

The effect of filter size on VBM analyses of DT-MRI data

Derek K. Jones,^{a,b,*} Mark R. Symms,^c Mara Cercignani,^d and Robert J. Howard^e

^aNeuroimaging Research Group, Centre for Neuroimaging Sciences, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK

^bSection of Tissue Biophysics and Biomimetics, Laboratory of Integrative Medicine and Biophysics, National Institute of Child Health and Development, Bethesda, MD 20892-5772, USA

^cDepartment of Clinical and Experimental Epilepsy, Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK

^dDepartment of Neuroinflammation, Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK

^eSection of Old Age Psychiatry, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK

Received 5 October 2004; revised 17 December 2004; accepted 8 February 2005
Available online 9 April 2005

Voxel-based morphometry (VBM) has been used to analyze diffusion tensor MRI (DT-MRI) data in a number of studies. In VBM, following spatial normalization, data are smoothed to improve the validity of statistical inferences and to reduce inter-individual variation. However, the size of the smoothing filter used for VBM of DT-MRI data is highly variable across studies. For example, a literature review revealed that Gaussian smoothing kernels ranging in size (full width at half maximum) from zero to 16 mm have been used in DT-MRI VBM type studies. To investigate the effect of varying filter size in such analyses, whole brain DT-MRI data from 14 schizophrenic patients were compared with those of 14 matched control subjects using VBM, when the filter size was varied from zero to 16 mm. Within this range of smoothing, four different conclusions regarding apparent patient control differences could be made: (i) no significant patient-control differences; (ii) reduced FA in right superior temporal gyrus (STG) in patients; (iii) reduced FA in both right STG and left cerebellum in patients; and (iv) reduced FA only in left cerebellum in patients. These findings stress the importance of recognizing the effect of the matched filter theorem on VBM analyses of DT-MRI data. Finally, we investigated whether one of the underlying assumptions of parametric VBM, i.e., the normality of the residuals, is met. Our results suggest that, even with moderate smoothing, a large number of voxels within central white matter regions may have non-normally distributed residuals thus making valid statistical inferences with a parametric approach problematic in these areas.

Published by Elsevier Inc.

Keywords: Voxel-based morphometry; DT-MRI; Filter size; SPM; Fractional anisotropy

Introduction

Diffusion tensor magnetic resonance imaging (DT-MRI) (Basser et al., 1994) is unique in providing information about organization and structural integrity of tissue in vivo. It has been used increasingly to study a host of brain conditions including white matter diseases, neuropsychiatric conditions, ischemic syndromes, neurodevelopment, and aging (Horsfield and Jones, 2002; Lim and Helpert, 2002; Moseley, 2002; Neil et al., 2002; Sotak, 2002). An unresolved question is how best to compare DT-MRI data from two or more groups of subjects. One approach involves making a priori hypotheses about the specific anatomical location of patient-control differences and defining a region of interest (ROI) within which to make measurements. This approach is appropriate when the prior information is sufficiently robust to allow prediction of the location and extent of the expected differences. However, in many cases, and particularly in psychiatric disorders, the spatial location and extent of differences is not known a priori. Consequently, more global search strategies, which here we refer to collectively as whole brain voxel-based approaches, are often used. Typically, image data sets are spatially normalized to a common template under the assumption that a particular voxel address in each image corresponds to the same anatomical structure across subjects. Tests are performed on a voxel-by-voxel basis, so that essentially every location within the brain is checked for patient-control differences. This overcomes the issue of having to predict the spatial location of putative differences since one can explore the whole of the brain. There are, however, a few disadvantages compared with the ROI approach. First and foremost, since one is performing many more statistical comparisons by performing statistical tests in every voxel, one increases the chance of Type I error simply due to multiple comparisons and hence the statistical power of the analysis is reduced. Bonferroni-type corrections across all voxels analyzed would be far too conservative and so other approaches are used. By far the most commonly used

* Corresponding author. Neuroimaging Research Group, Centre for Neuroimaging Sciences, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK. Fax: +44 207 919 2116.

E-mail address: d.jones@iop.kcl.ac.uk (D.K. Jones).

Available online on ScienceDirect (www.sciencedirect.com).

approach is based on the theory of Gaussian random fields developed by Worsley et al. (1992) and implemented in SPM (Ashburner and Friston, 2000; Friston et al., 1995). In this approach, the image data are processed with a single low pass filter (Gaussian kernel). This step confers a few important benefits. First, and foremost, it improves the signal to noise ratio (SNR) which will enhance detectability of genuine patient-control differences. Second, it helps to ensure that the assumptions underlying the theory of Gaussian random fields (i.e., Gaussian distribution and homoscedasticity of residuals to the linear model) are met so that the correction for multiple comparisons can be correctly performed. Third, it ‘cushions’ against imperfections in the spatial normalization of the constituent images in the data set (see Ashburner and Friston, 2001; Bookstein, 2001).

While whole brain voxel-based approaches circumvent the need to make a priori hypotheses concerning the location of differences, they do not strictly overcome a need to make such hypotheses about the spatial extent of any expected differences. This is an important point since the matched filter theorem (Rosenfeld and Kak, 1982) states that the width of the filter used to process the data should be tailored to the size of the difference one expects to see. Thus, when using a single low-pass filter to pre-process DT-MRI data prior to performing a voxel-wise comparison, one should specify a priori the spatial extent of the effect one expects to see. However, diffusion tensor imaging is a relatively new technique (Basser et al., 1994), and so experience of expected effects is limited. Unfortunately, while substantial research into the underlying physics of diffusion anisotropy has been carried out (see Beaulieu, 2002), a thorough understanding of the mechanism underlying diffusion anisotropy in human brain parenchyma is incomplete. Furthermore, with one possible recent exception (Kumari et al., 2004), no theories explaining the observation of the uniformity of the trace of the diffusion tensor across brain parenchyma (Pierpaoli et al., 1996) currently exist. With this lack of understanding, it is clear that predicting a priori how a particular disease state will affect mean diffusivity and anisotropy is problematic.

This may explain the range of choice of smoothing filter sizes which have been used in the literature describing applications of voxel-based approaches to analyze data from DT-MRI. Table 1 shows details of the filter sizes that have been used in voxel-based analyses of DT-MRI data. If one assumes that the matched filter theorem has been considered in designing the various studies, and that the FWHM of the smoothing kernel has been chosen to match the size of expected patient-control difference, then it is clear that there is no common consensus on the size of patient control differences one might expect to see, even within the same disease state. An often quoted ‘rule of thumb’ governing the size of the smoothing kernel is that a kernel of at least 2–3 times the voxel dimension should be used for VBM. However, this rule of thumb was empirically determined for fMRI and PET data to ensure that the strict requirements of Gaussian random field theory (Worsley et al., 1992), namely that the residuals to the general linear model have a Gaussian distribution and have spatially invariant smoothness, are met. Since there is no reason to expect that a rule developed for fMRI and PET data should be universal, and in particular should apply to DT-MRI data, we wished to investigate the influence of varying the filter size on the outcome of a VBM-type comparison of DT-MRI data and on the number of voxels with non-normally distributed residuals from two groups of subjects. In order to do this, we acquired DT-MRI data from 14 patients with schizophrenia and 14 age-matched healthy controls and in our analysis varied the size of the smoothing filter to cover the range reported in the literature.

Methods

Subjects

Schizophrenia subjects

Fourteen right-handed males with DSM-IV criteria schizophrenia were recruited from wards and clinics at the South London and Maudsley NHS Trust, London. Their median age was 34 years (range 22–53 years) with median IQ 110 (range 98–124), measured

Table 1
Examples of smoothing kernel sizes and voxel dimensions used in voxel-based analyses of DT-MRI data

Filter FWHM (mm)	Voxel dimensions of acquired data (mm)	Voxel dimensions of analyzed data (mm)	FWHM/voxel dimension of acquired data (x y z)	FWHM/voxel dimension of analyzed data (x y z)	Reference
None	1.72 × 1.72 × (5.00 + 1.00)	0.86 × 0.64 × (5.00 + 1.00)	NA	NA	Hubl et al., 2004 ^a
None	1.88 × 1.88 × (5.00 + 0.00)	1.88 × 1.88 × (5.00 + 0.00)	NA	NA	Ardekani et al., 2003 ^a
3 × 3 × 3	1.29 × 1.72 × (4.00 + 1.00)	0.86 × 0.64 × (4.00 + 1.00)	(2.33 1.74 0.60)	(4.66 3.48 0.60)	Park et al., 2004 ^a
4 × 4 × 4	1.88 × 1.88 × (5.00 + 1.50)	1.88 × 1.88 × (5.00 + 1.50)	(2.12 2.12 0.62)	(2.12 2.12 0.62)	Barnea-Goraly et al., 2004 ^b
4 × 4 × 4	1.88 × 1.88 × (5.00 + 1.00)	0.94 × 0.94 × (5.00 + 1.00)	(2.12 2.12 0.67)	(4.26 4.26 0.72)	Barnea-Goraly et al., 2003 ^c
5 × 5 × 5	1.88 × 1.88 × (2.80 + 0.00)	1.88 × 1.88 × (2.80 + 0.00)	(2.66 2.66 1.79)	(2.66 2.66 1.79)	Molko et al., 2004 ^d
6 × 6 × 6	3.00 × 3.00 × (3.00 + 0.00)	3.00 × 3.00 × (3.00 + 0.00)	(2.00 2.00 2.00)	(2.00 2.00 2.00)	Sach et al., 2004 ^e
6 × 6 × 6	3.00 × 3.00 × (5.00 + 0.00)	3.00 × 3.00 × (3.00 + 0.00)	(2.00 2.00 1.71)	(2.00 2.00 2.00)	Sommer et al., 2002 ^f
8 × 8 × 8	2.50 × 2.50 × (5.00 + 0.00)	1.88 × 1.88 × (5.00 + 0.00)	(3.20 3.20 1.60)	(4.27 4.27 1.60)	Eriksson et al., 2001 ^g
9 × 9 × 9	1.29 × 1.72 × (4.00 + 1.00)	0.86 × 0.64 × (4.00 + 1.00)	(6.98 5.23 1.80)	(10.46 14.06 1.80)	Park et al., 2004 ^a
10 × 10 × 10	2.50 × 2.50 × (5.00 + 0.00)	1.88 × 1.88 × (5.00 + 0.00)	(4.00 4.00 2.00)	(5.32 5.32 2.00)	Eriksson et al., 2001 ^f
12 × 12 × 12	1.88 × 1.88 × (5.00 + 0.00)	0.94 × 0.94 × (5.00 + 0.00)	(6.38 6.38 2.40)	(12.8 12.8 2.40)	Burns et al., 2003 ^a
16 × 16 × 16	2.50 × 2.50 × (5.00 + 0.00)	1.88 × 1.88 × (5.00 + 0.00)	(6.40 6.40 3.20)	(8.53 8.53 3.20)	Foong et al., 2002 ^a

Note that in acquisitions with non-zero slice gaps, we have taken the voxel dimension to be equal to the slice thickness plus the slice gap (shown in the second column as (slice thickness + slice gap)—as this is the resolution of the data in the slice direction. The fourth and fifth columns show the ratio of the FWHM of the filter to the dimension of the voxel along each of the three axes.

The disease/condition studied are indicated by the following superscripts: a = schizophrenia; b = autism; c = fragile X syndrome; d = Turner syndrome; e = amyotrophic lateral sclerosis; f = stuttering; g = epilepsy.

using the National Adult Reading Test. The median illness duration was 8 years (range 1–25 years) and all were in remission and being treated with antipsychotic medication. Exclusion criteria included history of head injury, neurological symptoms, speech or hearing difficulties, fulfillment of DSM-IV criteria for abuse of illicit drugs or alcohol during their lifetime, and any contraindications to MRI scanning, including metal implants and claustrophobia.

Comparison subjects

Comparison subjects were matched for gender, handedness, age (median 34 years, range, 19–57 years), and IQ (median 109, range 99–123), had no medical/psychiatric disorders, no family history of psychiatric disorder, and were not receiving medication. Subjects from both groups gave written consent after the procedure had been fully explained. The study was approved by the local Research Ethics Committee.

Data acquisition

Data were acquired using a GE Signa 1.5 T LX MRI system (General Electric, Milwaukee, WI), and an acquisition optimized for diffusion tensor MRI of white matter, providing whole head coverage with isotropic image resolution (2.5 mm × 2.5 mm × 2.5 mm). The acquisition was peripherally gated to the cardiac cycle using a device placed on the subjects' forefinger. Full details are provided elsewhere (Jones et al., 2002a). Following correction for image distortions introduced by the diffusion-weighting gradients, the diffusion tensor was determined in each voxel (Basser et al., 1994) and images of (a) T₂-weighted intensity (i.e., the image intensity with no diffusion-gradients applied), (b) mean diffusivity, and (c) fractional anisotropy (Basser and Pierpaoli, 1996) were computed for each subject.

Analysis

The DT-MRI acquisition and analysis allows images of the T₂-weighted intensity, mean diffusivity, and fractional anisotropy to be created which are in perfect registration with each other. Therefore, the T₂-weighted image for each subject was used to co-register the data sets to the T₂-weighted EPI template that is supplied as part of the SPM99 package (Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK) (Friston et al., 1995).

First, the data were masked using an automated masking procedure based on a combination of the software package 'BET-Brain Extraction Tool', part of the Functional Software Library package (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Oxford University, Oxford, UK) and an intensity thresholding procedure. Full details can be found in Jones et al. (2002b).

Following masking, each subject's T₂-weighted image was initially co-registered with the T₂-weighted EPI template using an affine transformation with 12 degrees of freedom. The 28 co-registered T₂-weighted images were then averaged and smoothed with an 8-mm Gaussian kernel to create a 'group template'. The purpose of generating a template from the data sets themselves was to reduce any bias in template selection. Each individual subject's T₂-weighted image was then co-registered to the group template using an affine transformation with 12 degrees of freedom together with a series of non-linear warps characterized by a set of 7 × 8 × 7 basis functions (in the *x*, *y*, and *z* directions). The transformations

required to do this were subsequently applied to the mean diffusivity and fractional anisotropy images.

Two contrasts were used to detect whether each voxel had a higher or lower mean diffusivity or fractional anisotropy in the patient group compared with the group of control subjects. Increases or decreases were deemed to be significant at the individual voxel level when the *P* value assuming a Gaussian distribution of the data was less than 0.05 after correcting for multiple comparisons. Patient-control comparisons were performed after smoothing the data with Gaussian kernels with FWHMs ranging from 3 mm to 16 mm in steps of 1 mm, as well as a comparison of the data without any smoothing. Patient/control differences were considered significant only by voxel height (i.e., the *P* value in each individual voxel, after multiple comparison correction), as opposed to using cluster-level statistics.

To test for non-normality of the residuals at different levels of smoothing, we utilized the 'SPMd' toolbox developed by Luo and Nichols (2003a,b) to compute the Shapiro–Wilk statistic (Shapiro and Wilk, 1965) in each image voxel. With the Shapiro–Wilk (SW) statistic, the null hypothesis is that the residuals follow a normal distribution, i.e., for an alpha value of 0.05, if the *P* value is less than 0.05, then the null hypothesis that the residuals have a Gaussian distribution is rejected. Full details are provided in Luo and Nichols (2003a). Two approaches were used to characterize the non-Gaussian residual voxels. First, the number of voxels throughout the entire brain volume (expressed as a proportion of the total number of voxels within the brain) for which the *P* value for the SW test was less than 0.05 was computed for a range of smoothing kernels. Second, for the slices containing local maxima in any significant patient-control differences, we identified which particular voxels were deemed to have non-normally distributed residuals and plotted their spatial distribution together with the distribution of voxels with significant patient-control contrast. Plotting the data in this way allowed us to see whether there was any overlap in regions of significant patient-control differences and regions in which residuals were deemed to be non-normally distributed.

Results

Patient control differences

Mean diffusivity

No significant differences in mean diffusivity between the control and patient groups were observed with any of the smoothing filters used.

Fractional anisotropy

No significant differences in anisotropy with the 'anisotropy elevated in patients' contrast were revealed. However, with the 'anisotropy reduced in patients' contrast, the results were dependent on variations in the width of the smoothing kernel used. With no smoothing, and filter sizes of 3, 4, 5, and 6 mm, no significant patient-control differences were observed. With smoothing kernels having an FWHM of 7 and 8 mm, a single region centered at MNI coordinate [54, −50, 22] (right superior temporal gyrus) was present. The corrected *P* values for the most significant voxel were *P* = 0.035 and *P* = 0.007 with the 7 and 8 mm filter sizes, respectively. With a 9-mm kernel, a second focus appeared at [−28, −44, −18] (left cerebellum). As the size of the smoothing

kernel was increased further, this latter region increased in size, while the former diminished in size to the point where, with a Gaussian kernel of FWHM of 15 mm, it disappeared. This is shown in Figs. 1 and 2. Note that the largest cluster extent at MNI coordinate [54, -50, 22] occurred with a smoothing kernel of around 11 mm.

Normality of residuals

Fig. 3 shows how the number of voxels with non-Gaussian distributed residuals in the fractional anisotropy data (expressed as a proportion of the total number of voxels analyzed) depends on the kernel size for 4 mm through 16 mm. Unsurprisingly, increasing the level of smoothing with a Gaussian kernel reduces the proportion of voxels in the brain that the SW test deems to have non-normally distributed residuals. Figs. 4 and 5 show the spatial distribution of voxels with non-Gaussian residuals in the slices containing the local maxima of patient-control differences in superior temporal gyrus and cerebellar region, respectively, at the different levels of smoothing.

Discussion

This study shows that, using a single low-pass filtering approach (e.g., Barnea-Goraly et al., 2003; Burns et al., 2003; Eriksson et al., 2001; Leung et al., 2004; Sommer et al., 2002), it

would be possible to draw four different conclusions from the DT-MRI data we collected from our patients and controls.

- (i) With filter sizes less than or equal to 6 mm, we would conclude that there were no patient-control differences in FA.
- (ii) With filter sizes having FWHM of 7 and 8 mm, we would conclude that there is a single focus of reduced FA in patients compared with controls in the vicinity of the right superior temporal gyrus (STG) in patients.
- (iii) With filter sizes of 9–14 mm, we would conclude that there are two foci of reduced FA in patients compared with controls, i.e., in the vicinity of the right STG and in the left cerebellum.
- (iv) With filter sizes greater than 14 mm, we would conclude that patient control differences are confined to the left cerebellum.

Interestingly, the first domain (filter sizes up to 6 mm in width) encompasses the range in which the FWHM is set to ‘at least twice the voxel dimension’. As stated in the Introduction, there has long been a rule of thumb in the SPM community that the FWHM should be at least twice the voxel dimension in order to get statistically robust results, although this result appears to be empirical rather than theoretical (Ashburner, personal communication via SPM mailing list, 2004). If we stick with this rule of thumb, however, we fail to see any patient control differences.

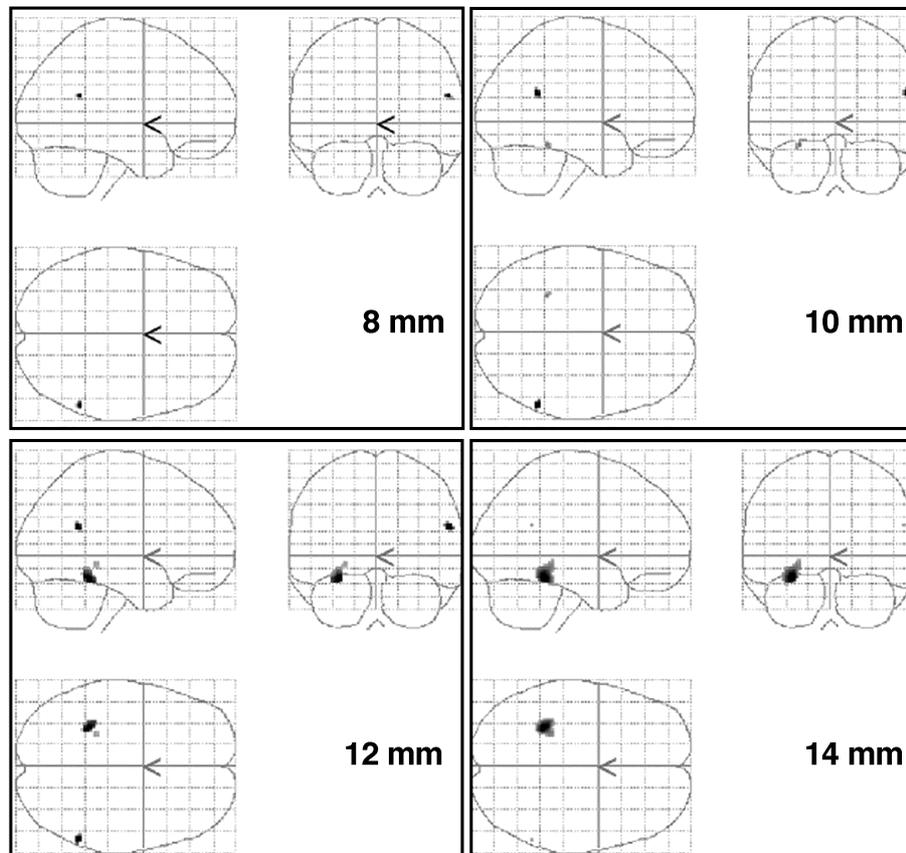


Fig. 1. Significant patient < control differences obtained with smoothing kernels with FWHM = 8, 10, 12 and 14 mm. Note the foci are significant at the voxel level at $P < 0.05$ after correction for multiple comparisons.

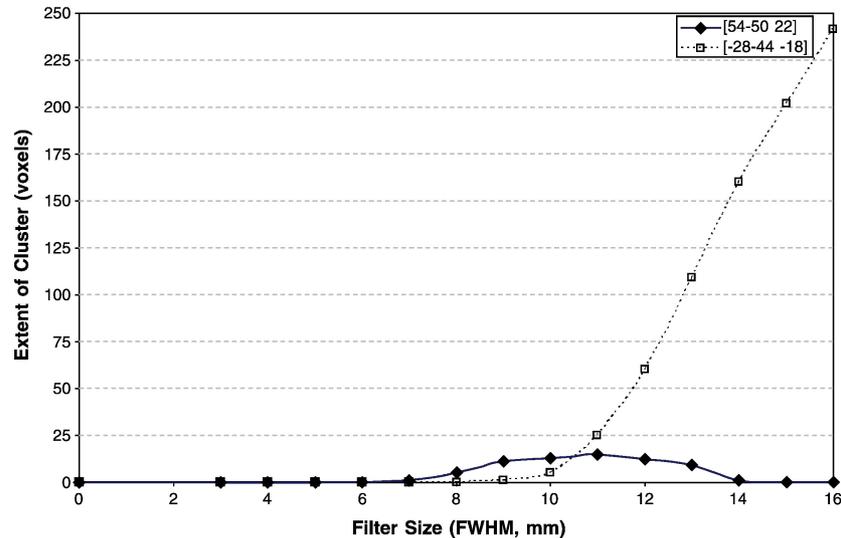


Fig. 2. Plot of size of clusters at the locations shown in Fig. 1, as a function of size of the Gaussian smoothing kernel. Note the appearance and subsequent disappearance of the significant reduction in FA centered at MNI coordinate [54, -50, 22].

If we apply a rule that the FWHM should be 3 times the voxel dimension (i.e., 7.5 mm), then we only observe the STG focus. It appears then, that unless we have a particular hypothesis about the spatial extent of a patient-control difference, a single low-pass filtering approach where the width of the filter has been chosen according to an ‘FWHM must be x times the voxel dimension’-type rule may be insufficient to identify patient control differences, at least in our data set. Indeed Fig. 2 suggests that there are real patient-control differences in FA, and that we are most ideally sensitized to detect them when using a filter size of around 11 mm (which is just over 4 times our voxel dimensions).

A review of Table 1 reveals that the rule of filter size being at least twice the voxel dimension is followed in less than 50% of the studies, if one considers all three dimensions. It is also clear that there is no generally agreed standard concerning the width of the Gaussian filter that one should use when analyzing DT-MRI data using VBM. We note that in none of the cited

studies was a clear hypothesis made about the expected extent of any differences, nor a justification for the choice of smoothing kernel used.

An interesting question is whether the effect of smoothing is in the reduction of noise or in the enhancement of signal. We believe that both mechanisms are at work here. There is no doubt that increasing the size of the smoothing kernel will reduce the noise level. However, if that were the sole mechanism at work, then one would expect the size of the superior-temporal gyrus patient-control difference to continually increase in size as the kernel size increase. The ‘peaking’ of the cluster region, however, at a smoothing of approximately 11 mm, and subsequent decline of the volume with further smoothing, reinforces the notion that the matched filter theorem is also a key factor in signal detectability.

Again, we wish to re-iterate that the ‘rule of thumb’ relating smoothing kernel sizes to voxel dimensions was originally

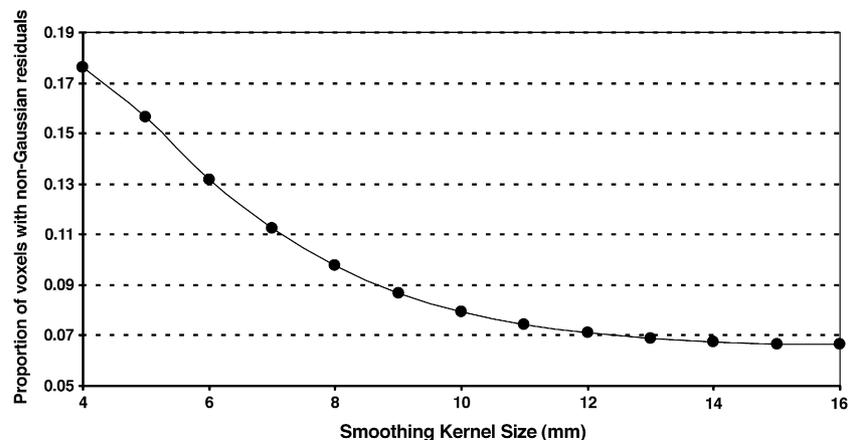


Fig. 3. Plot of number of voxels with non-Gaussian residuals (expressed as proportion of number of voxels analyzed) for smoothing kernels in the range 4–16 mm. Voxels were deemed to have non-Gaussian residuals if the P value from the Shapiro–Wilk test was less than 0.05.

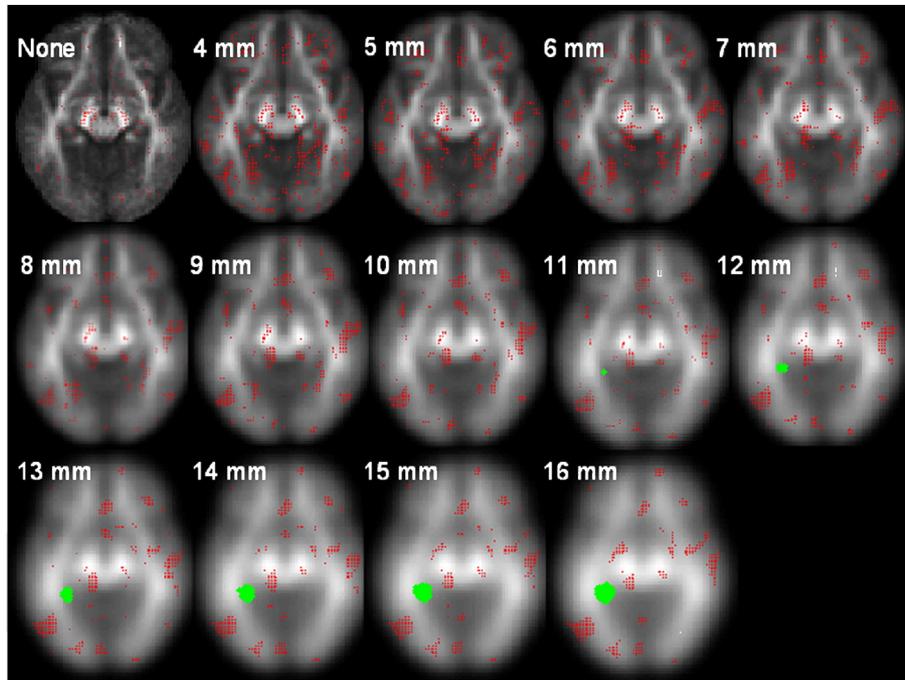


Fig. 4. Spatial distribution of voxels in which residuals are deemed to be non-normally distributed ($P = 0.05$, Shapiro–Wilk) in the axial slice containing the significant superior temporal gyrus patient-control difference in FA. Non-normal residual voxels are highlighted with a red dot. The green dots show voxels in which FA was deemed to be significantly lower in patients compared with controls. The gray-scale image shows the mean of the FA images from the 14 controls obtained at the different levels of smoothing.

established for fMRI and PET data to ensure that the statistical requirements of Gaussian random field theory are met, including the requirement for the residuals to the fitted model to be normally distributed. Figs. 3–5 show just how well this assumption is met at

different levels of smoothing. The Shapiro–Wilk (SW) test performed on our data, with an alpha value of 0.05, suggests that there are many voxels within central white matter structures in which the residuals do not follow a Gaussian distribution. Most

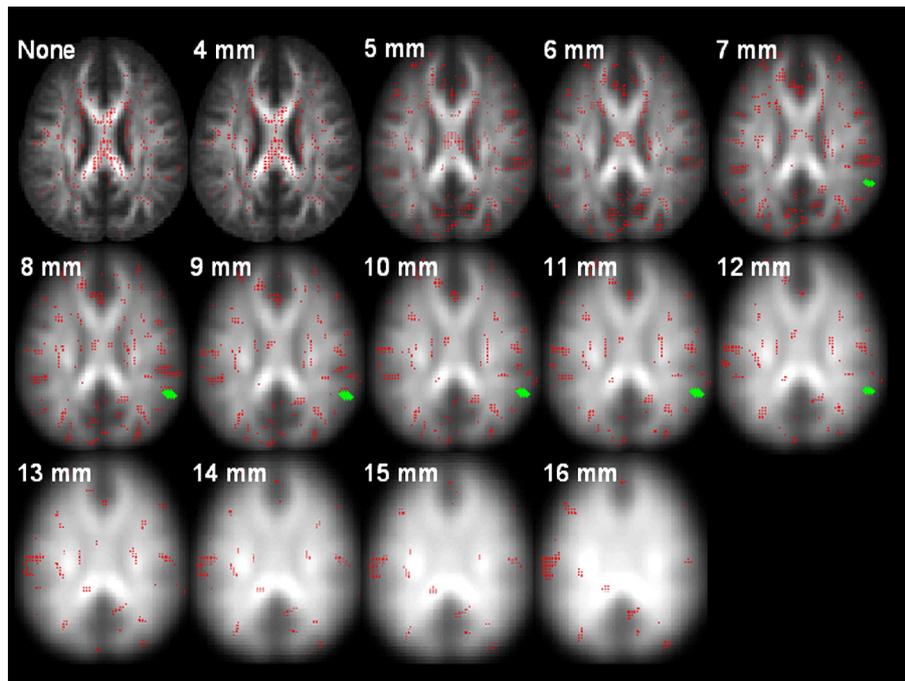


Fig. 5. Spatial distribution of voxels in which residuals are deemed to be non-normally distributed ($P = 0.05$, Shapiro–Wilk) in the axial slice containing the significant cerebellar patient-control difference in FA. Non-normal residual voxels are highlighted with a red dot. The green dots show voxels in which FA was deemed to be significantly lower in patients compared with controls. The gray-scale image shows the mean of the FA images from the 14 controls obtained at the different levels of smoothing.

worrisome is the large number of voxels in the corpus callosum in Fig. 4, and in the cerebral peduncles in Fig. 5, that the SW test deems to have non-Gaussian-distributed residuals, even for moderate levels of smoothing (typical of that employed in the literature). Similar patches of non-Gaussianity of residuals were found in most of the major white matter tracts including the genu and splenium of corpus callosum, cingulum, and internal and external capsules (not shown). To the best of our knowledge, this is the first time that the distribution of the residuals in VBM analyses of FA data has been analyzed. Although these results are to be regarded as preliminary given the relatively small sample size, they do suggest that in future, VBM-type studies of this type, in which significant findings are located, or a priori hypotheses predict changes in, the corpus callosum and corticospinal tracts should verify that one of the main assumptions underlying the approach (i.e., the Gaussian distribution of the residuals) is met. Interestingly, the two regions of significant patient/controls differences found in the present study do not superimpose suggesting that, in these areas at least, the assumption regarding Gaussianity of residuals is valid.

Based on the findings of this study, there appear to be two sensible approaches to analyzing data sets such as the one considered here. The first is to generate a well-formulated hypothesis upfront about the spatial extent of any expected patient-control differences with a clear explanation of the reasoning for such a hypothesis. This should be accompanied, especially in light of a negative finding, by a statement that the findings only pertain to the length scale that was examined. However, this can prove extremely difficult for certain disease states, such as schizophrenia, since very little is known about the white matter abnormalities, despite several DT-MRI studies published to date.

The second approach is to search the data with a number of length scales for smoothing. In this work, we adopted something of a ‘brute force’ approach and compared results using the range of filter sizes reported in the literature. One cannot simply do this type of analysis, however, until the maximum sensitivity is found and then report the P value obtained with this filter size. This is because we have performed multiple comparisons of the data that have not been corrected, thus the chance of Type I error increases.

The concept of searching scale space in the context of voxel-based analyses of image data, while not considered in DT-MRI to date, has been around for over a decade. Poline and Mazoyer (1994) described a 4D approach to searching PET data for activations. This approach was later refined in order to obtain a unified P value for the range of local foci identified by Worsley et al. (1996). In this work, we have not implemented this latter approach since our aim was simply to demonstrate that the matched filter theorem is just as relevant to studies of DT-MRI data as it is to PET and fMRI studies. We do note, however, that in our ‘brute force’ approach, we considered 12 different smoothing kernels in the interval 4–16 mm. Siegmund and Worsley (1995) have shown that the scale space is stationary in log space, hence a search of scale space in the range used in this study does not need to be at regular intervals, but as the sizes of the smoothing kernel increases, the interval between samples of scale space can increase, without loss of information. In other words, fewer samples of scale space are needed to recover the same information obtained in this study.

Alternative multi-scale approaches to VBM analyses that have not yet been applied to DT-MRI data include adaptive filtering approaches (see, for example, Davatzikos et al., 2001) and

analyzing the data in wavelet space (e.g., Brammer, 1998; Ruttimann et al., 1998). These are topics suitable for future research. We also note that the voxel-based morphometry approach is most often applied to data sets based on T_1 -weighted ‘structural’ scans, in which the image intensity is reasonably homogenous within white matter. As discussed earlier, the image intensity in mean diffusivity maps is fairly uniform within brain parenchyma. In contrast, different architectural paradigms give rise to a wide range of fractional anisotropy values (Pierpaoli et al., 1996) and so fractional anisotropy is very heterogeneously distributed throughout white matter. Thus, not only the effect of smoothing, but also the effects of registration on the data sets must therefore be considered when interpreting the results. Future investigations should focus on the effects of the co-registration techniques employed with such heterogeneous data to see how dependent the results are on the method used.

An underlying assumption within SPM is that the data are normally distributed. While this assumption may be valid for mean diffusivity data (e.g., Pajevic and Basser, 2003), it cannot be assumed to be correct for FA data. Indeed, our preliminary findings indicate that there are many places where the residuals are not normally distributed. The extent to which this affects the outcome of analyses of anisotropy data based on the approach discussed here, and hence the statistical validity of any inferences made, is open to debate and future research. It is not our intention here to comment on whether or how SPM should be used to compare data sets from different groups. The interested reader is referred elsewhere for such discussions Ashburner and Friston, 2001; Bookstein, 2001; Crum et al., 2003; Davatzikos, 2004; Salmond et al., 2002). Rather, it was our intention to survey the literature for the range of smoothing sizes used, to discover the effect of utilizing different filtering sizes, and to highlight the importance of considering the scale of the examination. Certainly, we encourage reporting whether or not smoothing was performed prior to analysis and, if appropriate, the size of the smoothing kernel used, which some groups have failed to do (e.g., Agartz et al., 2001; Buchsbaum et al., 1998) so that the results may be compared with other studies more meaningfully. We also encourage justification for the choice of filter size, if a single smoothing kernel is used.

We wish to comment on the interpretation of an apparently significant change in fractional anisotropy as obtained by the method reported here. Regardless of whether the results obtained here had revealed a significant difference in anisotropy between patients and controls, there would be no evidence to suggest that the ‘information pathway’ would be affected to any extent in the region of reduced anisotropy. In other words, DT-MRI cannot provide any information about functional connectivity. Furthermore, it is by no means clear that a reduction in anisotropy would indicate that two regions are more poorly structurally connected. For example, an increase in the connectivity through a region may lead to more intra-voxel averaging of fiber orientation in the voxel-averaged estimation of anisotropy, thereby leading to a reduction in FA (Pierpaoli et al., 1996). Indeed, to determine whether any finding derived from DT-MRI is a correlate of increased or decreased anatomical connectivity, a postmortem examination is needed. Therefore, in interpreting the results of VBM analysis of DT-MRI data we prefer to limit ourselves to saying that there is a significant difference in fractional anisotropy as determined by VBM of DT-MRI.

Finally, we note that alternative strategies to performing voxel-based analyses of neuroimaging data based on randomization/

permutation testing have been proposed (e.g., Bullmore et al., 1999; Holmes et al., 1996; Nichols and Holmes, 2001). While these approaches do not require smoothing to ensure that the data fulfil the assumptions underpinning Gaussian random field theory, the issue of optimally sensitizing the detection through the matched filter theorem may be equally as applicable as it is to the parametric methods and is worthy of further investigation.

Conclusion

Whole brain voxel-based approaches have been used in a number of studies to compare DT-MRI data obtained from patients with those obtained from controls. A wide range of smoothing kernels has been used in these studies. We have demonstrated the importance of considering, justifying, and reporting the length-scale of the statistical testing that is being performed. These issues are not new to neuroimaging but have yet to be considered and addressed fully in the DT-MRI arena. The different results we obtained here from the same data set suggest that three research groups analyzing the same set of schizophrenic patients, each employing a reasonable ‘rule of thumb’ for selecting the size of the smoothing kernel, would draw different conclusions about the presence and location of the white matter “abnormality” seen in schizophrenia.

We have also shown that, with the current data set at least, even with moderate levels of smoothing, many voxels within major white matter fasciculi may have non-normally distributed residuals—thereby breaking one of the central assumptions underlying statistical parametric methods. This initial investigation into the quantity and spatial distribution of voxels with non-normally distributed residuals will hopefully prompt future investigators to examine this issue in their own data and to select the level of smoothing appropriately.

Acknowledgments

This work was supported in part by an Advanced Training Fellowship grant from The Wellcome Trust, UK (Grant 067437/Z/02.A) and also grant (Grant number 054030/2/98). We are grateful to Dr. Sukhwinder S. Shergill, Prof. SCR Williams, Dr. A Simmons (Institute of Psychiatry), and Dr. MA Horsfield (University of Leicester). We would like to thank the anonymous referees for their careful review of this work.

References

- Agartz, I., Andersson, J.L.R., Skare, S., 2001. Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. *NeuroReport* 12, 2251–2254.
- Ardekani, B.A., Nierenberg, J., Hoptman, M.J., Javitt, D.C., Lim, K.O., 2003. MRI study of white matter diffusion anisotropy in schizophrenia. *NeuroReport* 14, 2025–2029.
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry: the methods. *NeuroImage* 11, 805–821.
- Ashburner, J., Friston, K.J., 2001. Why voxel-based morphometry should be used. *NeuroImage* 14, 1238–1243.
- Barnea-Goraly, N., Eliez, S., Hedeus, M., Menon, V., White, C.D., Moseley, M., Reiss, A.L., 2003. White matter tract alterations in fragile X syndrome: preliminary evidence from diffusion tensor imaging. *Am. J. Med. Genet., Part. B* 118B, 81–88.
- Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., Reiss, A.L., 2004. White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biol. Psychol.* 55, 323–326.
- Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissue elucidated by quantitative-diffusion-tensor MRI. *J. Mag. Reson., B* 111, 209–219.
- Basser, P.J., Mattiello, J., Le Bihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267.
- Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system—A technical review. *NMR Biomed.* 15, 435–455.
- Bookstein, F.L., 2001. “Voxel-based morphometry” should not be used with imperfectly registered images. *NeuroImage* 14, 1454–1462.
- Brammer, M.J., 1998. Multidimensional wavelet analysis of functional magnetic resonance images. *Hum. Brain Mapp.* 6, 378–382.
- Buchsbaum, M.S., Tang, C.Y., Peled, S., Gudbjartsson, H., Lu, D., Hazlett, E.A., Downhill, J., Haznedar, M., Fallon, J.H., Atlas, S.W., 1998. MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. *NeuroReport* 16, 425–430.
- Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imag.* 18, 32–42.
- Burns, J., Job, D., Bastin, M.E., Whalley, H., MacGillivray, T., Johnstone, E.C., Lawrie, S.M., 2003. Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *Br. J. Psychiatry* 182, 439–443.
- Crum, W.R., Griffin, L.D., Hill, D.L.G., Hawkes, D.J., 2003. Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage* 20, 1425–1437.
- Davatzikos, C., 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 23, 17–20.
- Davatzikos, C., Li, H.H., Herskovits, E., Resnick, S.M., 2001. Accuracy and sensitivity of detection of activation foci in the brain via statistical parametric mapping: a study using a PET simulator. *NeuroImage* 13, 176–184.
- Eriksson, S.H., Rugg-Gunn, F.J., Symms, M.R., Barker, G.J., Duncan, J.S., 2001. Diffusion tensor imaging in patients with epilepsy and malformations of cortical development. *Brain* 124, 617–626.
- Foong, J., Symms, M.R., Barker, G.J., Maier, M., Miller, D.H., Ron, M.A., 2002. Investigating regional white matter in schizophrenia using diffusion tensor imaging. *NeuroReport* 13, 333–336.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Holmes, A.P., Blair, R.C., Watson, J.D.G., Ford, I., 1996. Non-parametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* 16, 7–22.
- Horsfield, M.A., Jones, D.K., 2002. Application of diffusion weighted and diffusion tensor MRI to white matter diseases. *NMR Biomed.* 15, 570–577.
- Hubl, D., Koehnig, T., Strik, W., Federspiel, A., Kreis, R., Boesch, C., Maier, S.E., Scroth, G., Lovblad, K., Dierks, T., 2004. Pathways that make voices—White matter changes in auditory hallucinations. *Arch. Gen. Psychiatry* 61, 658–668.
- Jones, D.K., Williams, S.C.R., Gasston, D., Horsfield, M.A., Simmons, A., Howard, R., 2002a. Isotropic resolution diffusion tensor imaging with whole brain acquisition in a clinically acceptable time. *Hum. Brain Mapp.* 15, 216–230.
- Jones, D.K., Alexander, D.C., Griffin, L.D., Catani, M., Horsfield, M.A., Howard, R.J., Williams, S.C.R., 2002b. Spatial normalisation and averaging of diffusion tensor MRI data sets. *NeuroImage* 17, 592–617.
- Kumari, A., Tuch, D.S., Sorenson, A.G., 2004. Why is the trace of the diffusion tensor constant across brain? Book of Abstracts: 11th Annual Meeting of the International Society for Magnetic Resonance in Medicine. ISMRM, Berkeley, CA, p. 1237.
- Leung, L.H., Ooi, G.C., Kwong, D.L., Chan, G.C., Cao, G., Khong, P.L.,

2004. White matter diffusion anisotropy after chemo-irradiation: a statistical parametric mapping study and histogram analysis. *NeuroImage* 21, 261–268.
- Lim, K.O., Helpert, J.A., 2002. Neuropsychiatric applications of DTI—A review. *NMR Biomed.* 15, 587–593.
- Luo, W.-L., Nichols, T.E., 2003a. Diagnosis and exploration of massively univariate neuroimaging models. Technical Report.
- Luo, W.-L., Nichols, T.E., 2003b. Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage* 19, 1014–1032.
- Molko, N., Cachia, A., Riviere, D., Mangin, J.-F., Bruandet, M., Le Bihan, D., Cohen, L., Dehaene, S., 2004. Brain anatomy in Turner syndrome: evidence for impaired social and spatial-numerical networks. *Cereb. Cortex* 14, 840–850.
- Moseley, M.E., 2002. Diffusion tensor imaging and aging—A review. *NMR Biomed.* 15, 553–560.
- Neil, J., Miller, P., Mukherjee, P.S., Hüppi, P.S., 2002. Diffusion tensor imaging of normal and injured developing human brain—A technical review. *NMR Biomed.* 15, 543–552.
- Nichols, T.E., Holmes, A.P., 2001. Nonparametric analysis of PET functional neuroimaging experiments: a primer. *Hum. Brain Mapp.* 15, 1–25.
- Pajevic, S., Basser, P.J., 2003. Parametric and non-parametric statistical analysis of DT-MRI data. *J. Magn. Res.* 161, 1–14.
- Park, H.J., Westin, C.F., Kubicki, M., Maier, S.E., Niznikiewicz, M., Baer, A., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., Shenton, M.E., 2004. White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. *NeuroImage* 23, 213–223.
- Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A., Di Chiro, G., 1996. Diffusion tensor MR imaging of the human brain. *Radiology* 201, 637–648.
- Poline, J.-B., Mazoyer, B.M., 1994. Enhanced detection in brain activation maps using a multifiltering approach. *J. Cereb. Blood Flow Metab.* 14, 639–642.
- Rosenfeld, A., Kak, A.C., 1982. *Digital Picture Processing 2*. Academic Press, Orlando, FL, p. 42.
- Ruttimann, U.E., Unser, M., Rawlings, R.R., Rio, D., Ramsey, N.F., Mattay, V.S., Hommer, D.W., Frank, J.A., Weinberger, D.R., 1998. Statistical analysis of functional MRI data in the wavelet domain. *IEEE Trans. Med. Imag.* 17, 142–154.
- Sach, M., Winkler, G., Glauche, V., Liepert, J., Heimbach, B., Koch, M.A., Büchel, C., Weiller, C., 2004. Diffusion tensor MRI of early upper motor neuron involvement in amyotrophic lateral sclerosis. *Brain* 127, 340–350.
- Salmund, C.H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2002. Distributional assumptions in voxel-based morphometry. *NeuroImage* 17, 1027–1030.
- Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52 (3–4), 591–611.
- Siegmund, D.O., Worsley, K.J., 1995. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Stat.* 23, 608–639.
- Sommer, M., Koch, M.A., Paulus, W., Weiller, C., Buchel, C., 2002. Disconnection of speech-relevant areas in persistent developmental stuttering. *Lancet* 360, 380–383.
- Sotak, C.H., 2002. The role of diffusion tensor imaging in the evaluation of ischemic brain injury—A review. *NMR Biomed.* 15, 561–569.
- Worsley, K.J., Evans, A.C., 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, K.J., Marrett, S., Neelin, P., Evans, A.C., 1996. Searching scale space for activation in PET images. *Hum. Brain Mapp.* 4, 74–90.