NeuroImage 52 (2010) 158–171

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Determining the optimal level of smoothing in cortical thickness analysis: A hierarchical approach based on sequential statistical thresholding

Jorge L. Bernal-Rusiel, Mercedes Atienza, Jose L. Cantero *

Laboratory of Functional Neuroscience, Spanish Network of Excellence for Research on Neurodegenerative Diseases (CIBERNED), University Pablo de Olavide, Seville, Spain

ARTICLE INFO

Article history:
Received 10 November 2009
Revised 25 March 2010
Accepted 26 March 2010
Available online 1 April 2010

Keywords:
Cortical thickness
Spatial smoothing
Spherical wavelets
Cluster size inference
False discovery rate
Alzheimer's disease

ABSTRACT

The extent of smoothing applied to cortical thickness maps critically influences sensitivity, anatomical precision and resolution of statistical change detection. Theoretically, it could be optimized by increasing the trade-off between vertex-wise sensitivity and specificity across several levels of smoothing. But to date neither parametric nor nonparametric methods are able to control the error at the vertex level if the null hypothesis is rejected after smoothing of cortical thickness maps. To overcome these drawbacks, we applied sequential statistical thresholding based on a simple hierarchical model. This methodology aims at controlling erroneous detections; firstly at the level of clusters, over smoothed statistical maps; and secondly at the vertex level, over unsmoothed statistical maps, by applying an adaptive false discovery rate (FDR) procedure to clusters previously detected. The superior performance of the proposed methodology over other conventional procedures was demonstrated in simulation studies. As expected, only the hierarchical method yielded a predictable false discovery proportion near the predefined FDR q-value for any smoothing level at the same time as being as sensitive as the others at the optimal setting. It was therefore the only method able to approximate the optimal size of spatial smoothing when the true change was assumed unknown. The hierarchical method was further validated in a cross-sectional study comparing moderate Alzheimer's disease (AD) patients with healthy elderly subjects. Results suggest that the extent of cortical thinning reported in previous AD studies might be artificially inflated by the choice of inadequate smoothing. In these cases, interpretation should be based on the location of local maxima of suprathreshold regions rather than on the spatial extent of the detected signal in the statistical parametric map.

Introduction

Thickness is a descriptor of the mammalian neocortex that provides relevant information on the integrity of cortical columns and morphological correlates of higher cognitive functions (Von Economo, 1929; Jones and Peters, 1984). Computational neuroanatomy techniques are contributing substantially to *in vivo* measurements of human cortical thickness facilitating a topographical and quantitative description of atrophy patterns associated with prevalent neurological (Butman and Floeter, 2007; Charil et al., 2007; Biega et al., 2006; Singh et al., 2006; Lerch et al., 2005; Thompson et al., 2001) and psychiatric disorders (Makris et al., 2007; Shaw et al., 2006; Lyoo et al., 2006; Kuperberg et al., 2003). In addition, these techniques have been successfully applied to detect cortical descriptors that differentiate healthy aging (Sowell et al., 2003) from incipient neurodegenerative processes (Apostolova and Thompson, 2008). Thus, a different rate and topographical pattern of cortical thinning have been described in AD patients when compared with healthy elderly subjects, which seems relevant for the diagnosis and early detection of this prevalent neurodegenerative disease (Dickerson et al., 2009; Singh et al., 2006; Lerch et al., 2005; Thompson et al., 2003).

Cortical thickness maps are obtained by computational techniques either involving surfaces or voxels, or even a mixture of both. Our study is focused on surface-based methods that provide measures of the distance between surface-based models of the gray matter/white matter and the gray matter/CSF boundaries (Lerch and Evans, 2005; Fischl and Dale, 2000). Although the spatial resolution of these measures has drastically increased in the last few years, spatial blurring required by statistical testing results in decreased anatomical precision of cortical thinning estimations (Han et al., 2006; Lerch and Evans, 2005). Smoothing is typically performed by linearly filtering the cortical thickness maps with surface-based Gaussian kernel approximations after the maps are resampled to an average surface. These filters therefore act as low-pass spatial frequency filters along the average cortical manifold (Hagler et al., 2006; Chung et al., 2005). As a general rule, the larger the extent of smoothing the lower the spatial resolution of thickness measurements and accuracy at identifying cortical thinning. Note however that an appropriate level of smoothing can significantly increase the sensitivity of subsequent statistical analyses by increasing the signal-to-noise ratio in the statistical map.

* Corresponding author. Laboratory of Functional Neuroscience, University Pablo de Olavide, Ctra. de Utrera, km. 1, 41013-Seville, Spain. Fax: +34 954 349151.
E-mail address: jlcanlor@upo.es (J.L. Cantero).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.03.074

Cortical surfaces are typically mapped onto the sphere to establish an intrinsic 2D spherical coordinate system. To take advantage of this natural surface parameterization, previous studies have approximated the Gaussian kernel filtering with heat kernel smoothing over the sphere (Chung et al, 2007). We further found that non-linear spherical wavelet-based denoising schemes improve the trade-off between anatomical precision and thinning detection achieved with surface-based approximations to Gaussian kernel smoothings (Bernal-Rusiel et al., 2008). The lower smoothing introduced by wavelet-based spatial filters together with their adaptive properties likely account for their better performance when compared with the Gaussian smoothing operators. In the present study, we smooth individual cortical thickness maps over the spherical average surface by using both the previously developed spherical wavelet-based denoising and the approximation to Gaussian kernel smoothing given by the iterative nearest neighbor averaging algorithm (Hagler et al., 2006; Han et al., 2006).

According to the matched filter theory (Pratt, 1991), the optimal extent of spatial smoothing can be determined by enhancing the matching with the putative area of change. However, it seems difficult to optimize this criterion in exploratory studies. Results from real (Han et al. 2006; Singh et al., 2006) and simulation studies (Bernal-Rusiel et al., 2008; Lerch and Evans, 2005) suggest that increasing smoothness of thickness maps not only enhances sensitivity but also reduces specificity and image resolution. Here sensitivity is defined as the probability of correctly identifying a vertex showing true thinning, while specificity is the probability of correctly rejecting a vertex that did not change. It seems, therefore, that improving the trade-off between sensitivity and specificity over the range of detection might be a preliminary approach to determine the optimal smoothing to apply in cortical thickness analysis. But computation of sensitivity and specificity requires knowing the number of true positive and true negative vertices detected respectively, which can only be determined in simulation studies. Alternatively, estimation of this trade-off could be based on estimation of the proportion of false positives among all the detected vertices and the number of true null hypotheses. Unfortunately, neither parametric random field methods (Worsley et al., 1996; Friston et al., 1994) nor nonparametric permutation-based tests (Hayasaka and Nichols, 2003) are able to control the error at the vertex level if the omnibus null hypothesis is false and cortical thickness maps were previously smoothed. The same constraint is applicable to vertex-wise FDR procedures. In fact, the smoothing typically extends the signal present in one particular vertex to many null vertices resulting in artificial inflation of the proportion of detected vertices that contain true signal (Chumbley and Friston, 2009).

Here we propose a simple hierarchical model to overcome this drawback when applying either linear (e.g., Gaussian kernel approximation) or non-linear (e.g., wavelet-based) smoothing to cortical thickness maps. More specifically, our approach, firstly, controls for false positives at the level of clusters, via either random field theory or permutation-based inference, and then at the level of vertices (within each significant cluster detected in the previous step) by applying an adaptive FDR procedure. We confirmed the superior performance of the proposed methodology (for both Gaussian and spherical wavelet smoothing) over other statistical thresholding approaches by means of simulation studies. We further validated the method in a cross-sectional study comparing moderate AD patients with healthy elderly subjects.

Material and methods

Subjects

The simulation study was performed on 66 healthy elderly subjects (age: 59–94 yr, 50 women) selected from the OASIS database. Inclusion criteria consisted of Mini-Mental State Examination (MMSE) with scorings ≥ 29 (high level of functioning), and Clinical Dementia Ratings (CDR) of 0 (no dementia). They were randomly assigned into the control (no changes in cortical thickness were introduced) or experimental group (included hybrid changes in cortical thickness). Gender (25 females and 8 males in each group) and age were balanced between the two groups (control: 74.7 ± 8.9 yr; experimental: 74.7 ± 8.9 yr).

The proposed hierarchical methodology was further validated with two different smoothing methods (Gaussian and spherical wavelets), and its performance compared with five traditional statistical approaches (FDR and vertex- and cluster-based inference procedures by using either random field or permutation testing) in 53 mild-to-moderate AD patients (74.9 ± 9 yr, 37 females) selected from the OASIS database, and a subset of 53 well-matched healthy elderly subjects (75 ± 7 yr, 37 females) used in the simulation study.

MRI data acquisition

Four high resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired in each subject during the same session on a 1.5-T Vision™ scanner (Siemens, Erlangen, Germany). MP-RAGE parameters were empirically optimized for gray/white contrast (repetition time = 9.7 ms, echo time = 4 ms, flip angle = 10°, inversion time = 20 ms, delay time = 200 ms, 256×256 (1×1 mm) in-plane resolution, 128 sagittal 1.25 mm slices without gaps, time per acquisition = 6.6 min).

Cortical surface reconstruction and cortical thickness estimation

Volumetric segmentation and cortical surface reconstruction were performed automatically by using the Freesurfer toolkit (http://surfer.nmr.mgh.harvard.edu/). The analysis procedure was described in detail in prior studies (Fischl and Dale, 2000; Dale et al., 1999; Fischl et al., 1999a). Briefly, the processing steps for each subject included motion correction and averaging of the MP-RAGE volumes to generate a single volume with improved signal-to-noise ratio. Next, Talairach transformation and non-uniformity correction algorithms were applied before skull stripping through a hybrid watershed/surface deformation procedure (Segonne et al., 2004). After that, white matter was segmented and its boundary was tessellated to generate the gray/white surface. This surface was first deformed outwards to the location in the volume with the highest intensity contrast between the gray matter and the CSF and, later, it was refined to generate the pial surface (Dale et al., 1999). Cortical surfaces were constrained to a spherical topology by applying an automatic topology correction algorithm (Fischl et al., 2001). Once geometrically accurate and topologically correct models of both gray/white and pial surfaces were obtained, the cortical thickness map was generated in the native brains rather than in the Talairach space in order to increase the power of statistical tests (Ad-Dab'bagh et al., 2005). Cortical thickness at each vertex was defined as the average of the shortest distances between vertices of the gray/white matter boundary and the pial surface computed in both directions (Fischl and Dale, 2000).

Given the non Euclidian intrinsic geometry of cortical manifolds, spherical mapping of cortical surfaces were used to project the cortical thickness maps onto a spherical coordinate system for which a suitable parameterization, surface registration and basis were previously provided (Fischl et al., 1999a; Van Essen et al., 1998). Every thickness map was then resampled to the spherical average surface using the correspondence given by the spherical registration algorithm which aligns each individual cortical folding pattern with the average folding pattern (Fischl et al., 1999b). Finally, cortical thickness maps were smoothed by applying linear techniques like iterative nearest neighbor averaging (Hagler et al., 2006; Han et al., 2006) and non-linear spherical wavelet-based denoising schemes (Bernal-Rusiel et al., 2008).

The iterative nearest neighbor averaging is a particular case of the general heat kernel smoothing algorithm (Chung et al., 2005) for two-

dimensional Riemann manifolds when contributions of the neighbors of a given vertex to its post-iteration value are equally weighted. It therefore might provide less isotropic smoothing over the sphere than the iterative weighted averaging algorithm (Chung et al., 2005). However, as the average surface is tessellated with an approximately regular high resolution mesh (7th order subdivided icosahedron), differences between the iterative weighted averaging and the iterative nearest neighbor averaging algorithms should be negligible (Hagler et al., 2006). In addition, iterative nearest neighbor averaging algorithms are easier to implement and computationally more efficient than iterative weighted averaging procedures.

In the simulation study, the optimal Gaussian smoothing was obtained by testing different kernels ranging from 1 to 40 mm, in steps of 1 mm. Computation time was reduced because the iterative nearest neighbor averaging algorithm applied to each level of smoothing benefited from iterations computed for previous kernels. The optimal wavelet was obtained by varying two parameters of the filtering process: i) the basis dilation factor d which scales the initial support of the filters in harmonic space (and therefore in the real space), and ii) the number of the finest scales n_s at which the wavelet coefficients thresholding procedure was applied (Bernal-Rusiel et al., 2008).

For the wavelet-based smoothing, cortical thickness maps were re-interpolated from the mesh of the spherical average surface (7th order subdivided icosahedron, 163,842 vertices) to a spherical equiangular grid of 2048^2 samples. Fast nearest neighbor interpolation based on Delaunay triangulations was always applied (Watson, 1992). The equiangular grid is a mesh for the 2-sphere (S^2) defined as:

$$\wp = \left\{(\theta_m, \varphi_n) \in S^2 : \theta_m = \frac{(2m+1)\pi}{4B}, \varphi_n = \frac{n\pi}{B}\right\}; m, n \in T \equiv \{k \in N : k < 2B\}.$$

Any function $f \in L^2(S^2)$ of bandwidth $\{B \in 2N\}$ can be sampled without information loss on this grid (Driscoll and Healy, 1994). Resulting maps were then filtered by using a pyramidal decomposition based on axis-symmetric spherical discrete wavelet frames (Wiaux et al., 2008), and again re-interpolated back to the original icosahedron mesh before statistical analysis. A dyadic discretization was adopted for the scales of dilation in harmonic space on S^2 (Starck et al., 2006). The same analysis and synthesis filters (defined in harmonic space) were set at each scale by using a self-invertible filter bank. The different steps of the decomposition and signal reconstruction for admissible low-pass filter H and high-pass filter G were described in detail in Bernal-Rusiel et al. (2008).

In particular, the scale discretized axis-symmetric filters used here are defined in harmonic space by equations:

$$H^2 = \frac{\phi^2(l)}{\phi^2(2^{-1}l)} \quad \text{if } l < l_c/2 \quad \text{(low-pass filter)}$$

$$H^2 = 0 \quad \text{otherwise}$$

$$G^2 = 1 - H^2 \quad \text{(high-pass filter)}$$

where ϕ is an axis-symmetric scaling function of the positive real frequencies on the sphere and is given in harmonic space by:

$$\phi^2(l) = e^{-l^2}$$

The cut-off frequency is given by $l_c = 2d/\sqrt{\pi}$, where d is the initial basis dilation factor and the B (bandwidth parameter) frequency samples are taken uniformly within the interval $[0, l_c]$.

Simulating changes in cortical thickness

Performance of cluster size methods depends on parameters such as smoothness and threshold (Hayasaka and Nichols, 2003) whose optimal values are determined, in turn, by the size and strength of the signal. Accordingly, both the spatial extent and the strength of thickness reduction were modified in parallel in the present study.

In particular, cortical thickness of different regions was artificially reduced in half of the subjects (experimental group) once individual cortical thickness maps were resampled to the average surface. Regions of change were selected by comparing the regional mean cortical thickness of healthy elderly subjects with AD patients, after automatic segmentation with Freesurfer. For this analysis, the unsmoothed thickness maps in subject space were used and multiple comparisons over the whole brain were corrected with FDR (q-value = 0.01). Significant areas of change were grouped into more compact regions labeled as frontal, cingulate, temporal, parietal and occipital cortices (see Table 1).

As illustrated in Fig. 1, synthetic changes were either restricted to a single cluster within each cortical region (*concentrated simulation*) or dispersed in several clusters within each of the above-mentioned regions (*distributed simulation*). The number of vertices modified in each cortical region as well as the extent of change and the mean percentage range of thickness reduction are detailed in Table 2.

Determining the optimal smoothing and optimal threshold

Sensitivity and specificity for a given cortical thickness analysis can be defined in more detail as follows:

$$sensitivity = \frac{V_P - V_{0P}}{v - v_0} \quad \text{and} \quad specificity = 1 - \frac{V_{0P}}{v_0}$$

where the random numbers V_{0P} and V_P refer to the number of unobserved incorrect rejections of the null hypothesis (false positives) and the number of observed rejections respectively. Additionally, v_0 is the unknown number of true null hypotheses and v is the total number of tests.

Table 1
Mean variations in regional cortical thickness when comparing unsmoothed data from healthy elderly subjects to AD patients.

Region of change	Thinning (mm)			
	Mean (SEM)	Atrophy (%)	*t*-test	*P*
Left frontal	.08 (.02)	3.4	3.69	.001
Lateral orbitofrontal gyrus	.08 (.03)	3.5	2.68	.010
Pars opercularis	.08 (.03)	3.5	3.03	.005
Pars triangularis	.10 (.03)	4.4	3.45	.001
Middle frontal gyrus (caudal)	.10 (.03)	4.3	3.34	.001
Superior frontal gyrus	.07 (.03)	2.7	2.73	.010
Precentral gyrus	.09 (.03)	3.7	3.03	.005
Left cingulate (isthmus)	.13 (.04)	5.4	3.47	.001
Left temporal	.12 (.03)	4.6	4.21	.0001
Entorhinal cortex	.23 (.08)	7.7	2.90	.005
Superior temporal sulcus	.10 (.03)	4.5	3.30	.001
Transverse temporal gyrus	.13 (.04)	5.8	2.97	.005
Middle temporal	.11 (.03)	4.2	3.48	.001
Inferior temporal gryus	.11 (.04)	4.0	2.90	.005
Left parietal	.09 (.02)	4.3	4.22	.0001
Postcentral gyrus	.07 (.02)	3.7	3.42	.001
Precuneus	.09 (.03)	4.0	3.27	.001
Supramarginal gyrus	.12 (.03)	5.0	4.57	.0001
Superior parietal lobule	.08 (.03)	3.9	2.94	.005
Inferior parietal lobule	.11 (.03)	4.6	4.00	.0001
Left occipital (cuneus)	.08 (.02)	4.4	3.45	.001

Note. SEM = standard error of the mean.

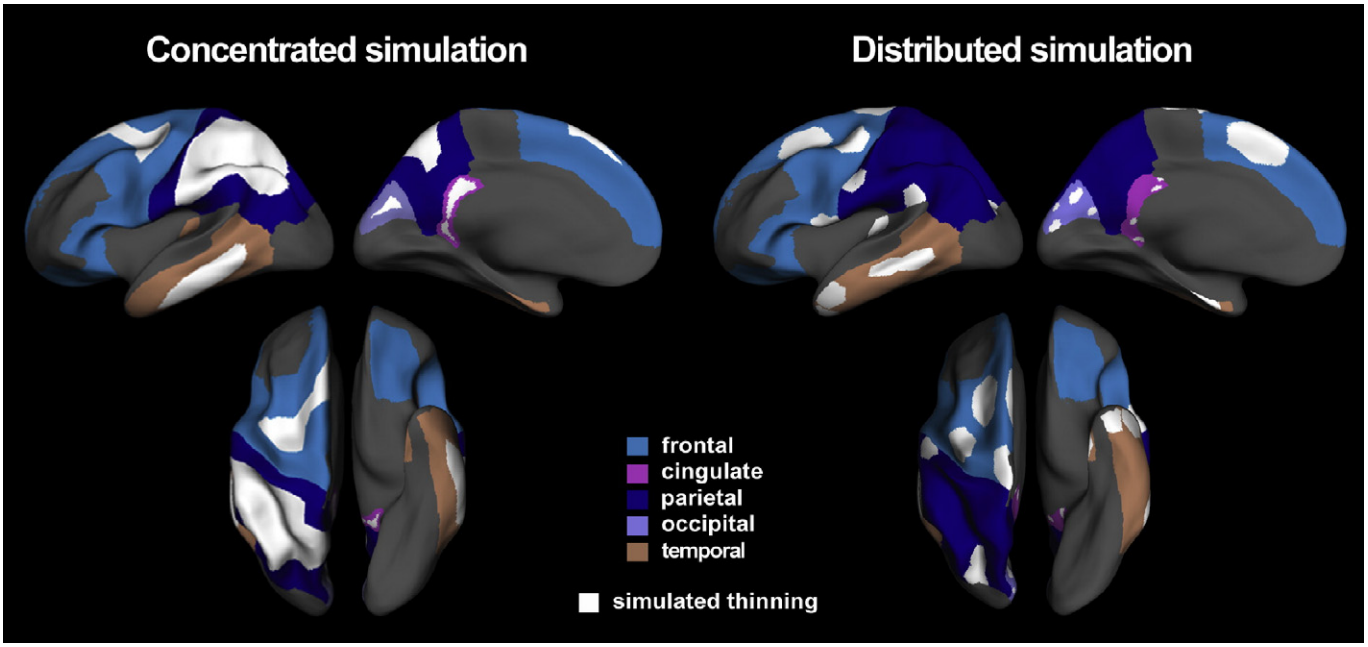

Fig. 1. Cortical regions used for simulating thinning in the experimental group. The main difference between concentrated and distributed simulations was the compactness of changes introduced within each cortical region: changes in the concentrated simulation were restricted to a single cluster whereas changes in the distributed simulation were spread in different independent clusters. Clusters of simulated thinning appear white filled in both simulations. Further information about the spatial extent of change and percent range of thinning introduced in each cortical region in both concentrated and distributed simulations is recollected in Table 2.

Let

$$t = (V_P - V_{0P})\left(1 - \frac{V_{0P}}{v_0}\right)$$

be the trade-off between the number of true discoveries and specificity, and denote by $FDP = \frac{V_{0P}}{V_P}$ the random false discovery proportion. If we assume that $V_P > 0$ (i.e., some thinning is detected in the statistical map) then t can be expressed as

$$t = V_P(1 - FDP)\left(1 - \frac{V_P FDP}{v_0}\right) \quad \text{(I)}$$

We define the filtering level α as either being an element of the positive integers Z^* or an element of $Z^* \times Z^*$. In the former case, it indicates the FWHM of a Gaussian kernel approximation for the linear filtering, whereas in the latter it refers to the pair (d, n_s) applied in the non-linear filtering. Then t is considered as a function of both α and the cluster-forming threshold u_c applied in cluster size inference analyses (Hayasaka and Nichols, 2003):

$$t = f(\alpha, u_c)$$

Table 2
Descriptive parameters of regional changes introduced in the simulation study.

Region of change	#Modified vertices	Change size mm^2 (%)	Mean thinning (mm) and range of thinning (%)
Concentrated			
Left frontal	3540	1753.3 (10)	0.34 (12–15)
Left cingulate	714	286.9 (40)	0.35 (12–15)
Left temporal	2435	1297.9 (20)	0.39 (13–16)
Left parietal	19529	6828.9 (40)	0.35 (14–17)
Left occipital	203	114.8 (10)	0.30 (14–17)
Distributed			
Left frontal	7511	4136.9 (30)	0.42 (15–18)
Left cingulate	192	75.1 (10)	0.39 (17–20)
Left temporal	3885	2465.2 (40)	0.46 (16–19)
Left parietal	3379	1927.5 (10)	0.37 (15–18)
Left occipital	362	329.2 (30)	0.35 (17–20)

In the case of vertex-wise analyses, t is considered as only being a function of α.

In general, the optimal smoothing level α_{opt} was operationally defined as that ensuing from the highest t across filtering levels and cluster-forming thresholds used in the above simulations:

$$\left(\alpha_{opt}, u_{copt}\right) = \arg\max(t)$$

When synthetic changes were considered unknown (same scenario as in real data analysis), we applied an estimate of t by estimating FDP and v_0 in Eq. (I) (see Eq. (II) below). Computation of t leads to the same optima (α_{opt}, u_{opt}) as sensitivity by specificity while avoids introducing excessive errors on the estimate of t due to errors in the estimation of $v - v_0$.

Statistical inference methods

The univariate general linear model (one-sided two-sample Student's T statistic) was applied at each cortical vertex to generate statistical parametric maps of cortical thickness. Here we propose a hierarchical methodology to estimate the optimal level of smoothing associated with the best trade-off given by Eq. (I). Following the original scheme of Benjamini and Heller (2007), our approach protects against false positives by controlling the error at the cluster level before proceeding to vertex-wise FDR. However, rather than controlling the FDR on clusters as Benjamini and Heller did, we controlled the family-wise error (FWE) by applying either random field theory (RFT) or permutation testing (PT) over the smoothed statistical maps. Next, we apply an adaptive multiple testing procedure over the unsmoothed statistical maps to control false detections at the vertex level (vertex-wise FDR) within the clusters. This approach ensures localizing power at least at the cluster level at the same time as reducing the probability of false positive clusters, which may increase global vertex-wise FDR in the second step of the hierarchical method. Results of this procedure were further contrasted with cluster- and vertex-wise parametric and nonparametric traditional testing and with vertex-wise FDR separately, in simulated and real datasets.

Cluster- and vertex-based inference by applying RFT- and PT-based methods

Cluster size inference techniques are among the most widely used statistical approaches to detect changes in functional neuroimaging, while their use is rare in anatomical studies. These methods test whether or not the spatial extent of a given cluster, which is obtained after thresholding the statistical map with an arbitrary cluster-forming threshold, is unusually large by chance alone. These tests show higher sensitivity than vertex-wise thresholding methods, particularly when changes are spatially extended (Friston et al., 1996). Unlike RFT-based methods that may require a high level of smoothing and a high statistical threshold to perform well (Friston et al., 1994), PT only requires weak distributional assumptions (Nichols and Holmes, 2001; Holmes et al., 1996) and provides a good estimation for any setting including low cluster-forming thresholds and low smoothed maps (Hayasaka and Nichols, 2003).

We approximated the true null distribution of the cluster size by applying parametric RFT as implemented in the SurfStat toolbox (http://www.math.mcgill.ca/keith/surfstat/), and nonparametric PT as implemented in Freesurfer release 4.0.5 (http://surfer.nmr.mgh.harvard.edu/). Three cluster-forming significance thresholds were investigated corresponding to p-values of 0.05, 0.01 and 0.005. As RFT methods require a high statistical threshold, only the two latter were used. Cluster- and vertex-wise corrected p-values were obtained based on the distributions of the largest cluster size and the vertex maximum significance respectively (10,000 permutations were computed for PT as recommended in http://www.fmrib.ox.ac.uk/fsl/randomise/index.html). The FWE rate was set at 0.05 in all cases.

Vertex-based thresholding by FDR

Vertex-wise FDR procedures have gained in popularity in the last few years due to their properties to provide sensitive statistical thresholds for determining changes in cortical thickness maps (Singh et al., 2006; Lerch et al., 2005). Instead of controlling the FWE, these methods focus on a less stringent measure of the Type I Error. FDR is the mathematical expectation of the false discovery proportion:

$$FDR = E(FDP) = E\left(\frac{V_{0P}}{V_P}\right)$$

The procedure of Benjamini and Hochberg (1995) provides a conservative control of FDR at the level q

$$FDR \le \frac{v_0}{v} q \le q.$$

The first inequality becomes equality when the distributions of the statistical tests are continuous. In the analysis of the entire cortex, the proportion $\frac{v_0}{v}$ is assumed to be close to 1 and reasonably ignored. Here v_0 is estimated by using an adaptive two-stage procedure (Benjamini et al., 2006) which allows for tight control of FDR within small clusters of contiguous vertices.

FDR analysis can lead to highly tolerant statistical thresholds even for low smoothed maps if the signal vertices constitute a large part of the cluster. This feature could be exploited to improve sensitivity by subdividing the global analysis into several smaller regional analyses where a significant number of vertices contain signal. Another good property of the FDR analysis relevant to our study is that besides controlling the expectation of the FDP it gives bounds for the variance:

$$\mathrm{Var}(FDP) = E\left(FDP^2\right) - FDR^2$$

As $0 \le FDP \le 1$ then

$$E\left(FDP^2\right) \le E(FDP) \quad \text{and} \quad \mathrm{Var}(FDP) \le FDR - FDR^2.$$

Therefore, if the q-value for the FDR is small enough the distribution of the FDP is sharp around its mean which will allow us to estimate its value at a given realization by its expected value. In the present study, the FDR was controlled at a 0.05 level, meaning that on average only 5% of the significant vertices are false positives and that the variance of this estimate is below 0.0475.

The FDR procedure suffers from a fundamental drawback. It assumes that signal only exists in a discrete subset of vertices within a cluster and that thickness at each vertex within the cluster either changes or not. When smoothing is applied the signal spreads to everywhere within the cluster becoming continuous rather than discrete (Chumbley and Friston, 2009). In addition, smoothing introduces change into vertices that have not truly changed but are neighbors of vertices showing true thinning. The resulting increase in false discoveries cannot be controlled by vertex-based FDR analysis. Therefore, conventional vertex-wise FDR procedures should not be used after spatial smoothing because it drastically increases the true FDP above the expected value, as will be shown in simulation analysis.

Hierarchical cluster- and vertex-based statistical thresholding (HT)

Estimation of the optimal smoothing in real data relies on estimation of Eq. (I). Parameters unknown in this equation are the global FDP and the global number of true null vertices v_0. FDR could be used as an estimator of FDP. But the FDR value is also unknown because, as mentioned above, traditional thresholding methods do not provide any control for this error after smoothing which can lead to disproportionate error rates (Logan et al., 2008). Alternatively, some regions could be initially delimited by applying cluster-based inference methods to the smoothed maps and then used for regional FDR analyses on the unsmoothed map. However, a minimal spatial smoothing is required to reduce errors derived from registration and cortical thickness measurements. This error source impedes applying FDR directly over unsmoothed statistical maps. To overcome this limitation, raw individual thickness maps were denoised (as a preprocessing step) by using sharp wavelets with a dilation factor of 3 and only thresholding the last finest scale of the decomposition. The high resolution intrinsic to the nonlinear spherical registration algorithm and the submillimeter precision of the thickness maps justify the low smoothing level (Fischl and Dale, 2000; Fischl et al., 1999b).

The global FDP of Eq. (I) can be estimated by the global FDR because the HT methodology (schematically illustrated in Fig. 2) significantly reduces the number of false positive vertices resulting from spatial smoothing. The global FDR (FDR_{global}), in turn, is approximated by the regional FDR (FDR_{regional}):

$$\hat{FDP} = FDR_{\mathrm{global}} \approx FDR_{\mathrm{regional}} \approx q$$

where FDR_{regional} is the vertex-wise FDR value controlled within the clusters at the level q by the adaptive FDR procedure (Benjamini et al., 2006).

FDR_{global} is expected to be slightly more liberal when compared with FDR_{regional}. However, the contribution of this error source is negligible due to (i) the strong control that FWE impose for false positive clusters, and (ii) the weak FWE control that FDR procedures impose for false positive vertices when the null hypothesis is true for a cluster erroneously detected (Nichols, 2007).

To estimate v_0, let n be the number of clusters detected by cluster size inference and m_i the number of vertices in the cluster i, $i = 1 \ldots n$. We estimate the number of null vertices $\left\{\hat{v}_{01}, \hat{v}_{02}, \ldots, \hat{v}_{0n}\right\}$ for each cluster in the first stage of the adaptive procedure applied to the unsmoothed statistical map inside each cluster (Benjamini et al., 2006). Then, the global v_0 can be estimated as:

$$\hat{v}_0 = v - \sum_{i=1}^{n} m_i - \hat{v}_{0i}.$$

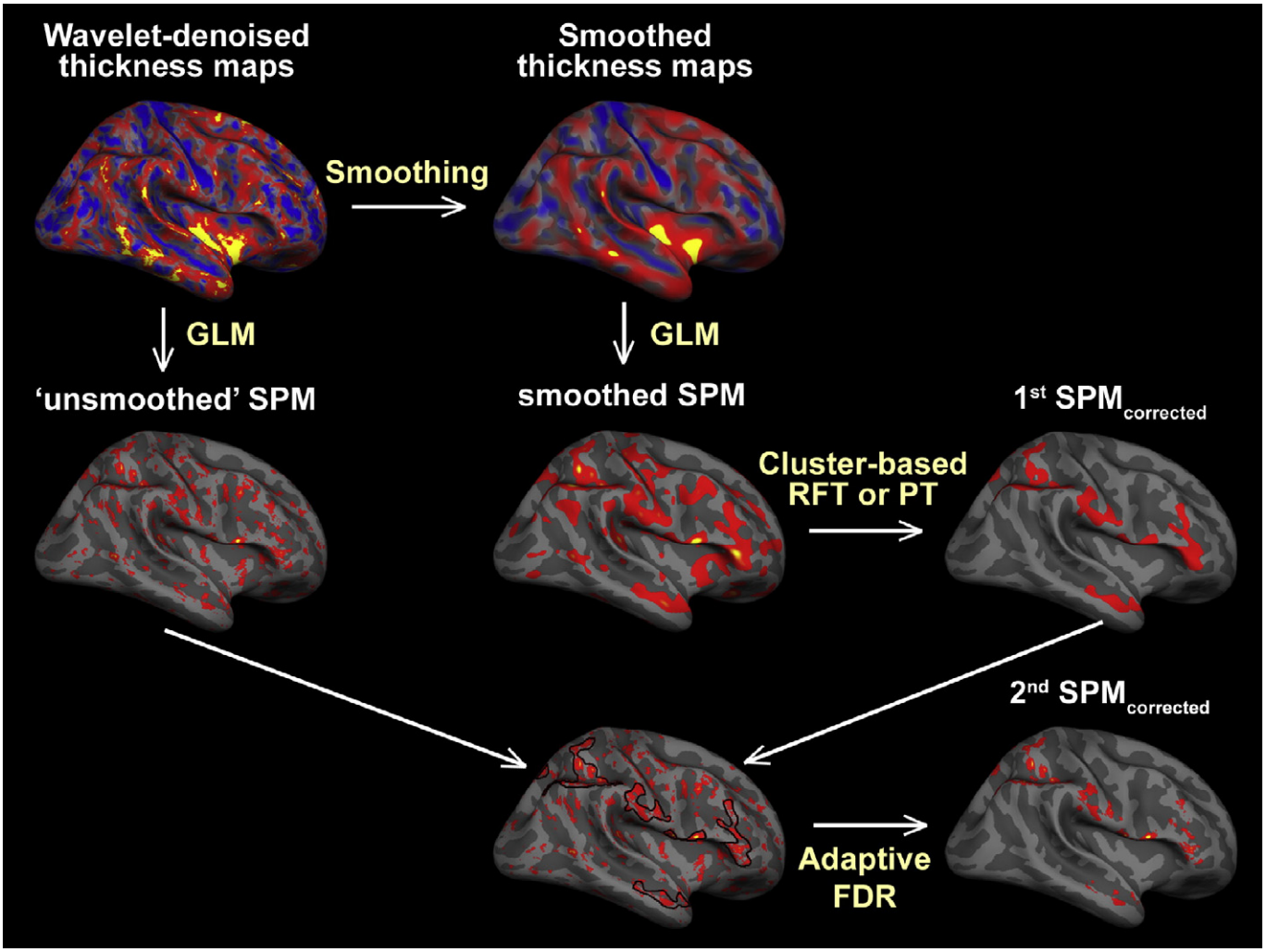

Fig. 2. Flow chart illustrating a schematic representation of the sequential hierarchical thresholding (HT) procedure. Raw individual thickness maps are firstly denoised using sharp spherical wavelets and subsequently smoothed with either Gaussian or spherical wavelet-based filters resulting in individual "unsmoothed" and smoothed thickness maps, respectively. Next, cluster size thresholding is applied to smoothed maps to determine cluster-like regions of interest of significant variation. Finally these clusters are transferred to the "unsmoothed" statistical map and adaptive false discovery rate (FDR) procedure is applied at each cluster separately to control the vertex-wise false detection.

Eq. (I) can now be estimated as:

$$\hat{t} = V_P(1-q)\left(1-\frac{V_P q}{\hat{v}_0}\right). \quad \text{(II)}$$

The optimal smoothing level can be estimated now by the following equation:

$$\left(\hat{\alpha}_{\text{opt}}, \hat{u}_{\text{copt}}\right) = \arg\max\left(\hat{t}\right) \quad \text{(III)}$$

Note that errors in the estimation of v_0 have a minimal impact on Eq. (II) because $V_P q \ll \hat{v}_0$ and the factor $\left(1-\frac{V_P q}{\hat{v}_0}\right) \approx 1$. In fact, $\hat{v}_0$ can be substituted by v (the total number of vertices) as in the FDR procedure of Benjamini and Hochberg (1995).

The HT methodology applied to cortical thinning detection has the power to decrease false positive discoveries resulting from spatial smoothing without a huge loss of sensitivity, as typically occurs with other approaches (e.g., Lerch and Evans, 2005). Even though vertex-wise thinning detection in the HT approach is always performed on slightly smoothed maps, sensitivity is comparable to that achieved with either cluster-based inference or traditional vertex-wise FDR obtained from smoothed statistical maps, as will be demonstrated with simulated data. The high sensitivity obtained with the HT methodology is caused by different reasons: i) because the multiple comparison problem that FDR procedures face is much less severe within the cluster; ii) because the proportion of vertices exhibiting thinning within the clusters is high and the chance for false positives low; iii) because a larger absolute value of false positives can be tolerated when detections within the clusters increase (Langers et al., 2007; Nichols, 2007); and iv) because the application of a sequential two-stage adaptive FDR procedure provides a tighter control of the FDR (Benjamini et al., 2006).

An important property of the HT methodology is that under the null hypothesis, the probability of finding one or more erroneous clusters is much less than the FWE of the cluster-based inference alone; whereas under the alternative hypothesis, the strong control of the FWE provides localizing power at least at the level of cluster. If cluster and vertex-wise statistics are independent and/or if independent data is used for cluster and FDR inference, then the FWE revealed by the HT will be the product of the FWE at the two levels. In these cases, the FDR procedure within the clusters will yield a FDR_{regional} under the alternative hypothesis extremely close to the pre-imposed q-value (Benjamini and Heller, 2007) which will improve t estimation by Eq. (II). Although results extracted from the two inference steps of the HT procedure are based on different features – spatial extent of

the cluster in the smoothed data and height of the unsmoothed vertex *T* statistic – they may covary. However, a good estimation of α_{opt} is still expected if the two statistics shows low correlations. The precision of this estimation was tested with the concentrated and distributed simulations described above, while the null simulations explained in the next section were used to estimate the actual FWE obtained with the HT methodology.

In the following, we will refer to the HT methodology as RFT_{HT} or PT_{HT} depending on whether the cluster-based inference is achieved by means of RFT- or PT-based methods, respectively.

Null simulations

Determining the optimal level of smoothing based on the statistical results requires testing a range of kernels, while testing several cluster-forming thresholds is optional. In any case, the problem of multiple comparisons may inflate the false positive rate. We tested these hypotheses with null simulations. For each simulation, we generated 50 white Gaussian noise spherical maps (6th order subdivided icosahedron, 40,962 vertices), and applied a two-sample *t*-test with equal sample sizes after smoothing the maps. Next, the RFT_{HT} method was performed over the resulting statistical map. To test the severity of the multiple comparison problem mentioned above, we first tuned the number of smoothing kernels (1 vs. 20 kernels from 1 to 20 mm) for one cluster-forming threshold (0.01), and next tested the effect of increasing the number of cluster-forming thresholds (2 or 3) while fixing the number of smoothing kernels (20). For each value of these parameters (number of smoothing kernels and number of cluster-forming thresholds), a total of 10,000 independent simulations were generated to compute the FWE.

Additionally, experiments with a single cluster-forming threshold (0.01) and a single level of smoothing (repeated separately for smoothing kernels from 2 to 12 mm, in steps of 2 mm) allowed us to test whether the FWE for the HT methodology was not far above 0.0025 as predicted if sequential test statistics were independent.

Results

Assessing performance of different statistical approaches on simulated cortical thinning

Before adding synthetic changes to the cortical maps, thickness was compared between groups to ensure that no bias was accidentally introduced. As expected, no group differences in thickness were found for any thresholding method after applying either Gaussian or wavelet-based smoothing.

Performance of all statistical approaches at detecting group differences in simulated thinning in the whole brain varied with the level of smoothing. Fig. 3 illustrates how the trade-off between sensitivity and specificity increases up to a kernel width and then declines. This behavior is particularly evident in the concentrated simulation (Fig. 3, left panel). When comparing the different statistical methods, inference obtained with vertex-wise techniques was the less sensitive approach while the HT method yielded the best trade-off.

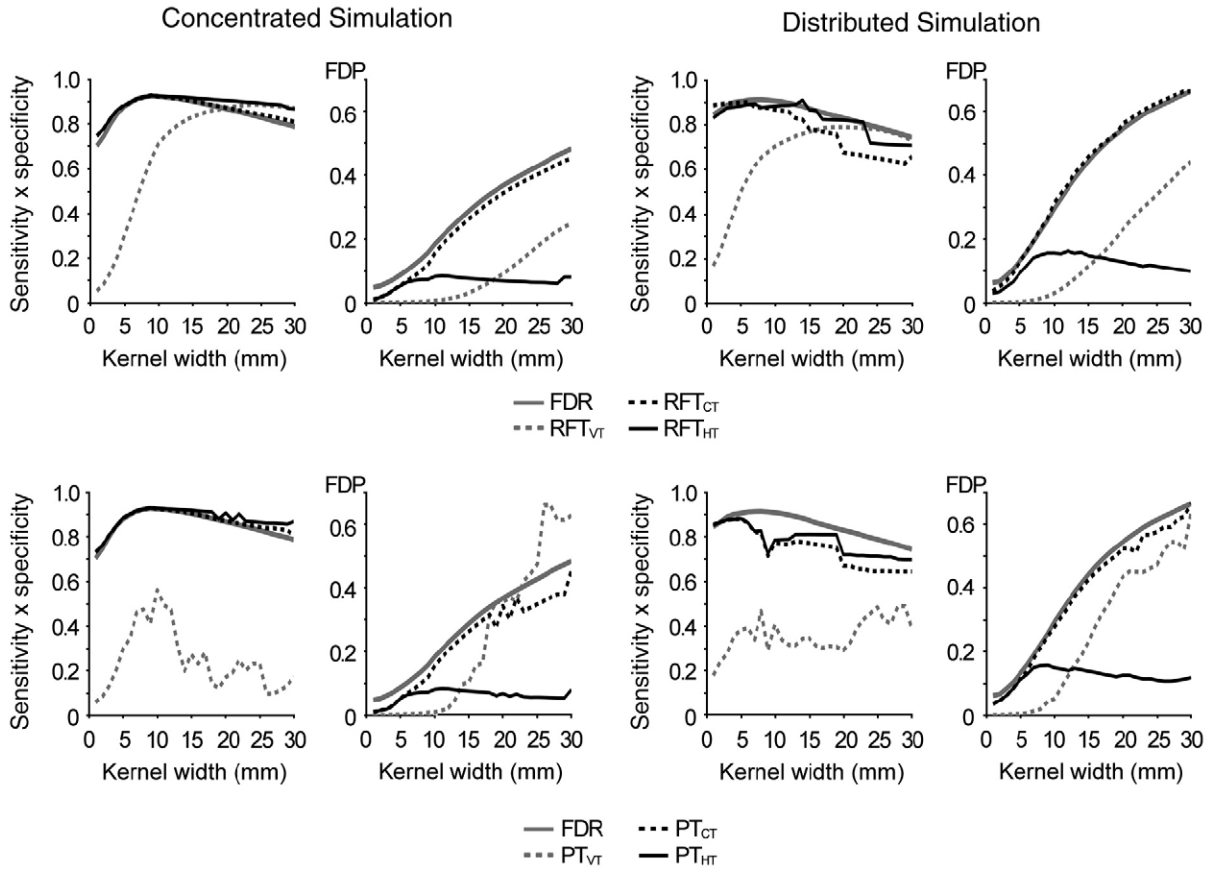

Fig. 3. Concentrated (left panel) and distributed (right panel) simulations of cortical thinning. Testing (i) trade-off between sensitivity and specificity and (ii) false discovery proportion (FDP) against a range of Gaussian smoothing widths (from 1 to 40 mm, only displayed until 30 mm) with different statistical methods: false discovery rate (FDR), cluster thresholding- (CT) and vertex-wise thresholding (VT) methods, and the hierarchical thresholding (HT) procedure. Results obtained with FDR and random field theory- (RFT) based methods are shown in the top panel; whereas results revealed by FDR and permutation testing- (PT) based methods are shown in the bottom panel.

Fig. 3 additionally shows the rapid increase of the FDP as smoothing gets wider. This effect is observed for all inference methods except for the HT, and appears more remarkable in the distributed simulation. From a kernel width of 2-mm forward, FDP was above 0.05 for the FDR and cluster-wise methods in both simulations. When the Gaussian kernel approximation was set at 20 mm, a common setting used in cortical thickness analysis (Charil et al., 2007; Singh et al., 2006; Lerch et al., 2005; Chung et al., 2003), FDP was above 0.3 in the concentrated simulation and above 0.5 in the distributed simulation. Interestingly, the HT approach kept the FDP close to its expected value (0.05) for any level of smoothing (below 0.09 in the concentrated simulation and below 0.17 in the distributed simulation). These results reveal that the HT methodology reaches the sensitivity of FDR and cluster-based procedures at the same time as keeping the FDP at a low level with independence from the Gaussian smoothing width.

Determining optimal smoothing and detection parameters on simulated cortical thinning

Optimal level of smoothing also varied across the different thresholding methods. Tables 3 and 4 show the t values and the FDP obtained with the optimal smoothing and optimal cluster-forming threshold (or imposed q-value for the FDR of 0.05) for each statistical method in both the concentrated and distributed simulations. In general, larger smoothing kernels were required for less sensitive methods (e.g., vertex-wise methods). This was true for Gaussian and wavelet-based spatial filters. More specifically, when the cluster-forming thresholding was similar in the two methods, for narrow optimal Gaussian kernel approximations the optimal wavelet filters showed smaller dilation factors (d) and fewer numbers of thresholded-finest decomposition levels (n_s) in contrast to relatively wide kernels. Interestingly, the FDP was closer to the expected value when wavelet-based smoothing was applied to distributed simulations of cortical thinning; whereas it was much higher in the distributed than in the concentrated simulation when using Gaussian filters. The latter result suggests that conventional statistical inference methods may increase their specificity after applying wavelet smoothing to real cortical thickness data.

The better performance of the HT methodology over the others was mainly evident when simulated group differences were assumed to be unknown. We refer to this analysis as "realistic data" in Tables 3 and 4. The kernel width was determined on the basis of the maximum value for $\hat{t}$ (Eq. II). The HT methodology was the only procedure that always yielded FDPs equal to or less than 0.16. Therefore, the higher sensitivity often shown by FDR and cluster-based methods when compared to the HT approach is only apparent, because more than ~45% of detected vertices were false positives. Also note that the HT approach is the only method for which $\hat{t}$ reasonably approximates t and for which $\hat{\alpha}_{opt}$ is close to its true optimal value. This issue is critical to approximate optimization of the Gaussian kernel and detection parameters in real data analyses where the true changes are *a priori* unknown.

Tables 5 and 6 include the number and proportion of true positive vertices detected by each statistical procedure at the optimal kernel for regions where concentrated and distributed thinning was simulated, respectively. Both concentrated and distributed synthetic changes introduced within the occipital region (cuneus) were equally neglected by all methods except for the FDR procedure; whereas thinning of the cingulate region (isthmus) was especially neglected in the distributed simulation. The high sensitivity of FDR was however

Table 3
Optimal smoothing in concentrated simulations based on whole brain analyses.

Smoothing procedure								
Statistical thresholding	Concentrated simulation				Realistic data			
Gaussian	$p_{cluster}$	FWHM	t	FDP	$p_{cluster}$	FWHM	$\hat{t}$	FDP
FDR	_	9	24871.8	.16	_	40	58038.4	.58
RFT_{CT}	.01	9	24939.4	.12	.01	39	51890.5	.53
RFT_{VT}	_	24	23823.5	.16	_	40	36307.3	.35
RFT_{HT}	.01	9	24899.8	.07	.01	11	25969.6	.08
PT_{CT}	.01	9	24938.8	.12	.01	40	52666.8	.53
PT_{VT}	_	10	15092.7	.01	_	40	21416.2	.63
PT_{HT}	.05	6	24917.3	.06	.05	5	26847.5	.11
Wavelet	$p_{cluster}$	d , n_s	t	FDP	$p_{cluster}$	d , n_s	$\hat{t}$	FDP
FDR	_	36 , 1	24888.1	.15	_	36 , 3	56266.8	.56
RFT_{CT}	.01	16 , 2	24957.8	.11	.01	28 , 3	45443.2	.46
RFT_{VT}	_	94 , 1	23863.7	.17	_	36 , 3	31605.2	.28
RFT_{HT}	.01	41 , 1	24908.1	.10	.01	19 , 2	25921.3	.08
PT_{CT}	.01	8 , 3	24961.5	.12	.01	36 , 3	51579.8	.61
PT_{VT}	_	18 , 2	16794.4	.01	_	34 , 3	20478.2	.51
PT_{HT}	.05	24 , 1	24923.9	.10	.05	9 , 2	26924.6	.11

Note. $p_{cluster}$ = cluster-forming p-value; t = true positives × specificity; FDP = false discovery proportion; $\hat{t}$ = estimate of t; d = dilation factor; n_s = number of thresholded finest scales. FWHM = full width at half maximum for the Gaussian kernel approximation.

Table 4
Optimal smoothing in distributed simulations based on whole brain analyses.

Smoothing procedure								
Statistical thresholding	Distributed simulation				Realistic data			
Gaussian	$p_{cluster}$	FWHM	t	FDP	$p_{cluster}$	FWHM	$\hat{t}$	FDP
FDR	_	8	14374.9	.22	_	40	51984.3	.74
RFT_{CT}	.01	6	14199.8	.16	.01	40	43831.1	.73
RFT_{VT}	_	19	12427.8	.20	_	40	25579.7	.57
RFT_{HT}	.005	14	14341.2	.16	.005	14	16421.8	.16
PT_{CT}	.01	5	13875.8	.11	.05	40	64990.9	.80
PT_{VT}	_	35	8655.8	.59	_	40	22653.8	.66
PT_{HT}	.01	5	13878.7	.11	.005	7	15179.7	.13
Wavelet	$p_{cluster}$	d , n_s	t	FDP	$p_{cluster}$	d , n_s	$\hat{t}$	FDP
FDR	_	30 , 1	14366.5	.16	_	15 , 3	27720.4	.49
RFT_{CT}	.01	3 , 1	14275.5	.12	.01	15 , 3	24964.4	.49
RFT_{VT}	_	79 , 1	12440.6	.16	_	30 , 2	13610.3	.14
RFT_{HT}	.005	59 , 1	14344.5	.10	.01	8 , 3	16409.6	.18
PT_{CT}	.005	19 , 1	13895.9	.16	.05	15 , 3	21676.6	.61
PT_{VT}	_	23 , 1	7974.3	.01	_	12 , 3	8225.0	.30
PT_{HT}	.005	19 , 1	13896.1	.10	.01	27 , 1	15289.5	.14

Note. $p_{cluster}$ = cluster-forming p-value; t = true positives × specificity; FDP = false discovery proportion; $\hat{t}$ = estimate of t; d = dilation factor; n_s = number of thresholded finest scales. FWHM = full width at half maximum for the Gaussian kernel approximation.

Table 5
Detection results by different statistical thresholding methods for each region where concentrated changes were simulated (whole brain analysis).

Statistical thresholding (FWHM)	Regions where concentrated changes were simulated				
	Frontal	Cingulate	Temporal	Parietal	Occipital
FDR (9) TP no. (%)	3507 (99.1)	598 (83.5)	2222 (91.2)	19082 (97.7)	169 (83.2)
RFT_{CT} (9) TP no. (%)	3429 (96.9)	603 (84.4)	2225 (91.4)	19094 (97.8)	0
RFT_{VT} (24) TP no. (%)	2992 (84.5)	521 (73)	2077 (85.3)	18983 (97.2)	0
RFT_{HT} (9) TP no. (%)	3179 (89.8)	603 (84.4)	2162 (88.8)	19054 (97.6)	0
PT_{CT} (9) TP no. (%)	3429 (96.9)	603 (84.4)	2225 (91.4)	19094 (97.8)	0
PT_{VT} (10) TDV no. (%)	1623 (45.8)	256 (35.8)	976 (40.1)	12243 (62.7)	12 (5.9)
PT_{HT} (6) TP no. (%)	3185 (90)	631 (88.4)	2172 (89.2)	19214 (98.4)	0

Note. TP = true positive vertices.

Table 6
Detection results by different statistical thresholding methods for each region where distributed changes were simulated (whole brain analysis).

Statistical thresholding (FWHM)	Regions where distributed changes were simulated				
	Frontal	Cingulate	Temporal	Parietal	Occipital
FDR (8)					
TP no. (%)	7323 (97.5)	24 (12.5)	3601 (92.7)	3301 (97.7)	299 (82.6)
RFT_{CT} (6)					
TP no. (%)	7239 (96.4)	24 (12.5)	3609 (92.9)	3306 (97.8)	0
RFT_{VT} (19)					
TP no. (%)	7011 (93.3)	0	2871 (73.9)	2851 (84.4)	0
RFT_{HT} (14)					
TP no. (%)	7367 (98.1)	22 (11.5)	3665 (93.3)	3314 (98.1)	0
PT_{CT} (5)					
TP no. (%)	7197 (95.8)	0	3580 (92.1)	3287 (97.3)	0
PT_{VT} (35)					
TP no. (%)	5445 (72.5)	64 (33.3)	2376 (61.2)	1507 (44.6)	0
PT_{HT} (5)					
TP no. (%)	7197 (95.8)	0	3580 (92.1)	3287 (97.3)	0

Note. TP = true positive vertices.

Table 7
Region-based optimal smoothing of distributed simulation and realistic data in occipital and cingulate cortices.

Smoothing procedure						
Statistical thresholding	Occipital			Cingulate		
Distributed simulation						
Gaussian	$p_{cluster}$	FWHM	TP (%)	$p_{cluster}$	FWHM	TP (%)
FDR	_	2	85.2	_	1	97.1
RFT_{CT}	.01	5	70.8	.005	25	16.1
RFT_{VT}	_	9	78.0	_	26	14.4
RFT_{HT}	.01	7	76.4	.01	27	100
Wavelet	$p_{cluster}$	d , n_s	TP (%)	$p_{cluster}$	d , n_s	TP (%)
FDR	_	3 , 1	90.6	_	3 , 1	100
RFT_{CT}	.01	3 , 1	91.2	.005	3 , 1	100
RFT_{VT}	_	33 , 1	79.4	_	4 , 1	100
RFT_{HT}	.01	3 , 1	91.2	.005	3 , 1	100
Realistic data						
Gaussian	$p_{cluster}$	FWHM	TP (%)	$p_{cluster}$	FWHM	TP (%)
FDR	_	36	17.6	_	40	7.7
RFT_{CT}	.01	15	40.3	.01	40	8.7
RFT_{VT}	_	9	78.0	_	40	9.0
RFT_{HT}	.01	5	73.1	.01	27	100
Wavelet	$p_{cluster}$	d , n_s	TP (%)	$p_{cluster}$	d , n_s	TP (%)
FDR	_	28 , 2	30.1	_	3 , 1	100
RFT_{CT}	.01	14 , 3	32.4	.01	81 , 1	11.3
RFT_{VT}	_	81 , 1	29.8	_	4 , 1	100
RFT_{HT}	.01	22 , 1	72.3	.005	3 , 1	100

Note. $p_{cluster}$ = cluster-forming p-value; TP (%) = percent of true positive vertices; d = dilation factor; n_s = number of thresholded finest scales. FWHM = full width at half maximum for the Gaussian kernel approximation.

misleading if we consider the high proportion of false positives detected at the optimal kernel (between around 15% and 20%) when compared to the HT methodology (below 10%) (see Tables 3 and 4). All methods yielded a satisfactory level of global thinning detection within the remaining cortical regions in both simulations. However, of relevance is that the HT approach besides achieving similar sensitivity as the cluster-based methods, it reduced the probability that the difference between groups could have arisen by chance in the majority of cases (as derived from analyses over null simulations, see next section).

The poor performance of the cluster-based methods within occipital and cingulate regions may be accounted for by the difficulty of defining the initial cluster-forming threshold (e.g., Smith and Nichols, 2009; Hayasaka and Nichols, 2003), particularly when both spatially reduced and extended regions are simultaneously affected by significant thinning reduction. In agreement with this notion, region-based analyses increased thinning detection within these particular regions in the simulated data (Table 7). However, only HT yielded a good estimate of thinning detection in the two regions independently of whether linear or non-linear smoothing procedures were applied. Indeed, the HT approach found 100% of true positive vertices in the two regions after estimating the optimal wavelet smoothing and a lower (but still relatively high) proportion (around 70%) when estimating the optimal FWHM for the Gaussian kernel approximation.

Results of null simulations

Determining the optimal level of smoothing based on statistical results requires testing a broad range of kernels, which leads to a problem of multiple comparisons. The false positive rate may be inflated even more if the experimenter further decides to test several cluster-forming levels. Two results support this notion, i) RFT_{HT} applied over null simulations revealed an increase of the FWE when testing 20 smoothing kernels as compared with only one kernel (see Table 8); and ii) the FWE grew up with increasing the number of cluster-forming thresholds. Importantly, when the cluster-forming threshold was fixed, the false positive rate resulting from testing 20 smoothing kernels was maintained below 0.05 (expected FWE for the cluster-based inference alone when applied with a single cluster-forming threshold and a single level of smoothing). These findings suggest that although the method is valid for optimizing the smoothing level, one must be cautious when testing several cluster-forming thresholds.

Additionally, null simulations allowed us to determine how far is the actual FWE from 0.0025 which would be expected if cluster and vertex-wise tests statistics were independent. However, as indicated in Table 8, the FWE was above of this expected value for several smoothing kernels, suggesting some correlation between the two statistical tests. Indeed, the FWE was up to three times 0.0025 (for a kernel of 6 mm) indicating that a similar relation can happen under the alternative hypothesis between the actual FDP and its expected value of 0.05 (e.g., FDP = 0.16 at the optimal setting in the distributed simulation). This error source for the estimation of the optimal level of smoothing (by means of Eq. (III)) did not prevent in practice a good estimation of the optimal smoothing level α_{opt}, as shown in the analyses over simulated signals. Note that this estimation was not possible with the other inference methods because of their increasing FDP with increasing smoothing level, leading to very inflated values of FDP for high smoothing levels.

Validation of optimal smoothing and detection parameters with HT methodology in real population data

The HT model revealed that the best trade-off between sensitivity and specificity with Gaussian smoothing was obtained with kernels smaller than 7 mm for the two cortical hemispheres (see Table 9). Wavelet-based smoothing provided comparable results, but permutation tests required less smoothing than random field methods to

Table 8
Results of null simulations with FWE rate for the cluster size inference controlled at the 0.05 level and vertex-wise FDR controlled at a 0.05 q-value.

Cluster-forming p-value	Smoothing range (mm)	FWE
.01	2	.0057
	4	.0064
	6	.0076
	8	.0066
	10	.0062
	12	.0069
.01	1:20	.0422
.01; .005	1:20	.0586
.01; .005; .001	1:20	.1

Note. FWE = family-wise error for the RFT_{HT}; mm=millimeters.

Table 9
Estimate of global optimal smoothing by HT in real data.

Smoothing procedure					
Statistical thresholding	$p_{cluster}$	Left hemisphere		Right hemisphere	
Gaussian		FWHM	$\hat{t}$	FWHM	$\hat{t}$
RFT_{HT}	.01	6	9752.3	7	11391.0
PT_{HT}	.05	2	8539.4	3	10880.1
Wavelet		d , n_s	$\hat{t}$	d , n_s	$\hat{t}$
RFT_{HT}	.01	27 , 1	10155.1	30 , 1	11853.6
PT_{HT}	.05	7 , 1	8485.6	8 , 1	10885.8

Note. $p_{cluster}$ = cluster-forming p-value; $\hat{t}$ = estimate of t (t = true positives × specificity); d = dilation factor; n_s = number of thresholded finest scales. FWHM = full width at half maximum for the Gaussian kernel approximation.

obtain optimal results. The reason for this is that PT uses more liberal cluster-forming thresholds at the optimal setting ($p = 0.05$) as compared to RFT ($p = 0.01$).

These results were compared to those yielded by FDR and by cluster- and vertex-based inference methods with a Gaussian kernel approximation of 20 mm, because this setting has frequently been used in prior studies of cortical thickness (e.g., Charil et al., 2007; Singh et al., 2006; Lerch et al., 2005; Chung et al., 2003). In general, the extent of detection was much smaller when applying the HT methodology (7.6%) than either FDR (56.2%) or cluster-based random field test (40.6%), but higher as compared with vertex-wise inference methods (3.4%). Differences in thinning detection for each statistical thresholding method are shown in Fig. 4.

Results obtained with the HT methodology on real thickness data showed that bilateral entorhinal cortex was the region most extensively affected (thinning of 56.3% and 73.4% of the left and right entorhinal cortex, respectively), followed by parahippocampal cortex mainly in the left hemisphere (24.7%), the pars triangularis and pars opercularis of the inferior frontal gyrus mainly in the right side (22.6%), and the right superior parietal lobe (21.1%). Conventional methods pointed to other structures that were similarly or more affected than regions of the medial temporal lobe as the bilateral inferior frontal gyrus (pars triangularis), the left inferior parietal gyrus, and the left supramarginal gyrus. In all these cortical regions, RFT_{HT} produced approximately 75% less detection on average relative to FDR or cluster-based tests. Interestingly, in the left inferior temporal gyrus, where cluster-based tests failed to detect changes, the HT methodology showed significant thinning in 11.6% of the region. As revealed by simulation analyses, the reduced number of vertices where the HT localized significant thinning suggests that decreased estimation of sensitivity was compensated by keeping the FDP close to its expected value. But even more important is the fact that HT was the only procedure able to approximate an optimal level of smoothing to apply in cortical thickness analysis. This fact is illustrated in Fig. 5, where the performance of FDR, cluster- and vertex-wise methods is overlapped for the left and right cortical hemisphere to show how the kernel width and $\hat{t}$ increase in parallel; whereas HT yields the expected inverted-U shape relationship between the size of smoothing kernel (or dilation factor for the wavelets) and the performance of the method.

Discussion

The main objective of the present study was to determine the optimal level of smoothing able to provide the best trade-off between sensitivity and specificity at detecting significant variations in cortical thickness maps. To achieve this goal we propose a sequential hierarchical methodology combining cluster- and vertex-based thresholding methods. The performance of hierarchical thresholding (HT) was compared with other widely used statistical inference procedures in both simulated and real data from healthy elderly subjects and AD patients. In most of the situations, the HT methodology performed better when compared to either cluster- and vertex-based FWE or vertex-based FDR procedures. The enhanced performance was independent of the spatial smoothing algorithm (Gaussian or wavelet), and of whether correction for multiple comparisons were based on RFT or PT methods. HT might be used not only to optimize the extent of smoothing but also the initial cluster-forming threshold for the cluster-based inference. However, including several cluster-forming thresholds in the numerical optimization may severely increase the FWE rate as suggested by the null simulation results.

Results from concentrated and distributed simulations showed that application of conventional thresholding methods (FDR, RFT and PT) tends to overestimate the extent of thinning as smoothing increases. This fact is particularly evident when distributed changes of cortical thickness were reproduced, due to inflated rates of false discoveries. In agreement with these findings, prior studies have noted that conventional thresholding methods are not able to provide any control of voxel-wise error measure with increasing smoothing under the alternative hypothesis (Chumbley and Friston, 2009; Logan et al, 2008). They only provide reliable information about the local maxima of suprathreshold regions but not about individual vertices in the statistical parametric map. Therefore, interpretation in these cases should be based on the location of those maxima rather than on the spatial extent of the detected signal. However, experimental reports based on coordinates of local maxima in smoothed SPMs suffer from spatial imprecision due to shifting and/or merging of adjacent signal peaks or clusters (Reimold et al., 2006; Mikl et al., 2008). Our HT approach provides a solution to the spatial precision problem by limiting the proportion of vertex-wise false discoveries among all the discoveries inside the spatial support of a significant cluster. Controlling the vertex-wise error is also critical if we are pursuing the optimization of the trade-off between sensitivity and specificity, two measures directly related to the signal detection theory. This optimization may further increase the anatomical precision of the observed cortical thinning pattern. Inspired by the hierarchical thresholding scheme of Benjamini and Heller (2007), our approach firstly controls the error at the level of clusters and then at the level of vertices within significant clusters detected in the first step. Application of this procedure to simulated thinning showed that the true global FDP with optimal settings was always not far from the pre-imposed q-value for the $FDR_{regional}$, suggesting that the HT methodology may be effective to approximately optimize the above-mentioned criterion by means of Eq. (III) (see Materials and methods).

Smoothing in cortical thickness analysis is typically applied to enhance both the quality of the raw image and the sensitivity of statistical inference (Smith and Nichols, 2009; Han et al. 2006; Lerch and Evans, 2005). However, the HT method is only devoted to optimize smoothing for the latter purpose. In fact, the denoising smoothing level is fixed in the HT method as a preprocessing step to reduce measurement noise and registration variability after resampling the maps to the average surface. We refer to the statistical map resulting from this preprocessing step as the "unsmoothed statistical map" which dictates the resolution of our findings. Consequently, any null vertices becoming non-null at this step are considered as part of the true signal for further analyses. That first smoothing level introduced by the HT method is small because of the high resolution of the spherical registration algorithm and the submillimeter precision of the thickness maps. The additional smoothing required by the statistical inference has a twofold purpose. On the one hand, each vertex finds support from its neighbors which are part of the same signal cluster and, on the other, it seeks to reduce the effective number of comparisons (Smith and Nichols, 2009). Maximum sensitivity (detection of weak signals) usually needs larger smoothing levels on the statistical map which, in turn, not only introduces an uncontrolled number of false positives but also erroneously extends

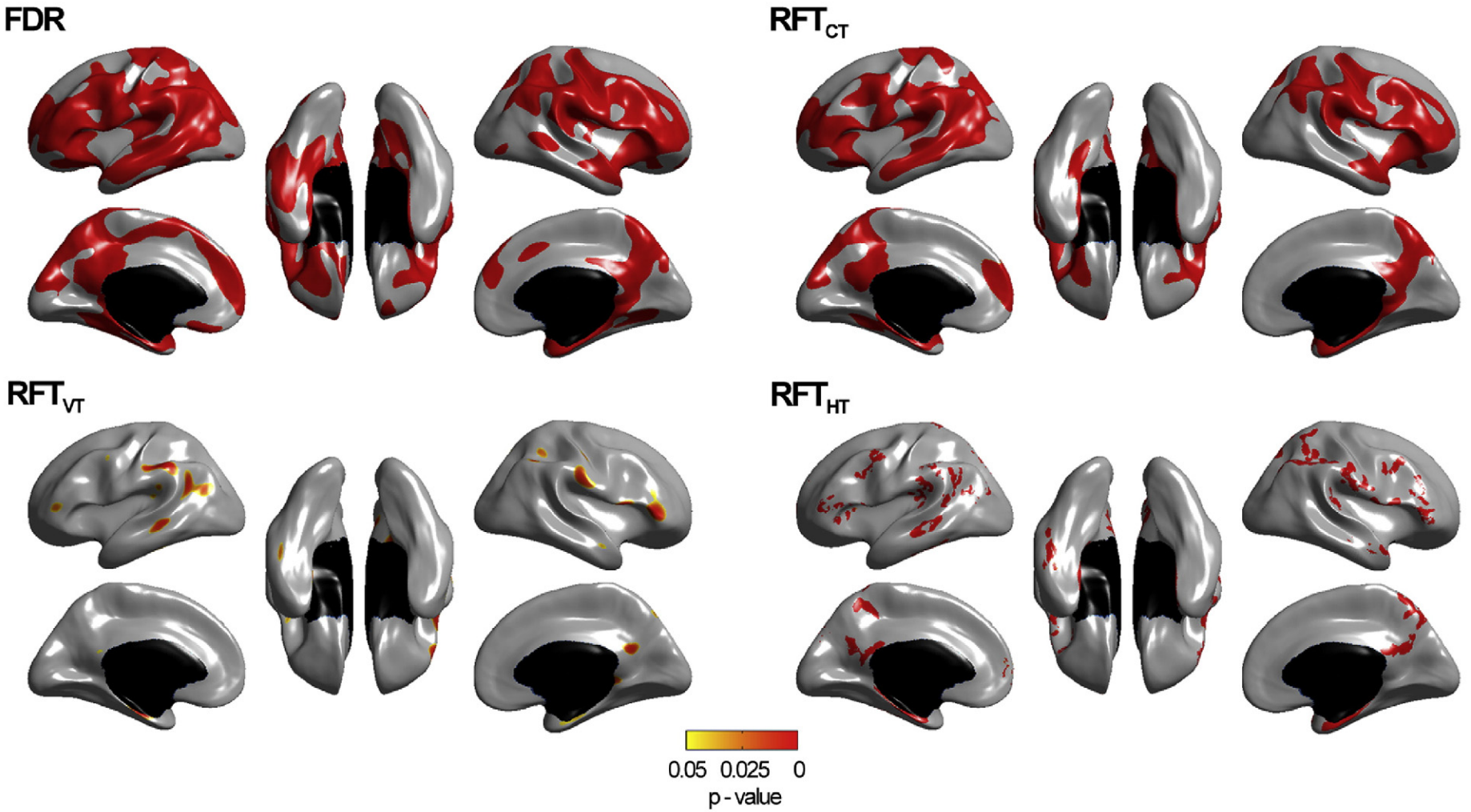

Fig. 4. *P*-value maps resulting from cortical thickness comparison between healthy elderly subjects ($N = 53$) and mild-to-moderate Alzheimer's patients ($N = 53$). Variation in thickness are displayed after applying false discovery rate (FDR) (left upper corner), cluster thresholding-based methods (right upper corner), and vertex-wise thresholding methods (left bottom corner) based on random field theory (RFT_{CT} and RFT_{VT}) with a kernel width of 20 mm. Results obtained with the hierarchical thresholding procedure based on RFT (RFT_{HT}) at the optimal wavelet parameters (left hemisphere: dilation factor of 27 at the finest scale, right hemisphere: dilation factor of 30 at the finest scale) and cluster-forming threshold ($p = 0.01$) are shown in the right bottom corner.

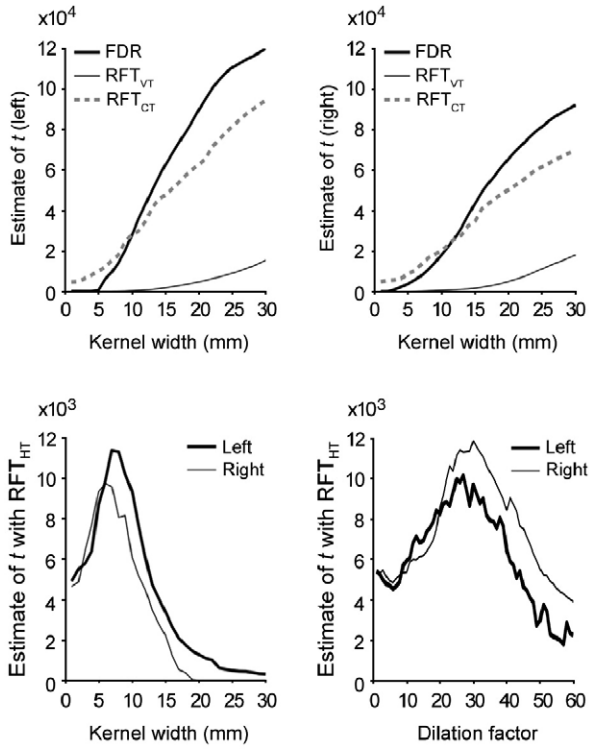

Fig. 5. Estimate of t (see Eq. (II) in Methods) provided by FDR, cluster- and vertex-wise inference methods based on random field theory (RFT) for the left and right hemisphere (top panel) and by the hierarchical thresholding (HT) procedure (bottom panel) across different widths of the Gaussian kernel approximation in the whole brain analysis of the real data set.

local thinning to other cortical regions. To counteract this effect, the HT methodology allocates clusters obtained over the smoothed statistical map into the "unsmoothed statistical map".

Our results show that although the final vertex-wise detection within the clusters was performed over the "unsmoothed statistical map", sensitivity was still high enough due to the better performance of FDR procedures applied to separated regions as well as to their lower conservativeness for low smoothed signals when compared with other approaches like e.g., RFT based on maximum vertex statistic (Langers et al, 2007; Genovese et al., 2002). It is important to note that the improved change detection is achieved without the cost of significant loss in image resolution. Sensitivity was further improved for those clusters with low strength signals by employing the recently developed two-stage adaptive FDR procedure (Benjamini et al., 2006), which mainly provides a tighter control of FDR (making the FDR value closer to the chosen q-value).

Results from the comparison between smoothing methods suggest that the non-linear spherical wavelet spatial filtering is a good alternative to surface-based Gaussian smoothing approximations even when using traditional statistical methods. Indeed, the former improved the optimization criterion for both whole brain and region-based statistical analyses. This result is mainly due to adaptive properties of spherical wavelets that increase sensitivity and decrease the blurring effect typically introduced by the Gaussian approximation which, in turn, improved the specificity of the cortical thinning detection, as was previously demonstrated in simulated data under several signal-to-noise ratios (Bernal-Rusiel et al., 2008). In agreement with this, as the bandwidth goes to infinity, the approximation to Gaussian kernel converges to the sample mean of data on a cortical manifold (Chung et al., 2005), thus eliminating any spatial variation in the individual cortical thickness map. The suppression of apparently undersized but relevant details in thickness maps constitutes an important motivation for using non-linear smoothing methods based on spherical wavelet transformations (Bernal-Rusiel et al., 2008; for an fMRI study using 3D-Euclidean wavelets see Wink and Roerdink (2004)). As smoothness of the resulting statistical map is estimated from the residuals of the general linear model (Worsley et al., 1999), RFT can also be applied after wavelet-based smoothing of the individual cortical thickness maps. In fact, RFT enhanced cortical thinning detection in our simulation experiments when combined with the wavelet-based filtering.

The enhanced performance of the wavelet-based approach over the Gaussian smoothing cannot be accounted for by the fact that artificial thinning was introduced in a discrete manner, by directly reducing cortical thickness in several clusters of contiguous vertices of the unsmoothed maps. On the contrary, iterative nearest neighbor smoothing (the approximation of Gaussian kernel smoothing on the surface proposed by Hagler et al., 2006 and Han et al., 2006), as an inherently discrete procedure, should benefit more from this kind of simulations than the wavelet-based filtering, which was implemented on harmonic space by using the fast Spherical Fourier Transform (Healy et al., 2003). In addition, our discrete simulations were specifically designed to keep the inter-subject variability of the real data at each vertex, becoming similar to what is typically observed in unsmoothed real data sets. The adequate design of these simulations is also supported by the fact that the performance of HT resulted in inverted-U curves with increasing smoothing when applied to real data, reproducing the pattern of results obtained in simulated data. Continuous simulations where cortical thinning is directly introduced in the T1 image (Lerch and Evans, 2005) may provide a more realistic view, but they preclude exact definition of the region of change.

Although wavelet filtering outperforms the Gaussian spatial smoothing in cortical thickness maps, it is still suboptimal and should be an issue of further improvement. This algorithm has been proved to be asymptotically optimal only for orthonormal (or biorthogonal) wavelet basis (Donoho, 1995; Donoho and Johnstone, 1994) whereas our representation is a redundant tight frame. Better results can be achieved by devising optimal wavelet coefficient thresholding techniques for the redundant spherical basis, as recently shown in a study on redundant Euclidean's space systems (Raphan and Simoncelli, 2008).

Results from comparing whole brain and region-based analyses confirm that the size of the search region as well as the extent and strength of the signal clusters is an important constraint influencing the level of smoothing required for optimizing sensitivity (see also Bernal-Rusiel et al., 2008). These parameters further influence the tuning of the cluster-forming threshold which, in turn, will affect the sensitivity of the hierarchical method in detecting reliable cortical thinning. The trade-off seems to be between enhanced detection of wide diffuse changes and localization of sharp focal changes (Smith and Nichols, 2009; Hayasaka and Nichols, 2003). As shown in the two simulations (performed at the whole brain level), no significant thinning was detected by the HT method in occipital and cingulate regions because clusters within these regions were extremely small relative to others located within the remaining regions. However, these changes in thickness were accurately identified by HT when region-based analysis was applied. But this choice only represents a good alternative to the whole brain analysis when the null hypothesis is false somewhere at each region of interest and therefore avoids an uncontrolled global number of false positives which may considerably inflate differences in cortical thickness.

Results from null simulations indicated some correlation between the sequential test statistics. In this case, FDR_{regional} is above the q-value because the null p-values within selected clusters may actually have a distribution that is stochastically smaller than the uniform (0,1). The p-values for the individual locations selected for FDR analysis at the second stage are not longer marginal ones but rather conditional on being in a significant cluster (Benjamini and Heller, 2007). Although this issue requires future research, it is important to note that it does not prevent,

in practice, a good estimation of the optimal smoothing level α_{opt}, as shown in the simulation analysis, which is not feasible with other inference methods. Importantly, the increase in the vertex-wise Type 1 error is limited to clusters whose FWE is strongly controlled by our method. This procedure avoids isolated false positive vertices in random regions (as in simple vertex-wise FDR analysis). In addition, HT has demonstrated to have higher power than conventional FDR on the unsmoothed map (even applying kernels as small as 5 mm) when the signal-to-noise ratio in the SPM is low (see Fig. 5).

The HT methodology is only applicable for signals whose spatial pattern consists of compact clusters in the statistical map. For example, it does not seem applicable to neither fMRI nor PET mainly because, as stated by Chumbley and Friston (2009), the statistical signal provided by these recording techniques is continuous (it extends over the whole brain) and all voxels are indeed non-null. In these cases, FDR is zero and it does not make sense to talk about active or inactive voxels. This is the reason why topological features (clusters and peaks) are tested with RFT or PT methods instead.

The pattern of cortical thinning in AD has been extensively described by using different computational neuroanatomy techniques (e.g., Frisoni et al., 2007; Lerch et al., 2005; Thompson et al., 2003; Baron et al., 2001). Some regions appear systematically damaged in AD patients in most of the cortical thickness studies (e.g., medial temporal structures, posterior cingulate gyrus, precuneus, and temporoparietal association cortex). On the contrary, some regions of change differ from one study to another (e.g., occipital lobe, and orbitofrontal cortex). The extent of cortical thinning and the severity of atrophy also seem to be highly variable across studies. The present study suggests that both the level of spatial smoothing and the statistical inference method used in each study might account for these apparently contradictory results. Thus, our findings revealed the artificial extent of cortical thinning pattern in AD patients yielded by FDR and by cluster- and vertex-based inference tests after applying a commonly used Gaussian kernel width of 20 mm when compared to the hierarchical method at the optimal wavelet parameters (see Fig. 4). These results suggest that cortical atrophy patterns reported in AD patients might be partially inflated by excessive smoothing and by the failure of inference methods to control the proportion of false discoveries. Further research is needed to confirm this hypothesis with different AD datasets and different computational neuroanatomy techniques to obtain cortical thickness measurements.

Acknowledgments

The authors want to thank the three anonymous reviewers of this paper for their helpful comments and insightful suggestions. We also wish to thank Dr. Randy Buckner (Harvard University), the Neuroinformatics Research Group (Washington University School of Medicine), and the Biomedical Informatics Research Network for making available the OASIS MRI dataset used in the present study. This research was supported by research grants from the Spanish Ministry of Science and Innovation (SAF2008-03300) and the Regional Ministry of Innovation, Science and Enterprise, Junta de Andalucía (CTS-4604) to JLC.

References

Ad-Dab'bagh, Y., Singh, V., Robbins, S., Lerch, J., Lyttelton, O., Fombonne, E., Evans, A.C., 2005. Native space cortical thickness measurement and the absence of correlation to cerebral volume. Proceedings of the 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto.

Apostolova, L.G., Thompson, P.M., 2008. Mapping progressive brain structural changes in early Alzheimer's disease and mild cognitive impairment. Neuropsychologia 46, 1597–1612.

Baron, J.C., Chételat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., Eustache, F., 2001. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. Neuroimage 14, 298–309.

Benjamini, Y., Heller, R., 2007. False discovery rates for spatial signals. J. Am. Stat. Assoc. 102, 1272–1281.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507.

Bernal-Rusiel, J.L., Atienza, A., Cantero, J.L., 2008. Detection of focal changes in human cortical thickness: spherical wavelets *versus* Gaussian smoothing. Neuroimage 41, 1278–1292.

Biega, T.J., Lonser, R.R., Butman, J.A., 2006. Differential cortical thickness across the central sulcus: a method for identifying the central sulcus in the presence of mass effect and vasogenic edema. Am. J. Neuroradiol. 27, 1450–1453.

Butman, J.A., Floeter, M.K., 2007. Decreased thickness of primary motor cortex in primary lateral sclerosis. Am. J. Neuroradiol. 28, 87–91.

Charil, A., Dagher, A., Lerch, J.P., Zijdenbos, A.P., Worsley, K.J., Evans, A.C., 2007. Focal cortical atrophy in multiple sclerosis: relation to lesion load and disability. Neuroimage 34, 509–517.

Chumbley, J.R., Friston, K.F., 2009. False discovery rate revisited: FDR and topological inference using random fields. Neuroimage 44, 62–70.

Chung, M.K., Worsley, K.J., Robbins, S., Paus, P., 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.

Chung, M.K., Dalton, K.M., Shen, L., Evans, A.C., Davidson, R.J., 2007. Weighted Fourier representation and its application to quantifying the amount of gray matter. IEEE Trans. Med. Imaging 26, 566–581.

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

Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L., 2009. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb. Cortex 19, 497–510.

Donoho, D.L., 1995. De-noising by soft-thresholding. IEEE Trans. Inf. Theory 41, 613–627.

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.

Driscoll, J.R., Healy, D.M., 1994. Computing Fourier transforms and convolutions on the 2-sphere. Adv. Appl. Math. 15, 202–250.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U.S.A. 97, 11050–11055.

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.H., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans. Med. Imaging 20, 70–80.

Frisoni, G.B., Pievani, M., Testa, C., Sabattoli, F., Bresciani, L., Bonetti, M., Beltramello, A., Hayashi, K.M., Toga, A.W., Thompson, P.M., 2007. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain 130, 720–730.

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, 214–220.

Friston, K.J., Holmes, A., Poline, J.-B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4, 223–235.

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, D.J., Saygin, A.P., Sereno, M.I., 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. Neuroimage 33, 1093–1103.

Han, X., Jovicich, J., Salat, D., Van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32, 180–194.

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

Healy, D., Rockmore, D., Kostelec, P., Moore, S., 2003. FFTs for the 2-sphere — improvements and variations. J. Fourier Anal. Appl. 9, 341–385.

Holmes, A.P., Blair, R.C., Watson, J.D.G., Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. J. Cereb. Blood F Met. 16, 7–22.

Jones, E.G., Peters, A., 1984. Cerebral cortex: Cellular Components of the Cerebral Cortex. Plenum Press, New York. Volume 1 (580 pp.).

Kuperberg, G.R., Broome, M.R., McGuire, P.K., David, A.S., Eddy, M., Ozawa, F., Goff, D., West, W.C., Williams, S.C., Van der Kouwe, A.J., Salat, D.H., Dale, A.M., Fischl, B., 2003. Regionally localized thinning of the cerebral cortex in schizophrenia. Arch. Gen. Psychiatry 60, 878–888.

Langers, D.R.M., Jansen, J.F.A., Backes, W.H., 2007. Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. Neuroimage 38, 43–56.

Lerch, J.P., Evans, A.C., 2005. Cortical thickness analysis examined through power analysis and a population simulation. Neuroimage 24, 163–173.

Lerch, J.P., Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J., Evans, A.C., 2005. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cereb. Cortex 15, 995–1001.

Logan, B.R., Geliazkova, M.P., Rowe, D.B., 2008. An evaluation of spatial thresholding techniques in fMRI analysis. Hum. Brain Mapp. 29, 1379–1389.
Lyoo, I.K., Sung, Y.H., Dager, S.R., Friedman, S.D., Lee, J.Y., Kim, S.J., Kim, N., Dunner, D.L., Renshaw, P.F., 2006. Regional cerebral cortical thinning in bipolar disorder. Bipolar Disord. 8, 65–74.
Makris, N., Biederman, J., Valera, E.M., Bush, G., Kaiser, J., Kennedy, D.N., Caviness, V.S., Faraone, S.V., Seidman, L.J., 2007. Cortical thinning of the attention and executive function networks in adults with attention-deficit/hyperactivity disorder. Cereb. Cortex 17, 1364–1375.
Mikl, M., Marecek, R., Hlustík, P., Pavlicová, M., Drastich, A., Chlebus, P., Brázdil, M., Krupa, P., 2008. Effects of spatial smoothing on fMRI group inferences. Magn. Reson. Imaging 26, 490–503.
Nichols, T., 2007. False discovery rate procedures. In: Friston, K.J., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (Eds.), Statistical Parametric Mapping: the Analysis of Functional Brain Images. Academic Press, London, pp. 246–252.
Nichols, T.E., Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.
Pratt, W.K., 1991. Digital Image Processing. John Wiley and Sons, Inc., New York.
Raphan, M., Simoncelli, E.P., 2008. Optimal denoising in redundant representations. IEEE Trans. Image Process. 17, 1342–1352.
Reimold, M., Slifstein, M., Heinz, A., Mueller-Schauenburg, W., Bares, R., 2006. Effect of spatial smoothing on t-maps: arguments for going back from t-maps to masked contrast images. J. Cerebr. Blood F. Met. 26, 751–759.
Segonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.
Shaw, P., Lerch, J., Greenstein, D., Sharp, W., Clasen, L., Evans, A., Giedd, J., Castellanos, F.X., Rapoport, J., 2006. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. Arch. Gen. Psychiatry 63, 540–549.
Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E., Kabani, N.J., 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. Brain 129, 2885–2893.
Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98.
Sowell, E.R., Peterson, B.S., Thompson, P., Welcome, S.E., Henkenius, A.L., Toga, A.W., 2003. Mapping cortical changes across the human life span. Nat. Neurosci. 6, 309–315.
Starck, J.L., Moudden, Y., Abrial, P., Nguyen, M., 2006. Wavelets, ridgelets and curvelets on the sphere. Astron. Astrophys. 446, 1191–1204.
Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W., 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. Cereb. Cortex 11, 1–16.
Thompson, P.M., Hayashi, K.M., Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. 23, 994–1005.
Van Essen, H., Drury, S., Miller, J., Miller, M., 1998. Functional and structural mapping of human cerebral cortex: solution are in the surfaces. Proc. Natl. Acad. Sci. USA 95, 788–795.
Von Economo, C., 1929. The Cytoarchitectonics of the Human Cerebral Cortex. Oxford University Press, London. (186 pp.).
Watson, D.F., 1992. Contouring: a Guide to the Analysis and Display of Spatial Data. Pergamon (Elsevier Science, Inc.), Tarrytown, NY.
Wiaux, Y., McEwen, J.D., Vandergheynst, P., Blanc, O., 2008. Exact reconstruction with directional wavelets on the sphere. Mon. Not. R. Astron. Soc. 000, 1–18.
Wink, A.M., Roerdink, J.B.T.M., 2004. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. IEEE Trans. Med. Imaging 23, 374–387.
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.