

A New Statistical Image Analysis Approach and Its Application to Hippocampal Morphometry

Mark Inlow^{1*}, Shan Cong², Shannon L. Risacher¹, John West¹, Maher Rizkalla², Paul Salama², Andrew J. Saykin¹, Li Shen^{1*}, and for the ADNI**

¹ Radiology and Imaging Sciences, Indiana University School of Medicine, IN, USA

² Electrical and Computer Engineering, Purdue University Indianapolis, IN, USA

Abstract. In this work, we propose a novel and powerful image analysis framework for hippocampal morphometry in early mild cognitive impairment (EMCI), an early prodromal stage of Alzheimer’s disease (AD). We create a hippocampal surface atlas with subfield information, model each hippocampus using the SPHARM technique, and register it to the atlas to extract surface deformation signals. We propose a new alternative to standard random field theory (RFT) and permutation image analysis methods, *Statistical Parametric Mapping (SPM) Distribution Analysis or SPM-DA*, to perform statistical shape analysis and compare its performance with that of RFT methods on both simulated and real hippocampal surface data. The major strengths of our framework are twofold: (a) SPM-DA provides potentially more powerful algorithms than standard RFT methods for detecting weak signals, and (b) the framework embraces the important hippocampal subfield information for improved biological interpretation. We demonstrate the effectiveness of our method via an application to an AD cohort, where an SPM-DA method detects meaningful hippocampal shape differences in EMCI that are undetected by standard RFT methods.

1 Introduction

The hippocampus plays an important role in learning and memory, and is a widely studied brain structure in Alzheimer’s Disease (AD). Hippocampal measures extracted from magnetic resonance imaging (MRI) scans have been shown as effective biomarkers for detecting the status of AD or mild cognitive impairment (MCI, a prodromal stage of AD) [4, 8, 9, 14]. Investigation of hippocampal

* Correspondence to Mark Inlow (minlow@iupui.edu) or Li Shen (shenli@iu.edu). This work was supported by NIH R01 LM011360, U01 AG024904, R01 AG19771, P30 AG10133, UL1 TR001108, R01 AG 042437, and R01 AG046171; DOD W81XWH-14-2-0151, W81XWH-13-1-0259, and W81XWH-12-2-0012; and NCAA 14132004.

** Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This is the author's manuscript of the article published in final edited form as:

Inlow, M., Cong, S., Risacher, S. L., West, J., Rizkalla, M., Salama, P., ... Adni, F. T. (2016). A New Statistical Image Analysis Approach and Its Application to Hippocampal Morphometry. In *Medical Imaging and Augmented Reality* (pp. 302–310). Springer, Cham. https://doi.org/10.1007/978-3-319-43775-0_27

morphometry as an early biomarker for detecting early MCI (EMCI) is a significant but yet under-explored topic. Cong et al. [2] studied this topic using random field theory [12, 13] and surface-based morphometry, but identified no significant difference between healthy control (HC) and EMCI participants.

To bridge this gap, we propose a novel and powerful image analysis framework for hippocampal morphometry in EMCI. We create a hippocampal surface atlas with subfield information using the method described in [2]. We model each hippocampus using the SPHARM technique [7] and register it to the atlas for subsequent analyses. We propose analyzing the resulting data using a new approach, *Statistical Parametric Mapping (SPM) Distribution Analysis or SPM-DA*. SPM-DA methods can provide greater power by more fully exploiting SPM information than current random field theory (RFT) and permutation methods which only use information provided by SPM peaks and/or clusters. In addition, SPM-DA methods use permutation inference to allow the use of algorithms with mathematically intractable distributions and to avoid restrictive RFT assumptions which, if violated, can reduce power. Thus the major strengths of this work are twofold: (a) SPM-DA provides potentially more powerful algorithms than current RFT and permutation methods for detecting weak signals and (b) the work embraces, rather than ignores, the important hippocampal subfield information for improved interpretation of the identified pattern.

We compare the performance of RFT Peak and RFT Peak methods with that of a specific SPM-DA histogram method by analyzing simulated and real (Alzheimer’s Disease Neuroimaging Initiative (ADNI) [10]) data. We also conduct a subfield analysis of the ADNI data to fully demonstrate the utility of our hippocampal morphometric analysis framework.

2 Materials and Methods

ADNI Data: The real data used in this study were downloaded from the ADNI database [10]. One goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. We downloaded baseline 3T MRI scans of 172 HC, 267 early MCI (EMCI), and 140 late MCI (LMCI), along with demographic and diagnostic information.

Hippocampal Surface Modeling and Alignment: Hippocampal segmentation is conducted by FIRST [3], a surface registration and segmentation tool developed as part of the FMRIB Software Library (FSL). FreeSurfer is not used here since it tends to yield noisy hippocampal boundary not suitable for shape analysis [2]. Topology fix is performed on the binary segmentation results to make sure that each hippocampal surface has spherical topology. The SPHARM method [7] is used to model each surface. FreeSurfer is used for hippocampal subfield segmentation, because this function is unavailable in FIRST. Following [2], we create a hippocampal surface atlas labelled with five subfields: hippocampal tail, CA1, CA2-3, CA4-DG, and SUB (containing both presubicu-

lum and subiculum). Each hippocampal surface is a SPHARM reconstruction registered to the atlas by aligning its first order ellipsoid. Surface signals are extracted as the deformation along the surface normal direction of the atlas.

RFT Surface Analysis: Using random field theory (RFT) implemented in SurfStat [12], the surface signals $N_{i,j}$ are analyzed using the regression model

$$N_{i,j} = \beta_0 + \beta_{1,j}I_i + \beta_{2,j}\text{age}_i + \beta_{3,j}\text{gender}_i + \epsilon_{i,j}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

in which I_i is the group indicator, e.g., 1 if EMCI and 0 if HC, n is the number of subjects, m is the number of surface vertices. The SPM consisting of the t statistics for testing $H_0: \beta_{1,j} = 0, j = 1, \dots, m$, is then analyzed using both peak amplitude and cluster size statistics as implemented by SurfStat [12, 13].

SPM Distribution Analysis (SPM-DA): The SPM-DA method investigated here captures the information provided by the SPM statistics by estimating their distribution with a frequency histogram. The histogram bin boundaries are chosen so that each bin is equally likely under the null (permutation) distribution. The bin frequencies are then analyzed to detect departures from count uniformity. In these analyses two regression models are employed,

$$F_i = \beta_u x_{u,i} + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

$$F_i = \beta_l x_{l,i} + \epsilon_i, \quad i = 1, \dots, n, \quad (2)$$

in which F_i denotes the frequency of the i th of $n = 12$ bins. For the first model in Eq (1), we let $x'_u = (0, 0, 0, 0, 0, 0, 0, 1, 2, 3, 4, 5)'$ be our predictor. Thus, the coefficient β_u will be positive when there is an overabundance of positive SPM statistics (right-tail values) indicating a positive relationship between image values and the predictor of interest. Similarly, for the second model in Eq (2), we let $x'_l = (5, 4, 3, 2, 1, 0, 0, 0, 0, 0, 0, 0)'$ be our predictor. Thus, the coefficient β_l will be positive when there is an overabundance of negative (left-tail) values indicating a negative relationship. Shown in Figure 2 are a few examples of the Eq (1) predictor data (i.e., the solid line) and the bin counts computed from the real hippocampal data (i.e., the “•” values).

To detect a relationship between the image and the predictor of interest generating the SPM, the following compositive hypotheses are tested:

$$H_0: \beta_u \leq 0 \text{ and } \beta_l \leq 0 \text{ versus } H_1: \beta_u > 0 \text{ or } \beta_l > 0.$$

Let $\hat{\beta}_u$ and $\hat{\beta}_l$ denote the least squares estimates of β_u and β_l from the unpermuted data. The corresponding one-sided p-values, $p_u = P(\beta_u \geq \hat{\beta}_u)$ and $p_l = P(\beta_l \geq \hat{\beta}_l)$, are combined using Bonferroni to get the p-value for testing H_0 vs. H_1 , $p = 2 \min(p_l, p_u)$. Simulation is used to compute p_u and p_l by randomly permuting the predictor (after it’s orthogonalized with respect to covariates if they are present [11]) with respect to the surface data, recomputing the SPM, and then computing the corresponding permutation coefficient estimates $\hat{\beta}_u^*$ and $\hat{\beta}_l^*$. This process is repeated N times and then p_u is estimated by

$$p_u = (\# \text{ of } \hat{\beta}_u^* \text{'s} \geq \hat{\beta}_u) / N; \quad (3)$$

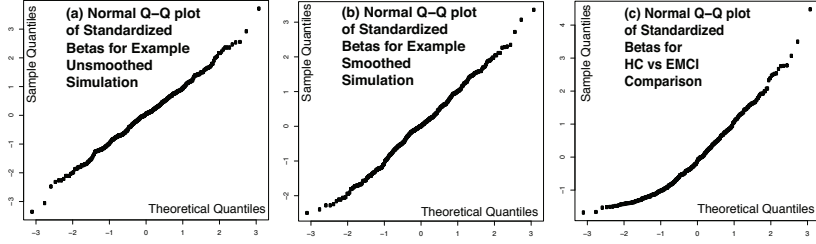


Fig. 1. Normal Q-Q plot of standardized betas for (a-b) example unsmoothed and smoothed simulations and (c) HC vs EMCI comparison.

p_l is estimated similarly. The only requirement for p-values p_u and p_l to be valid is the usual permutation assumption of exchangeability. Exchangeability is satisfied much more readily than the stringent RFT assumptions [11].

If, in addition, the distributions of the permutation coefficient estimates $\hat{\beta}_u^*$ and $\hat{\beta}_l^*$ are normal, as will often be the case for large samples [6] (e.g., those shown in Figure 1(a-b)) then p_u and p_l can be computed using the t distribution:

$$p_u = P\left(t_{N-1} \geq \frac{\hat{\beta}_u - \bar{X}}{S\sqrt{1+1/N}}\right) \quad (4)$$

in which \bar{X} and S are the sample mean and sample standard deviation of the N $\hat{\beta}_u^*$'s. The factor $\sqrt{1+1/N}$ in the denominator is needed since $\text{Var}[\hat{\beta}_u - \bar{X}] = \sigma^2(1+1/N)$ under the null hypothesis. Using this approach small p-values can be accurately estimated with N as small as 30 or so. This procedure, implemented in R, is used to analyze the hippocampal surface normals and the simulated data described below. The results are compared with the SurfStat RFT results.

Simulation Studies: SPM-DA and RFT peak and cluster methods are compared using two simulation studies. For both studies random data on a hippocampal template surface with 652 vertices are generated for 72 subjects according to the model

$$\begin{aligned} S_{i,j} &= \beta x_i + \epsilon_{i,j}, \quad i = 1, \dots, 72, \quad j = 1, \dots, 126, \\ &= \epsilon_{i,j}, \quad i = 1, \dots, 72, \quad j = 127, \dots, 652, \end{aligned}$$

in which $S_{i,j}$ represents the surface value at location j for subject i . Both studies simulate two-sample data with x_i equal to -1 for $i = 1, \dots, 36$ and 1 for $i = 37, \dots, 72$. Thus the signal, which extends across 126 contiguous locations, is constant with a magnitude determined by β . For both studies, values for β are 0, 1/12, 1/6, and 1/3. In the first study the random errors $\epsilon_{i,j}$ are independent normal ($\mu = 0, \sigma^2 = 1$) pseudorandom numbers. In the second study the $\epsilon_{i,j}$ are also independent normal ($\mu = 0, \sigma^2 = 1$) but are smoothed prior to the signal being added using the heat kernel smoothing method [1] applied to the hippocampal surface atlas. The resulting data sets are analyzed using SPM-DA

Table 1. Simulation results: The number of rejections (out of 100 runs) at $\alpha = 0.05$ and 0.01 for the SPM-DA (SDA), RFT Peak (RFP), and RFT Cluster (RFC) methods.

Signal Strength	Unsmoothed Data						Smoothed Data					
	$\alpha = 0.05$			$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.01$		
	SDA	RFP	RFC	SDA	RFP	RFC	SDA	RFP	RFC	SDA	RFP	RFC
0	5	1	0	1	0	0	3	2	0	0	1	0
1/12	92	7	1	65	3	0	83	10	0	60	3	0
1/6	100	57	6	100	17	5	100	47	51	100	17	43
1/3	100	100	5	100	100	5	100	100	49	100	100	44

Table 2. Statistical analysis results on real data using three approaches: SPM-DA, RFT Peak and RFT Cluster. P values are shown for pairwise comparison among three groups HC, EMCI and LMCI. N.S. indicates not significant. Note that 2E-04 is the smallest nonzero p value that can be obtained in our SPM-DA permutation tests.

Comparison	P from SPM-DA	Smallest P	
		RFT Peak	RFT Cluster
HC vs EMCI	9.20E-03	1.51E-01	N.S.
EMCI vs LMCI	<2E-04	9.46E-08	N.S.
HC vs LMCI	<2E-04	1.72E-08	N.S.

(programmed in R [5]) and RFT peak and cluster statistics as implemented by SurfStat [12, 13]. For each combination of β and choice of unsmoothed or smoothed random errors, 100 data sets are constructed and analyzed by SPM-DA and RFT methods to compare their power.

3 Results

In our simulation studies, the distributions of the permutation coefficient estimates by SPM-DA are always normal (see Figure 1(a-b) for a couple of examples). Thus, p_u and p_l are computed using the fast approach of Eq (4). However, in the real data study, the distributions of the permutation coefficient estimates are nonnormal (see Figure 1(c) for one example). In this case, we compute p_u and p_l using Eq (3) with $N=10,000$ permutations.

Table 1 presents the results of our simulation studies by providing the number of rejections (out of 100 runs) of the SPM-DA, RFT Peak, and RFT Cluster methods for significance levels $\alpha = 0.05$ and 0.01. For the null (signal strength=0) scenarios, all three methods have type I error rates at or below α . For all non-null scenarios the SPM-DA method dominates the RFT Cluster method, exhibiting substantially greater power at all signal strengths. It also dominates the RFT Peak method in all but the strongest signal case. In particular, its power is at least eight times greater than RFT Peak for the weakest signals.

Table 2 presents the results of analyzing the three hippocampal pairwise comparisons using the three methods. The SPM-DA method was the most powerful, detecting shape differences at level $\alpha = 0.01$ for all three comparisons in contrast to RFT Peak which detected two and RFT cluster which detected none. We believe that SPM-DA would yield smaller p-values than RFT Peak for the EMCI vs LMCI and HC vs LMCI comparisons if sufficient permutations, e.g. 10^9 , were

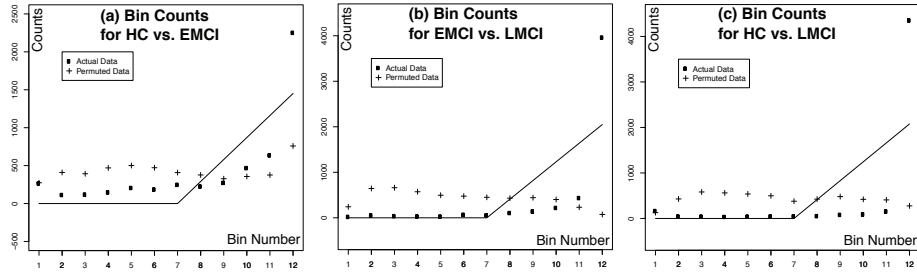


Fig. 2. Bin counts for (a) HC vs EMCI, (b) EMCI vs LMCI, and (c) HC vs LMCI. Our linear model in Eq (1) aims to use the values on the solid line to predict the “+” values (for permuted data) or the “•” values (for real data). Note that the count scales on the y-axis are different across these three cases, and the significance of the group difference is driven mainly by the “•” value on the 12th bin in each case.

Table 3. Comparison between RFT Peak and SPM-DA on the signal region size (i.e., number of vertices with $p < 0.05$ and number of vertices with bin value (bv) = 12 respectively) in each subfield. No data are shown for the RFT Cluster method, since no signals were identified in any RFT cluster analysis on real data.

Hemisphere		Left					Right				
Subfield		Tail	CA1	CA2-3	CA4-DG	SUB	Tail	CA1	CA2-3	CA4-DG	SUB
Total # of vertices		398	389	728	91	956	405	362	735	119	941
RFT Peak: # of vertices with $p < 0.05$	HC vs EMCI	0	0	0	0	0	0	0	0	0	0
	EMCI vs LMCI	48	15	39	0	92	15	0	49	3	351
	HC vs LMCI	64	92	204	27	501	94	116	184	28	450
SPM-DA: # of vertices with $bv = 12$	HC vs EMCI	87	113	260	85	572	128	273	263	80	393
	EMCI vs LMCI	291	389	441	49	896	242	295	450	97	805
	HC vs LMCI	283	389	507	91	914	278	362	509	117	903

used. The encouraging fact that the SPM-DA method was able to detect HC vs EMCI shape differences demonstrates the promise of SPM-DA for detecting early biomarkers in AD studies.

Figure 2 shows the Eq (1) predictor data (i.e., the solid line) and the bin counts generated by SPM-DA for each of the three comparisons. It is obvious that the shape differences were detected by the first regression model (see Eq (1)) in each case. In other words, SPM-DA detected trends toward an overabundance of SPM values in the upper tail of the distribution, indicating hippocampal atrophy in EMCI compared with HC, in LMCI compared with EMCI, and in LMCI compared with HC.

Figure 3(b) shows the surface map of the SPM values for HC vs EMCI, where the red color indicates the atrophy region in EMCI compared with HC. For comparison, Figure 3(a) shows the t-map of the SurfStat analysis (p-map not shown due to lack of signal). Although capturing a similar pattern, the RFT methods used by SurfStat cannot claim the group differences between HC and EMCI are significant. However, the RFT Peak method used by SurfStat was able to identify statistical shape differences between EMCI and LMCI (t-map

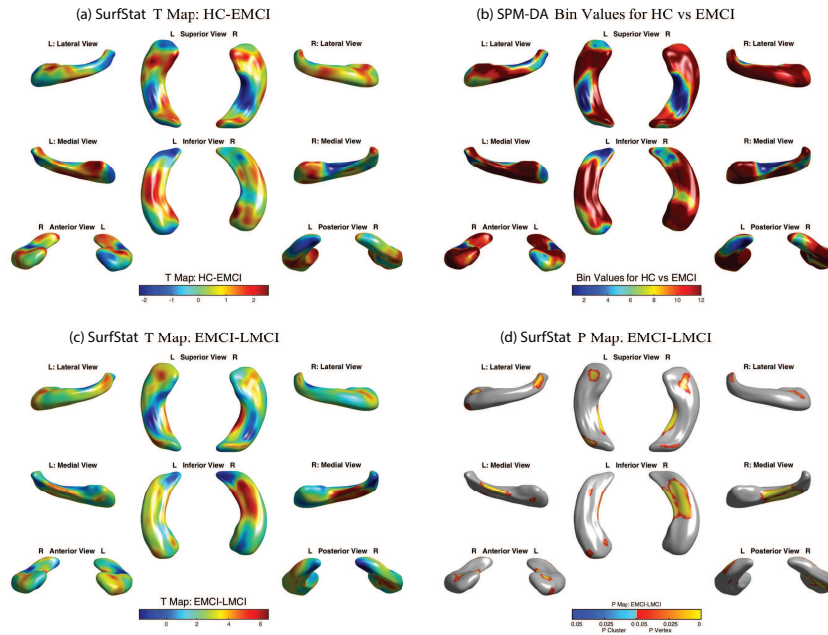


Fig. 3. (a) The SurfStat t-map of the diagnostic effect (HC-LMCI) on surface signals after removing the effects of age and gender. (b) The SPM-DA bin value map for the comparison of HC vs EMCI after removing effects of age and gender. (c-d) The SurfStat t-map and p-map of the diagnostic effect (EMCI-LMCI) on surface signals after removing the effects of age and gender.

and p-map shown in Figure 3(c-d)) and between HC and LMCI (t-map and p-map similar to Figure 3(c-d) and thus not shown).

Given that we have a surface atlas of hippocampal subfields, Table 3 shows the signal region size in each subfield according to RFT Peak and SPM-DA methods, i.e., number of vertices with $p < 0.05$ and number of vertices with bin value (bv) = 12 respectively. Below we summarize the amount of the subfield atrophy region detected by SPM-DA. (1) HC vs EMCI: EMCI demonstrated atrophy patterns compared with HC in 27% of tail, 51% of CA1, 36% of CA2-3, 79% of CA4-DG, and 51% of SUB. (2) EMCI vs LMCI: LMCI demonstrated atrophy patterns compared with EMCI in 66% of tail, 91% of CA1, 61% of CA2-3, 70% of CA4-DG, and 90% of SUB. (3) HC vs LMCI: LMCI demonstrated atrophy patterns compared with HC in 70% of tail, 100% of CA1, 69% of CA2-3, 99% of CA4-DG, and 98% of SUB.

4 Discussion

We have proposed a novel and powerful image analysis approach, *Statistical Parametric Mapping (SPM) Distribution Analysis or SPM-DA*, and applied it

to statistical shape analysis in hippocampal morphometry coupled with subfield information. We have compared its performance with that of standard random field theory (RFT) in surface-based morphometry. Our empirical studies on both simulated and real hippocampal data demonstrate that the SPM-DA method has greater power than either RFT Peak or RFT Cluster methods. Of particular importance to early MCI biomarker research, it has substantially greater power to detect weak signals, e.g., it was able to detect HC vs. EMCI differences undetected by RFT methods. These results provide proof of concept evidence for the core premise of SPM-DA, namely, that greater power can be achieved by more fully utilizing SPM distribution information. The specific method considered here used histograms to capture this information. Although this approach worked well in our proposed hippocampal morphometry analysis framework, one future direction is to explore other means of fully utilizing SPM distribution information. Given that SPM-DA is a generic approach, another future direction is to apply it to other image and/or shape analysis studies.

References

1. Chung, M.K., Robbins, S., et al.: Cortical thickness analysis in autism via heat kernel smoothing. *NeuroImage* 25, 1256–1265 (2005)
2. Cong, S., Rizkalla, M., et al.: Surface-based morphometric analysis of hippocampal subfields in mild cognitive impairment and Alzheimer’s disease. MWSCAS’15: IEEE 58th Int. Midwest Sym. on Circuits and Systems, pp. 813–816 (2015)
3. Patenaude, B., Smith, S.M., et al.: A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56(3), 907–922 (2011)
4. Pluta, J., Yushkevich, P., et al.: In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. *Journal of Alzheimer’s Disease* 31(1), 85–99 (2012)
5. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2015), <https://www.R-project.org/>
6. Sen, P.K., Singer, J.M.: Large Sample Methods in Statistics. An Introduction with Applications. Chapman & Hall, New York-London (1993)
7. Shen, L., Farid, H., McPeck, M.A.: Modeling three-dimensional morphological structures using spherical harmonics. *Evolution* 63(4), 1003–1016 (2009)
8. Shen, L., et al.: Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. *Brain Imaging Behav* 4(1), 86–95 (2010)
9. Testa, C., Laakso, M.P., et al.: A comparison between the accuracy of voxel-based morphometry and hippocampal volumetry in Alzheimer’s disease. *J Magn Reson Imaging* 19(3), 274–82 (2004)
10. Weiner, M.W., et al.: The Alzheimer’s Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement* 9(5), e111–94 (2013)
11. Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E.: Permutation inference for the general linear model. *Neuroimage* 92, 381–97 (2014)
12. Worsley, K.J.: SurfStat: <http://www.math.mcgill.ca/keith/surfstat>
13. Worsley, K., et al.: A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* 4, 58–73 (1996)
14. Yushkevich, P.A., Pluta, J.B., et al.: Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human brain mapping* 36(1), 258–287 (2015)