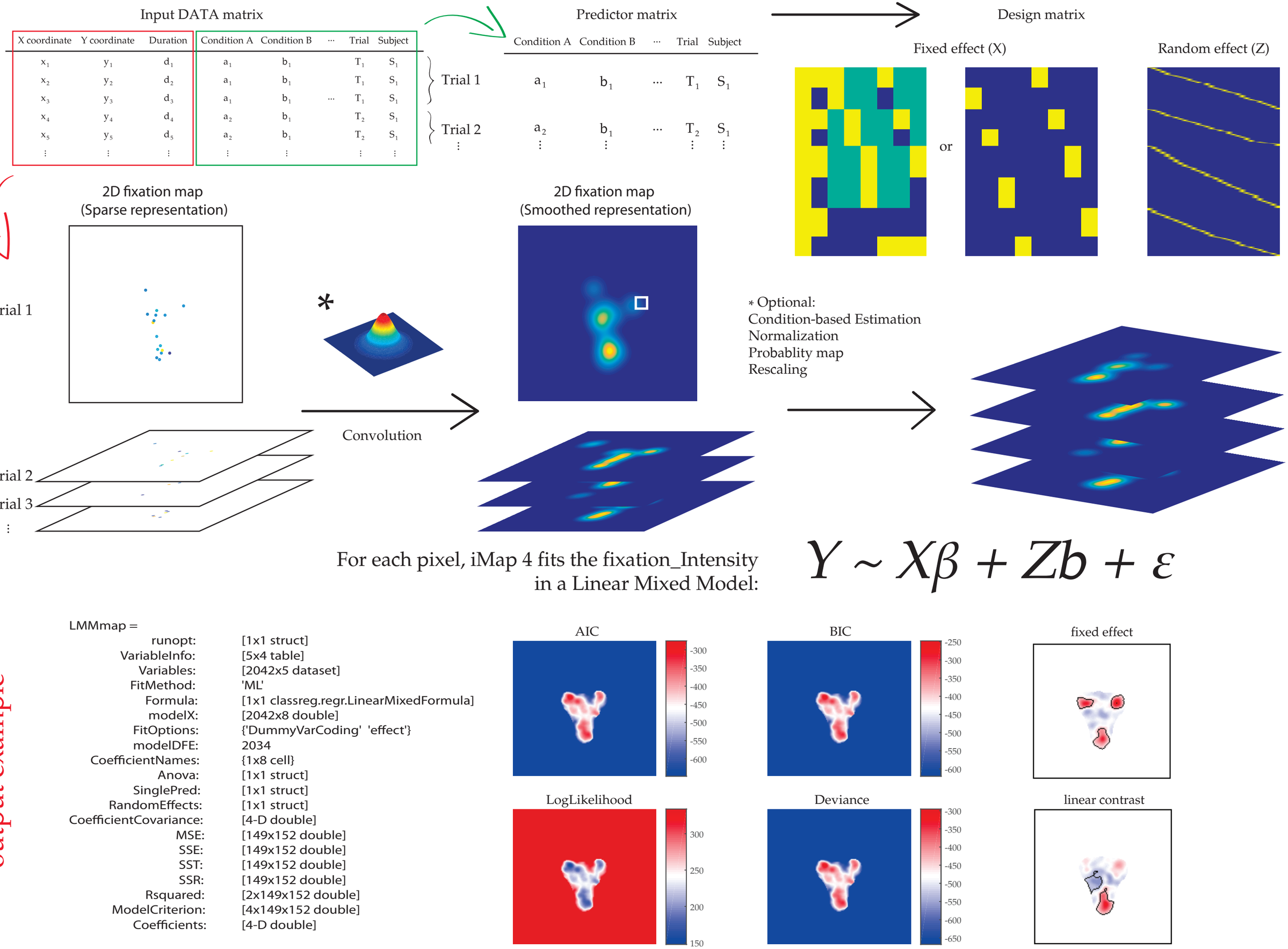

Input DATA matrix
X coordinate
Y coordinate
Duration
Condition A
Condition B
...
Trial
Subject
Trial 1
Trial 2
Predictor matrix
Design matrix
Fixed effect (X)
or
Random effect (Z)
2D fixation map
(Sparse representation)
2D fixation map
(Smoothed representation)
Trial 1
Trial 2
Trial 3
Convolution
* Optional:
Condition-based Estimation
Normalization
Probablity map
Rescaling
For each pixel, iMap 4 fits the fixation_Intensity
in a Linear Mixed Model:
$Y \sim X\beta + Zb + \varepsilon$
output example
LMMmap =
runopt: [1x1 struct]
VariableInfo: [5x4 table]
Variables: [2042x5 dataset]
FitMethod: 'ML'
Formula: [1x1 classreg.regr.LinearMixedFormula]
modelX: [2042x8 double]
FitOptions: {'DummyVarCoding' 'effect'}
modelDFE: 2034
CoefficientNames: {1x8 cell}
Anova: [1x1 struct]
SinglePred: [1x1 struct]
RandomEffects: [1x1 struct]
CoefficientCovariance: [4-D double]
MSE: [149x152 double]
SSE: [149x152 double]
SST: [149x152 double]
SSR: [149x152 double]
Rsquared: [2x149x152 double]
ModelCriterion: [4x149x152 double]
Coefficients: [4-D double]
AIC
BIC
fixed effect
LogLikelihood
Deviance
linear contrast

Figure 2

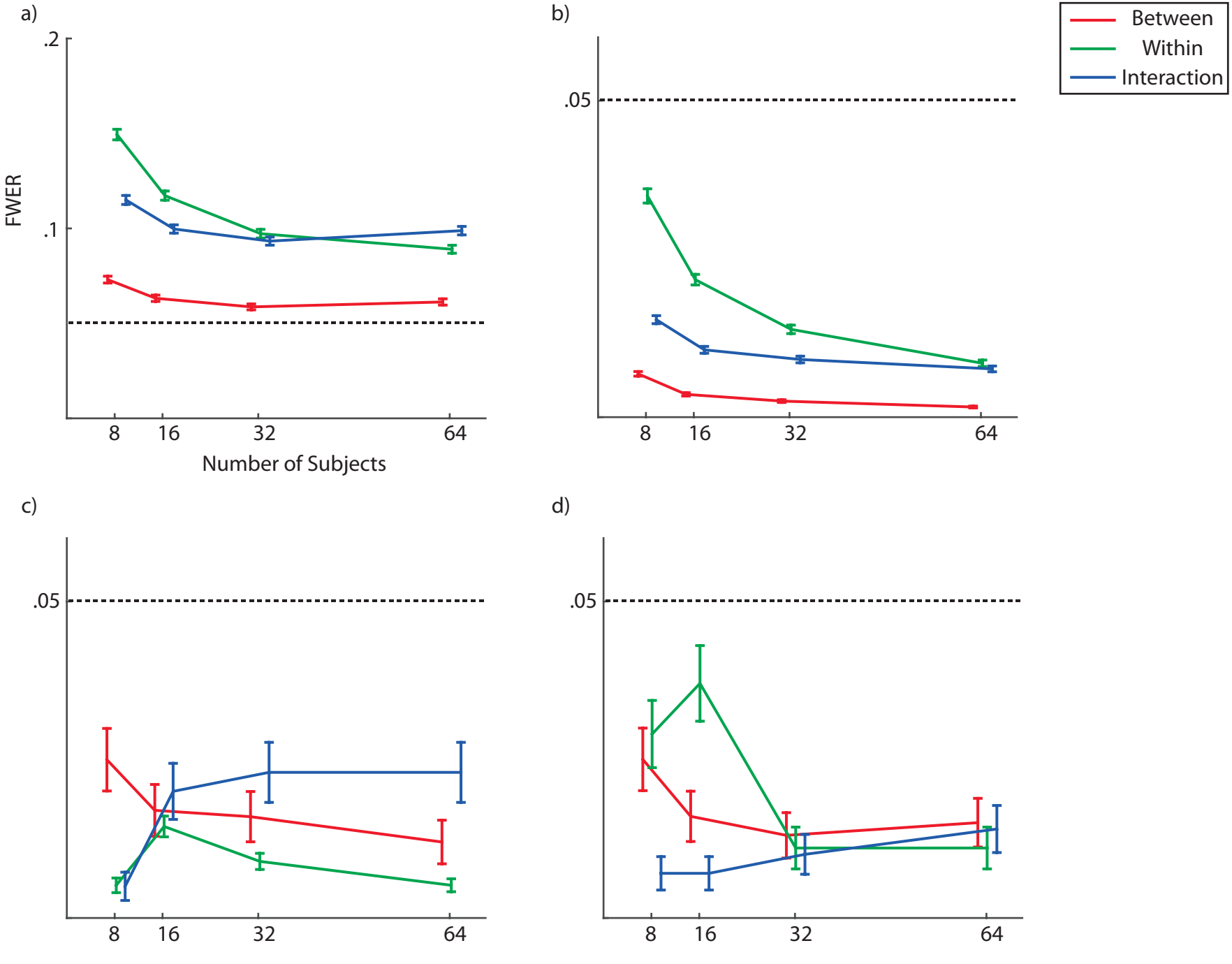
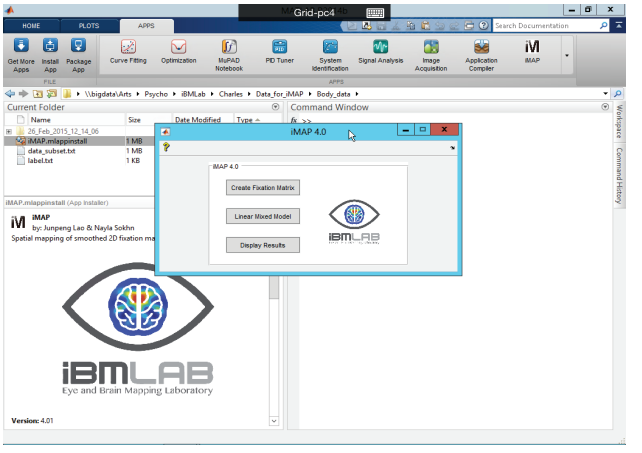
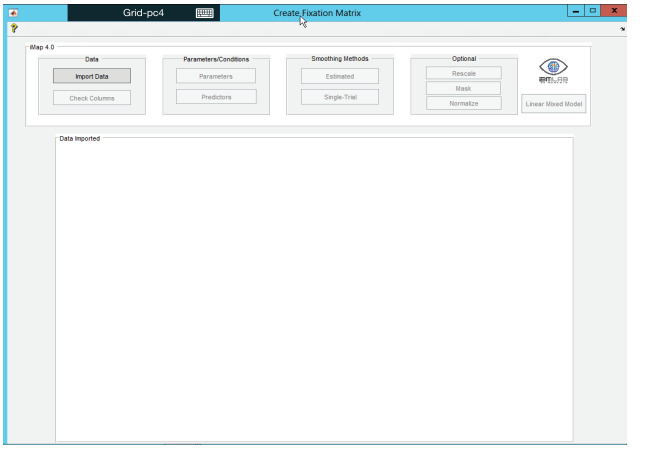

a)
b)
c)
d)
Between
Within
Interaction
FWER
.2
.1
.05
8
16
32
64
Number of Subjects

Figure 3

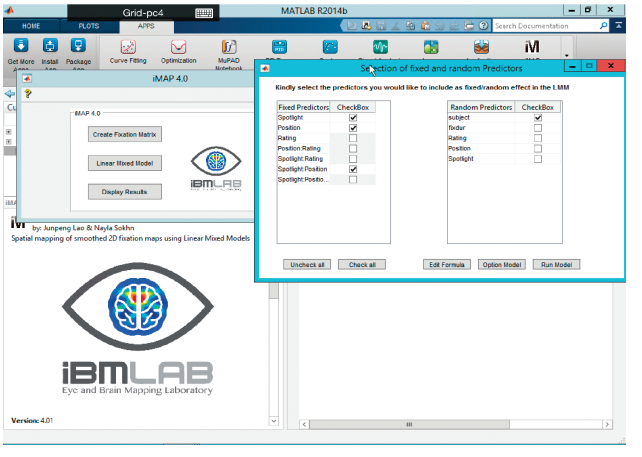
a)



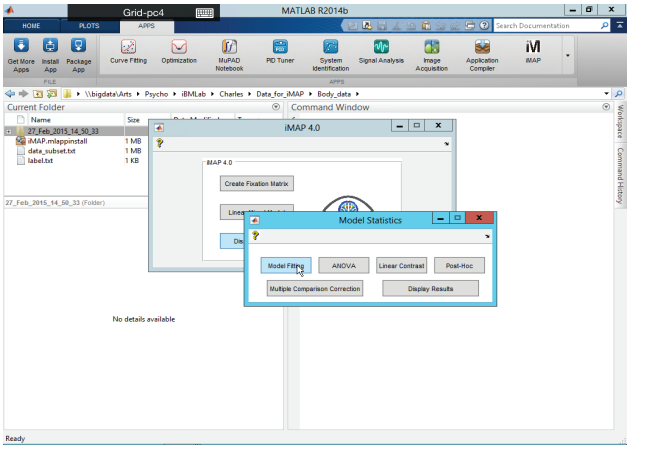
b)



c)



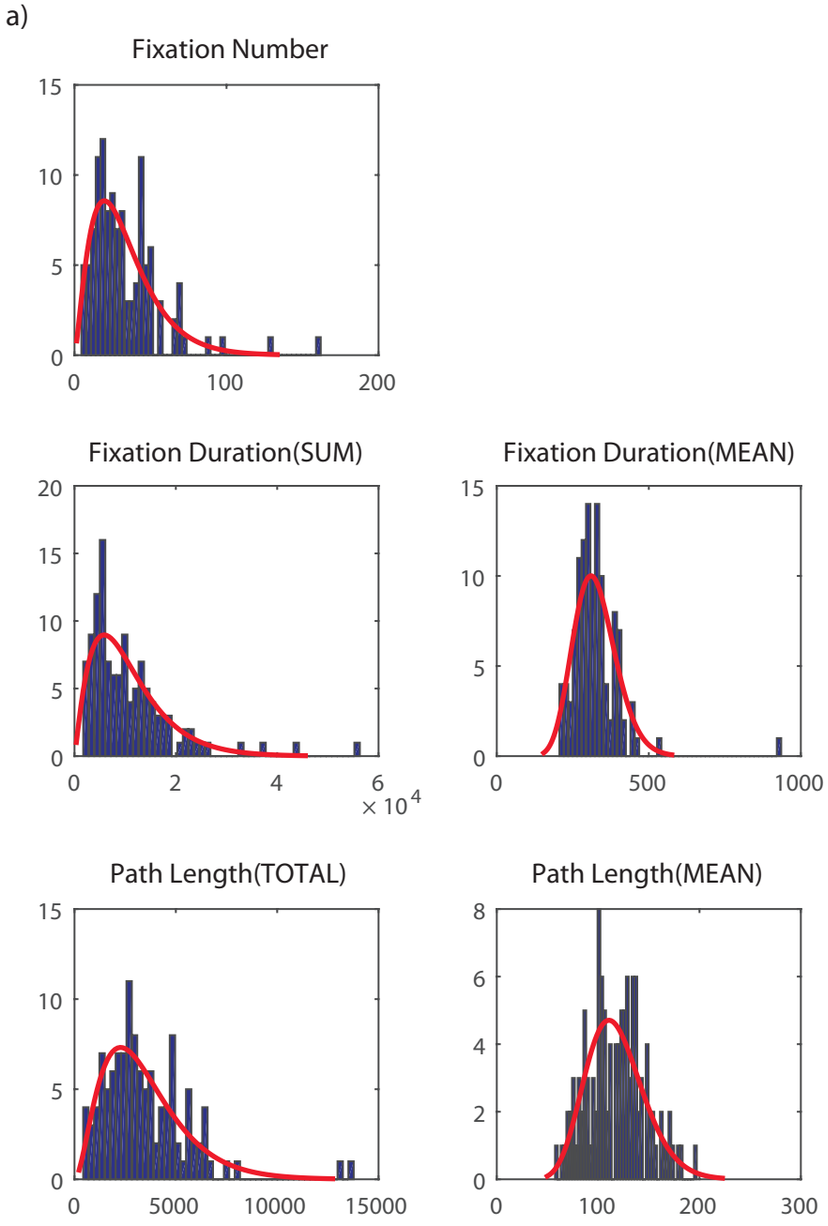
d)



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
```

Figure 4


a)
Fixation Number
Fixation Duration(SUM)
Fixation Duration(MEAN)
Path Length(TOTAL)
Path Length(MEAN)
× 10^4

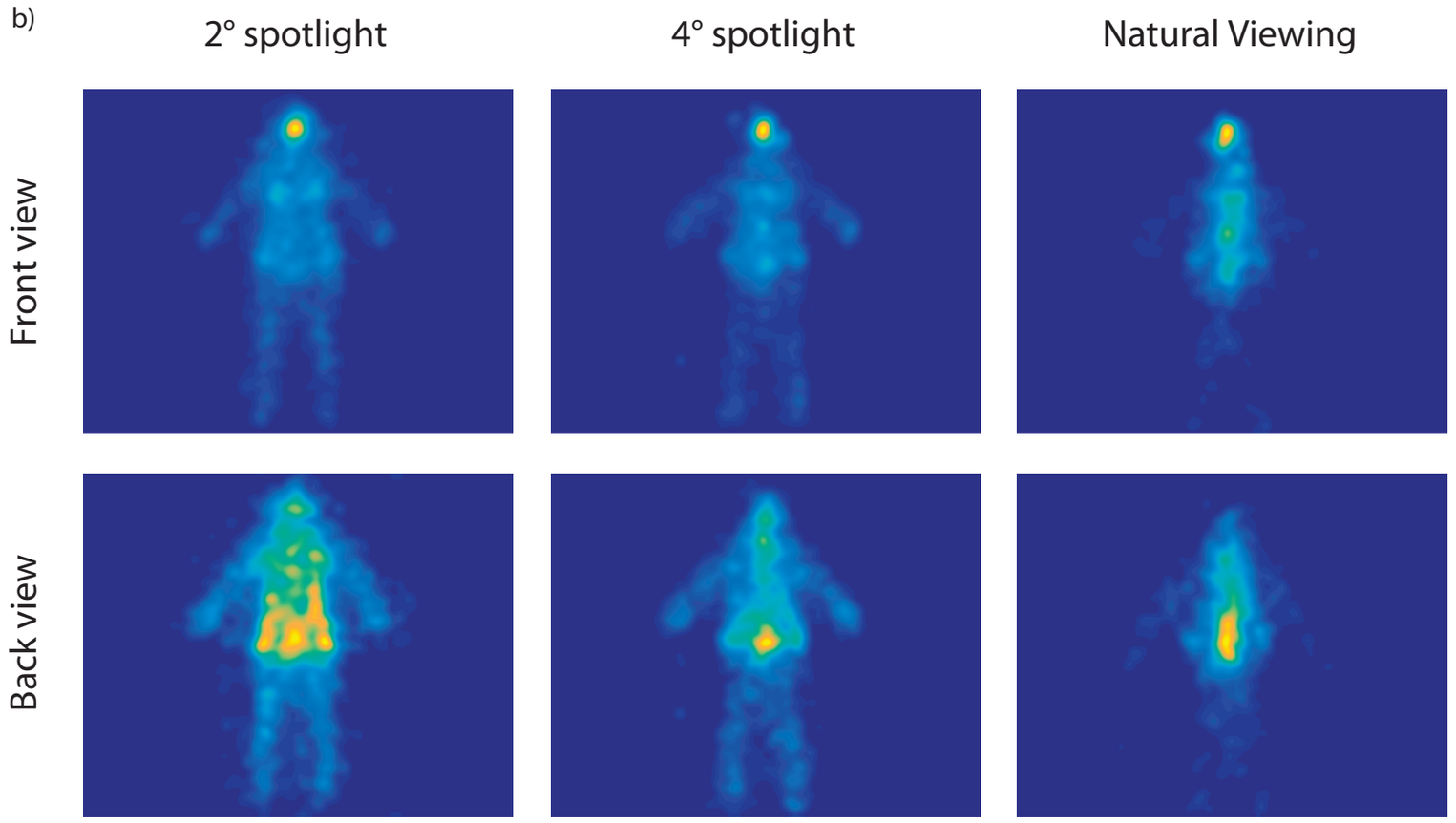
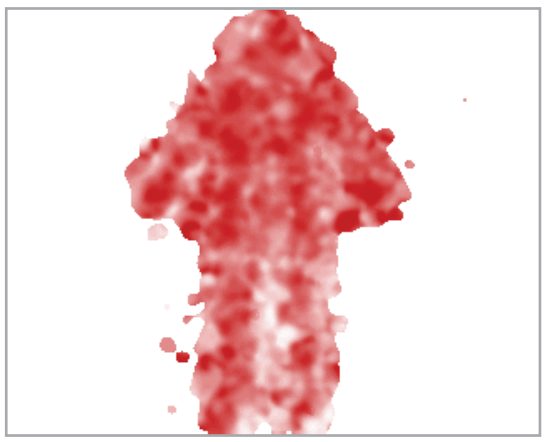
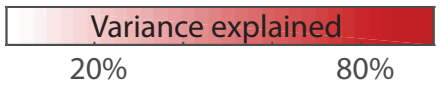

b)
2° spotlight
4° spotlight
Natural Viewing
Front view
Back view

Figure 5

a)

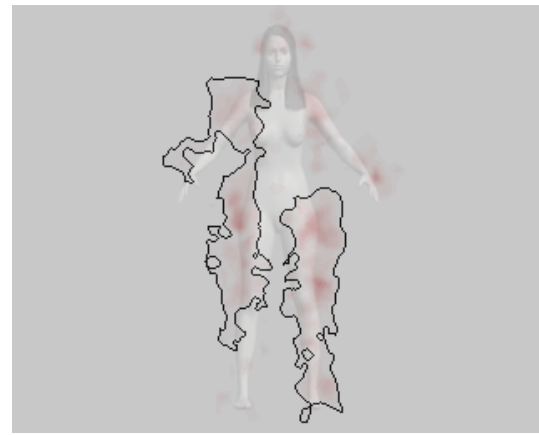
Ordinary R^2



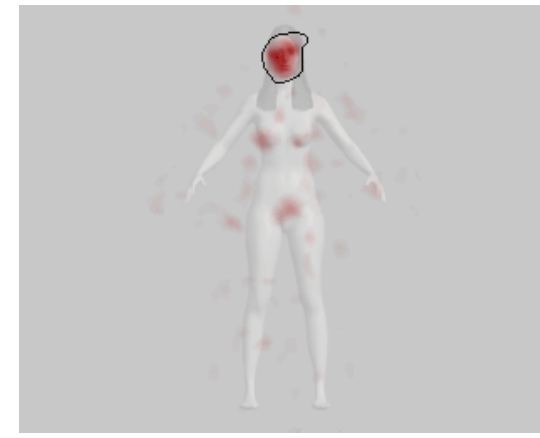


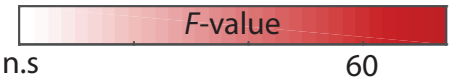
b)

Main effect of *Viewing Condition*



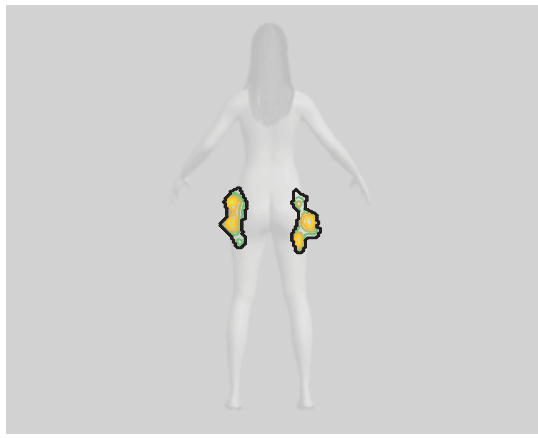
Main effect of *Body Orientation*

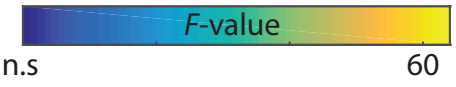




c)

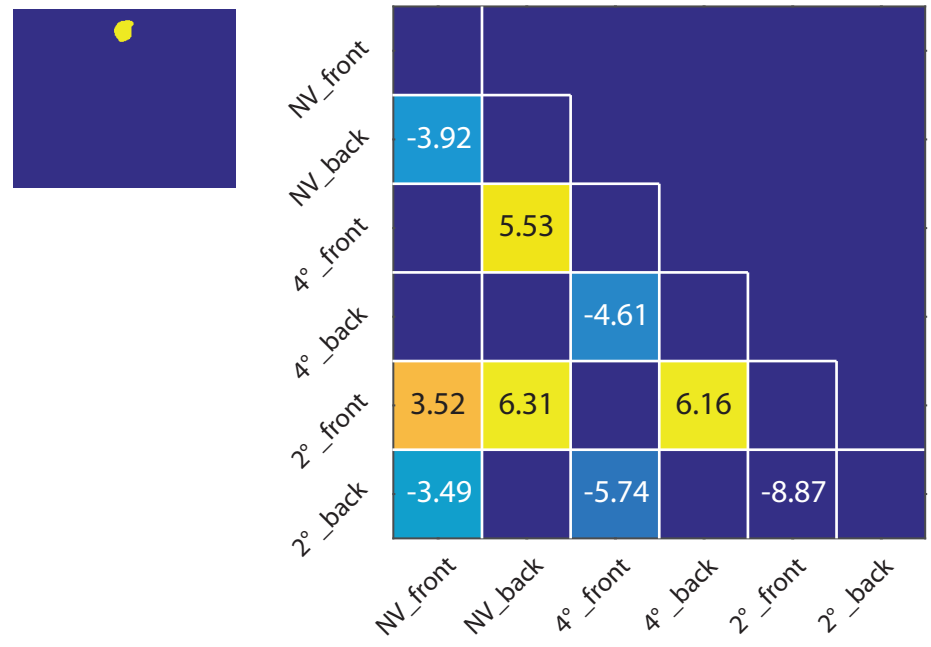
Linear contrast





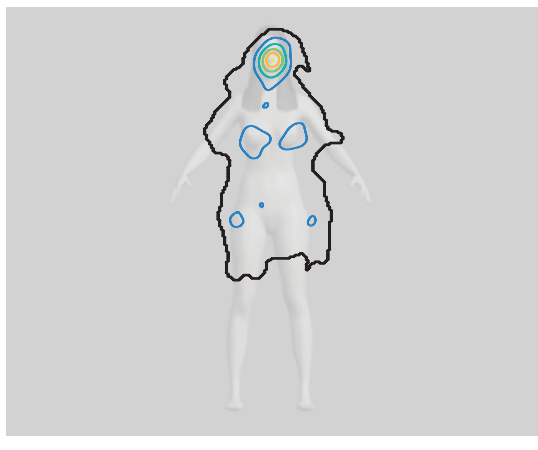
d)

Tvalue ($p < .05$ bonferroni corrected)

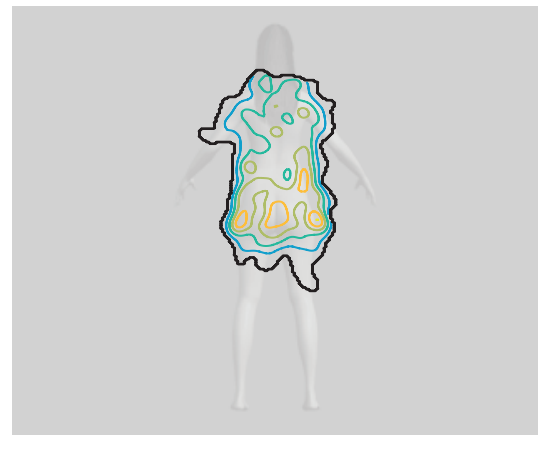


e)

2° spotlight front view



2° spotlight back view



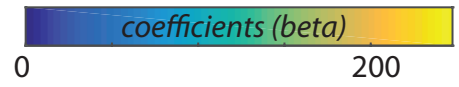


Figure 6

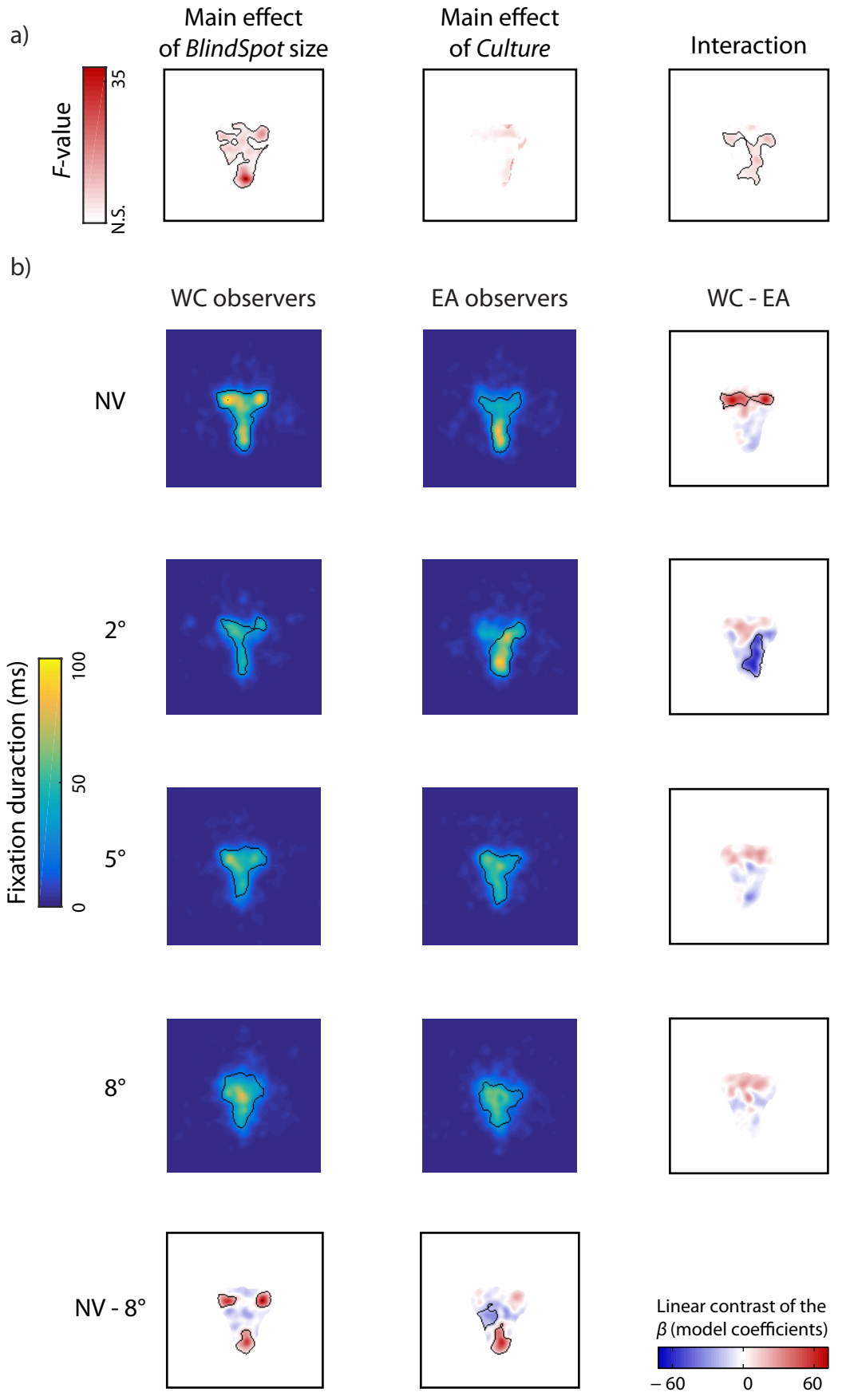
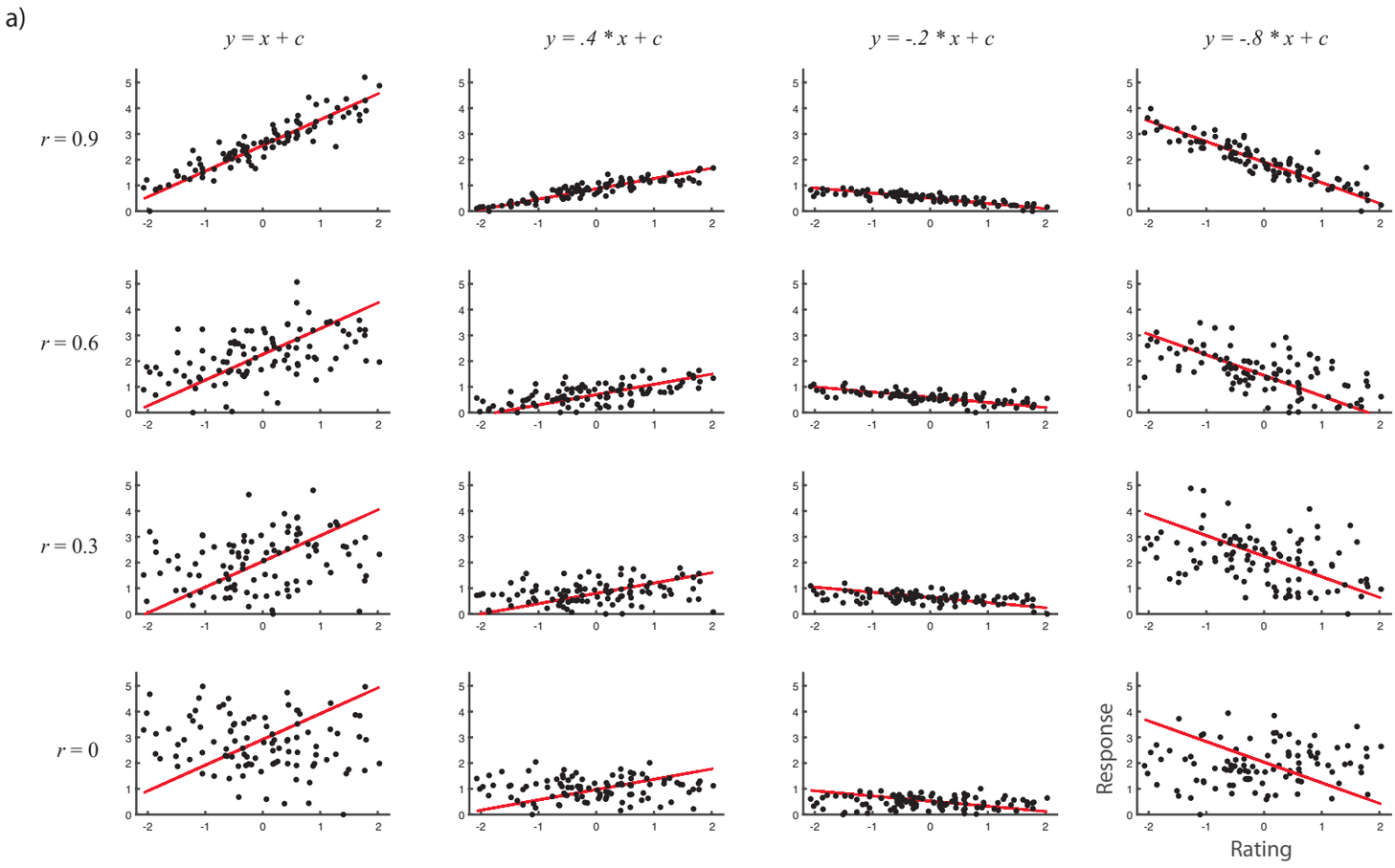
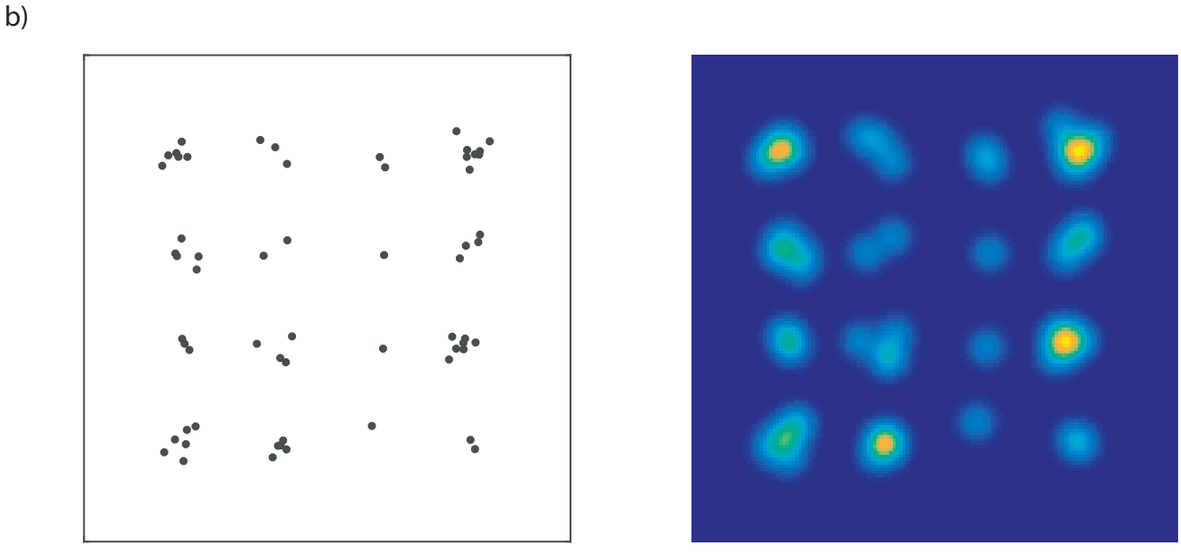
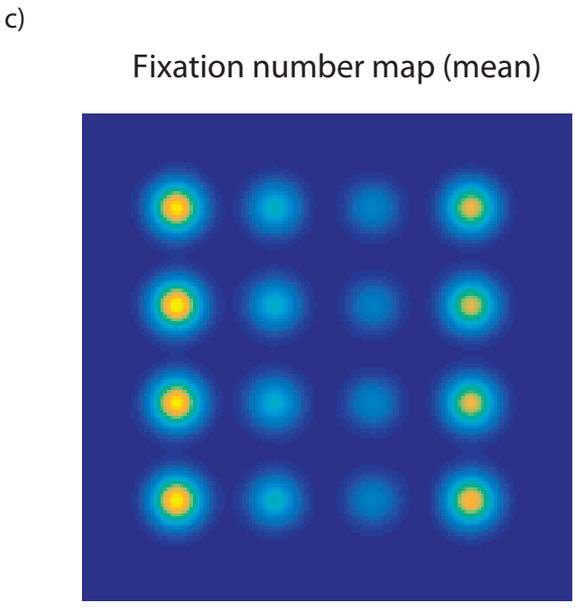
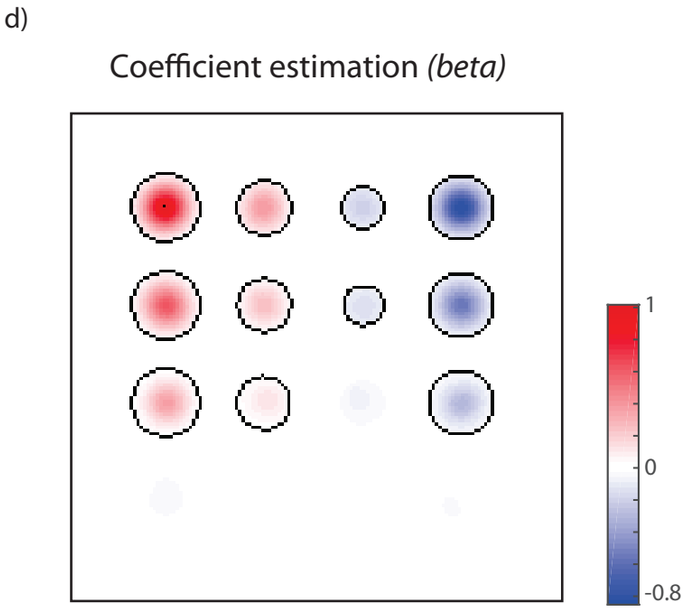

a)
Main effect of *BlindSpot* size
Main effect of *Culture*
Interaction
F-value
35
N.S.
b)
WC observers
EA observers
WC - EA
NV
2°
5°
8°
NV - 8°
Fixation duration (ms)
100
50
0
Linear contrast of the β (model coefficients)
– 60
0
60

Figure 7


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)
1
0
-0.8