ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage xx (2006) xxx – xxx

Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data

Donald J. Hagler Jr.*, Ayse Pinar Saygin, and Martin I. Sereno

University of California, San Diego, Department of Cognitive Science, 9500 Gilman Drive #0515, La Jolla, CA 92093-0515, USA

Received 23 September 2005; revised 8 July 2006; accepted 20 July 2006

Cortical surface-based analysis of fMRI data has proven to be a useful method with several advantages over 3-dimensional volumetric analyses. Many of the statistical methods used in 3D analyses can be adapted for use with surface-based analyses. Operating within the framework of the FreeSurfer software package, we have implemented a surface-based version of the cluster size exclusion method used for multiple comparisons correction. Furthermore, we have a developed a new method for generating regions of interest on the cortical surface using a sliding threshold of cluster exclusion followed by cluster growth. Cluster size limits for multiple probability thresholds were estimated using random field theory and validated with Monte Carlo simulation. A prerequisite of RFT or cluster size simulation is an estimate of the smoothness of the data. In order to estimate the intrinsic smoothness of group analysis statistics, independent of true activations, we conducted a group analysis of simulated noise data sets. Because smoothing on a cortical surface mesh is typically implemented using an iterative method, rather than directly applying a Gaussian blurring kernel, it is also necessary to determine the width of the equivalent Gaussian blurring kernel as a function of smoothing steps. Iterative smoothing has previously been modeled as continuous heat diffusion, providing a theoretical basis for predicting the equivalent kernel width, but the predictions of the model were not empirically tested. We generated an empirical heat diffusion kernel width function by performing surface-based smoothing simulations and found a large disparity between the expected and actual kernel widths.

Introduction

Analysis of fMRI data using cortical surface models offers at least three advantages over more conventional 3-dimensional analysis methods. First, cortical surface models allow for better visualization of activations, providing a more global view than single slices and a better view of the spatial extent of activation foci and their locations relative to each other and to sulcal/gyral landmarks (Dale and Sereno, 1993). Second, statistical methods for the analysis of single subject data can benefit from the exclusion of non-gray matter signals, and smoothing signals along the cortical surface, rather than in 3D results in superior resolution and sensitivity (Kiebel et al., 2000; Andrade et al., 2001; Formisano et al., 2004). Finally, group analysis with cortical surface models employs inter-subject alignment based on the patterns of sulci and gyri, as opposed to Talairach registration, which often ignores sulcal/gyral landmarks and tends to blur activity across neighboring banks of a sulcus (Fischl et al., 1999a,b).

In this paper, we describe methods that we have used and developed to facilitate cortical surface-based inter-subject analyses. A routine aspect of inter-subject fMRI analyses – for both 2D cortical surface and 3D volumetric methods – is spatial smoothing. Smoothing acts as a low-pass spatial frequency filter and thus improves the signal-to-noise ratio (SNR) by filtering out high spatial frequency noise (Petersson et al., 1999a). Smoothing increases the likelihood that inter-subject analyses will detect signals from foci that display variability in their precise cortical location across subjects. Spatial smoothing also ensures that the data more closely approximate a continuous field of random values, a necessary assumption of the random field theory used for multiple comparisons correction (Worsley et al., 1992, 1996, 1999; Petersson et al., 1999b; Andrade et al., 2001).

The vertices of a cortical surface mesh are not, however, arranged on a grid with regular spacing; instead distances and angles between neighboring vertices are slightly variable (Fischl et al., 1999a; Chung et al., 2003). Because of this, direct application of a Gaussian blurring kernel – as is done with 3D data sets – is computationally intensive for surface-based data. A computationally efficient method is to iteratively perform nearest-neighbor averaging; where each vertex's value is averaged with those of its neighbors. Slightly more complicated iterative smoothing algorithms have also been developed, which rely on a heat diffusion model and involve unequally weighting the contribution from each neighbor in the post-iteration value of a given vertex (Andrade et al., 2001; Chung et al., 2003, 2005). For example, in the latest and simplest version of this method, called heat kernel smoothing, the weights are calculated based on the distances between a vertex and each of its neighbors (Chung et al., 2005). The

* Corresponding author. Fax: +1 858 534 1128.
E-mail address: dhagler@cogsci.ucsd.edu (D.J. Hagler).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.07.036

effect of such iterative smoothing is quite similar to what would be expected if a Gaussian blurring kernel were used.

For heat kernel smoothing, the full-width-half-max (FWHM) size of the Gaussian blurring kernel equivalent to a particular number of iterations can be predicted from heat diffusion equations (Chung et al., 2003, 2005). The accuracy of those predictions, however, is limited by the extent to which the assumptions of the heat diffusion model are met. Specifically, the model assumes continuous diffusion, whereas iterative smoothing is inherently discrete and the vertices are spaced at discrete intervals. The bandwidth of the heat kernel determines the amount of smoothing performed at each iteration, and larger bandwidths make the continuous diffusion model less accurate (Chung et al., 2005). Similarly, the heat diffusion model is less accurate with increasing inter-neighbor distances. Thus, even though the heat diffusion model provides predictions of FWHM smoothness as a function of number of iterations and bandwidth, it is necessary to determine the accuracy of those predictions. To address this concern, we have made empirical measures of FWHM smoothness as a function of smoothing steps for both heat kernel smoothing and nearest-neighbor average smoothing.

Another important aspect of inter-subject analyses is the statistical correction for multiple comparisons. Multiple comparisons correction is desirable because of the large number of voxels, and thus independent statistical tests, involved in voxel-by-voxel analysis of fMRI images. An alternative to the overly conservative Bonferroni correction is cluster size exclusion (Worsley et al., 1992, 1999; Poline and Mazoyer, 1993; Friston et al., 1994; Forman et al., 1995; Andrade et al., 2001). Modeling the value (e.g., *t*-statistic) of each voxel or vertex as a normally distributed random variable, it is assumed that neighbors display some degree of covariance, or spatial smoothness. That is, the value at a voxel is likely to be somewhat similar to its neighbors, either because of spatial smoothing, spatially correlated noise – produced by the scanner or physiologically – or because of the spatial extent of actual brain activations and the resulting hemodynamic response (Kiebel et al., 1999). With smoother data, it is more likely that a suprathreshold voxel will have neighbors that are also suprathreshold, forming a cluster.

Cluster size limits can be generated, with random field theory (Worsley et al., 1996; Andrade et al., 2001), permutation tests (Hayasaka and Nichols, 2003, 2004; Hayasaka et al., 2004), or Monte Carlo simulations (Poline and Mazoyer, 1993; Forman et al., 1995), for any number of uncorrected *p*-values. A liberal *p*-value coupled with a very large cluster size limit is theoretically equivalent to a conservative *p*-value coupled with a small cluster size limit. In practice, however, the choice of the uncorrected *p*-value can have a strong influence on the number of clusters satisfying the cluster size limits. Furthermore, for the purpose of defining regions of interest (ROIs), the use of a conservative threshold will tend to underestimate the true size of a focus of activation, identifying only the very peak of activations; lower thresholds will tend to result in the joining of distinct clusters. This situation has motivated our development of a method for identifying ROIs using a sliding threshold followed by growth of clusters.

Because random cluster sizes depend on the smoothness of the images, an estimate of this is necessary. Actual data, however, contain regions of activation with some spatial extent, increasing measures of overall smoothness which would bias the cluster size thresholds to be overly conservative (Kiebel et al., 1999). For this reason, it would be more appropriate to measure the smoothness of random noise that shares the spatial correlations related to hemodynamic and scanner properties. As a proxy for actually measuring noise, smoothness can be measured from the normalized residual error of single subject GLM deconvolution (Kiebel et al., 1999). In order to determine a suitable method for estimating the intrinsic smoothness of a group analysis data set, we compared the smoothness of actual group averages, normalized residual error of group analysis, and the group average of simulated noise data sets.

Methods

Estimation of smoothness of fMRI statistics in 3D and on cortical surface meshes

Smoothness of fMRI statistics was estimated by comparing the local variance between neighboring vertices with the overall variance between all vertices (Forman et al., 1995; Worsley et al., 1999). AFNI's 3dFWHM (Ward, 2000b) was used to estimate the smoothness (i.e., full width half max (FWHM) Gaussian filter width) of 3-dimensional (3D) data sets. This program estimates smoothness separately for each of the *x*, *y*, and *z* dimensions with the following equation:

$$\mathrm{FWHM}_x = \mathrm{d}x \cdot \sqrt{\frac{-2\ln 2}{\ln\left(1 - \frac{\mathrm{var}(\mathrm{d}s)}{2\ \mathrm{var}(s)}\right)}} \quad (1)$$

where dx is the voxel size in the x dimension, var(ds) is the variance of the difference in signal between neighbors in the x dimension, and var(s) is the overall variance of the values at each voxel.

In a custom program, we adapted this method for use with statistics associated with cortical surface meshes generated by the FreeSurfer software package (Dale et al., 1999; Fischl et al., 1999a). Even though the cortical surface is a 2-dimensional sheet, vertices in a cortical surface mesh are not arranged in an orderly 2-dimensional grid. Instead, there are on average 6 neighbors connected to every vertex, arrayed like a pinwheel. For this reason, separate FWHM measures for the x and y dimensions are not easily calculated for surface-based statistics. Instead, a single FWHM measure was calculated using the following modification of Eq. (1):

$$\mathrm{FWHM}_{\mathrm{surf}} = \mathrm{d}v \sqrt{\frac{-2\ln 2}{\ln\left(1 - \frac{\mathrm{var}(\mathrm{d}s)}{2\ \mathrm{var}(s)}\right)}} \quad (2)$$

where dv is the average inter-neighbor distance, var(ds) is the variance inter-neighbor differences, and var(s) is the overall variance of the values at each vertex.

Stationary smoothness assumption

We chose to estimate smoothness of statistics globally; i.e., calculating the variance of inter-vertex or inter-voxel differences across all vertices or voxels. This effectively assumes that smoothness is uniform across the image. An alternative is to represent cluster sizes in terms of "resolution elements" or "resels"

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

(Worsley et al., 1992, 1999; Hayasaka et al., 2004). In this framework, a metric related to smoothness is calculated locally for each voxel or surface triangular face, and the resulting resel estimates are used to normalize the extent of clusters. It has been previously noted that such resel estimates are quite noisy (Hayasaka et al., 2004). Indeed, we found that the local variability of resel estimates is much greater than any regional differences in smoothness across the cortical surface. For this reason, we avoided the use of local smoothness estimates and instead made the simplifying assumption of uniform smoothness.

Smoothing statistics on cortical surface meshes

We applied an iterative smoothing algorithm in which, after each iteration, the new value for a given vertex is the average of its previous value and the values of its nearest neighbors. Chung et al. describe a "heat kernel smoothing" algorithm based on heat diffusion models that is similar to iterative nearest neighbor averaging, except that the contributions from each of the neighbors are unequally weighted in a way that depends on the distance from the central vertex (Chung et al., 2005):

$$W_\sigma(p,q_i) = \frac{e^{-\frac{d^2(p,q_i)}{2\sigma^2}}}{\sum_{j=0}^{m} e^{-\frac{d^2(p,q_j)}{2\sigma^2}}} \quad (3)$$

where for a given vertex p with a set of m neighbors q_i, with q_0 being the central vertex p itself, $W_\sigma(p,q_i)$ is the weight assigned to neighbor q_i. $d^2(p,q_i)$ is the squared 3D Euclidean distance between the central vertex p and the neighbor q_i, and σ is the bandwidth of the blurring kernel. Based on heat diffusion equations, Chung et al. note that, for a particular bandwidth and number of smoothing steps, the predicted FWHM smoothness is:

$$\mathrm{FWHM}(k,\sigma) = 2\sqrt{\ln 4}\sqrt{k\sigma} \quad (4)$$

As this method of iterative smoothing is a discrete approximation of a continuous diffusion process, Eq. (4) will not hold true for larger bandwidths (Chung et al., 2005).

Choice of cortical surface mesh for use in measurements

In the FreeSurfer software package, different surface meshes are generated at successive stages of the process of generating the complete cortical surface model. Because smoothness estimates depend on accurate inter-vertex distances, it is important to choose the appropriate cortical surface model for these measurements. The original – or "orig" – surface traces the edges of MRI voxels along the gray/white matter boundary, and thus appears blocky. The orig surface is then smoothed, generating the "smoothwm" (smoothed white matter) surface; with its smoothed edges, the smoothwm surface more closely matches the actual location of the gray/white matter boundary. As a result, inter-vertex distances are more accurate for the smoothwm surface (dv_{avg}=0.8 mm); for this reason, measurements were made using this surface. Other surfaces that involve inflation or flattening are inappropriate for these measurements as these processes slightly warp the inter-vertex distances in order to unfold the cortical surface. Inter-vertex distances and areas are also not accurate for the average unit sphere brain used to register the brains of multiple subjects (Fischl et al., 1999a,b).

Acquisition and analysis of fMRI data

Our methods for cortical surface-based group analysis of fMRI data have been described previously (Hagler and Sereno, 2006), but they are briefly summarized here. For each subject, cortical surface models were generated from high-resolution (1 mm^3) structural MRI scans using FreeSurfer (Dale et al., 1999; Fischl et al., 1999a). Twelve right handed subjects were scanned with a 4 T Varian MRI scanner while performing two working memory tasks in a random-order block design, with each stimulus condition contrasted to a third, passive condition (EPI T2*-weighted gradient echo pulse sequence, 8′15″ scan time, TR=3000 ms, TE=26.3 ms, flip angle=90°, bandwidth=125 kHz, 64×64 matrix, 36 axial slices, 3.75×3.75×3.8 mm voxels). Single subject data were analyzed using AFNI's 3dDeconvolve, generating general linear model (GLM) test correlation coefficients corresponding to the area under the hemodynamic response curve (Ward, 2000a). Because we did not use single subject *t*-statistics, we do not need to be concerned about bias from temporal autocorrelations. Before deconvolution, raw data were motion corrected using AFNI's 3dvolreg and then normalized by the mean value for each voxel. A second order polynomial baseline fit was used and motion estimates were used as nuisance regressors in the baseline GLM model. For volumetric analysis, GLM coefficients for each subject were warped into standard Talairach space using AFNI's adwarp before calculating group means and *t*-statistics. For cortical surface-based analysis, 3D GLM coefficients for a particular subject were sampled onto their cortical surface mesh, and then, after smoothing on the surface, those values were resampled onto an icosahedral sphere using a sulcal alignment with an average icosahedral (Fischl et al., 1999a,b). Group means and *t*-statistics were then calculated for each icosahedral vertex and sampled back onto the surface of a single subject.

Sampling 3D data onto cortical surface meshes

For each subject, 3D fMRI data sets were manually aligned with the high resolution structural MRI volumes used to create the cortical surface meshes. GLM statistics from the 3D data sets were "painted" onto the cortical surface by assigning to each surface vertex the value from the voxel at the corresponding 3D coordinates (Dale et al., 1999). We also tested a "vertex search" method; for each vertex, voxel values are tested at multiple points along the vertex's normal vector (i.e., perpendicular to the cortical surface) and the maximum value is assigned to the vertex. The idea of this method is to find the most significantly activated voxel within the gray matter. Another potential advantage of this method is that it can compensate for slight misalignments between functional and structural images, caused, for example, by distortion in EPI fMRI images. We used a relatively conservative search distance of 2 mm from the white/gray matter boundary.

Generation of realistic noise data sets

We generated realistic noise data sets for several subjects by shuffling the 3D GLM coefficient voxels contained within brain

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

mask volumes defined by the raw fMRI data itself. After shuffling, the resulting data sets had the same mean and variance, but the average 3D FWHM was reduced from ~5 mm to ~2 mm. Much of the smoothness in the original data is likely due to the widespread areas of activation (Kiebel et al., 1999); however, it is possible that some of the greater spatial correlation is related to the intrinsic smoothness of fMRI images. In an attempt to more accurately model the original data, though possibly erring conservatively, we applied a 5 mm FWHM Gaussian blurring kernel to the shuffled 3D data; this increased the estimated FWHM to ~5 mm, matching the original data.

Group analysis residual error calculation

Residual error of group analysis was calculated as the difference between single-subject data (GLM coefficients re-sampled to the spherical average surface) and the group mean. The inter-subject standard deviation of the residual error was then used to normalize the residuals (Kiebel et al., 1999; Worsley et al., 1999). FWHM smoothness was calculated from the normalized residuals using a modification of Eq. (2), wherein var(ds) represents the inter-neighbor variance of the normalized residual error across surface vertices and subjects, and var(s) represents the overall variance of values across vertices and subjects.

Random field theory estimates and Monte Carlo simulations of cluster size

Cluster size limits for use with multiple comparisons correction were estimated using random field theory (Worsley et al., 1996; Andrade et al., 2001) and validated with Monte Carlo simulations (Forman et al., 1995). Cortical surface cluster size limits were estimated with random field theory using the stat_thresh program from Keith Worsley's fmristat. For these estimates, we made the simplifying assumption of spatially uniform smoothness, using FWHM smoothness values calculated from the simulated noise group analysis. Cluster size limits were also generated with Monte Carlo simulations using a method similar to that used in AFNI's AlphaSim (Ward, 2000b).

Group analysis of fMRI data typically results in t-statistics. We simulated t-statistics by generating $N=12$ cortical surface data sets with normally distributed random values. Iterative smoothing steps were applied to achieve a range of smoothness levels. For each level of smoothness, and for each surface vertex, a t-statistic was calculated and statistical thresholds were applied corresponding to various uncorrected probability (p) values assuming a t-distribution with degrees of freedom equal to N-1. Clusters were defined as those areas of contiguous vertices with supra-threshold values; the area of the largest cluster was then calculated. By repeating this process for 500 iterations, a histogram of maximum cluster area was generated. A corrected p-value, or alpha, for each cluster size was also calculated; the largest cluster size that corresponds to a desired alpha could then be used as a cluster size limit.

Generating t-statistics in this way is computationally intensive, mostly because of the surface smoothing applied to N data sets for 500 iterations. With large numbers of subjects, t-statistics can be approximated with normally distributed z-statistics. Because it requires considerably less computation to generate, smooth, and threshold a single data set per iteration, and because this is the

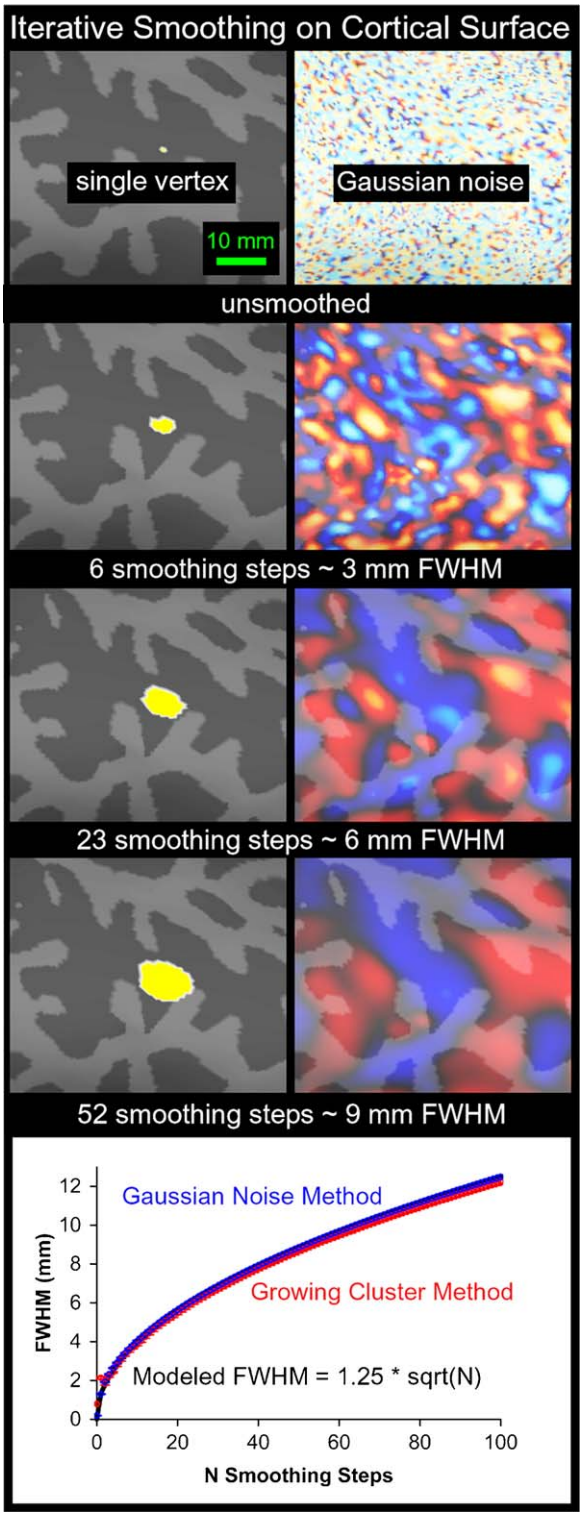

Fig. 1. Iterative smoothing of simulated data on the cortical surface. Images on the left illustrate a cluster growing from a single "seed" vertex as smoothing is applied. For each level of smoothing, the cluster includes those vertices with values at least half the maximum value (the current value of the seed vertex). Images on the right show the effect of smoothing on Gaussian noise (normally distributed random values) at each vertex. Below is a graph plotting the full-width-half-max (FWHM) distance as a function of the number of smoothing steps for each of these estimation methods. Average values ($n=12$ subjects) were plotted with error bars representing the standard deviation.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

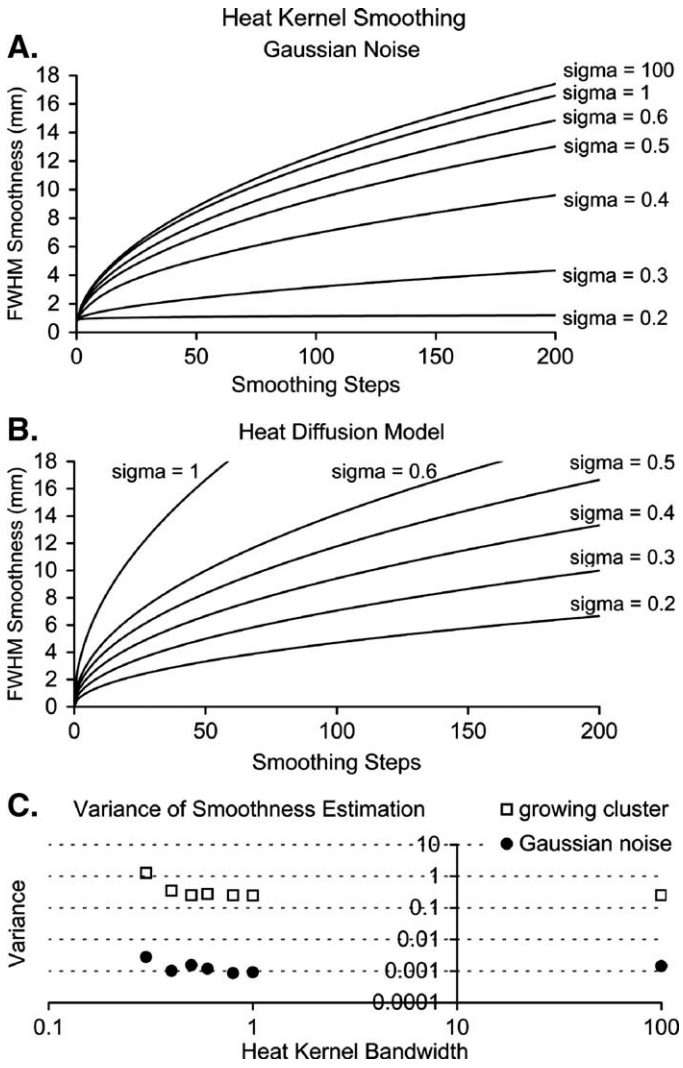

Fig. 2. Heat kernel smoothing. (A) FWHM smoothness as a function of iterations – estimated with the Gaussian noise method – is plotted for several heat kernel bandwidths. (B) FWHM smoothness as function of iterations predicted by heat diffusion model for different heat kernel bandwidths (see Eq. (4)). (C) Variance of smoothness estimates does not decrease with smaller bandwidths.

method used by AFNI's AlphaSim, we also generated cluster size thresholds this way.

The surface-based smoothing, cluster, and simulation programs described in this paper, written in C and C++, are freely available for download at http://kamares.ucsd.edu/~dhagler/dhagler-tools.tar.gz.

Results

Estimation of effective Gaussian filter width due to iterative smoothing

Two different methods were used for estimating the effective Gaussian filter width as a function of the number of smoothing steps. First, the effective FWHM distance was calculated directly from a single cluster of surface vertices. Starting with zero values at each cortical surface vertex, a single vertex was randomly selected, given a constant starting value, and then iteratively smoothed using nearest-neighbor averaging, creating a cluster of vertices with non-zero values. With more smoothing, this cluster grows with an approximately circular shape when viewed on an inflated or flattened representation of the cortical surface (Fig. 1). A threshold was applied equal to half the value of the seed vertex (i.e., the maximum value of the cluster) and the surface area of the remaining cluster was used, assuming a roughly circular shape, to calculate the cluster's diameter (i.e., FWHM distance). The second method involved estimating the Gaussian FWHM from a simulated noise data set after iterative smoothing. Random values with a normal distribution were generated at each vertex of a cortical surface. After each iteration of smoothing, the Gaussian filter width was estimated using Eq. (2). Both of these methods were applied 100 times, and the results were separately averaged. These simulations were repeated using the cortical surface models of 12 subjects, and the results were again averaged. Fig. 1 illustrates these two methods and plots FWHM as a function of smoothing steps for both methods. A small degree of variability was observed between estimates obtained using different subjects' surface meshes (error bars in Fig. 1 represent standard deviation).

One peculiarity of the cluster area FWHM measure should be noted; namely, there is a local maximum for $N=1$ smoothing step. This is explained by the fact that after a single smoothing step, the

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

seed vertex and each of its neighbors have the same value, including all of them in the suprathreshold cluster. After two smoothing steps, however, the cluster area FWHM closely matches the Gaussian noise measure. Their strong correlation is reflected in a correlation coefficient of 0.9993.

Both functions are well described as proportional to the square root of the number of smoothing steps:

$$\mathrm{FWHM}_{\mathrm{surf}} \approx k\sqrt{N} \quad (5)$$

where N is the number of iterative smoothing steps and k is a constant. Eq. (5) allows for easy determination of the appropriate number of smoothing steps necessary to simulate a blurring kernel with a desired FWHM. A linear best fit was found for FWHM as a function of $\sqrt{N}$, constraining the y-intercept to 0, and k was set equal to the slope. Because the cluster size simulations described below employed the Gaussian noise method for estimating FWHM, k was determined from those values ($k=1.25$, least squares linear fit $R^2=0.9998$).

Evaluation of diffusion-based heat kernel smoothing

Smoothing simulations similar to those described above were performed using Chung et al.'s heat kernel smoothing (Chung et al., 2005). For each vertex, weights were assigned to neighboring vertices according to Eq. (3). Smoothness was estimated as a function of both number of iterations and the bandwidth of the heat kernel (Fig. 2A). As bandwidth increases, the heat diffusion model becomes less valid (Chung et al., 2005), and for large bandwidths (e.g., 100), this method becomes indistinguishable from nearest-neighbor averaging (compare Figs. 1 and 2A). A comparison of the empirical smoothness estimates with those predicted by the heat diffusion model (Eq. (4)), however, shows that even for small bandwidths the heat diffusion model does not accurately predict the degree of smoothing caused by the heat kernel method (Eq. (3)) (Fig. 2B).

Despite this discrepancy between the measured and predicted smoothness curves, the potential advantage of heat kernel smoothing is that it may provide more spatially uniform smoothing; however, this method does increase the number of iterations required – and hence computing time – to achieve a particular level of smoothness, particularly for small bandwidths (Fig. 2A). Because the variability of the inter-vertex distances of the cortical surface meshes models we used was relatively small (SD ~0.2 mm, with mean distance ~0.8 mm), it may be reasonable to make the simplifying assumption of equal inter-vertex distances, as the simple average method does.

If heat kernel smoothing improves the spatial uniformity of smoothing, then the variance of FWHM estimates should be reduced relative to the simple average method, particularly when FWHM is estimated with the cluster area method, which is more susceptible to local variations in inter-vertex distances. To test this, we calculated the variance of smoothness estimates (both cluster area and Gaussian noise FWHM methods) for different heat kernel bandwidths, with the number of smoothing steps for each bandwidth adjusted to obtain ~4 mm FWHM smoothness (Fig. 2C). As expected, the variance is greatly increased for the cluster area method relative to the Gaussian noise method (~100-fold increase). Comparing across different bandwidths, we found that using a smaller bandwidth did not significantly reduce the measurement variance. The variances of the area FWHM were actually higher for lower bandwidths (F-test $p<0.0001$ for bandwidth 0.3 compared to 100). This increased variance is not due to rounding errors, as there is no variance between estimates when the same noise data set is used for each trial.

Smoothness estimates of group analysis statistics

Smoothness estimates from single subject residual variance have been used previously for multiple comparisons correction for single subject statistics (Kiebel et al., 1999; Worsley et al., 1999). For group analysis, a similar method can be used on the inter-subject residual variance (Worsley et al., 1999; Hayasaka et al., 2004). We compared estimates from this method with smoothness estimates derived from a group analysis of simulated noise data sets. Furthermore, we carried out our group analysis of simulated noise data sets in parallel with group analysis of actual fMRI data, using both 3D and cortical surface-based methods.

By measuring the FWHM smoothness of group t-statistics for shuffled data sets and normalized residual error from group analysis of actual data (see methods), we were able to determine the appropriate range of FWHM values to use for cluster size simulations as a function of the blurring kernel applied to the single subject data. In doing these simulations, we were also able to measure the FWHM smoothness of single subject statistics for multiple levels of smoothing, comparing the properties of iterative smoothing on the cortical surface and Gaussian smoothing in 3D; this was done for real data, normalized residual error, and realistic noise.

The effect of smoothing on single subject statistics and group analysis statistics, both with and without shuffling, is illustrated in Fig. 3. Graphs of corresponding FWHM measurements are shown in Fig. 4. Resampling voxels to 1 mm^3, which accompanied Talairach registration, by itself reduced FWHM smoothness (Figs. 4A, B). This effect was even greater for sampling onto a high resolution (inter-vertex distance ≈0.8 mm) cortical surface mesh. 3D Talairach registration and resampling often include some form of interpolation, but this was not done to the make the comparison between 3D and 2D registration methods more fair. Sampling to the surface with the normal search painting method (see Methods) had no immediate effect on the smoothness of either actual data or shuffled statistics (Fig. 5). For both 3D and surface smoothing, smoothness increased linearly as a function of blurring kernel width. The slopes of these smoothness functions were decreased for shuffled data relative to the original data (Figs. 4A, B), demonstrating the extent to which real activation foci contribute to smoothness estimates (Kiebel et al., 1999).

The slopes of the smoothness functions for data sampled to the surface with and without normal search were quite similar, with minimal difference for shuffled data and a small increase in smoothness for actual data painted with vertex search (Fig. 5). The increased smoothness of data with normal search likely reflects additional activation that is painted to the surface—that would otherwise be missed due to slight misalignments.

Smoothness of group statistics increased similarly as a function of blurring kernel, although with some differences compared to single subject statistics. The smoothness of surface-based group t-statistics from actual data increased with a slope very similar to that of single subject data. For 3D smoothing of data, the smoothness function noticeably deviated from linearity, taking on a slightly sigmoid shape (Fig. 4D). Smoothness of group

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

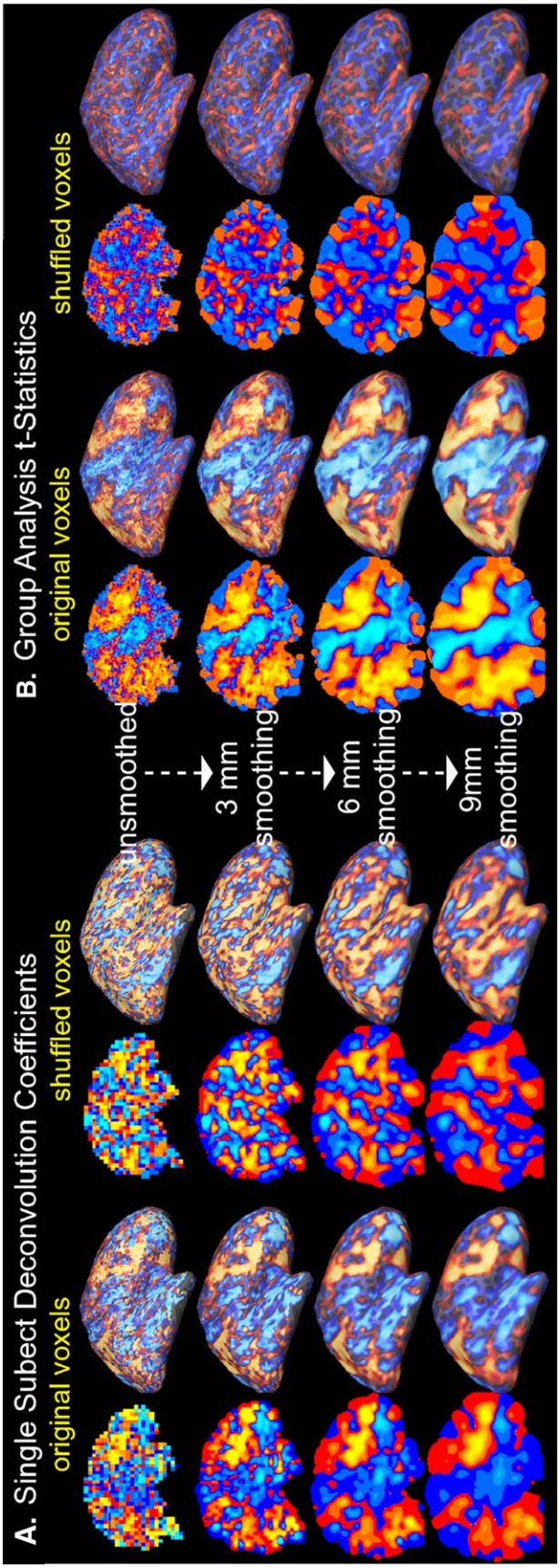

Fig. 3. 3D and surface smoothing applied to real and simulated noise data sets. In panel A, single subject GLM deconvolution coefficients (generated with AFNI's 3dDeconvolve) are shown. Odd columns show the data (subject performing a working memory task) from the original 3D voxels for a single Talairach-transformed sagittal slice ($x = -42$ mm), while even columns show the same data sampled onto the cortical surface model. The third and fourth columns show the same data but after shuffling the 3D voxels within the brain to simulate noise. The first row shows these data without smoothing, and subsequent rows show the data after 3, 6, or 9 mm FWHM smoothing, either in 3D or on the cortical surface. In panel B, the results of group analysis (t-statistics, $n = 12$ subjects) after varying levels of smoothing applied to each subjects' data are shown. All data are unthresholded and scaled liberally to enable viewing of even the smallest values.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

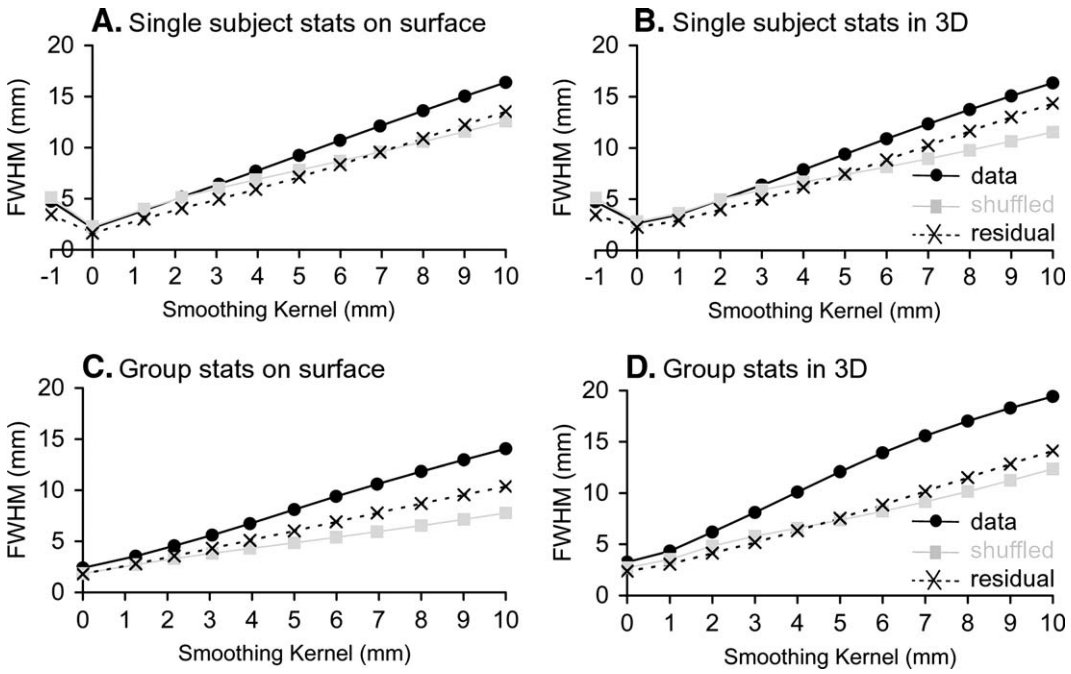

Fig. 4. Smoothness measurements as a function of 3D or surface smoothing. FWHM Gaussian smoothness was estimated from 3D and cortical surface data by comparing the variance of inter-neighbor differences to overall variance (Eqs. (1) and (2) in text). These measurements are plotted as functions of the FWHM of the blurring kernel applied. Results are shown for single subject deconvolution coefficients (A and B, average of measurements from 12 subjects) and group analysis *t*-statistics (C,D). In panels A and B, the leftmost data points are measurements from the original 3D voxels, prior to Talairach registration or resampling to the cortical surface. Residuals in panels A and B are the differences between the GLM model and the single subject time series data, normalized by the standard deviation across time. Residuals in panels C and D are the differences between each subject's GLM coefficients and the group mean, normalized by the standard deviation across subjects.

t-statistics from shuffled data increased much more slowly than group *t*-statistics from actual data (Figs. 4C, D). The smoothness function of residual error was generally similar to that of shuffled data, particularly at an applied smoothness of 4 mm FWHM. With

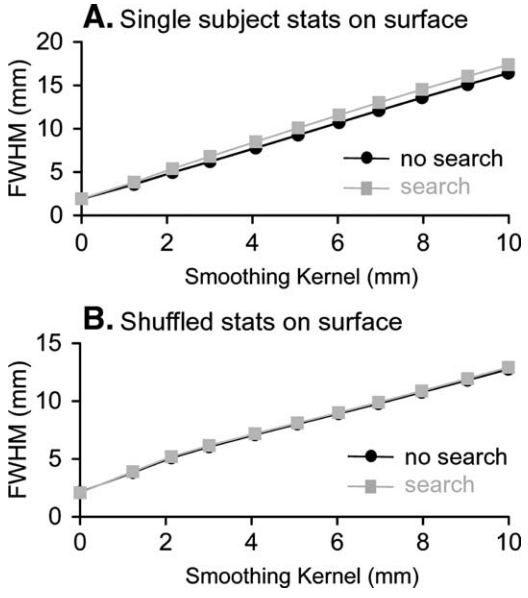

Fig. 5. Smoothness measurements as a function of surface smoothing for data painted to surface with and without normal vector search, for both actual and shuffled single subject GLM coefficients.

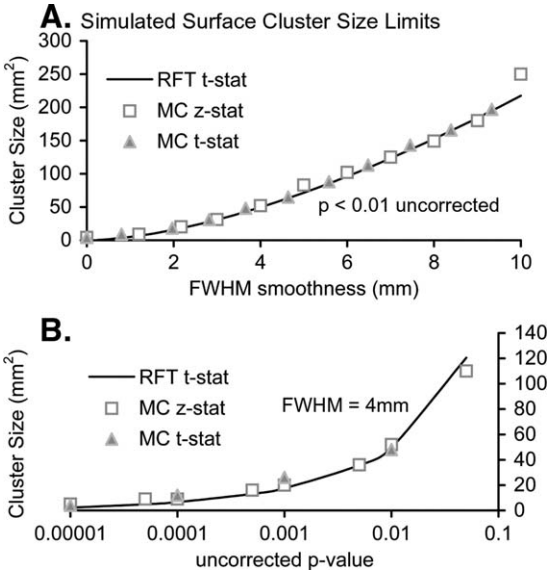

Fig. 6. Cluster size threshold estimates. Cortical surface cluster size (mm^2) is plotted in panel A as a function of FWHM smoothness. The uncorrected *p*-value was held constant at $p<0.01$. In panel B, cluster size is plotted as a function of uncorrected *p*-value, with smoothness FWHM=4 mm. Cluster sizes plotted correspond to the maximum cluster size that would occur by chance 5% of the time with the given uncorrected *p*-value and FWHM smoothness according to either random field theory (RFT) using a *t*-distribution, Monte Carlo simulations with a *z*-distribution, or Monte Carlo simulations with a *t*-distribution.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

larger blurring kernels, normalized residuals were smoother than the shuffled data sets.

A comparison of surface-based (Fig. 4C) and 3D (Fig. 4D) group statistics, shows that 3D averaging results in substantially more blurring than surface-based averaging. For example, if a 4 mm blurring kernel is applied to 3D single subject data, the group *t*-statistics exhibit ~6.5 mm FWHM intrinsic smoothness (shuffled stats or residual error, Fig. 4D). If 10 steps of surface smoothing are applied to single subject statistics, corresponding to a 4 mm FWHM blurring kernel, the intrinsic smoothness of the group *t*-statistics is ~4.5 mm FWHM (Fig. 4C).

Surface-based cluster size thresholding

Having arrived at smoothness estimates unbiased by real brain activations, we used these smoothness values to estimate cluster size thresholds. Cluster size thresholds were estimated directly, assuming a *t*-distribution, with random field theory using Keith Worsley's fmristat, and then compared to cluster sizes generated empirically with Monte Carlo simulations. Additionally, we simulated both *t*-statistics, which is computationally intensive (Hayasaka and Nichols, 2003), and *z*-statistics, as is done by AFNI's AlphaSim (Ward, 2000b). We find a close agreement between these different methods for generating cluster size thresholds (Fig. 6). An example of the application of such thresholds is shown in Fig. 7. We performed a surface-based inter-subject *t*-test after applying 4 mm surface-based smoothing to the single subject GLM coefficients, which resulted in 4.5 mm FWHM smoothness of the group *t*-statistics of shuffled data or the normalized residuals. Two separate thresholds – corresponding to uncorrected *p*-values of 10^{-2} and 10^{-3} – were applied to the group analysis *t*-statistics, setting sub-threshold values to zero. Clusters were found among the remaining values and those clusters smaller than the cluster size limit for the given uncorrected *p*-value – corresponding to a corrected alpha of 0.05 – were also set equal to zero.

Defining ROIs with sliding threshold clustering

The set of clusters identified with different uncorrected *p*-values were naturally quite different in spatial extent. More conservative thresholds excluded weaker, but presumably real activations while more liberal thresholds resulted in the joining of multiple foci into larger clusters (Fig. 7). We developed a method for defining surface-based ROIs by using a sliding threshold clustering scheme in which clusters were first identified using a low *p*-threshold and the corresponding large cluster size limit (Fig. 7B). Smaller clusters within those larger clusters were then identified by applying successively higher *p*-thresholds—along with their smaller cluster size limits (Fig. 7C). After identifying multiple clusters in this way, overlap between clusters was resolved by excluding the larger cluster in a pair of overlapping clusters. Finally, the remaining

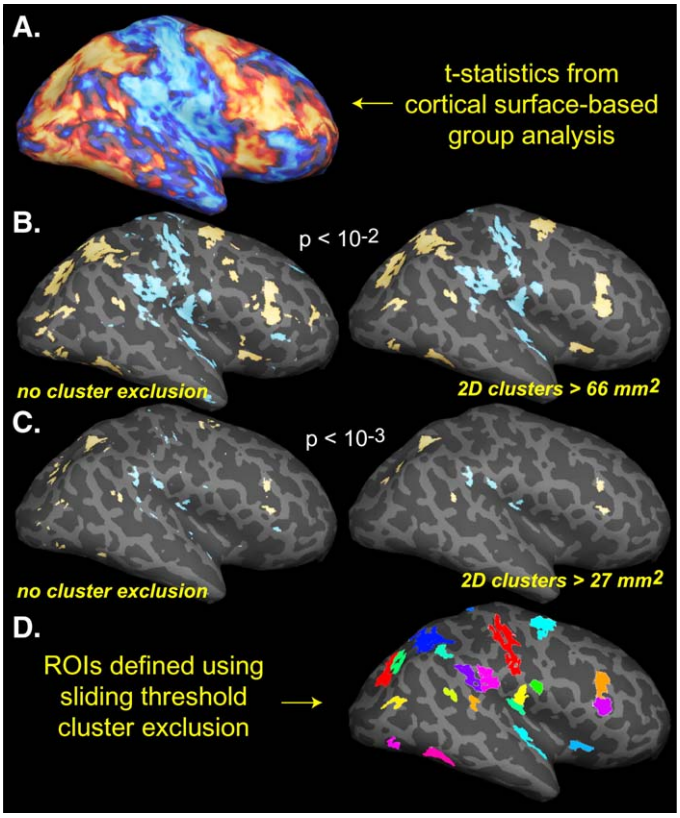

Fig. 7. Cluster size exclusion applied to surface-based group statistics. Group analysis *t*-statistics (real data, $n=12$ subjects) are displayed on a cortical surface mesh. (A) Unthresholded *t*-statistics. (B) *t*-statistics thresholded at $p<10^{-2}$, before (left) and after (right) clusters smaller than the cluster size threshold were excluded. In panel C, a threshold of $p<10^{-3}$ was applied. In panels B and C, the multiple comparison corrected *p*-value was 0.05 after cluster exclusion for the surface-based statistics. (D) Multiple ROIs defined with sliding threshold cluster exclusion and subsequent cluster growth. Automatically generated cortical surface ROIs are shown in different colors.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

clusters were allowed to grow radially outward until they reached another cluster or they reached the edge of the original low threshold boundary (Fig. 7D).

Discussion

The introduction of cortical surface-based methods has provided the means for improved visualization and analysis of fMRI data (Dale and Sereno, 1993; Sereno et al., 1995; Dale et al., 1999; Fischl et al., 1999a; Kiebel et al., 2000; Andrade et al., 2001; Formisano et al., 2004). We have implemented cluster-based multiple comparisons correction for cortical surface meshes generated with FreeSurfer and developed a sliding threshold clustering method for defining surface-based ROIs. To enable the use of these methods, we adapted cluster size simulation methods used in 3D analysis to cortical surface meshes and we performed other simulations to inform the choice of key parameters in cluster size simulations. We simulated the effects of iterative smoothing on the cortical surface, empirically deriving Gaussian filter width as a function of smoothing steps. This relationship guided the selection of the number of smoothing steps with which to precede group averages. Using this method, we also empirically tested the predictions of the heat kernel diffusion model of cortical surface smoothing. Furthermore, we compared the effects of 3D and surface smoothing on simulated noise and actual data and tested methods for estimating the intrinsic smoothness of group analysis statistics.

Surface-based smoothing

We propose using a simple method of iterative smoothing: averaging the values of neighboring vertices. The advantage of this method is that it is uncomplicated and computationally efficient. We showed that iterative smoothing based on models of diffusion does not seem to provide advantages that would outweigh the increased computation they require (Andrade et al., 2001; Chung et al., 2003, 2005). These diffusion smoothing methods may provide better spatial uniformity of smoothing when the surface meshes used have large variability in inter-vertex distance, but not with the relatively uniform, high-resolution cortical surface meshes typically used in MRI research. In addition, because of the approximations required for representing continuous diffusion as a discrete, iterative process, the FWHM smoothness achieved with a particular number of smoothing steps is not necessarily modeled accurately by the diffusion-derived equations, especially at the high bandwidths used by Chung et al. Finally, we find that use of a high bandwidth is only minimally different from simple averaging. Thus it remains necessary, as we have done, to measure the actual FWHM smoothness resulting from a particular set of smoothing parameters.

Sampling 3D data onto a cortical surface mesh

We sampled, or "painted" 3D GLM coefficients onto cortical surface meshes using a simple method of assigning values to surface vertices based on the voxel within which they lie, and compared this to a searching method in which for each vertex, the maximum value was chosen from the voxels a short distance along the vertex's normal vector. If a particular patch of gray matter is imaged with 2 or more voxels that include signal from white matter and/or CSF (partial voluming), this search method will choose the voxel with the largest signal and presumably the largest fraction of gray matter. The search method can also compensate for slight misalignments – e.g., due to image distortion – between the EPI images and the high-resolution structural images used to create the cortical surface meshes. Accurate shimming and robust methods for correcting EPI image distortion are important for reducing registration errors, but residual distortions typically remain. We found that normal search painting slightly increased the FWHM smoothness of painted data, but had no effect on simulated noise data sets.

One could also paint time series data, rather than GLM coefficients, onto the cortical surface and then carry out GLM deconvolution (Andrade et al., 2001). Using normal search painting in this case would not be recommended as there is no reason to expect that the higher amplitude voxel at a given time point is the one containing the largest fraction of gray matter. If data are painted without normal search, it makes no difference whether values are sampled to the surface before or after GLM tests, as the values are not changed by the resampling. There could be a very slight difference if spatial smoothing was applied on the surface before, rather than after, single subject GLM deconvolution. We found, however, that smoothing 3D time courses before, rather than after, GLM deconvolution had a minimal difference on the resulting statistics; neither order appeared to provide any systematic advantage (data not shown). Because painting time series data onto a cortical surface mesh are much more time consuming than painting a few GLM coefficients, it is more efficient to paint and smooth after deconvolution.

Estimating the intrinsic smoothness of group analysis statistics

Cluster size limits for inter-subject analyses derived from random field theory or Monte Carlo simulations require an estimate of the smoothness of group analysis statistics. Because actual data presumably contain real activations, direct estimation of smoothness will result in overly conservative cluster size thresholds. Our approach to estimating the intrinsic smoothness of group analysis statistics was based on simulation of an actual group analysis. As expected, the smoothness estimates of the group *t*-statistics were substantially lower for the simulated noise analysis, resulting in lower cluster size thresholds. We found that this method was similar to estimation of smoothness from normalized residual error; because of the considerable computation involved in shuffling the voxels for each subject and computing a separate group analysis, it is preferable to estimate smoothness from the normalized residuals. The two methods provide the same result when a 4 mm blurring kernel is used (Fig. 4). The source of the disagreement between these methods with larger blurring kernels is unclear. It is possible that the greater smoothness of the residuals is due to anatomical correlations between subjects; some of the smoothness related to cortical activation could also be retained in the residual error, particularly if the response is not perfectly modeled by the GLM equations.

Using these methods for estimating the intrinsic smoothness of group statistics, we found that cortical surface-based group analysis resulted in substantially less additional blurring than 3D analysis via Talairach registration. One practical consideration of this difference is that lower smoothness estimates allow for smaller cluster size thresholds for multiple comparison correction. This difference also appears to suggest higher effective resolution for cortical surface-based group analysis.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036

Sliding threshold clustering

Because the sliding threshold clustering method applies multiple statistical tests, each with its own likelihood of detecting false clusters, an additional multiple comparisons problem arises, perhaps suggesting a Bonferroni-type correction such as requiring that the false positive rate, or alpha, for each test be divided by the number of tests. Alternatively, False Discovery Rate methods could be used to exclude some of the less significant clusters (Genovese et al., 2002). It should be recognized, however, that the tests are not truly independent. If a cluster of a particular size and intensity passes one test, whether it is true or false, it is likely that the same cluster will pass another test with a slightly higher *p*-threshold but a lower cluster size limit. We have, however, avoided the need for additional correction by applying a single cluster threshold test with a relatively low uncorrected *p*-threshold – and the corresponding large cluster size limit – and then limiting the purpose of the sliding threshold method to finding smaller clusters within those larger clusters. Thus, the risk is not that we increase the rate of false positives, but that we incorrectly subdivide a cluster, a perhaps more acceptable risk.

Restricting analysis to ROIs can be another way to control for the multiple comparisons problem. ROIs can be created on the basis of anatomical markers – such as sulci and gyri – but such regions may not directly correspond to functional subdivisions of the brain. Using clusters of activation in response to a given stimulus as the ROIs can provide a way to more meaningfully restrict the analysis. For example, we have previously used activations from block-design experiments to guide the creation of ROIs with which to analyze phase-encoded retinotopic mapping data (Hagler and Sereno, 2006). The problem, however, is that it is difficult to choose a single threshold that results in clusters that adequately reflect the population of cortical areas of interest. Clusters formed with a low threshold will tend to cover too large a region, often including several local maxima within a single cluster. Higher thresholds, however, may exclude areas with weaker, but true, activation; furthermore, the extent of the remaining clusters is usually quite small, making the ROI analysis too restricted, ignoring potentially important regions of activation surrounding the peaks. The advantage of the sliding threshold method is that a balance is struck between these two extremes without requiring the subjective choice of the single threshold that results in the optimal set of ROIs.

Acknowledgments

The authors thank Moo Chung for helpful clarifications of his heat kernel smoothing method. Support contributed by: NSF BCS 0224321 and NIMH NRSA 5F32MH066578-02.

References

Andrade, A., Kherif, F., Mangin, J.F., Worsley, K.J., Paradis, A.L., Simon, O., Dehaene, S., Le Bihan, D., Poline, J.B., 2001. Detection of fMRI activation using cortical surface mapping. Hum. Brain Mapp. 12, 79–93.

Chung, M.K., Worsley, K.J., Robbins, S., Paus, T., Taylor, J., Giedd, J.N., Rapoport, J.L., Evans, A.C., 2003. Deformation-based surface morphometry applied to gray matter deformation. NeuroImage 18, 198–213.

Chung, M.K., Robbins, S.M., Dalton, K.M., Davidson, R.J., Alexander, A.L., Evans, A.C., 2005. Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage 25, 1256–1265.

Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207.

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.

Formisano, E., Esposito, F., Di Salle, F., Goebel, R., 2004. Cortex-based independent component analysis of fMRI time series. Magn. Reson. Imaging 22, 1493–1504.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1, 210–220.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.

Hagler Jr., D.J., Sereno, M.I., 2006. Spatial maps in frontal and prefrontal cortex. NeuroImage 29, 567–577.

Hayasaka, S., Nichols, T.E., 2003. Validating cluster size inference: random field and permutation methods. NeuroImage 20, 2343–2356.

Hayasaka, S., Nichols, T.E., 2004. Combining voxel intensity and cluster extent with permutation test framework. NeuroImage 23, 54–63.

Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 22, 676–687.

Kiebel, S.J., Poline, J.B., Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. NeuroImage 10, 756–766.

Kiebel, S.J., Goebel, R., Friston, K.J., 2000. Anatomically informed basis functions. NeuroImage 11, 656–667.

Petersson, K.M., Nichols, T.E., Poline, J.B., Holmes, A.P., 1999a. Statistical limitations in functional neuroimaging: II. Signal detection and statistical inference. Philos. Trans. R Soc. Lond., B Biol. Sci. 354, 1261–1281.

Petersson, K.M., Nichols, T.E., Poline, J.B., Holmes, A.P., 1999b. Statistical limitations in functional neuroimaging: I. Non-inferential methods and statistical models. Philos. Trans. R. Soc. Lond., B Biol. Sci. 354, 1239–1260.

Poline, J.B., Mazoyer, B.M., 1993. Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters. J. Cereb. Blood Flow Metab. 13, 425–437.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893.

Ward, B.D., 2000a. Deconvolution Analysis of FMRI time series data. AFNI 3dDeconvolve Documentation, Medical College of Wisconsin.

Ward, B.D., 2000b. Simultaneous Inference for FMRI Data. AFNI 3dDeconvolve Documentation, Medical College of Wisconsin.

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., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., Evans, A.C., 1999. Detecting changes in nonisotropic images. Hum. Brain Mapp. 8, 98–101.

Please cite this article as: Donald J. Hagler et al., Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data, NeuroImage (2006), doi:10.1016/j.neuroimage.2006.07.036