Combining Suprathreshold Average Voxel Intensity and Cluster Extent with Permutation Test Framework

Hui Zhang[1],Jun Ding[1], Satoru Hayasaka[2,3], and Thomas E. Nichols[1]

[1]Department of Biostatistics, University of Michigan, [2]Wake Forest University School of Medicine, Department of Public Health Science
[3]Wake Forest University School of Medicine, Department of Radiology

Introduction

Neuroimaging inferences are generally based on the size of a cluster or the voxel intensities within a cluster. Cluster size is sensitive for spatially extended signals while methods based on intensity are sensitive when signal magnitudes are large. Previous methods have combined both cluster size and maximum intensities within clusters[1,2,3,4]. In our previous work we found that the cluster mass was generally most sensitive [3,4]. As the cluster mass is the product of the cluster size statistic and the suprathreshold average intensity statistic, cluster mass can be seen as a method that combines two statistics with the "product" combining function. The goal of this work is to see if combining *P-values*–cluster size P-values and suprathreshold average intensity P-values–is more sensitive than combining statistics.

Methods: Combining Functions

When combining P-values we have a choice of combining either corrected or uncorrected P-values. Let uP^t and uP^c be the uncorrected P-values for suprathreshold average intensity (AvgT) and cluster size, respectively; let cP^t and cP^c be the corresponding FWE-corrected P-values. For each cluster we compute Tippet's and Fisher's combining functions based on the uncorrected P-values

$$\mathrm{W}^T = 1 - \min(\log_{10}\mathrm{uP}^t, \log_{10}\mathrm{uP}^c) \quad (1)$$
$$\mathrm{W}^F = -2(\log_{10}\mathrm{uP}^t + \log_{10}\mathrm{uP}^c) \quad (2)$$

as well as the analogous calculations for corrected P-values; observe that Tippet is equivalent to picking the better of the two P-values. Cluster mass W^M is computed as the sum of T values within a cluster above the cluster-defining threshold. Corrected P-values are found with permutation by building the maximum distribution [6]; uncorrected P-values are found by building the distribution of all clusters found in all permutations. Note that the combined P-values (W^F & W^T) represent an arbitrary statistic, upon which either corrected or uncorrected inferences can be made. This is a source of confusion: For, say, Tippet's combining function, corrected P-values can be based on either (uP^t, uP^c) or (cP^t, cP^c). Considering also the partial (non-combined) tests, we have five cluster test statistics to compare: AvgT, cluster size, W^T, W^F & W^M.

Methods: Simulation

We use 2,000 realizations of a 15-subject dataset using Gaussian noise images (48×48×32 voxels) and add a sphere-shaped signal with uniform intensity to each dataset, varying signal diameter and intensity. A P=0.01 cluster defining threshold and nominal $\alpha = 0.05$ level is used and for each realization a permutation test (1,000 perms) is performed. The rejection rate of each type of test is recorded, though here we only show results for corrected inferences based on (cP^t, cP^c).

Methods: Application to fMRI Data Analysis

We use a second level fMRI data set on working memory dataset with 12 subjects [7]. The analysis is based on contrast images for item recognition versus control, and all of the statistics described above are computed, in addition to MaxT and Tippet/Fisher combining with MaxT (to compare with previous work on MaxT). The total number of permutation we used is 4096 and the cluster defining threshold is P=0.001 with 11 degree of freedom.

Results: Simulation

The rejection rates of the partial, and combined tests from the simulations are shown in Figure 1. It shows that when the diameter is small, for example, 6, the

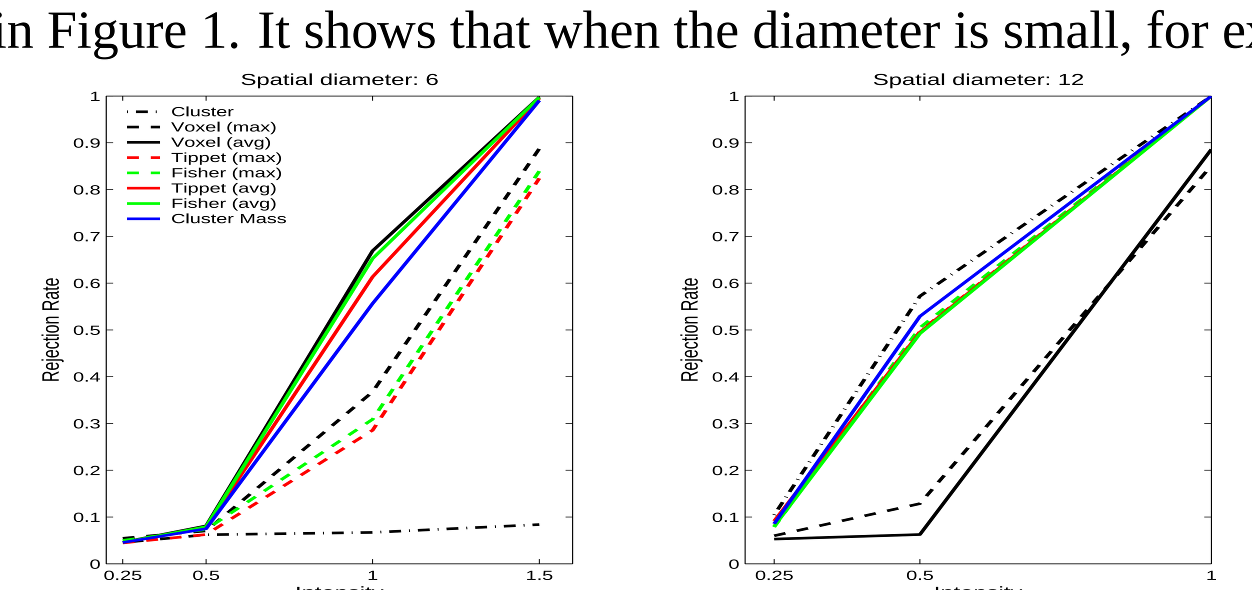

Figure 1: Rejection rates versus signal intensity

AvgT, Tippet and Fisher with AvgT methods have the best rejection rate, and cluster size is very insensitive. For larger diameters cluster size is the most sensitive, but the combining methods are all nearly as good. In general, for small diameter and moderate intensity (0.5-1.5), the AvgT, Tippet and Fisher using AvgT statistics are the best performers.

Results: Application to fMRI Data Analysis

The combined inferences based on (cP^t, cP^c) were consistently more sensitive than those based on (uP^t, uP^c), and so we only show results for combining corrected P's. The results for the 5 largest clusters are shown Table 1 & 2, and Figure 2 shows the rejection regions implied by the different statistics considered. Table 1 shows results using AvgT and Table 2 shows results with MaxT. The statistic that is uniformly the worst is partial AvgT, but the combining Fisher combining method is often most sensitive and never far from optimal. The MaxT methods appear to be similar or a touch more sensitive than the AvgT methods in this data.

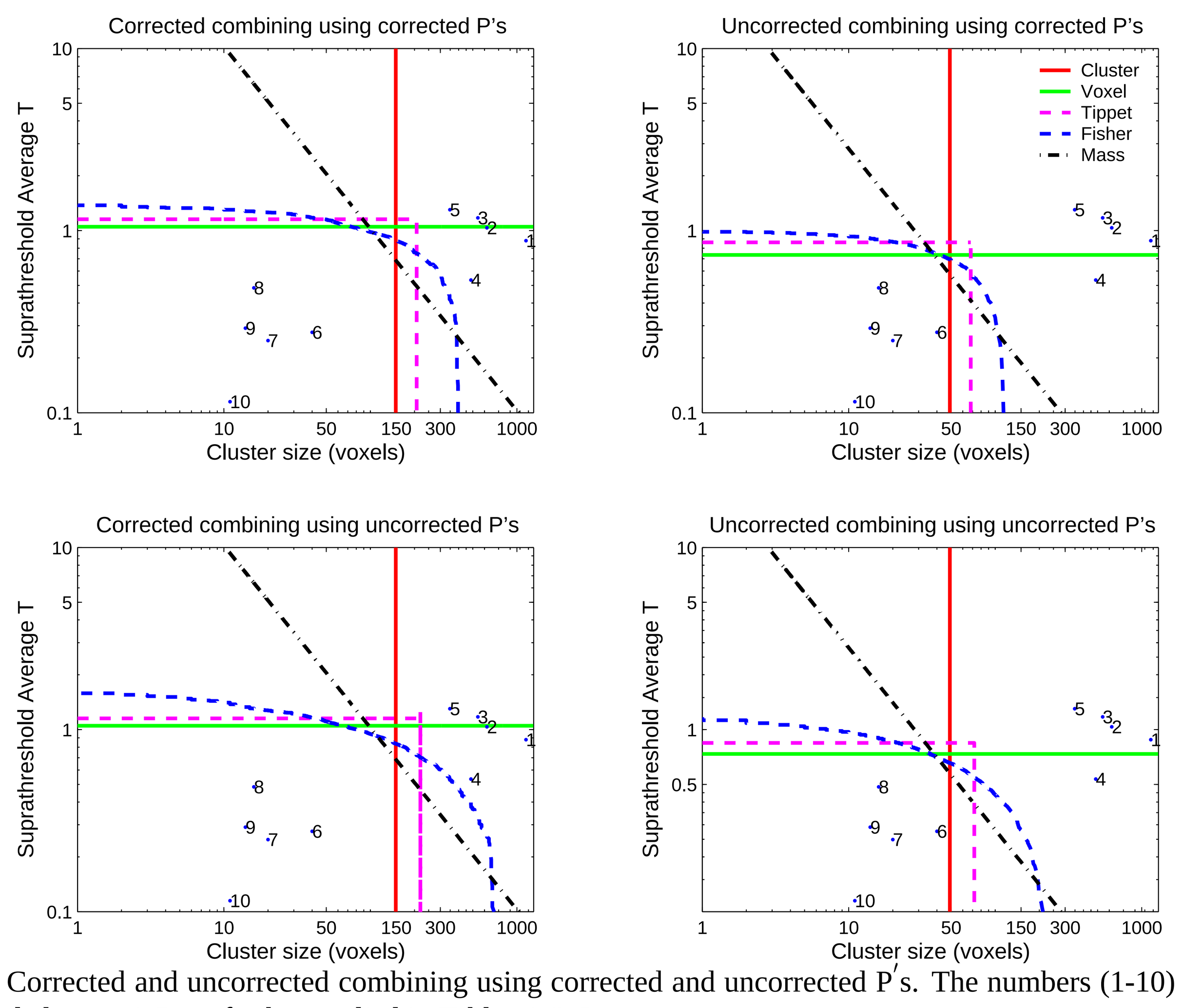

Figure 2: Corrected and uncorrected combining using corrected and uncorrected P′s. The numbers (1-10) refer to clusters (from largest to smallest) and clusters 1-5 are further studied in Tables 1 & 2

Table 1: Cluster size & AvgT: combining with corrected P′s

			Corrected P′s					Uncorrected P′s				
Cluster	Size	AvgT	Partial		Combined			Partial		Combined		
i			Cluster	AvgT	W^T	W^F	W^M	Cluster	AvgT	W^T	W^F	W^M
1	1154	4.9052	0.0002	0.1362	0.0005	0.0002	0.0002	0.0000	0.0226	0.0001	0.0000	0.0000
2	625	5.0607	0.0037	0.0562	0.0078	0.0024	0.0012	0.0006	0.0087	0.0012	0.0005	0.0002
3	542	5.1987	0.0039	0.0249	0.0078	0.0010	0.0012	0.0008	0.0037	0.0014	0.0002	0.0002
4	485	4.5595	0.0051	0.5259	0.0103	0.0244	0.0107	0.0010	0.1425	0.0017	0.0041	0.0020
5	349	5.325	0.0098	0.0105	0.0193	0.0010	0.0027	0.0018	0.0015	0.0032	0.0002	0.0006

Table 2: Cluster size & MaxT: combining with corrected P′s

			Corrected P′s					Uncorrected P′s				
Cluster	Size	MaxT	Partial		Combined			Partial		Combined		
i			Cluster	MaxT	W^T	W^F	W^M	Cluster	MaxT	W^T	W^F	W^M
1	1154	7.3582	0.0002	0.0803	0.0002	0.0005	0.0002	0.0000	0.0131	0.0001	0.0001	0.0000
2	625	9.3708	0.0037	0.0117	0.0068	0.0017	0.0012	0.0006	0.0018	0.0012	0.0004	0.0002
3	542	10.1937	0.0039	0.0051	0.0073	0.0005	0.0012	0.0008	0.0008	0.0013	0.0001	0.0002
4	485	6.3076	0.0051	0.2305	0.0098	0.0220	0.0107	0.0010	0.0458	0.0017	0.0037	0.0020
5	349	13.1468	0.0098	0.0002	0.0002	0.0005	0.0027	0.0018	0.0000	0.0001	0.0001	0.0006

Conclusions

Our simulations and real data show that no single method is optimal for all types of signals, however the partial methods (cluster size specifically) vary considerably in their sensitivity, while the combining methods are usually not far from optional. The main result from this work is that combining with corrected P-values appears to be more sensitive than combining with uncorrected P-values. Combining based on AvgT, while slightly less optimal based on our real data, was found to be more sensitive in our simulations. In future we will pursue a random field theory result for a combining method using AvgT.

Acknowledgments & References

This project is supported by the NIH's Human Brain Project. The authors would like to thank Dr. Christy Marshuetz for providing the dataset.

References :[1] Poline *et al.* Neuroimage 5:83-96 (1997).[2] Bullmore *et al.* IEEE Trans Med Img 18:32-42 (1991). [3] Hayasaka & Nichols. OHBM 2004, Poster: WE 191. [4] Hayasaka *et al.* Neuroimage 23:54-63 (2004). [5] Pesarin. Multivariate permutation Tests. Wiley. (2001). [6] Holms *et al.* JCBFM 16:7-22 (1996). [7] Marsheutz *et al.* J Cog Neuro 12(S2):130-144 (2000).