

# Robust Estimation of the Probabilities of 3-D Clusters in Functional Brain Images: Application to PET Data

Anders Ledberg\*

*Division of Human Brain Research, Department of Neuroscience,  
Karolinska Institute, Stockholm, Sweden*

---

**Abstract:** Recently, we presented a method (the CS method) for estimating the probability distributions of the sizes of supra threshold clusters in functional brain images [Ledberg A, Åkerman S, Roland PE. 1998. Estimating the significance of 3D clusters in functional brain images. *NeuroImage* 8:113–128]. In that method, the significance of the observed test statistic (cluster size) is assessed by comparing it with a sample of the test statistic obtained from simulated statistical images (SSIs). These images are generated to have the same spatial autocorrelation as the observed statistical image (t-image) would have under the null hypothesis. The CS method relies on the assumptions that the t-images are stationary and that they can be transformed to have a normal distribution. These assumptions are not always valid, and thus limit the applicability of the method. The purpose of this paper is to present a modification of the previous method, that does not depend on these assumptions. This modified CS method (MCS) uses the residuals in the linear model as a model of a dataset obtained under the null hypothesis. Subsequently, datasets with the same distribution as the residuals are generated, and from these datasets the SSIs are derived. These SSIs are t-distributed. Thus, a conversion to normal distribution is no longer needed. Furthermore, no assumptions concerning the stationarity of the statistical images are needed. The MCS method is validated on both synthetical images and PET images and is shown to give accurate estimates of the probability distribution of the cluster size statistic. *Hum. Brain Mapping* 9:143–155, 2000.

© 2000 Wiley-Liss, Inc.

**Key words:** PET; statistics; linear model; cluster size; Monte Carlo

---

## INTRODUCTION

The detection of regional changes in tissue blood flow and blood oxygenation level is one of the main goals of functional brain imaging. This goal is often accomplished in two steps: In the first step, a linear

model is fitted to each and every voxel. Then, hypotheses are tested on the fitted parameters. Each tested hypothesis results in a statistical image (SI). These images are usually so called t-images, where the “t” refers to the fact that the obtained statistics (one for each voxel) are distributed as Student’s *t*. In the second step, the purpose is to decide which regional changes in the SIs are “true” changes and which are the result of random fluctuations. This way of analyzing functional brain imaging data was introduced by Friston and colleagues [Friston et al., 1990, 1991] and is usually referred to as statistical parametric mapping.

---

\*Correspondence to: Anders Ledberg, Division of Human Brain Research, Department of Neuroscience, Karolinska Institute, Berzelius väg 3, 5th floor, S-171 77 Stockholm, Sweden. email: Anders.Ledberg@neuro.ki.se.

Received for publication 1 July 1999; accepted 10 November 1999

To find the “true” changes, one usually derives a test statistic from the SI (e.g., global maximum or cluster size) and then compares the value of this statistic with values of the same statistic obtained under the null hypothesis. However, the exact probability distributions (pds) of these statistics are usually unknown, even under the assumption that the data is normally distributed. One therefore has to resort to some kind of estimation of the pd. This is a problem faced by all existing methods of statistical analysis of functional imaging data. Furthermore, the way in which this estimation is made constitutes the main differences between the different methods. One commonly used approach is to estimate the smoothness of the SI and then to use results from random field theory to obtain expressions for the pds of the different statistics: global maxima [Worsley et al., 1992, 1996] and cluster size (for Gaussian SI) [Friston et al., 1994; for a review, see Cao, 1997]. Recent developments in this field have enabled an approximate expression for the pd of cluster sizes also for t-images [Cao, 1997, 1999]. For a stationary t-image, this expression only is dependent on the smoothness and the degrees of freedom (df) of the t-image. However, since the smoothness usually is unknown, it has to be estimated, and the pds of the test statistics are dependent on this estimate. Another approach to obtain probabilities of different test statistics is the nonparametric method described in Holmes et al. [1996]. Here, a randomization approach is used to test hypotheses on the test statistics, and this approach is basically free from assumptions.

The fact that the pd of the test statistic is dependent on parameters that are difficult to estimate makes the statistical analysis of functional imaging data different from the traditional usage of statistical tests where the pd of the test statistic is known and often even tabulated. It is therefore an important task to find good estimators of the pd of test statistics in functional brain image analysis.

In a recent paper, we described a method (referred to as the CS method) that can be used to estimate the pd of cluster sizes [Ledberg et al., 1998]. This method utilizes a linear model to test hypotheses and thus to generate SIs (t-images). These images are subsequently converted to normal distributed images (i.e., Gaussianized). The cluster size (i.e., the number of connected voxels above a certain threshold) is the statistic used to detect the regional changes. The pd of this statistic is estimated with Monte Carlo simulations. In these simulations, a large number of random images (referred to as simulated statistical images, SSIs) are generated to represent SIs under the null hypothesis. These simulations are based on the as-

sumption that the Gaussianized t-images can be approximated by stationary normally distributed images. This is a good approximation for t-images with many degrees of freedom but is less good for t-images with few df and thus limits the usage of the CS method. Furthermore, the approach adopted to generate the SSIs, although technically correct is difficult to implement due to the relatively small size (i.e., number of voxels in each dimension) of the images, especially when the images are heavily smoothed.

The purpose of the present paper is to present a more robust method for generating the pd of the cluster size statistic under the null hypothesis. This method requires the statistical images to be neither normally distributed nor stationary. The main steps of the proposed method are: (i) fitting a linear model to the data, (ii) testing hypotheses on the parameters in the model, (iii) generating many SSIs having the same pd as the observed SI would have under the null hypothesis, and (iv) searching these generated SSIs for clusters above a certain threshold in order to estimate the pd of the cluster size statistic. The difference between this modified CS method (henceforth MCS method) and the CS method of Ledberg et al. [1998] concerns the generation of the SSIs (i.e., step iii). The SSIs are now generated from datasets having the same distribution as the residuals in the linear model. The variates in each and every voxel of these SSIs are t-distributed; thus, the original t-image need not be Gaussianized. Furthermore, the SIs need not be stationary for this approach to be applicable. Detailed descriptions of each step is given in the next section. The method is validated on two different sets of images: synthetical and real PET images. A comparison between the MCS and the CS method also is made. The validation shows that the MCS method gives reliable estimates of the pds, and the comparison with the CS method shows that the MCS method is indeed a significant improvement of the CS method.

## THEORY

### Notation

Let the following boldface upper case letters denote matrices and boldface lower case letters denote vectors. Sometimes, where appropriate, vectors will be referred to as images and the components of the vectors as voxels of the image. Let  $\mathbf{A}'$  denote the transpose of  $\mathbf{A}$ , let  $\mathbf{A}^-$  denote a generalized inverse of  $\mathbf{A}$ . Let  $\otimes$  denote the Kronecker product. Let  $N(\mathbf{M}, \Phi_n \otimes \Sigma_v)$  represent the multivariate normal distribution with mean value matrix  $\mathbf{M}$  and covariance matrix  $\Phi_n$

$\otimes \Sigma_v$ , where  $\Phi_n$  is a symmetric nonnegative definite  $n \times n$  matrix and  $\Sigma_v$  is a symmetric nonnegative definite  $v \times v$  matrix. This notation means that if  $\mathbf{U} = (\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_n)'$  is a  $n \times v$  random matrix, then to say that  $\mathbf{U}$  is distributed as  $N(\mathbf{M}, \Phi_n \otimes \Sigma_v)$  means that the  $nv \times 1$  vector  $(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)'$  is normally distributed with mean vector  $(\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_n)'$  and covariance matrix  $\Phi_n \otimes \Sigma_v$ , where  $\mathbf{m}_i$  is the  $i$ th row of the  $n \times v$  matrix  $\mathbf{M}$ . Let  $t_r$  represent the univariate Student's  $t$  distribution with  $r$  df. An image in which the variables in each and every voxels are distributed as  $t_r$  is called a  $t$ -image and the values in the voxels in this image are called  $t$ -values. With the df of a  $t$ -image is meant the df of the  $t$ -distributed values in each voxel. Let  $\text{diag}(\mathbf{A})$  be a matrix operation that sets all the off diagonal terms of  $\mathbf{A}$  to zero. Finally some notation concerning clusters. If  $\mathbf{t}$  is a  $t$ -image with a certain df, let  $\text{clu}(\mathbf{t}, \alpha)$  denote the sets of connected voxels in  $\mathbf{t}$  in which all voxels have a  $t$ -value equal to or higher than the  $t$ -value corresponding to a  $p$ -value of  $\alpha$ . Note that this  $t$ -value is dependent on the df of the image. For example, if  $\mathbf{t}$  has 53 df and  $\alpha = 0.01$  then  $\text{clu}(\mathbf{t}, \alpha)$  are the sets of connected voxels in  $\mathbf{t}$  in which all voxels have a  $t$ -value equal to or higher than 2.399.  $\alpha$  and the corresponding  $t$ -value will be referred to as the "threshold" of  $\mathbf{t}$  interchangeably. These sets of connected voxels are called "clusters," and the size of a cluster is the number of voxels it consists of. Let  $s(\text{clu}(\mathbf{t}, \alpha))$  denote the set of sizes of the clusters in  $\mathbf{t}$  above the threshold  $\alpha$  and let  $\max\{s(\text{clu}(\mathbf{t}, \alpha))\}$  denote the maximum of these sizes.

### Fitting a linear model to the data

Consider a set of functional images obtained in a number of subjects. As an example take six subjects, ten scans per subject, five obtained under the condition TEST and five under CONTROL. All images are assumed to be in the same anatomical format so that voxel-wise comparisons make sense. Let  $\mathbf{Y}$  be a  $n \times v$  matrix where  $n$  is the number of scans (i.e., 60 in the example) and  $v$  is the number of voxels in each scan. That is, each scan is represented as a row in  $\mathbf{Y}$ . Let  $\mathbf{X}$  be a  $n \times p$  design matrix; thus, there are  $p$  parameters in the model. A linear model for the data is then

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} \quad (1)$$

Here  $\mathbf{B}$  is a  $p \times v$  matrix that contains the parameters for each voxel and  $\mathbf{E}$  is a  $n \times v$  matrix of error terms. The df of this model is  $\gamma = n - \text{rank}(\mathbf{X})$ . This model is equivalent to the design model described in Ledberg

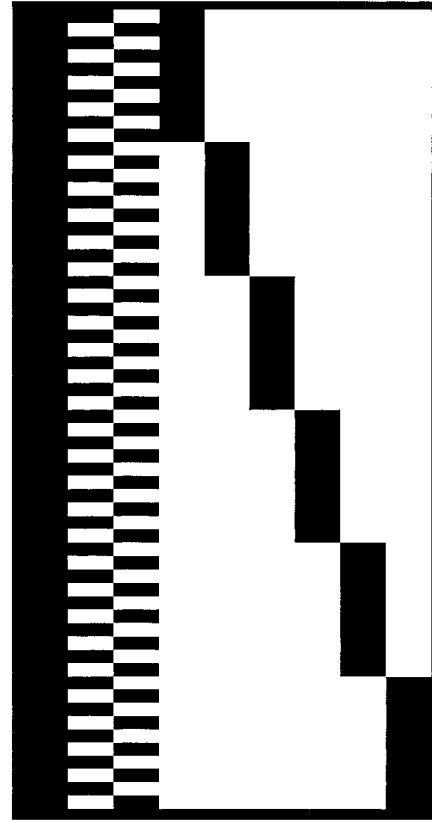


Figure 1.

This figure shows the design matrix used. The columns refer to mean value (one column), task factor (two columns) and subject factor (six columns). The rows represents the scans.

et al. [1998] but formulated for all voxels at the same time. Figure 1 shows the design matrix for the example case. Only two factors, condition (two levels) and subject (six levels), are used in this case. Let  $\hat{\mathbf{B}}$  be the matrix containing the parameter estimates (i.e.,  $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ) and let  $\mathbf{R} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}$  be the matrix containing the residuals. To proceed, the following assumption is needed:  $\mathbf{E}$  is distributed as  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$  and  $\text{rank}(\Sigma) = v$ .

### Testing hypotheses on the estimated parameters

Almost all hypotheses of interest can be formulated as linear combinations of the estimated parameters, so called contrasts. In the example, consider the contrast TEST-CONTROL. That contrast would be implemented by  $\mathbf{h}\hat{\mathbf{B}}$ , with  $\mathbf{h} = [0 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$ . The variance of this contrast can be estimated as  $\widehat{\text{var}}(\mathbf{h}\hat{\mathbf{B}}) = \text{diag}\{\mathbf{R}'\mathbf{R}(\mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}')\}/\gamma$  [Graybill, 1976; Muirhead, 1982]. If  $\mathbf{h}\hat{\mathbf{B}}$  is divided with the square root of this estimated variance we get a  $\mathbf{S}\mathbf{I} \mathbf{t}$  say,

$$\mathbf{t} = (\mathbf{h}\hat{\mathbf{B}})(\widehat{\text{var}}(\mathbf{h}\hat{\mathbf{B}}))^{-1/2} \quad (2)$$

Such a SI is referred to as a t-image, because the variable in each and every voxel is distributed as  $t_\gamma$ .

To detect the significant regional changes in  $\mathbf{t}$ ,  $s(\text{clu}(\mathbf{t}, \alpha))$  is compared to the pd of  $\max\{s(\text{clu}(\mathbf{t}_{\text{null}}, \alpha))\}$  where  $\mathbf{t}_{\text{null}}$  denotes a t-image obtained under the null hypothesis. A cluster of size  $s$  is classified as significant at the 0.05 level if

$$p(\max\{s(\text{clu}(\mathbf{t}_{\text{null}}, \alpha))\} \geq s) < 0.05 \quad (3)$$

The next issue then concerns how to estimate the pd of  $\max\{s(\text{clu}(\mathbf{t}_{\text{null}}, \alpha))\}$ . This pd will be estimated by generating many images having the same pd as  $\mathbf{t}$  would have had under the null hypothesis and sampling these simulated t-images for the largest clusters. How these SSIs are generated is the issue of the next section.

### Generating the images

This is the most important step in the MCS method and is also the step where the MCS method differs from the CS method. The generation of the images is based on a very simple idea described below. Some of the more technical aspects are dealt with in the Appendix.

The main idea is to generate a set of images having the same distribution as the real images would have under the null hypothesis and then to derive a t-image from these images. The original data matrix  $\mathbf{Y}$  should not be a realization of the null hypothesis (because there should be differences between TEST and CONTROL). Instead, the residuals  $\mathbf{R}$  will be used as a model for a dataset obtained under the null hypothesis. This is intuitively appealing since if  $\mathbf{B} = \mathbf{0}$  in the linear model (Eq. 1), then  $\mathbf{R} \approx \mathbf{Y}$ . The issue is then to generate datasets having the same distribution as  $\mathbf{R}$ . To do this, one can use the fact that distributions of the type  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$  are invariant under left multiplication with orthogonal matrices [i.e., if  $\mathbf{Z}$  is distributed as  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$ ] and  $\Gamma$  is an orthogonal  $n \times n$  matrix then  $\Gamma\mathbf{Z}$  will also be distributed as  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$ . However, the situation is slightly more complicated since  $\mathbf{R}$  is distributed as  $N(\mathbf{0}, \mathbf{M} \otimes \Sigma)$  where  $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ , so it has to be transformed into a distribution of the  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$  type for the invariance argument to work. How this can be done is shown in the Appendix. Thus, a matrix,  $\tilde{\mathbf{R}}$  say, distributed as  $N(\mathbf{0}, \mathbf{I}_\gamma \otimes \Sigma)$  is generated. From this matrix, new matrices,  $\mathbf{R}_i$  say, with the same distribution, are derived as follows:  $\tilde{\mathbf{R}}_i = \Gamma_i \tilde{\mathbf{R}}$  where  $\Gamma_i$  is a random

orthogonal matrix of size  $\gamma \times \gamma$ . From these  $\tilde{\mathbf{R}}_i$  images, t-images are generated by, for each column, dividing the estimated mean value with the estimated standard deviation. This will generate a t-image with  $\gamma - 1$  df.

### Estimating the probability distribution

The df of the simulated t-images generated as described above, is  $\gamma - 1$ , whereas the df of the original t-images is  $\gamma$ . However, by using thresholds corresponding to the same  $\alpha$ , this discrepancy should not matter. Thus, each of the simulated t-images is searched for clusters above the threshold given by  $\alpha$  at  $\gamma - 1$  df, whereas the real t-images are searched for clusters above the threshold given by  $\alpha$  at  $\gamma$  df. The estimation of the pd of  $\max\{s(\text{clu}(\mathbf{t}, \alpha))\}$  from the simulated t-images is done as follows: the probability of obtaining a cluster of size  $s$  say, is estimated by dividing the number of simulated t-images in which a cluster of size  $s$  or larger occurred with the total number of simulated t-images [see Eq. 11 in Ledberg et al., 1998].

## VALIDATION

In this section, the MCS method is validated and compared to the CS method. This validation/comparison is made on two different sets of data. The first set consists of synthetical (i.e., computer-generated) noise images, and the second set consists of real PET data. Because the true pd of the cluster size statistic is unknown, the validity of the pd estimated with MCS method will be investigated by a comparison with a reference pd derived by a randomization procedure. The choice of reference pd is important, and is the issue of the next section.

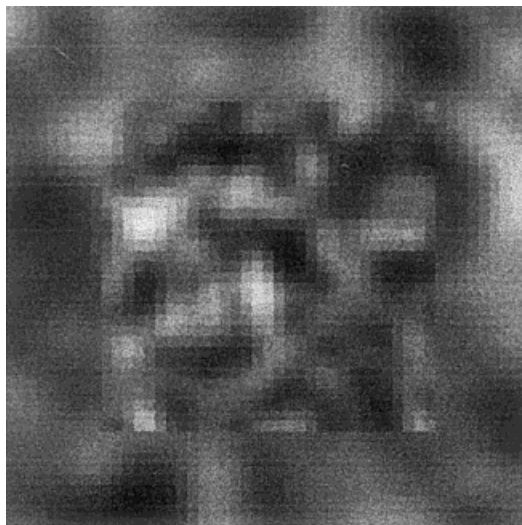
### Generating the reference distribution

To see whether the MCS method gives an accurate estimation of the pd of the cluster size statistic, one would like to know the true distribution. Because this distribution is unknown, it also needs to be estimated by some means. I have chosen to use a nonparametric randomization procedure to derive the reference distribution. A randomization procedure is not dependent on any strong assumptions about the data (assumptions that might be wrong) and is in this respect, a good way to generate the reference distribution. A general introduction to randomization tests is given in Edgington [1980], and for an applications to PET images, see Holmes et al. [1996] and Ledberg et al.

[1998].<sup>1</sup> As argued in Ledberg et al. [1998], a randomization procedure might lead to conservative estimates of the pd of the cluster size statistic. This is because some of the randomizations will be close to the original ordering of the scans, and thus the derived t-image will contain signal. One way around this problem is to use a step-down procedure [Holmes et al., 1996]. In the step-down procedure, one starts with a randomization on the whole image (the first step). The voxels belonging to significant clusters (if any) are then removed and a new set of randomizations is performed on the remaining voxels (the second step). The significant clusters (if any) are again removed and a new set of randomizations are performed (the third step). In principle, one can continue as long as there are any significant clusters in the original t-image.

### Synthetical images

Three different sets of synthetical images of size  $64 \times 64 \times 64$  voxels were generated, with the value of each voxel distributed as  $N(0, 1)$ . The images in the first set were smoothed with a Gaussian kernel of three voxels FWHM and those in the second set with a kernel of six voxels FWHM. The images in the third set were smoothed with a Gaussian kernel of three voxels FWHM in the central part and with a kernel of six voxels FWHM in the peripheral parts. The purpose of using two filters of different widths in the same image was to generate nonstationary images (Fig. 2). To the first two sets of images, three linear models of 14, 30, and 62 df, respectively, were applied. For each combination of smoothness and df, the pd of the cluster size statistic at two different thresholds was determined with the MCS method using 5000 SSIs. The reference distribution was calculated by the randomization procedure. The results are shown in Figures 3 and 4. Because in this case there are no signals in the images, the randomizations will give the correct distribution without using the step-down approach. However, for low df, only very few randomizations can be made, which makes the estimates noisy (as can be seen in the figures). It is clear from the figures that the MCS method gives accurate estimates for  $p$ -values above 0.01 for both filter widths and all df. For smaller  $p$ -values, the lack of fit between the MCS and the



**Figure 2.**

This figure shows a section from one of the nonstationary images used in the validation. The central part was filtered with a Gaussian filter with three voxels FWHM and the peripheral part with a Gaussian filter with six voxels FWHM.

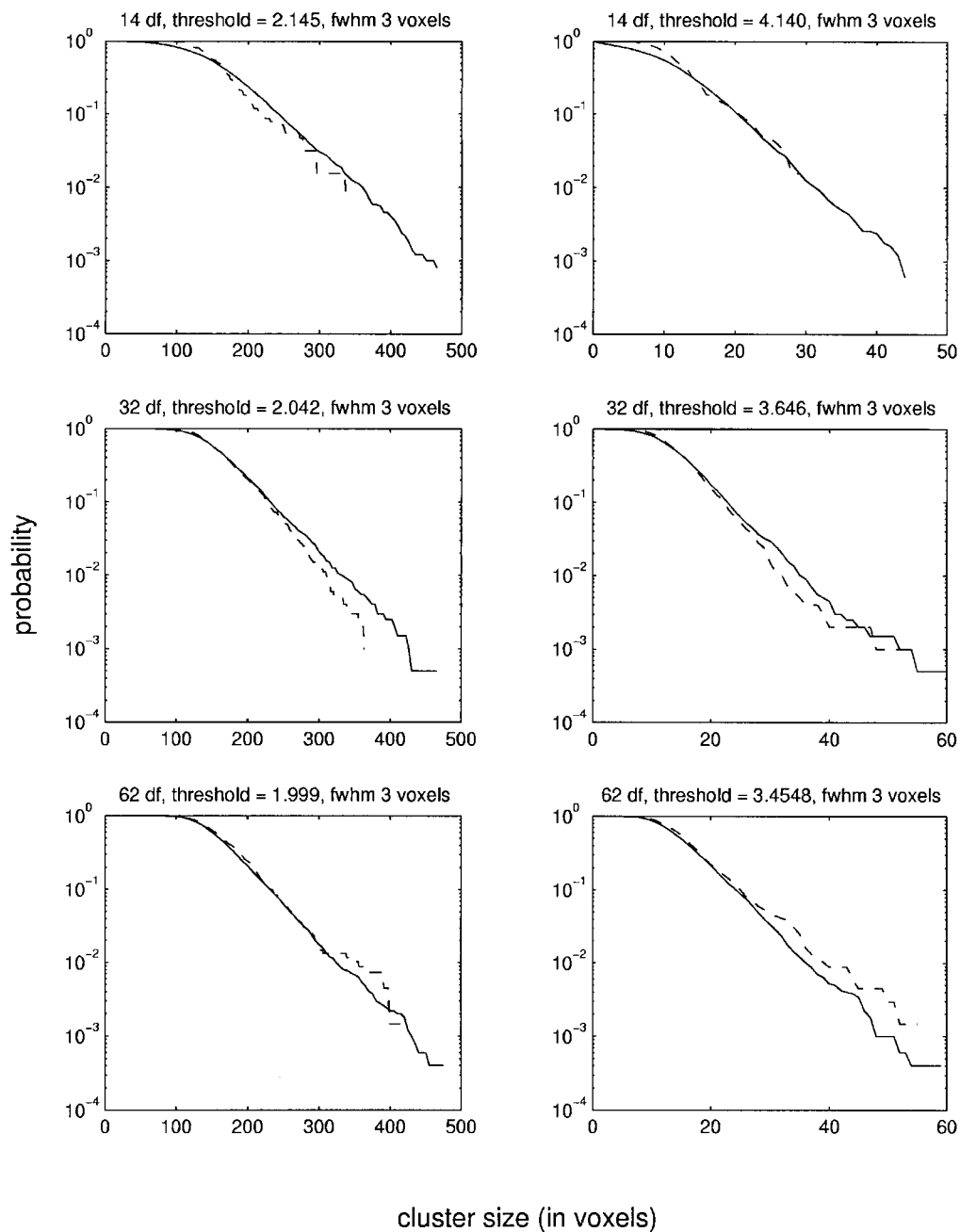
randomization method is probably because of the low number of randomizations used. This is because estimating the tails of the pd requires many simulations/randomizations. To verify that the two estimates do converge, a large number of simulations/randomizations were made on a set of smaller images. These images were the central  $39 \times 39 \times 39$  voxels of the images comprising the three voxel FWHM dataset. 15000 simulations/randomizations were made on these images with a model of 62 df. The results are shown in Figure 5. It is clear that the two estimates are very similar if enough simulations/randomizations are made.

For the third set of images, the nonstationary set, a model of 62 df was applied. The pd of the cluster size statistic was estimated with the MCS method as well as with the CS method. The reference distribution was calculated with the randomization procedure. The result is shown in Figure 6. It is clear that the MCS method gives very accurate estimates also on these nonstationary images. However, the CS method performed less well, as expected.

### Real PET data

To evaluate the performance of the MCS method on real PET data, it was applied to the same dataset used in the evaluations in Ledberg et al. [1998]. A short description of the experiment and image preprocessing is given on the next page.

<sup>1</sup>Note that, in this paper, the randomization procedure is used as a tool to estimate the "true" underlying pd. This differs from the usage of the same procedure in randomization tests where one is not interested in any underlying pd and where the "distribution" of the test statistic obtained by the randomizations always (by definition) is the "true" distribution.



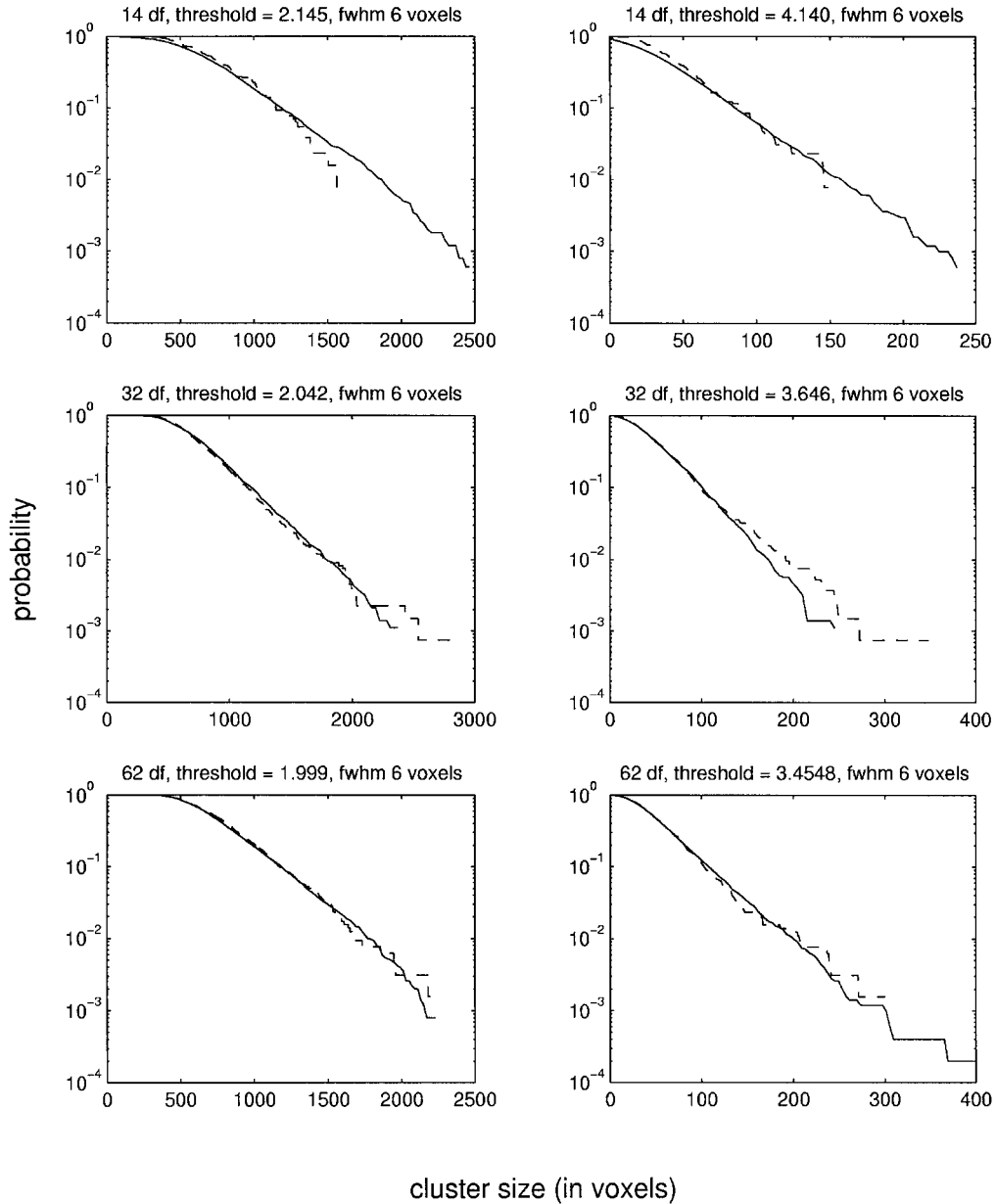
**Figure 3.**

This figure shows the estimated probabilities of the cluster size statistic. The data are from the synthetic images filtered with a Gaussian filter of three voxels FWHM. Two different thresholds were used for each df. The dashed line shows the reference

distribution as determined by randomizations. The solid line shows the probabilities estimated with MCS method. Please note that a logarithmic scale is used for the y-axis.

Six subjects participated in this study. They all gave informed consent. The study was approved by the local Ethics Committee and the Radiation Safety Committee of the Karolinska Hospital. Each subject was scanned in a PET scanner (SIEMENS-CTI ECAT

EXACT HR) during ten injections of 15-O labeled butanol. The experimental paradigm consisted of two conditions each repeated five times. The conditions will be referred to as TEST and CONTROL. For each subject, the ten injections were divided into five blocks



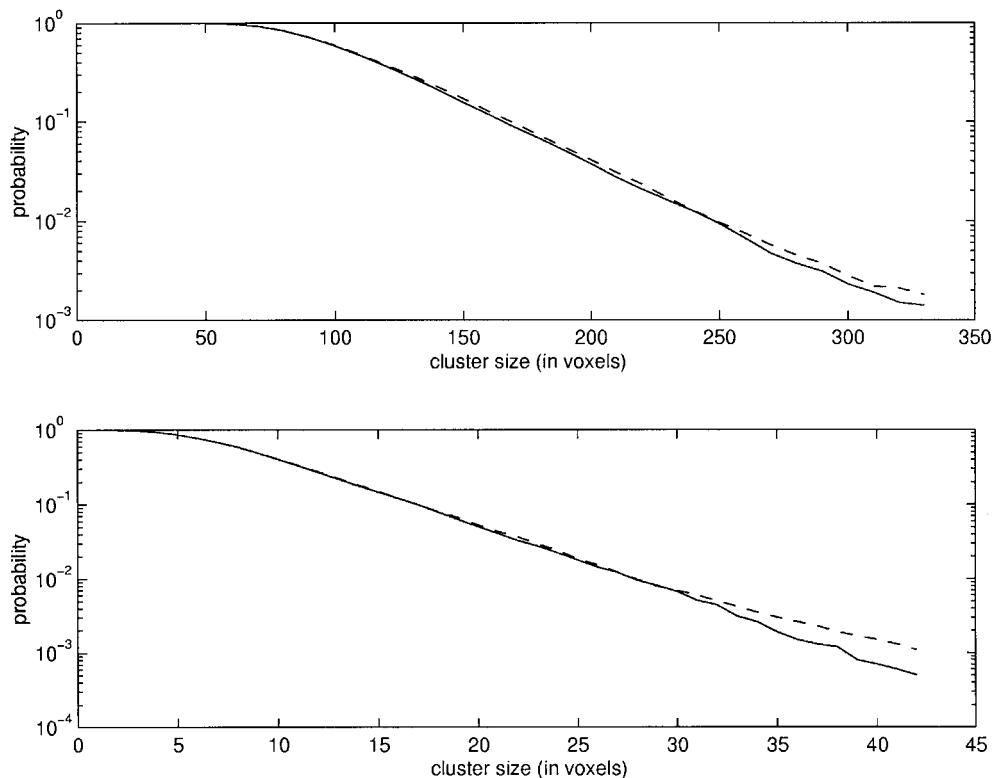
**Figure 4.**

This figure shows the estimated probabilities of the cluster size statistic. The data are from the synthetic images filtered with a Gaussian filter of six voxels FWHM. Two different thresholds were used for each df. The dashed line shows the reference

distribution as determined by randomizations. The solid line shows the probabilities estimated with the MCS method. Please note that a logarithmic scale is used for the y-axis.

of two injections each. In each block, the order between TEST and CONTROL was randomized with the restriction that each block had to contain one TEST and one CONTROL condition. The TEST consisted in a two alternative forced choice tactile discrimination paradigm [Roland and Mortensen, 1987]. The subjects discriminated which of two aluminum parallelepipeda was the most oblong. They made the discrimi-

nation using active palpation with their right hand. In CONTROL, the subjects were lying in the PET scanner with their eyes closed and were instructed not to do or think anything in particular. Arterial radioactivity was continuously monitored during all injections. All data was acquired with the PET scanner operating in 3-D mode. The images were reconstructed with a ramp filter having a cutoff frequency of 0.5 to a voxel size of



**Figure 5.**

This figure shows the estimated probabilities of the cluster size statistic on a subset of the three voxel FWHM dataset. 15,000 simulations/randomization were made. The solid and dashed lines refer to the MCS and randomization methods, respectively. The

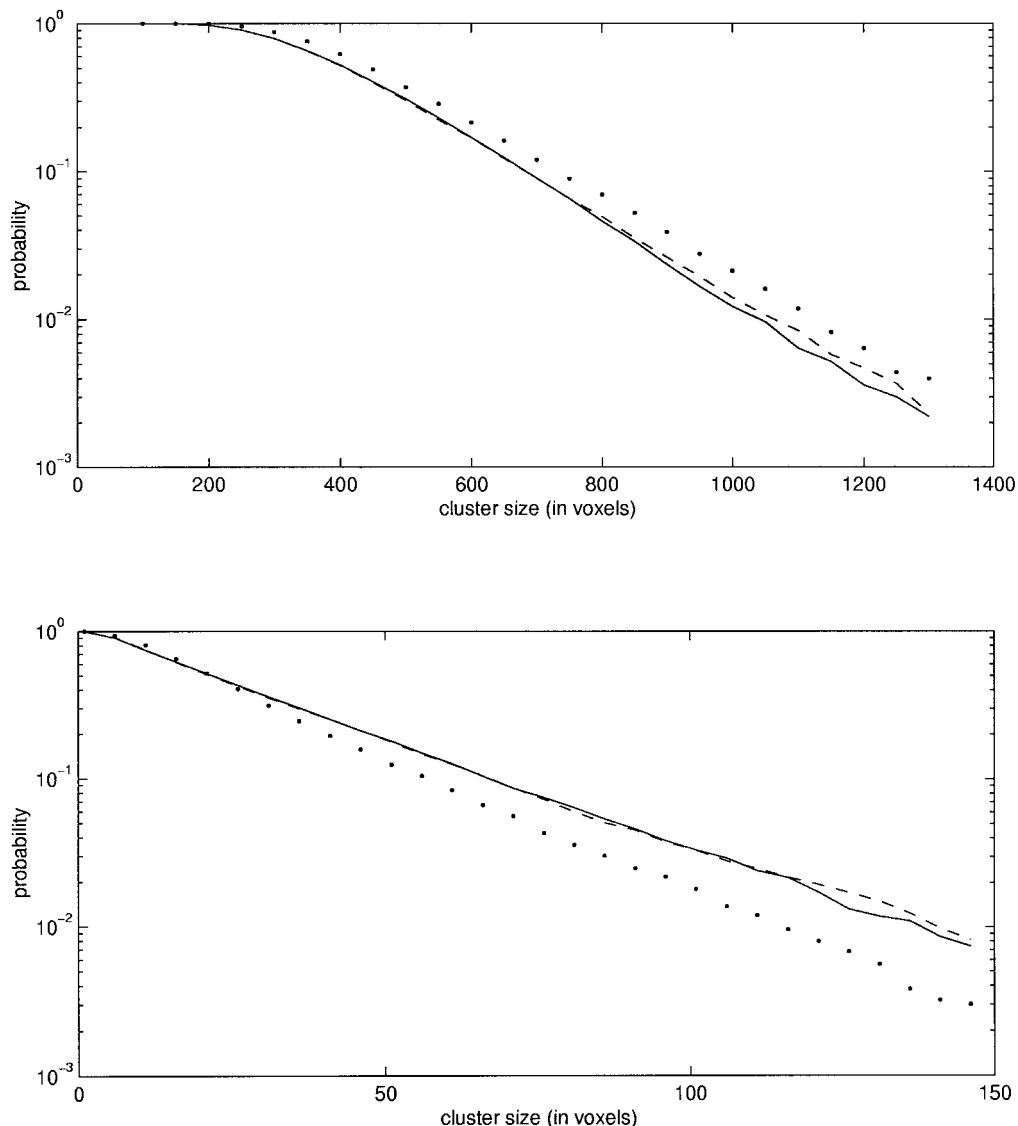
top figure shows cluster sizes above a threshold of 2.0 and the lower figure above thresholds of 3.457 and 3.454 for the MCS and randomization methods, respectively.

$2 \times 2 \times 3.125$  mm. In all images, only data in voxels contained inside the brain were used for further analysis, all other voxels were set to zero. This zeroing was done to prevent filtering in spurious data located outside the brain. The images were then filtered with a Gaussian filter of 6 mm FWHM in all dimensions. For each individual, the images from injections two to ten were aligned to the image from the first injection using the AIR-software [Woods et al., 1992]. Regional cerebral blood flow (rCBF) was calculated using an autoradiographic method [Meyer, 1989]. Subsequently, we used information obtained from individual MR scans to reformat all images into a standard anatomical space [Roland et al., 1994] with cubic voxels of  $2 \times 2 \times 2$  mm. The global cerebral blood flow was normalized to 50 ml/100 gmin by multiplication by a constant.

The design matrix used to model the data is shown in Figure 1. The pd of the cluster size statistic for this model was estimated with the MCS and CS methods at two different thresholds corresponding to  $\alpha = 0.01$  and  $\alpha = 0.0005$ . The reference pd in this case was also generated using the randomization procedure. Unlike

the above case of synthetic images, however, these real PET images do contain a strong signal that will make the distributions generated by the randomizations too conservative. However, by using the step-down procedure described above, this bias can be minimized. For the present dataset, three steps were used and, at each step, the significant clusters at 0.05 level were removed. The number of randomizations was 5000 at each step. The pd obtained in the third step was used as the reference distribution. Because this pd is obtained from a subset of the original voxels (the voxels of the significant clusters in the previous two steps have been removed), the MCS and CS methods were also applied to the same subset of voxels. Thus, the pd derived from the CS and MCS methods and the reference pd were generated from the same set of voxels, and because these voxels do not contain any signal, the distributions should be the same. For the MCS and CS methods, 5000 simulations were used to estimate the pds. The results are shown in Figure 7. It is clear that, at both thresholds, the MCS method gives reliable estimates of the probabilities for probabilities





**Figure 6.**

This figure shows the estimated probabilities of the cluster size statistic for the nonstationary dataset. In the upper figure, the threshold used was 2.0, and in the lower it was 3.457 and 3.454 for

the MCS and randomization methods, respectively. The solid, dashed, and dotted lines refer to the MCS, randomization, and CS method, respectively.

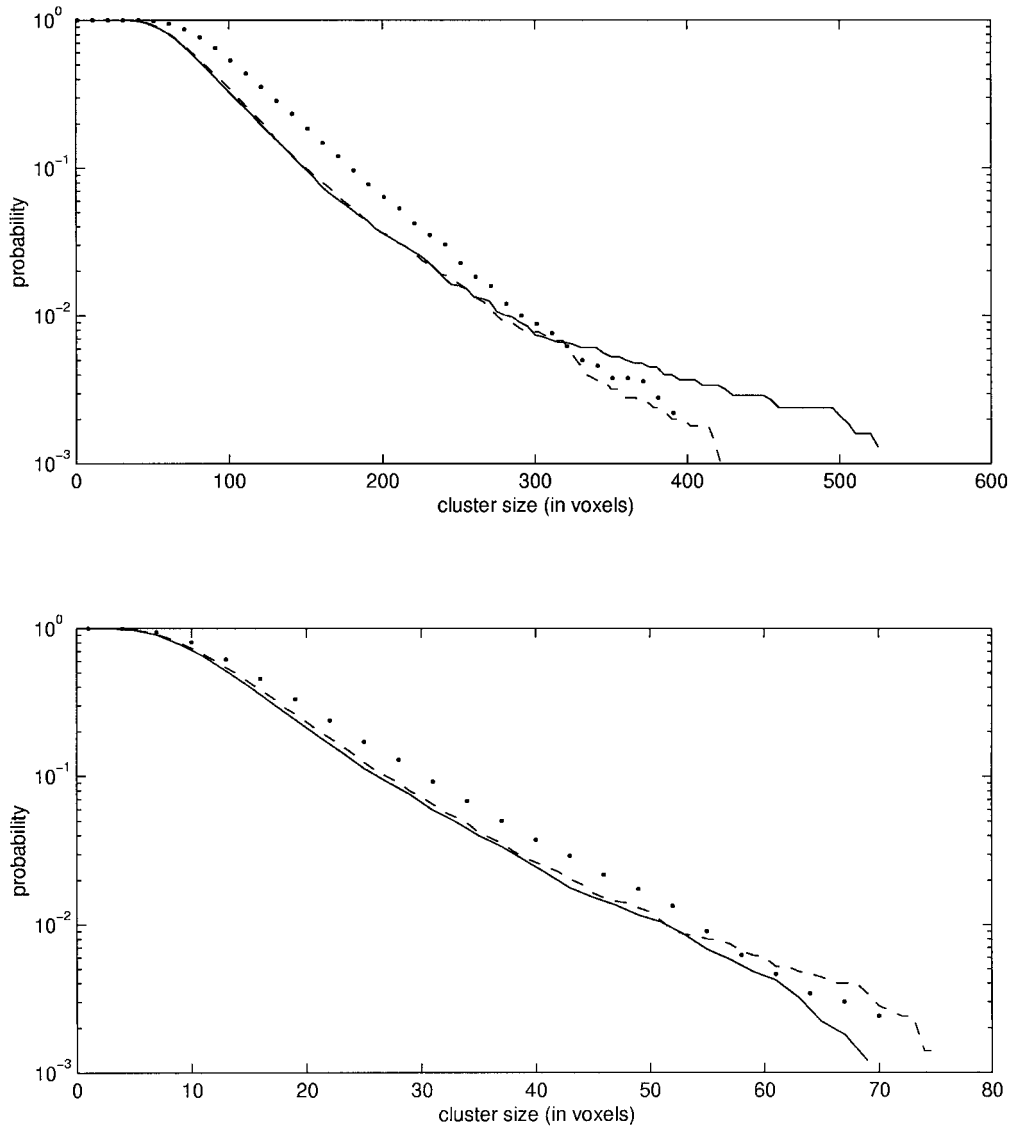
higher than 0.01. Furthermore, it is clear from Figure 7 that the MCS method is better than the CS method because it gives an estimate closer to the one obtained by the randomizations.

## DISCUSSION

### MCS vs. CS

This paper has presented a novel way of estimating the probabilities of cluster sizes in PET images. The method is a modification of a previously presented

method, the CS method of Ledberg et al. [1998]. Although very similar in spirit, there are some important differences between the two methods. In the CS method, it is assumed that the t-images obtained under the null hypothesis are stationary and that they can be converted to a normal distribution (Gaussianized). These assumptions are not always valid. Furthermore, the relatively small sizes (i.e., number of voxels in each dimension) of the images makes the CS method less well suited for images with a broad autocorrelation function (acf) (i.e., smooth images). This is because the estimator of the acf [Eq. 2 in Ledberg et



**Figure 7.**

This figure shows the estimated probabilities of the cluster size statistic for the real PET dataset. The thresholds in the upper figure are 2.399, 2.400, and 2.33 for the randomization, MCS, and CS methods, respectively, and in the lower figure 3.484, 3.488, and 3.29. The dashed line corresponds to the reference distribution as

determined with the randomization procedure (step three). The solid line shows the results of the MCS method and the dotted line shows the estimate obtained by the CS method. Please note that a logarithmic scale is used for the y-axis.

al., 1998] is noisy at large lags and smooth images will have a nonzero acf at large lags. Another drawback of the CS method (at least in its current implementation) is that it requires the user to set the value of a parameter used to scale the filter kernel used to generate the SSIs, so that the result can in this respect be dependent on the particular user. In the MCS method, on the other hand, the t-images need not be stationary, they are not converted into a normal distribution, and there are no parameters to be set by the user. The only

assumption needed in the MCS method is that the error terms in the model (Eq. 1) are distributed as  $N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$ .

The validation showed that the MCS method gives accurate estimates of the pd of the cluster size statistic independent of the df of the models and the smoothness of the images. Furthermore, it was shown that the MCS method also gives good estimates on nonstationary images. This was in contrast to the CS method that did not work so well on these images (Fig. 6). That the

CS method performed less well on the nonstationary images is not surprising because the estimator of the acf used assumes that the image is stationary. The impact of nonstationarity on the estimated pds will most likely vary depending on the source nonstationarity. Because it is unknown whether the SIs obtained in real PET experiments are stationary or not, one should perhaps be a little careful in using methods assuming stationarity. The comparison between the MCS and CS methods on the real PET data showed that the MCS method generated pds closer to the ones obtained by the randomizations. The CS method generated too conservative estimates, something that was observed also in Ledberg et al. [1998]. That the MCS method performed better on the real PET image might be because the assumption of stationarity is violated in these images. It could, however, have other explanations as well.

Because the MCS method gave better results than the CS method on both nonstationary and real PET data, it is indeed a significant improvement.

#### MCS vs. randomizations

Because both the randomization and MCS methods require quite massive computations to estimate the pds for each dataset and, as shown in the validation, gives very similar estimates of the pds, it might seem that the MCS method is superfluous: why not always use the randomizations? However, the randomization approach has two drawbacks: (i) The number of possible randomizations can be very low for designs with a low df. This implies that the estimates might be conservative. (ii) Not all designs are suitable for randomization tests. Furthermore, if the images contain a signal, especially a strong signal, the estimated pds will be too conservative unless a (time-consuming) step-down approach is used. The MCS method does not suffer from these shortcomings. On the other hand, the MCS method relies on the assumption that the error terms in the linear model have a multivariate normal distribution. This assumption is not needed in the randomization approach. Another difference between the MCS and randomization methods is that the SSIs generated in the MCS method have  $df = \lambda - 1$ , whereas the original t-images as well as the SSIs from the randomization have  $df = \lambda$ . However, this difference did not seem to matter if the appropriate thresholds were used, as shown in the validation.

Thus, the applicability of the MCS method complements that of the randomizations. The only limiting factor is the assumption of multivariate normality of the residuals. This assumption can probably be weak-

ened to any distribution invariant under left multiplication by orthogonal matrices.

#### Random field and nonstationarity

To be able to apply the distributional results from random field theory, the SI need to be stationary. However, a technique developed recently by Keith Worsley and colleagues [Worsley et al., 1999] might be a way out of this limitation. Given a nonstationary SI, the idea is to warp this image into a new space where it is stationary. In this new space, one calculates the threshold corresponding to a desired  $\alpha$ , applies this threshold, and then warps the image back to the original space.

#### Implementation

The MCS requires an estimation of the pd for each study where it is applied. These estimations are time consuming and require a computer with much memory. The current implementation required 512 Mbyte fast (100 MHz) RAM and the 5000 simulations on the real PET dataset took approximately 24 h on a computer with an Intel™ Pentium™ II processor of 400 MHz clock frequency. Probably, in studies with a very similar design and data preprocessing, the same cluster size thresholds will turn out to be significant. However, with small groups of subjects, some variance can be expected. The time to do one simulation is proportional to the square of the number of voxels times the number of scans (i.e., double the number of scans and the time it takes to make one simulation will increase by a factor of four). The software is available for Sun Solaris and Linux operative systems on request to the author.

#### CONCLUSIONS

The MCS method was shown to give accurate estimates of the probabilities of suprathreshold clusters in both stationary and nonstationary images. It was also shown to give “better” estimates than the CS method of Ledberg et al. [1998]. Thus, the MCS is a possible alternative to randomization tests in situations where the assumptions of the more theoretical approaches are violated.

#### ACKNOWLEDGMENTS

I would like to thank Dietrich von Rosen for useful comments concerning the statistical details and Christine Schiltz, Per E. Roland, Jonas Larsson, Karl Mag-

nus Petersson, and Rita Almeida for constructive criticisms on the many previous versions of this manuscript.

## REFERENCES

- Cao J (1997): The excursion set of random fields with applications to human brain mapping. PhD thesis, Department of Mathematics and Statistics, McGill University, Montreal.
- Cao J (1999): The size of the connected components of excursion sets of  $\chi^2$ ,  $t$ , and  $F$  fields. *Adv Appl Prob*. In press.
- Eaton M (1983): *Multivariate statistic: A vector space approach*. New York: Wiley.
- Edgington ES (1980): *Randomization tests*. New York: Dekker.
- Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RSJ (1990): The relationship between global and local changes in PET scans. *J Cereb Blood Flow Metab* 10:458–466.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991): Comparing functional (PET) images: The assessment of significant change. *J Cereb Blood Flow Metab* 11:690–699.
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994): Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapping* 1:210–220.
- Graybill F (1976): *Theory and application of the linear model*. North Scituate: Duxbury Press.
- Holmes AP, Blair RC, Watson JDG, Ford I (1996): Nonparametric analysis of statistic images from functional mapping experiments. *J Cereb Blood Flow Metab* 16:7–22.
- Ledberg A, Åkerman S, Roland PE (1998): Estimating the significance of 3D clusters in functional brain images. *NeuroImage* 8:113–128.
- Meyer E (1989): Simultaneous correction for tracer arrival delay and dispersion in CBF measurements by the H215O autoradiographic method and dynamic PET. *J Nucl Med* 30:1069–1078.
- Muirhead RJ (1982): *Aspects of multivariate statistical theory*. New York: Wiley.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992): *Numerical recipes in C (2nd Edition)*. Cambridge: Cambridge University Press.
- Roland PE, Mortensen E (1987): Somatosensory detection of microgeometry, macrogeometry and kinesthesia in man. *Brain Res Rev* 12:1–42.
- Roland PE, Graufelds CJ, Wählin J, Ingelman L, Andersson M, Ledberg A, Pedersen J, Åkerman S, Dabringhaus A, Zilles K (1994): Human brain atlas: For high-resolution functional and anatomical mapping. *Hum Brain Mapping* 1:173–184.
- Woods RP, Cherry SR, Mazziotta JC (1992): Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr* 16:620–633.
- Worsley KJ, Evans AC, Marrett S, Neelin P (1992): A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab* 12:900–918.
- Worsley KJ, Marret S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996): A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapping* 4:58–73.
- Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC (1999): Detecting changes in nonstationary images via statistical flattening. *NeuroImage* 9:s11.

## APPENDIX

Here, it is described how the simulated statistical images are generated both in the randomization and MCS methods. The assumption about the residuals in the randomization method is not really needed but it is used to show how closely related the two methods are. For proofs of some of the statements below, see Eaton [1983].

### A linear model for the data

Let  $\mathbf{Y}$  be a  $n \times v$  matrix containing the observations, with  $n$  scans and  $v$  voxels per scan ( $v \gg n$ ). Let

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} \quad (\text{A1})$$

be a linear model for the data in  $\mathbf{Y}$  where  $\mathbf{X}$  is the design matrix,  $\mathbf{B}$  is a matrix containing the parameters, and  $\mathbf{E}$  is a matrix of errors. The errors are assumed to be distributed as  $N(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$ . Let  $\mathbf{h}$  be a testable contrast, then an estimate of  $\mathbf{h}\mathbf{B}$  is given by

$$\mathbf{m}_{\text{original}} = \mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (\text{A2})$$

$\mathbf{m}_{\text{original}}$  is distributed as  $N(\mathbf{h}\mathbf{B}, \mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}'\boldsymbol{\Sigma})$  and an estimate of the variance of  $\mathbf{m}_{\text{original}}$  is given by

$$\widehat{\text{var}}(\mathbf{m}_{\text{original}}) = (\text{diag}(\mathbf{Y}'\mathbf{M}\mathbf{Y})\mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}')/\gamma \quad (\text{A3})$$

where  $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$  and  $\gamma$  is the degrees of freedom of the model (i.e.,  $n - \text{rank}(\mathbf{X})$ ). By dividing the estimate of  $\mathbf{h}\mathbf{B}$  with the square root of the estimated variance, we get a t-image:

$$\mathbf{t}_{\text{original}} = \mathbf{m}_{\text{original}}(\widehat{\text{var}}(\mathbf{m}_{\text{original}}))^{-0.5} \quad (\text{A4})$$

The two different approaches to generate simulated statistical images having a similar distribution to what  $\mathbf{t}_{\text{original}}$  would have under the null hypothesis are described below.

### Randomizations

Let  $\mathbf{P}_i$  be a  $n \times n$  permutation matrix. Assume that the permutation described by  $\mathbf{P}_i$  makes sense given the present design. Then let  $\mathbf{X}_i = \mathbf{P}_i\mathbf{X}$  where  $\mathbf{X}$  is the same design matrix as above. Then we have a new linear model for the data, namely,  $\mathbf{Y} = \mathbf{X}_i\mathbf{B}_i + \mathbf{E}_i$ . Using the same contrast  $\mathbf{h}$  as above we get

$$\mathbf{m}_{\text{rand}} = \mathbf{h}(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{Y} = \mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'_i\mathbf{Y} \quad (\text{A5})$$

and under the assumption that the residuals in the new model have the same distribution as the residuals in the old model, we have the following distribution of  $\mathbf{m}_{\text{rand}}$ :  $N(\mathbf{h}\mathbf{B}_i, \mathbf{h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}'\boldsymbol{\Sigma})$ . Thus,  $\mathbf{m}_{\text{rand}}$  has the same variance–covariance structure as  $\mathbf{m}_{\text{original}}$ . In complete analogy of how  $\mathbf{t}_{\text{original}}$  was formed, we can now form  $\mathbf{t}_{\text{rand}}$ . If the null hypothesis is true, the distribution of  $\mathbf{t}_{\text{rand}}$  will be the same as that of  $\mathbf{t}_{\text{original}}$ .

### Simulations<sup>2</sup>

Let  $\mathbf{Y}$ ,  $\mathbf{X}$ ,  $\mathbf{M}$ , and  $\mathbf{E}$  be as above. Let  $\mathbf{R}$  be the residuals of the model in Eq. (A1) (i.e.,  $\mathbf{R} = \mathbf{Y} - \mathbf{X}\mathbf{B} = \mathbf{M}\mathbf{Y}$ ).  $\mathbf{R}$  is distributed as  $N(\mathbf{0}, \mathbf{M} \otimes \boldsymbol{\Sigma})$ . Because  $\mathbf{M}$  is an orthogonal projection, it can be decomposed into the

product of two orthogonal matrices,  $\mathbf{M} = \mathbf{K}\mathbf{K}'$ , say with  $\mathbf{K}$  being of size  $n \times \gamma$ . It then follows that  $\mathbf{K}'\mathbf{R}$  is distributed as  $N(\mathbf{0}, \mathbf{I}_\gamma \otimes \boldsymbol{\Sigma})$  where  $\mathbf{I}_\gamma$  is the identity matrix of size  $\gamma \times \gamma$ . Let  $\boldsymbol{\Gamma}_i$  be a random orthogonal matrix of size  $\gamma \times \gamma$ . Because distributions of the type  $N(\mathbf{0}, \mathbf{I}_\gamma \otimes \boldsymbol{\Sigma})$  are invariant under left multiplication by orthogonal matrices, it follows that  $\boldsymbol{\Gamma}_i\mathbf{K}'\mathbf{R}$  is also distributed as  $N(\mathbf{0}, \mathbf{I}_\gamma \otimes \boldsymbol{\Sigma})$ . By dividing the mean of each column of  $\boldsymbol{\Gamma}_i\mathbf{K}'\mathbf{R}$  with the estimated standard deviation of each column and multiplying with the square root of the degrees of freedom, we obtain  $\mathbf{t}_{\text{simul}}$ . This image will have a distribution similar to what  $\mathbf{t}_{\text{original}}$  would have had under the null hypothesis, but the df of this image is  $\gamma - 1$ , whereas the df in  $\mathbf{t}_{\text{original}}$  is  $\gamma$ . However, by adjusting the threshold to yield the same  $\alpha$  for the two different df, this should not matter. The orthogonal matrices  $\boldsymbol{\Gamma}_i$  are generated by constructing an orthonormal basis for random normal matrices (distributed as  $N(\mathbf{0}, \mathbf{I}_\gamma \otimes \mathbf{I}_\gamma)$ ) using a singular value decomposition [Press et al., 1992].

<sup>2</sup>The derivation below was inspired by a suggestion from an anonymous reviewer. An earlier version of the SSI generation was based on a singular value decomposition of the residuals, an equivalent but less clear approach.