Post-print: Effects of spatial smoothing on inter-subject correlation based analysis of FMRI

Juha Pajula, Jussi Tohka

Department of Signal Processing, Tampere University of Technology, Finland

Official final publication is available on Magnetic Resonance Imaging:
http://www.sciencedirect.com/science/article/pii/S0730725X14001866

Citation:
Juha Pajula, Jussi Tohka, Effects of spatial smoothing on inter-subject correlation based analysis of FMRI, Magnetic Resonance Imaging, Available online 23 June 2014, ISSN 0730-725X, http://dx.doi.org/10.1016/j.mri.2014.06.001.
(http://www.sciencedirect.com/science/article/pii/S0730725X14001866)

Keywords: General linear model; Gaussian filter; FWHM; Filtering; Functional magnetic resonance imaging; Block-design

Effects of Spatial Smoothing on Inter-Subject Correlation Based Analysis of FMRI

Juha Pajula[a,1,*], Jussi Tohka[a,1,**]

[a]*Department of Signal Processing, Tampere University of Technology*
P.O. Box 553,
FIN-33101, Finland

Abstract

This study evaluates the effects of spatial smoothing on inter-subject correlation (ISC) analysis for FMRI data using the traditional model based analysis as a reference. So far within ISC analysis the effects of smoothing have not been studied systematically and linear Gaussian filters with varying kernel widths have been used without better knowledge about the effects of filtering. Instead with the traditional general linear model (GLM) based analysis, the effects of smoothing have been studied extensively.

In this study, ISC and GLM analyses were computed with two experimental and one simulated block-design datasets. The test statistics and the detected activation areas were compared numerically with correlation and Dice similarity measures, respectively. The study verified that 1) the choice of the filter substantially affected the activations detected by ISC analysis, 2) the detected activations according to ISC and GLM methods were highly similar regardless of the smoothing kernel and 3) the effect of spatial smoothing was mildly smaller on ISC than GLM analysis. Our results indicated that a good selection for full width at half maximum smoothing kernel for ISC was slightly larger than double the original voxel size.

Keywords: General linear model, Gaussian filter, FWHM, Filtering,

*Corresponding Author
**Principal corresponding author
Email addresses: juha.pajula@tut.fi (Juha Pajula), jussi.tohka@tut.fi (Jussi Tohka)
[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland

Functional magnetic resonance imaging, Block-design, Jackknife estimates, Wavelet denoising

Introduction

Typical stimulus paradigms used with the functional magnetic resonance imaging (FMRI) are strictly controlled having precisely defined task and baseline periods. This makes it possible to define a parametric model for the activation time-course. This kind of FMRI data can be analyzed using a general linear model (GLM), where the stimulus time course needs to be modeled parametrically. However, these strictly controlled stimulus paradigms are necessarily simplified and perhaps limited in their facilities to reveal brain activity that occurs in real-life, and therefore more naturalistic stimulus such as movies are increasingly used instead. However, the parametric model for multidimensional stimulus such as a movie is challenging to define and for this reason alternatives to the GLM based analysis are required. The inter-subject correlation (ISC) is one potential analysis method for these more naturalistic stimulus paradigms. The ISC analysis is a completely data-driven method, based on voxel-wise correlation between the corresponding time series of the subjects where the high correlations are typically interpreted as activations. ISC does not require any parametric model of the stimulus time course and has no assumptions about the smoothness of the data. Albeit ISC methods do not require a model for the stimulus, they require the data to be task based. In the absence of the stimulus, we would expect no significant correlations due to the between subject similarities in processing the stimulus.

Since its introduction [1], the ISC methodology has been further developed [2, 3, 4]. The ISC method of [4] was validated in our earlier study [5] by comparing its results with the results of GLM. In [5], the used data was based on simple block-design tasks, ideal for the GLM based analysis, which justified the use of GLM as a gold standard. The validation confirmed that ISC detects highly similar activation areas as GLM within strictly controlled research setups. This indicated that ISC could find reliable activations in more complex, naturalistic stimulus paradigms. In [5], the Gaussian filter of 5 mm full width at half maximum (FWHM) was used for spatial smoothing according to the requirements of the GLM analysis. However, it is well known that the selection of filter width has a strong effect on the activations detected by GLM, which lead to interesting follow-up questions: Does

the similarity between the analysis results of these methods depend on the applied filtering and how does the filtering influence the results of the ISC based analysis? This paper sets to answer these questions and evaluates the effects of spatial smoothing on the ISC analysis of FMRI data.

The effects of spatial smoothing in FMRI are known well within a GLM based analysis, but smoothing effects have not been studied within ISC. Typically, linear Gaussian filters with varying widths (see table 1) have been used in ISC studies. In table 1, the width of the filters varied from 5 to 12 mm and the ratio between voxel size and filter width from 1.67 to 4. A general rule of thumb in FMRI studies is that FWHM of the Gaussian filtering kernel is twice [6, 7] or three times [8, 9] the voxel size: For example, if the original voxel size is 2 mm, a kernel with at least 4 mm or 6 mm FWHM should be used. These considerations are based on the requirements of the random field theory (RFT) based multiple comparisons correction, which is commonly used to determine thresholds with the GLM based massively univariate analysis. However, in the ISC based analysis, non-parametric multiple comparison correction procedures, which do not have any formal requirements concerning data smoothness, are far more commonly used than RFT based multiple comparison correction.

In the GLM based analysis of FMRI data, spatial smoothing improves the signal to noise ratio (SNR), helps meeting the RFT requirements for smoothness, and increases the overlap of single subject statistics after spatial normalization in the group-level studies [6]. Nevertheless, using too large filters can cause several errors in statistical analysis: merge small activations together, spread the activation signal out from the original area and shift the activation centers to arbitrary positions [10]. Respectively, too small filters can compromise the spatial overlap of the activations of individual subjects in the group-level studies [10, 11], as well as they can lead to a low SNR possibly resulting in a loss in statistical power [12]. The optimal spatial filter width is study dependent and thus the original voxel size of the data, the region of interest and inter-subject variability should be considered when selecting it [13, 14, 15]. The selected filter width should also correspond to the assumed size of the region of interest [16].

In this paper, we studied the effects of spatial smoothing on the ISC analysis by comparing the analysis results between ISC and GLM methods in a similar manner as in [5]. Both ISC and GLM analyses were computed with various Gaussian filter widths and a wavelet denoising method in two experimental blocked-design (external order and hand imitation tasks from

the ICBM Functional Reference Battery) data sets and a simulated data set. Equal pre-processing was performed to experimental data within ISC and GLM methods as well as the resulting statistical maps were thresholded at the same significance level. This allows us to draw conclusions about the relative effects of the different filtering levels on the ISC and GLM analysis and contrast the differences (due to spatial filtering) in the ISC detected activations to these of GLM. We primarily consider Gaussian smoothing because the vast majority of the ISC based studies had applied it. However, we additionally consider a wavelet based spatial smoothing method to see if the conclusions reached for Gaussian smoothing extend to other denoising methods. The results of this work showed clearly how the choice of filter width can substantially affect activation areas detected by the ISC analysis. Interestingly, the effects of spatial smoothing differed between the GLM and ISC analysis methods.

Methods

A part of the materials and methods are as in [5], but the descriptions of the used data and main methods are briefly repeated here.

ICBM functional reference battery data

The used FMRI data included stimulations from the Functional Reference Battery (FRB) tasks developed by the International Consortium for Human Brain Mapping (ICBM) [17][2]. The data was obtained from ICBM database in the Image Data Archieve of the Laboratory of Neuro Imaging[3]. The ICBM project (Principal Investigator John Mazziotta, M.D., University of California, Los Angeles) is supported by the National Institute of Biomedical Imaging and BioEngineering. ICBM is the result of efforts of co-investigators from UCLA, Montreal Neurologic Institute, University of Texas at San Antonio, and the Institute of Medicine, Juelich/Heinrich Heine University - Germany.

The age range of the subjects was restricted to 20-38 years. In the ICBM database, this resulted in 41 right-handed subjects who had FMRI measurements from all five different FRB tasks: auditory naming (AN), external

[2] `http://www.loni.ucla.edu/ICBM/Downloads/Downloads_FRB.shtml`
[3] `http://www.loni.ucla.edu/ICBM`

ordering (EO), hand imitation (HA), oculomotor (OM) and verbal generation (VG). The image data was pre-screened before analysis to ensure the high quality of the data. According to pre-screening, FMRI data from four subjects were discarded because of a poor data quality for at least one task in the battery. Instead of using all five tasks in this study the analysis was limited to the tasks with the highest (HA) and lowest (EO) similarity between ISC and GLM on [5]. The detailed definitions of the tasks are available in [5] and in the FRB software package.

The selected data included measurements from 37 healthy right-handed subjects (19 men and 18 women; average age was 28.2 years from the range of 20-36 years). The functional data was collected with a 3 Tesla Siemens Allegra FMRI scanner and the anatomical T_1 weighted MRI data with an 1.5 Tesla Siemens Sonata scanner. The TR / TE times for the functional data were 4 s / 32 ms, flip angle 90 degree, pixel spacing 2 mm and slice thickness 2 mm. The parameters for the anatomical T1 data were 1.1 s / 4.38 ms, 15 degree, 1 mm and 1 mm, correspondingly.

Pre-processing

The pre-processing and the GLM part of statistical analysis were performed by using FSL (version 4.1.6) from Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Oxford University, Oxford, U.K. [18]. The data pre-processing was done in three phases. First, motion correction was performed using the FSL's MCFLIRT by maximizing the correlation ratio between each time point and the middle volume, using linear interpolation [19, 20]. Second, the Brain extraction tool (BET) [21] was applied to extract the brain volume from functional data. Finally, the images were temporally high-pass filtered with a cutoff period of 60 s and the spatial smoothing was applied with an isotropic three dimensional Gaussian kernel with the FWHM of 0 mm, 2 mm, 4 mm, 5 mm, 8 mm and 12 mm in each direction. As an alternative to Gaussian smoothing, we also applied wavelet denoising with Wavelet Denoising Toolbox [22, 23]. Following the recommendations in [22, 23], two decomposition levels and wavelet family Symlet 2 (sym2) were applied. The original data had 87 volumes with three stabilization volumes, which were discarded from the analysis. The brain extraction from the anatomical T_1 images was also performed by BET, but this was done manually and separately from the main procedure for each T_1 weighted images as the parameters of BET required individual tuning.

The image registration was performed in two phases using FSL Linear

Registration Tool (FLIRT) [19, 20]. First, the skull-stripped functional images were aligned (6 degrees of freedom, full search) to the skull-stripped high-resolution T_1 weighted image of the same subject, and then the results were aligned to the standard (brain only) ICBM-152 template (12 degrees of freedom, full search).

Simulated Data

In additon to experimental data, a set of simulated imaging data was generated. The data set contained 37 simulated functional images in the standard MNI-152 space. Every voxel in these images was either activated or non-activated. Voxels were selected as activated according to the binarized statistical maps of the AN task (thresholded with q=0.05 false discovery rate (FDR) corrected thresholds) from the GLM analysis in [5]. A hemodynamic signal was included in the time series of the activated voxels. The signal was selected to be exactly the same which was used as a model in the GLM analysis, i.e., a boxcar convolved with a canonical hemodynamic response function (HRF). Finally, pink 1/f noise was generated as described in [24][4] and added to every time series in the volume. The contrast to noise ratio of the noisy data was approximately 0.06 when computed against the amplitude of the boxcar before convolved with canonical HRF. The areas outside the activated regions contained only the noise signal. The simulation procedure was exactly the same for every 37 simulated images, that is, we ignored the anatomical and effect size variations between subjects. As the data was generated directly in MNI-152 coordinates no registration or motion correction was performed and pre-processing included only temporal and spatial filtering, exactly as described for experimental data.

Analysis Methods

The ISC analysis was computed with ISCtoolbox [4, 25]. The ISC is based on Pearson's correlations between the corresponding time series of all subject pairs and we denote the correlation coefficient between subjects i and j by r_{ij}. To obtain the final multi-subject test statistic, correlation values of all subject pairs r_{ij} are combined into a single ISC statistic by averaging:

$$\bar{r} = \frac{1}{\frac{m^2-m}{2}} \sum_{i=1}^{m} \sum_{j=2,j>i}^{m} r_{ij}, \tag{1}$$

[4] `https://ccrma.stanford.edu/~jos/sasp/Example_Synthesis_1_F_Noise.html`

where m is the number of subjects. The m was 37 in our study and thus the correlation coefficients were averaged from $(37^2 - 37)/2 = 666$ subject pairs. The statistical inference was accomplished by a fully non-parametric voxel-wise resampling test implemented in the ISCtoolbox. This test accounts for temporal correlations inherent to FMRI data (for details of the test, see [4]). The resampling distribution was approximated with 1,000,000 realizations and the resulting p-values were corrected voxel-wise over the whole brain using an FDR based multiple comparisons correction with independence or positive dependence assumption [26, 9]. The used threshold was $q = 0.001$.

The GLM was computed first with a single subject level and then the group-level statistic was computed from the single subject level results. Both phases were performed with FMRI Expert Analysis Tool (FEAT, version 5.98) from FSL [27, 28]. The particular implementation of the multi-subject GLM (FSL's FLAME using MCMC) was selected because it is widely used and properly evaluated. The details of the analysis procedure were as described in [5] and as earlier FDR was chosen also for GLM as it is essential to compare the detected activations at the same significance level. More detailed discussions about the choice of thresholding methodology are available in [5, 29].

Methods Comparison

The activation maps were compared with different numerical measures: Dice index between two sets of activated voxels, Pearson's correlations between test statistics and sensitivity and specificity measures between the detected activations and the ground truth of simulations. Dice index [30] measures the similarity of the detected activation areas between two thresholded and binarized activation maps and it is closely related to Kappa coefficient [5, 31]. The binarized maps were created by assigning the value of voxel as one if the statistic value passed the threshold and otherwise as zero. The Dice index between two sets B_1 and B_2 of voxels was then defined as in Equation 2:

$$d_{B_1,B_2} = \frac{2|B_1 \cap B_2|}{|B_1| + |B_2|}, \tag{2}$$

The numerator in Eq. (2) measures the size of common activation occurrence and the denominator measures the sizes of activated areas according to individual methods.

Pearson's correlations C were computed between $\bar{r}$-values of ISC and absolute Z-values of GLM [5].

The absolute value of the Z-statistic was taken before computing the correlation measure because it is expected that both large negative and large positive z-statistics (so called activations and de-activations) relate to high $\bar{r}$ -values (as activations and de-activations both lead to (large) positive inter-subject correlation values. It should be noted that large negative $\bar{r}$ -value should not be interpreted as activation or similarity, and therefore, we do not take absolute value of the $\bar{r}$ statistic before computing the correlation. The ISC statistic $\bar{r}$ has a maximal value of 1 while the Z-statistic is not bounded. This difference has very small practical meaning since the observed Z-statistic values were finite and the observed $\bar{r}$ values were considerably smaller than the maximal value of one [5].

With simulated data, the Dice index was computed between the ground truth of simulated activations and the binarized results of ISC and GLM. In addition, we computed the sensitivity (the percentage of the activated voxels detected as activated) and specificity (the percentage of non-activated voxels detected as non-activated) measures with the simulated data. The measures were computed for both methods against the ground truth of the simulation across all the filtering levels.

Certain numerical methods were used to compare and display the differences between the methods over the used filtering levels and, to simplify the inspection of the results, specific detection matrices were generated. The Dice index values were collected in the Dice matrix according to Equation 3:

$$D_{(k,l)} = \begin{bmatrix} d_{k,1} & d_{k,2} & \cdots & d_{k,k-1} & d_{k,l} \\ d_{k-1,1} & & & & d_{k-1,l} \\ \vdots & & \ddots & & \vdots \\ d_{2,1} & d_{2,2} & & & d_{2,l} \\ d_{1,1} & d_{1,2} & \cdots & d_{1,l-1} & d_{1,l} \end{bmatrix}, \tag{3}$$

where values $d_{k,l}$ ($k, l = 1, \ldots, n$ corresponding to the filtering levels) denote Dice indexes computed according to Equation 2 between the thresholded activation maps with filtering levels k and l. For experimental data n was 6 *(FWHM = 0 mm, 2 mm, 4 mm, 5 mm, 8 mm, 12 mm)* and for the simulated data it was 5 *(FWHM = 0 mm, 2 mm, 4 mm, 5 mm, 8 mm).* It should be noted that in the detection matrix the vertical indexes increases from bottom to top for more intuitive display of the results. Filtering with 12 mm FWHM was not applied for simulated data as it was considered unnecessary.

We computed these matrices over a single method resulting in the intra-

method Dice matrices as well as between the two methods resulting in the inter-method Dice matrix. For example, in the intra-method Dice matrix for ISC, $d_{3,2}$ refers to the Dice index between activations detected by ISC with 2 mm and 4 mm filtering. Similarly, in the inter-method Dice matrix, $d_{3,2}$ refers to the Dice index between activations detected by ISC with 4 mm and GLM with 2 mm.

When the matrix was computed over a single method, e.g., activations detected by ISC or GLM were compared with themselves over different filtering levels, the main cross diagonal $(d_{1,1}, d_{2,2}, ..., d_{n,n})$ contained only ones corresponding perfect similarity (since a set was compared with itself). When the matrix was computed over a single method, it was also symmetric over the main cross diagonal.

The voxel was defined as unique when it was detected as activated with a filtering level k but not with a level l. The numbers of unique voxels $v_{k,l}$ across all filtering levels were counted and collected in a unique voxel count matrix V in the same manner as matrix D. The matrix V was computed for both studied analysis methods independently. Now the main cross diagonal of the matrix V contained only zeros, which correspond to the full similarity in the sense of unique detected voxels (no unique voxels exists when the data is compared with itself).

In order to measure the development in detected activation clusters, the 26-connectivity measure [32] was used to search for the clusters in the binarized statistics.

Jackknife Estimates

We used the leave one out jackknife standard error estimator [33] to evaluate the variability of correlation and Dice measures of similarity with respect to subject sample. Denote the fMRI data from subject i as x_i, $i = 1, \ldots, n$, where $n = 37$ is the number of subjects. The jackknife estimate of standard error for a measure $\hat{\rho}$ is defined as

$$\hat{\sigma}_J = \sqrt{\frac{(n-1)}{n} \sum_{i=1}^{n} (\hat{\rho}_{(i)} - \hat{\rho}_{(\cdot)})^2}, \tag{4}$$

where $\hat{\rho}_{(i)} = \hat{\rho}(x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$ denotes the measure $\hat{\rho}$ computed without the data from i^{th} subject and $\hat{\rho}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\rho}_{(i)}$. In our case $\hat{\rho}$ can be either the correlation measure C between the test statistic values or Dice index between the two sets of voxels. In practice, the ISC and GLM analyses

were computed on each possible set of 36 subjects from the original set of 37 subjects and thereafter computing the Dice and correlation measures between the analysis results based on a set of 36 subjects.

Peak Shift Measurements

To quantify the changes in the coordinates of the peaks over the used filtering levels, we followed a technique adapted from [13]. First, three spatial locations of the peaks (i.e., local maxima of the test statistic) from the EO and HA tasks were manually selected so that the peaks in the Z-statistics of GLM and the $\bar{r}$ statistics of ISC matched. This was done by first selecting three clusters so that corresponding clusters were approximately on the same location after thresholding in both, ISC and GLM, statistics with non-smoothed data set. Then, local maxima were selected from each cluster so that the spatial distances between corresponding peaks in ISC and GLM statistics were as small as possible. Only a single peak per cluster and method was selected. These peaks were then tracked over the smoothing levels (2 mm to 12 mm filtering levels) so that, for each smoothing level, the local maximum nearest to the selected peak in non-smoothed data set was searched. Each search was further limited to a search area with the radius corresponding to the used FWHM. For example, if the data was smoothed by 4 mm FWHM kernel, it was assumed that the local maximum could move a maximum of 4 mm from the original location. This restriction was necessary to ensure the tracing of the same peak in practice. Finally, the absolute changes for each filtering level was computed as Euclidean distance between the original location and the new location and these distances were averaged over successive filtering levels.

Results

Simulated data

Table 2 presents the Dice indexes, sensitivities, and specifities of the ISC and GLM methods against the ground truth of the simulated data. With the simulated data, both methods performed poorly when no spatial filtering was applied. With both methods, the average Dice value over 2 mm to 8 mm (0.84 for GLM and 0.88 for ISC) against the ground truth was on the level of "Almost Perfect Agreement" (0.8 - 1.0) according to the categorization of Landis and Koch [34]. The sensitivity of GLM was on average higher (0.97

for GLM versus 0.83 for ISC) and vice versa the specificity of ISC was slightly higher on average (0.96 for GLM versus 0.99 for ISC).

Figure 1 presents the Dice matrices defined in Equation 3 for the simulated data (this figure concerns only differences of the activations detected by GLM and ISC methods and makes no reference to ground-truth of the simulation). The matrix in the panel (a) of Figure 1 presents the intra-method Dice similarity across the used filtering levels by combining two symmetric Dice matrices. The upper half of the matrix represents ISC (blue triangle) and the lower half represents GLM (red triangle). Both methods had the highest values in the first cross diagonal, which was natural as filters with similar FWHM should lead to similar detected activations. The matrix in the panel (b) of Figure 1 shows the Dice similarity between the methods. The highest similarity (0.90) was found when ISC had 8 mm filtering and GLM 4 mm filtering.

Experimental data

Figure 2 (a) presents the number of voxels that were detected as activated by both methods and the number of unique voxels detected as activated by one method but not by the other. According to the Figure, ISC was more conservative than GLM, because ISC found fewer unique voxels than GLM with all studied filtering methods.

Table 3 and Figure 2 (b) present Pearson's correlations C between the test statistics over different filtering levels. Table 4 and Figure 2 (c) present the Dice index values between activation maps of the ISC and GLM methods. The average correlation for the EO task was 0.70 with jackknife standard error estimates of 0.014 averaged over filtering levels and for the HA task the average correlation was 0.84 with average error standard estimate of 0.024. The average Dice measures were 0.55 for EO and 0.85 for HA with average standard error estimates of 0.13 and 0.064. Corresponding Dice measures were 0.54 (EO) and 0.60 (HA). For the wavelet denoised (WD) data the correlation values were 0.67 (EO) and 0.69 (HA) and Dice values were 0.65 (EO) and 0.68 (HA).

According to Landis and Koch [34] categorization, all Dice values for the EO task were categorized as "Moderate Agreement". Instead, most of the Dice values for the HA task were categorized as "Almost Perfect Agreement" (see [5] about the details of the categorization). The highest correlation (0.76) and Dice value (0.63) were reached using the largest 12 mm filter in the EO task. With the HA task, the highest correlation (0.89) was reached with

12 mm filter also, but the highest Dice similarity (0.88) was reached with 8 mm filter. The correlations C increased with the filter size and for both tasks the increase in C across filtering levels was several times larger than the standard error estimate. The Dice values for the HA task were rather stable across filtering levels (within one standard error margin). For the EO task, a trend of increasing Dice values could be noticed, however, these increases are small with respect to standard error estimates. The jackknife standard error estimates of the correlations were all small for all tested filtering levels in both tasks (the average standard error was 0.014 for EO and 0.024 for HA). For Dice measures, the standard errors were small for the HA task (average 0.064) but larger for the EO task (average 0.13).

Figure 3 shows the detected activation areas in the axial slice at the Z=32 mm (anteroposterior axis in MNI-152 coordinates) over ICBM-152 template. It can be observed that large smoothing kernels spread the activation area. The values of the test statistics along the lines from left to right at MNI coordinates Y=76 mm, Z=32 mm of the EO task and at Y=26 mm, Z=32 mm of the HA task are presented in Figure 4 for both methods. The corresponding lines are marked by the blue color in Figure 3.

In Figure 4, it is easy to notice how the shapes of activation profiles are similar between the methods. However, with ISC, the activation peaks were staying here approximately in a single position although smaller peaks were merged together (see, for example, 0 mm position in Figure 4 (d)). The peak shifting in this single direction for GLM is easy to notice here, for example, in Figure 4 (a) at the approximate x-coordinate location of -50 mm. Other effects of spatial smoothing on GLM analysis are noticeable in Figure 4 (b) which demonstrates how the small activation peaks were merged together (for example, compare the peaks in $\pm$65 mm positions).

To obtain a quantitative measure of peak movement the local maxima in three clusters were traced in both tasks. The average movement distances of local maxima are listed in table 5. The average movement distance for ISC method was 2.9 mm and 3.2 mm for GLM method (averaged over three peaks). The initial local maxima for EO task were in the vicinity of MNI coordinates (-2,18,44) in Paracingulate Gyrus, (50,12,32) in Precentral Gyrus (right) and (-46,6,30) in Precentral Gyrus (left). For HA task the initial peak locations were (2,-6,50) in Supplementary Motor Cortex, (-36,-4,62) in Precentral Gyrus (left) and (36,-22,66) in Precentral Gyrus (right). Supporting to the visual evidence in Figure 4 the peak movement across the filtering levels was slightly smaller in ISC analysis.

Figures 5 and 6 present the Dice and unique voxel count matrices relating to EO and HA tasks. The Dice matrices in panels (a) and (b) of these figures are as explained in the context of Figure 1. The matrices in panels (c) and (d) of these figures are the unique voxel count matrices. The values in the Dice matrix for EO task (Figure 5 (a)) increased towards main cross diagonal as it could be expected based on the simulations. The similarity between methods in Figure 5 (b) was not linear as the highest values between methods were not on the main cross diagonal of the matrix: The highest Dice similarity between methods (0.77) was reached when ISC had 8 mm filter and GLM had 2 mm filter in use. With the HA task, the results of inter-method Dice matrix differed slightly as the between methods similarity was close to linear in Figure 6 (b) (the highest values were on the main cross diagonal) and the highest similarity (0.88) was found when both methods had 8 mm filter in use.

When comparing the intra-method unique voxel count matrices in Figures 5 and 6 (c) and (d), it can be observed that ISC had lower values than GLM in the upper left triangle. This points out that, in comparison to GLM, there were fewer such voxels for ISC that were detected by smaller filters but not by larger filters. The significance of the unique voxel count proportion differences between ISC and GLM in Figures 5 (c-d) and 6 (c-d) were tested with Fisher's exact test. The test confirmed that all differences were significant (one-sided $p < 0.025$ in all cases).

The filter size had a clear effect on the size of detected activations. Figure 7 presents the number of detected clusters and the average size of the clusters over the different filtering levels. The lines in the figure present clusters computed from the voxels detected by ISC (turquoises) and GLM (grays) methods. The bars present the size of the detected clusters in voxels by ISC (green and blue) and GLM (yellow and red) for both tasks. As predicted, the size of clusters increased at the same time as their number decreased: The smaller clusters were merged together when the filter size was increased.

Discussion

In this study, the effects of smoothing on the ISC based analysis of fMRI data was evaluated by comparing the analysis results of ISC and GLM methods in standardized block design tasks, where one can expect the GLM analysis to be accurate. Both ISC and GLM analyses were performed using smoothing with varying FWHM kernel widths for the same real and simu-

lated datasets. Pre-processing steps and the FDR based false positive thresholding were the same for both methods. Our results indicated that the choice of filter width affects the ISC analysis results, as it is well demonstrated in the case of standard GLM analysis. Interestingly, the effects of filtering were slightly different between the ISC and GLM analyses.

According to the earlier research [6, 8, 9, 13, 10, 11, 16, 14, 15], multiple strategies exist for selecting the spatial filtering kernel for a FMRI study. These strategies are highly dependent on the used statistical analysis methods and reasons for spatial smoothing differ, for example, between the ISC and GLM analyses. First, ISC is computed over the group of subjects, which makes the registration to the common stereotactic space the most critical part of the pre-processing. With an inaccurate registration, ISC fails as the FMRI time-courses that are correlated do not match spatially. For this reason spatial smoothing is often used to counterbalance the inaccuracy of spatial registration. Second, ISC does not apply any parametric stimulus model and therefore it has no parametric assumptions about the data. The statistical inference in ISC is commonly based on non-parametric tests unlike in GLM, where satisfying the assumptions by the parametric model is one of the major reasons for spatial smoothing. For these reasons, spatial smoothing guidelines devised for GLM might be inapplicable to ISC.

Earlier we had demonstrated that ISC found the same activation areas as GLM in strictly controlled research setups [5]. In this paper, we showed that both methods found similar activations regardless of the used filtering method. This is visible in Figure 2 (b) and (c) where the correlations between test statistics and Dice values, were high and fairly stable across the filtering levels. This was also confirmed in Figure 3 where the detected activation areas changed in a similar manner with both methods. The wavelet denoising led to similar correlation values and Dice indeces as the Gaussian smoothing, which indicates that similarity between the ISC and GLM analysis did not depend on Gaussian smoothing. The maximal similarity between the methods was not achieved by using the same filtering in the EO task. The maximum Dice index was 0.77 when ISC used 8 mm filter and GLM used 2 mm filter (see Figure 5 (b)). Instead, for the HA task, the highest similarity (0.88) was found using 8 mm filter kernel with both methods (see Figure 6 (b)).

The widely used rule of thumb that the filter width should be at least twice the size of the voxel (here the voxel size was 2 mm) seemed to hold also with the ISC analysis: ISC analysis results lacked sensitivity and were noisy

with smaller filter widths. The unwanted smoothing effects such as merging different activation areas together, moving the centers of activations and spreading the activation in the surrounding areas were visible in the results of both ISC and GLM analysis methods. Based on Figure 4, it appears that movements of activation peaks were a more severe problem with GLM than with ISC. As an example, the peak-movement problem with GLM is visible with larger filters (8 mm and 12 mm) in Figure 4 (a) around the position of 50 mm in x-axis. With ISC in Figure 4 (c) and (d) the movement was mainly caused by the merger of smaller peaks and thus ISC suffered mainly from the spreading and merging effects. These observations are also supported by peak shift measurements in table 5, where the changes in peaks were slightly smaller with ISC than with GLM.

When examining the detected activations within a single method across filtering levels, the detected activations changed less with ISC than with GLM. For example, when successive filtering levels were compared (see values next to the cross diagonal in Figure 5 (a)) ISC had on average 3.7 % larger values with the smoothed data (2 mm - 12 mm kernels) in the EO task and, in the HA task (see Figure 6 (a)), the average values of ISC were 3.3 % larger. The unique voxel count matrices in Figures 5 (c) and (d) and Figures 6 (c) and (d) suggested that within ISC analysis there existed less voxels which were detected with smaller filter widths but not with larger filter widths. These observations combined with the peak shift measurements suggest that the variation on spatial filtering had smaller effects on the ISC analysis than on the GLM analysis. In consequence, ISC could be argued to tolerate slightly larger Gaussian smoothing kernels than GLM.

Caution is needed when investigating the values in the table 2, as the simulation was extremely simple and thus the results of the table are merely suggestive. The simulation was observed to be almost too perfect for GLM, which detected activation also from the areas where the signal had spread from the true activation area. However, also with simulated data, ISC seemed to be affected less by the change of filter size than GLM as was the case with experimental data. According to the Dice values in the table 2, the optimal filtering level with the simulated data was 5 mm for ISC (Dice value 0.91) and 4 mm for GLM (Dice value 0.92).

The spatial filtering was mandatory for both analysis methods: GLM was not able to perform whitening and the sensitivity of ISC was minimal with 0 mm filter with simulated data (see table 2). The main reason for GLM to fail on simulated data with 0 mm kernel was that GLM failed to

compute estimates for the whitening process and for this reason the GLM analysis was computed without whitening for 0 mm kernel. This led to a severe number of false positive voxels in the thresholded statistics. Because of the poor performance of both ISC and GLM, the results of 0 mm kernel in the table 2 are presented only as a reference and are not comparable to the other results. As both ISC and GLM were able to detect reasonable activations without filtering with the experimental data (see Figure 2), the simulation pointed out an important property of the pre-processing pipeline: the experimental data had multiple pre-processing steps, such as motion correction and image registration, which resulted in a certain amount of smoothing as a side product [35].

The ISC analysis requires spatial smoothing to ensure the spatial registration accuracy and decent SNR. Our results indicated that, for ISC, a good selection of FWHM for Gaussian smoothing kernel was slightly larger than double the original voxel size. Furthermore, this study verified that 1) the choice of the filter width substantially affected the activations detected by ISC analysis, 2) the detected activations according to ISC and GLM methods were highly similar regardless of the used spatial smoothing method and 3) the used Gaussian filter for the spatial smoothing can be slightly wider with ISC than with GLM analysis.

Acknowledgments

This research was supported by the Academy of Finland (grant numbers 130275 and 263785).

Data collection and sharing for this project was provided by the International Consortium for Brain Mapping (ICBM; Principal Investigator: John Mazziotta, MD, PhD). ICBM funding was provided by the National Institute of Biomedical Imaging and BioEngineering. ICBM data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

References

[1] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, R. Malach, Intersubject synchronization of cortical activity during natural vision., Science 303 (5664) (2004) 1634–1640.

[2] M. P. Hejnar, K. A. Kiehl, V. D. Calhoun, Interparticipant correlations: A model free fmri analysis technique, Human brain mapping 28 (9) (2007) 860–867.

[3] U. Hasson, O. Furman, D. Clark, Y. Dudai, L. Davachi, Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding, Neuron 57 (3) (2008) 452–462.

[4] J.-P. Kauppi, I. P. Jääskeläinen, M. Sams, J. Tohka, Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency, Frontiers in Neuroinformatics 4 (2010) 5.

[5] J. Pajula, J.-P. Kauppi, J. Tohka, Inter-subject correlation in fmri: Method validation against stimulus-model based analysis, PLoS ONE 7 (8) (2012) e41196. doi:10.1371/journal.pone.0041196.

[6] T. Ball, T. P. K. Breckel, I. Mutschler, A. Aertsen, A. Schulze-Bonhage, J. Hennig, O. Speck, Variability of fmri-response patterns at different spatial observation scales, Human brain mapping 33 (5) (2012) 1155–1171.

[7] K. Friston, A. Holmes, J.-B. Poline, C. Price, C. Frith, Detecting activations in pet and fmri: Levels of inference and power, NeuroImage 4 (3) (1996) 223 – 235. doi:10.1006/nimg.1996.0074.

[8] K. M. Petersson, T. E. Nichols, J.-B. Poline, A. P. Holmes, Statistical limitations in functional neuroimaging ii: Signal detection and statistical inference, Philosophical Transactions of the Royal Society of London.Series B 354,.

[9] T. Nichols, S. Hayasaka, Controlling the familywise error rate in functional neuroimaging: a comparative review, Statistical Methods in Medical Research 12 (5) (2003) 419–446.

[10] T. White, D. O'Leary, V. Magnotta, S. Arndt, M. Flaum, N. C. Andreasen, Anatomic and functional variability: The effects of filter size in group fmri data analysis, NeuroImage 13 (4) (2001) 577–588.

[11] P. Fransson, K. . Merboldt, K. M. Petersson, M. Ingvar, J. Frahm, On the effects of spatial filtering - a comparative fmri study of episodic

memory encoding at high and low resolution, NeuroImage 16 (4) (2002) 977–984.

[12] A. Scouten, X. Papademetris, R. T. Constable, Spatial resolution, signal-to-noise ratio, and smoothing in multi-subject functional mri studies, NeuroImage 30 (3) (2006) 787–793.

[13] M. Mikl, R. Mareček, P. Hluštík, M. Pavlicová, A. Drastich, P. Chlebus, M. Brázdil, P. Krupa, Effects of spatial smoothing on fmri group inferences, Magnetic resonance imaging 26 (4) (2008) 490–503.

[14] M. E. Shaw, S. C. Strother, M. Gavrilescu, K. Podzebenko, A. Waites, J. Watson, J. Anderson, G. Jackson, G. Egan, Evaluating subject specific preprocessing choices in multisubject fmri data sets using data-driven performance metrics, NeuroImage 19 (3) (2003) 988–1001.

[15] A. Weibull, H. Gustavsson, S. Mattsson, J. Svensson, Investigation of spatial resolution, partial volume effects and smoothing in functional mri using artificial 3d time series, NeuroImage 41 (2) (2008) 346 – 353. doi:10.1016/j.neuroimage.2008.02.015.

[16] S. Clare, Functional mri : Methods and applications, Ph.D. thesis, University of Nottingham (1997).

[17] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. L. Goualher, D. Boomsma, T. Cannon, R. Kawashima, B. Mazoyer, A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm), Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences 356 (1412) (2001) 1293–1322.

[18] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, P. M. Matthews, Advances in functional and structural mr image analysis and implementation as fsl, NeuroImage 23, Supplement 1 (2004) 208–219.

[19] M. Jenkinson, S. Smith, A global optimisation method for robust affine registration of brain images, Medical image analysis 5 (2) (2001) 143–156.

[20] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, NeuroImage 17 (2) (2002) 825–841.

[21] S. M. Smith, Fast robust automated brain extraction, Human brain mapping 17 (3) (2002) 143–155.

[22] S. Khullar, A. Michael, N. Correa, T. Adali, S. Baum, V. Calhoun, Wavelet-based denoising and independent component analysis for improving multi-group inference in fmri data, in: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, 2011, pp. 456–459. doi:10.1109/ISBI.2011.5872444.

[23] S. Khullar, A. Michael, N. Correa, T. Adali, S. A. Baum, V. D. Calhoun, Wavelet-based fmri analysis: 3-d denoising, signal separation, and validation metrics, NeuroImage 54 (4) (2011) 2867–2884. doi:http://dx.doi.org/10.1016/j.neuroimage.2010.10.063.

[24] J. O. Smith, Spectral Audio Signal Processing, `http://-ccrma.stanford.edu/~jos/sasp/`, Accessed 24.05.2013, online book.

[25] J.-P. Kauppi, J. Pajula, J. Tohka, A versatile software package for inter-subject correlation based analyses of fmri., Front Neuroinform 8 (2014) 5.

[26] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society. Series B (Methodological) 57 (1) (1995) 289–300.

[27] C. F. Beckmann, M. Jenkinson, S. M. Smith, General multilevel linear modeling for group analysis in fmri, NeuroImage 20 (2) (2003) 1052–1063.

[28] M. W. Woolrich, T. E. Behrens, C. F. Beckmann, M. Jenkinson, S. M. Smith, Multilevel linear modelling for fmri group analysis using bayesian inference., NeuroImage 21 (4) (2004) 1732–1747.

[29] B. R. Logan, M. P. Geliazkova, D. B. Rowe, An evaluation of spatial thresholding techniques in fmri analysis, Human brain mapping 29 (12) (2008) 1379–1389.

[30] L. R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.

[31] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, A. C. Palmer, Morphometric analysis of white matter lesions in mr images: method and validation, Medical Imaging, IEEE Transactions on 13 (4) (1994) 716–724.

[32] P. Bellec, V. Perlbarg, S. Jbabdi, M. Pélégrini-Issac, J.-L. Anton, J. Doyon, H. Benali, Identification of large-scale networks in the brain using fmri, NeuroImage 29 (4) (2006) 1231–1243.

[33] B. Efron, G. Gong, A leisurely look at the bootstrap, the jackknife, and cross-validation, The American Statistician 37 (1) (1983) pp. 36–48.

[34] J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data., Biometrics 33 (1) (1977) 159–174.

[35] S. Strother, S. L. Conte, L. K. Hansen, J. Anderson, J. Zhang, S. Pulapura, D. Rottenberg, Optimizing the fmri data-processing pipeline using prediction and reproducibility performance metrics: I. a preliminary group analysis, NeuroImage 23, Supplement 1 (2004) 196–207.

[36] S. M. Wilson, I. Molnar-Szakacs, M. Iacoboni, Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension, Cerebral Cortex 18 (1) (2008) 230–242.

[37] I. P. Jääskeläinen, K. Koskentalo, M. H. Balk, T. Autti, J. Kauramäki, C. Pomren, M. Sams, Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing., Open Neuroimag J 2 (2008) 14–19.

[38] Y. Golland, S. Bentin, H. Gelbard, Y. Benjamini, R. Heller, Y. Nir, U. Hasson, R. Malach, Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation, Cerebral Cortex 17 (4) (2007) 766–777.

[39] L. Nummenmaa, E. Glerean, M. Viinikainen, I. P. Jääskeläinen, R. Hari, M. Sams, Emotions promote social interaction by synchronizing brain activity across individuals, Proceedings of the National Academy of Sciences 109 (24) (2012) 9599–9604.

Tables

Table 1: Used filter widths with ISC analysis

Pixel Resolution	FWHM	Ratio	Article
3 mm	6 mm	2x	Hasson et al. [3]
3 mm	12 mm	4x	Hasson et al. [1]
3 mm	8 mm	2.67x	Wilson et al. [36]
3 mm	8 mm	2.67x	Jääskeläinen et al. [37]
3.44 mm	12 mm	3.49x	Hejnar et al. [2]
3 mm	8 mm	2.67x	Golland et al. [38]
3 mm	5 mm	1.67x	Kauppi et al. [4]
3 mm	8 mm	2.67x	Nummenmaa et al. [39]

Table 2: Comparison of the detected activations against the ground truth of the simulation

Filter	Dice		Specificity		Sensitivity	
(FWHM)	GLM	ISC	GLM	ISC	GLM	ISC
0 mm	0.42	0.00	0.9980	1.0000	0.2714	0.0015
2 mm	0.89	0.86	0.9792	0.9979	0.9432	0.7693
4 mm	0.92	0.88	0.9841	0.9992	0.9652	0.7890
5 mm	0.86	0.91	0.9654	0.9964	0.9779	0.8574
8 mm	0.70	0.87	0.9029	0.9815	0.9844	0.8896
Average	0.76	0.70	0.9659	0.9950	0.8284	0.6614
Average (2-8 mm)	0.84	0.88	0.9579	0.9938	0.9677	0.8263

Table 3: Correlation values (C) between ISC and GLM statistics and their jackknife standard error estimates ($\hat{\sigma}_{JC}$).

Filter	EO		HA	
(FWHM)	C	$\hat{\sigma}_{JC}$	C	$\hat{\sigma}_{JC}$
2 mm	0.67	0.013	0.80	0.024
4 mm	0.67	0.013	0.81	0.026
5 mm	0.69	0.012	0.83	0.028
8 mm	0.73	0.016	0.87	0.024
12 mm	0.76	0.017	0.89	0.020
Average	0.70	0.014	0.84	0.024

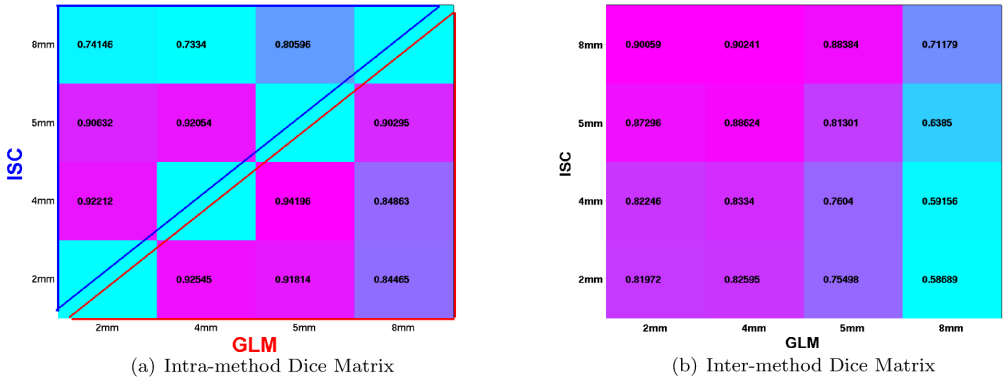
Table 4: Dice measurements (d) between the activation areas detected by ISC and GLM methods and their jackknife standard error estimates ($\hat{\sigma}_{Jd}$).

Filter	EO		HA	
(FWHM)	d	$\hat{\sigma}_{Jd}$	d	$\hat{\sigma}_{Jd}$
2 mm	0.48	0.15	0.84	0.075
4 mm	0.51	0.15	0.85	0.065
5 mm	0.55	0.13	0.86	0.048
8 mm	0.59	0.10	0.88	0.036
12 mm	0.63	0.11	0.84	0.097
Average	0.55	0.13	0.85	0.064

Table 5: Average ISC and GLM local maximum peak location changes over the tested gaussian denoising levels. Differences are in millimeleters of the euclidean distance Δ from the local maximum peak of the corresponding test with non-smoothed data. For EO task the selected peaks were from clusters in Paracingulate Gyrus (ParCin), Precentral Gyrus at Left (PreCL) and Precentral Gyrus at right (PreCR) areas. Correspondingly for the HA task the three peaks were selected from clusters in areas of Supplementary Motor Cortex (SMC), Precentral Gyrus Left (PreCL) and Precentral Gyrus Right (PreCR).

Task		EO			HA			
Method		ParCin	PreCR	PreCL	SMC	PreCL	PreCR	Average
Average	ISC	3.0	4.9	1.9	1.8	2.5	3.2	2.9
Δ (mm)	GLM	3.3	4.8	1.9	3.6	3.8	1.9	3.2

Figures



(a) Intra-method Dice Matrix

(b) Inter-method Dice Matrix

Figure 1: **Dice matrices for the simulated data.** The Dice matrices were constructed according to Equation 3. The intra-method Dice matrix (a) is constructed from two cross diagonal symmetric matrixes by selecting either upper (ISC, blue) or lower (GLM, red) triangle for the final matrix. The inter-method Dice matrix (b) contains the Dice similarity between the detected activations of ISC and GLM methods. The matrix in panel (a), shows that ISC had the lowest similarity (Dice value 0.73) between thresholded areas with 4 mm and 8 mm filters and GLM had the lowest similarity (Dice value 0.84) with 2 mm and 8 mm filters. The matrix in panel (b) shows that ISC and GLM had the most similar detection areas (Dice value 0.90) when ISC had 8 mm and GLM had 4 mm kernel in use.

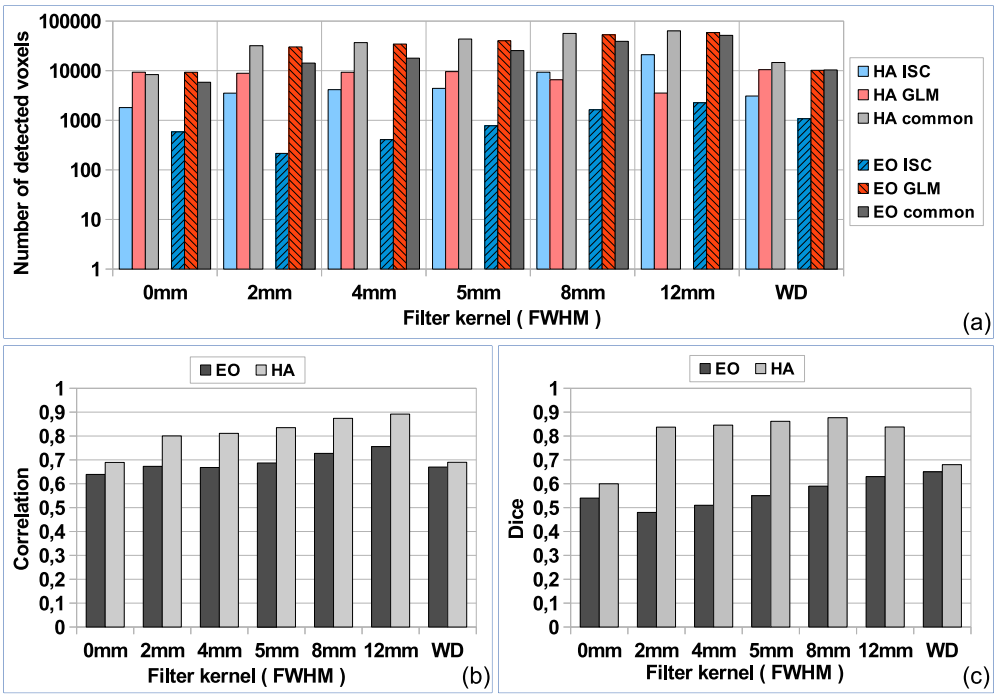
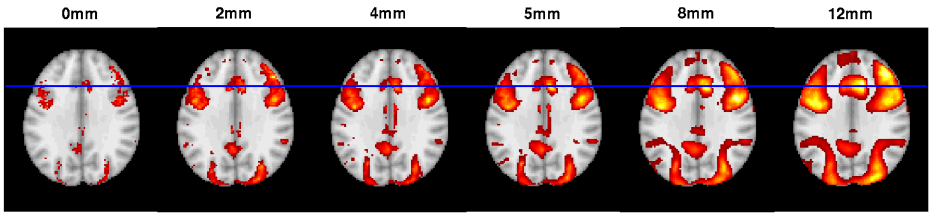
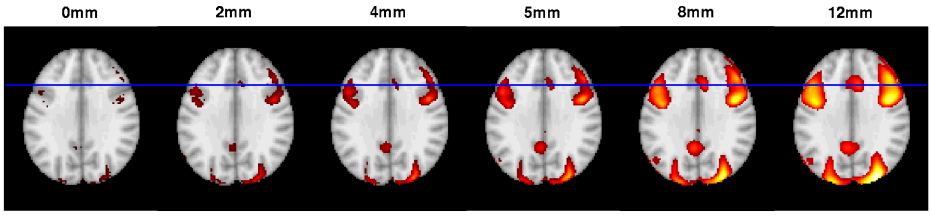


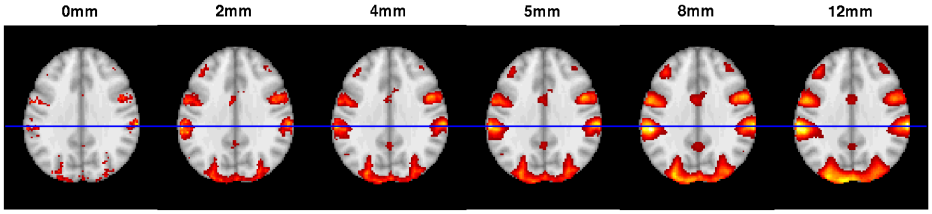
Figure 2: **Unique voxels (a), Correlation measures (b) and Dice measures (c) between ISC and GLM.** The panel (a) presents unique voxels between methods after FDR corrected (q=0.001) thresholding. The voxel is defined as unique if it is detected by a single analysis method (ISC or GLM) but not by both. The columns present the number of unique voxels, which were activated according to ISC (blue columns) or GLM (red columns) analysis only. The gray bars present the number of common voxels detected by both ISC and GLM methods. The abreviation WD refers to wavelet denoising. The panel (b) presents the correlation measures between ISC and GLM statistics and panel (c) presents the Dice values between ISC and GLM detections.



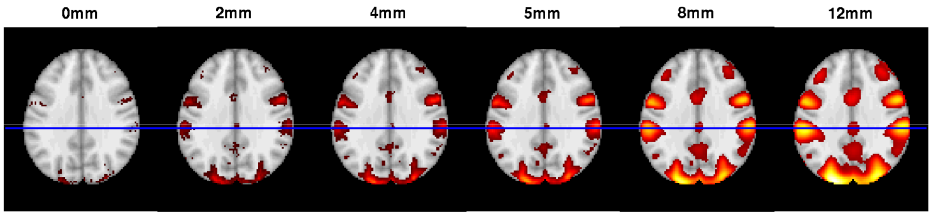
(a) GLM results from EO task



(b) ISC results from EO task

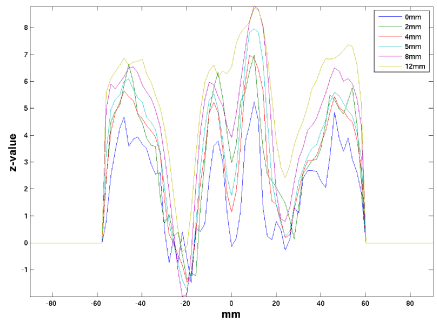


(c) GLM results from HA task

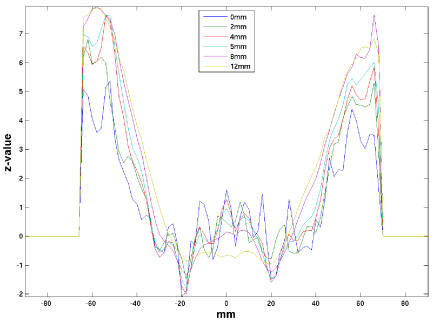


(d) ISC results from HA task

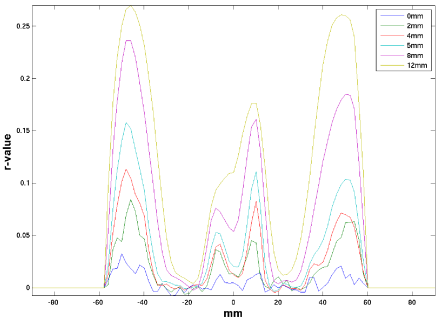
Figure 3: **Thresholded activation maps for the EO (panels a,b) and HA (panels c, d) task (FDR corrected, q=0.001).** The images in panels (a) and (c) present the results of the GLM analysis and the images in panels (b) and (d) present the corresponding results for the ISC method. The images are from the axial slice of Z=32 mm anteroposterior axis in MNI-152 coordinate system. The blue line in the images presents the position of line visualized in Figure 4. It can be observed that the ISC analysis was on average more conservative than the GLM analysis and the spreading effect was less obvious with ISC.



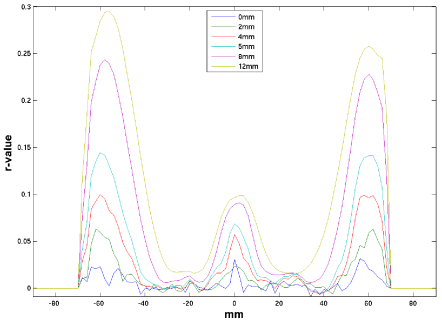
(a) GLM results from EO task



(b) GLM results from HA task

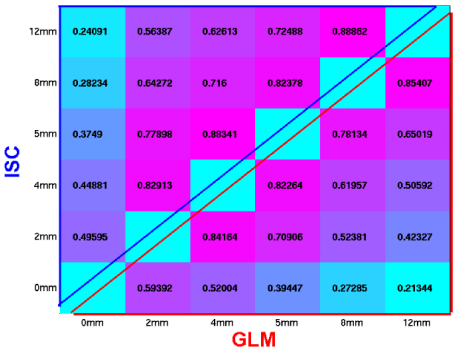


(c) ISC results from EO task

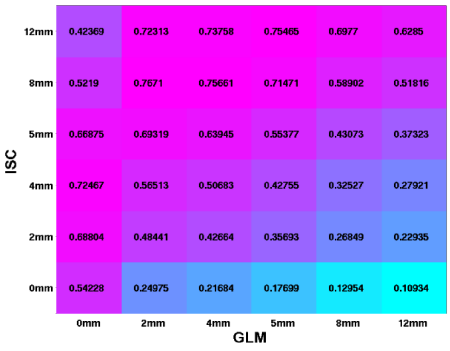


(d) ISC results from HA task

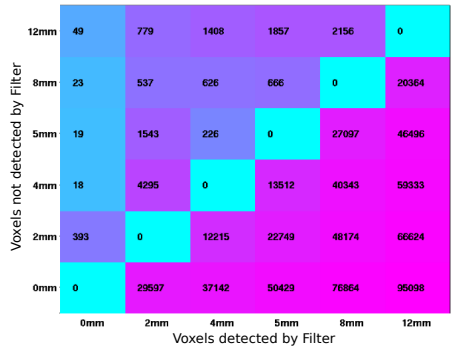
Figure 4: **The values of the test statistics for GLM (a, b) and ISC (c, d) tasks.** The images present the values of test statistics along the left to right lines, which are marked with the blue color in the image series of Figure 3. The images in panels (a) and (c) present the value of test statistics along the line from the EO task in the position Y = 24 mm , Z = 32 mm in MNI-152 coordinate system. Panels (b) and (d) presents the lines from HA task in the position of Y = -26 mm, Z = 32 mm in MNI-152 coordinate system. The shapes of lines are similar. In the case of ISC (panels (c) and (d)) the peaks remained in one position; instead with GLM (panels (a) and (b)) the peaks shifted across the filtering levels. The phenomenon is easy to notice for example from the line of the EO task (a) at the approximated x-coordinate of -50 mm.



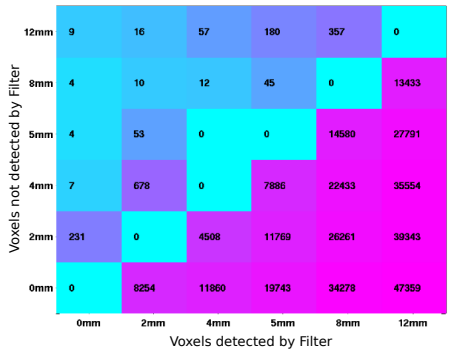
(a) Intra-method Dice Matrix



(b) Inter-method Dice Matrix

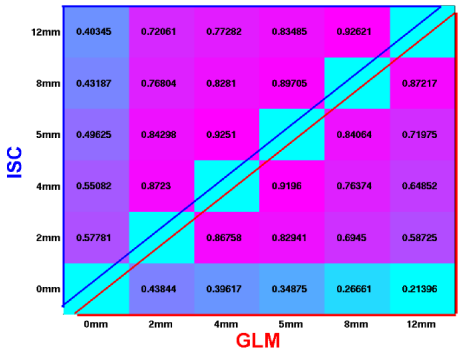


(c) Unique Voxel Count Matrix of GLM

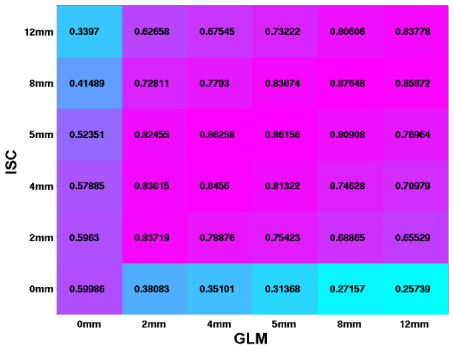


(d) Unique Voxel Count Matrix of ISC

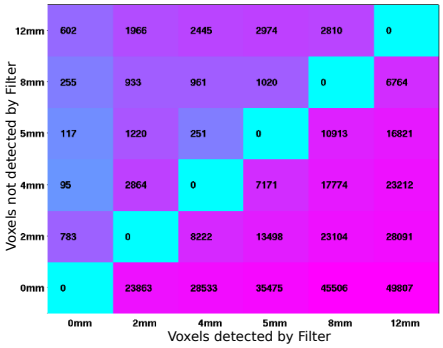
Figure 5: **The Dice and the unique voxel count matrices of the EO task.** See Figure 1 for the details of the Dice matrices in panels (a) and (b). In panel (a), intra-method Dice values are similar between the methods. In panel (b), the best similarity (0.77) between the methods was reached when ISC had a large 8 mm filter in use and GLM had a small 2 mm filter in use. The panels (c) and (d) present the unique voxel count matrices for GLM and ISC. When comparing the unique voxel counts of GLM (c) and ISC (d), it is noticeable that GLM had clearly more voxels which were detected by larger filters but not by smaller filters as indicated by relatively large values in the upper left triangle in the unique voxel count matrix in panel (c). This means that increased filter width led to the loss of some activation areas with GLM. Instead, ISC had relatively smaller values in the upper left triangle in the panel (d), which means that with ISC there was no such loss and the detected activation area merely grew with filter width.



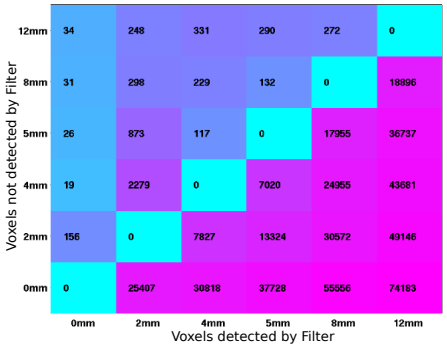
(a) Intra-method Dice Matrix



(b) Inter-method Dice Matrix



(c) Unique Voxel Count Matrix of GLM



(d) Unique Voxel Count Matrix of ISC

Figure 6: **The Dice and the unique voxel count matrices of the HA task.** See Figure 1 for the details of the Dice matrices in panels (a) and (b). In panel (a), intra-method Dice values of ISC were higher than the intra-method Dice values of GLM. The matrix in the panel (b) shows how the results of both methods were close the same with all filtering levels and the highest values were in the cross diagonal. The highest similarity (0.88) was reached when both methods had the filter kernel of 8 mm in use. When comparing the unique voxel counts of GLM in the panel (c) and ISC in the panel (d) the similar behavior was present as with the EO task in Figure 5: ISC had lower values in the upper triangle.

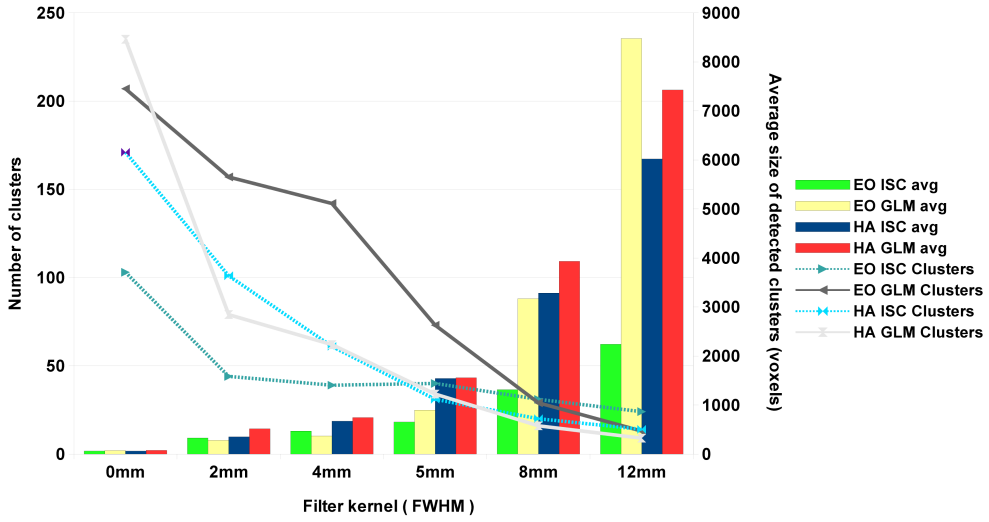


Figure 7: **The number of detected activation clusters versus the average sizes of clusters.** The bars present the average size of detected clusters in voxels and the lines present the number of the detected clusters. The size of clusters increases approximately at the same speed as the number of clusters decreases when filter kernel width increases. When comparing the number of voxels in clusters, cluster sizes and number of detected voxels in Figure 2 (a), one observation is that small clusters are merged together when filter kernel width increases. For example, the average cluster size with GLM in the EO task is almost three times larger with 12 mm filter than 8 mm filter.