On Non-Normality, Non-Parametric Tests and Pooling Permutations Over Space for Voxel-Based Morphometry

Anderson M. Winkler*, Thomas E. Nichols**, David C. Glahn*

*Department of Psychiatry and Research Imaging Center, University of Texas Health Science Center at San Antonio, USA
**GSK Clinical Imaging Centre, UK, FMRIB Centre, Oxford University, UK, and Department of Biostatistics, University of Michigan, USA

Background

Voxel-based morphometry (VBM) data have been shown to exhibit non-normality [1,2] and, as a result, large smoothing kernels are recommended to ensure valid inferences, among other reasons. Instead of altering the data to satisfy statistical assumptions, we argue that nonparametric methods [3, 4], which do not depend on normality, should be used. While permutation methods have been championed for their accurate FWE-corrected inferences, uncorrected inferences at each voxel are also valuable. In particular, uncorrected P-values are explicitly used to find FDR thresholds and implicitly used when defining clusters.

Here we assess the need for nonparametric P-values with VBM data, by comparing parametric and permutation P-values, using the standard method of a permutation-test per-voxel. Second, we evaluate pooling permutation distributions over space [5], a technique that would allow computing a tiny number of permutations (e.g. 10) and as is used by the BAMM/CAMBA software.

Method

T1-weighted images (see [6] for the acquisition protocol) from 25 healthy subjects were analyzed (12 males). These images were normalized, segmented, modulated and smoothed in SPM5, using 15 different Gaussian filters, with FWHM ranging from no smoothing to 15 mm (isotropic). The Shapiro-Wilk normality-test was applied for every voxel, at a significance level of 5%, uncorrected.

The 25 subjects' images were randomly split into 2 groups (sizes 13 and 12), and both parametric and non-parametric P-values were obtained using SnPM5b, for each smoothing level. For the unpooled nonparametric maps, 2000 permutations were used, and, from these, 10 were selected at random to be pooled across space to generate a pooled permutation distribution.

Results

Without smoothing, many voxels (68.6%) were detected as non-normal, as expected. However, no amount of smoothing was able to reduce non-normality to nominal rates. For example, after filtering with a 6 mm kernel, 21.1% and after 12 mm, 9.1% of voxels were detected as non-normal (nominal level of 5% expected) [Figure 1].

Fig 1: Proportion of non-normal voxels.

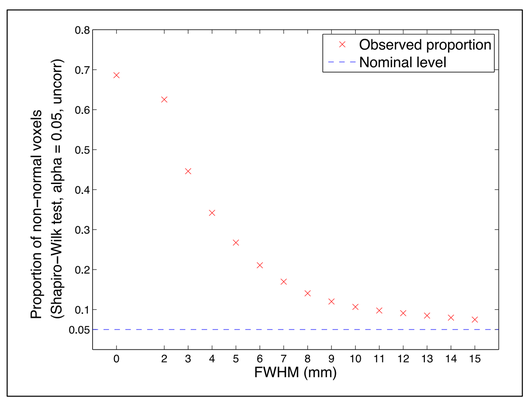
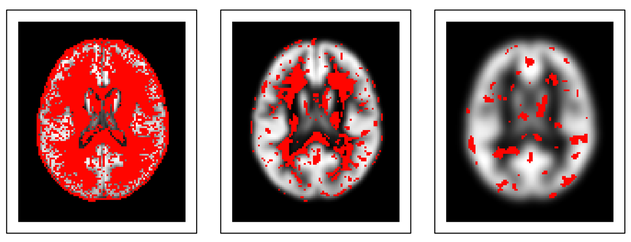
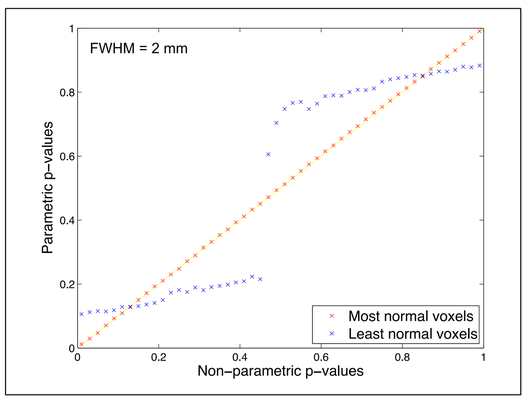


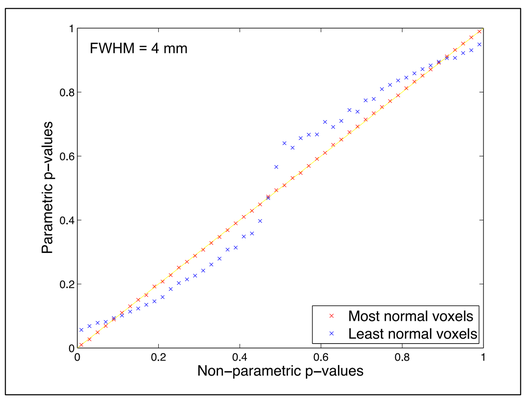
Fig 2: Non-normal voxels.

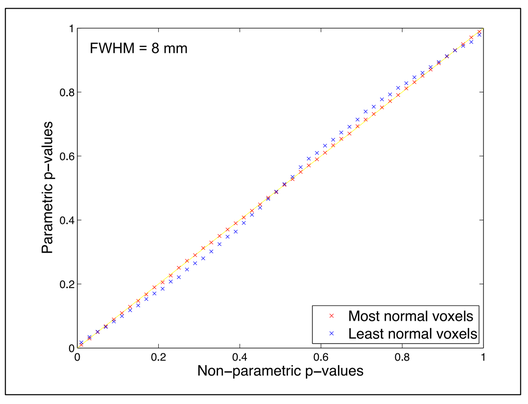


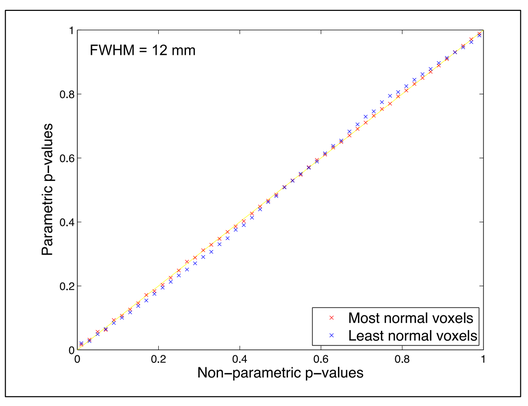
The parametric-vs-nonparametric P-value plots illustrate the impact of non-normality. While normally-distributed voxels have close agreement between parametric and nonparametric methods, non-normal voxels diverge from the identity line: for low smoothing, parametric P-values are conservative for small P, dramatically so for low smoothing; with increased smoothing, the effect reverses, with small P-values tending towards anticonservativeness and suggesting poor control of Type I error [Figure 3].

Fig 3: Parametric vs. Non-parametric.





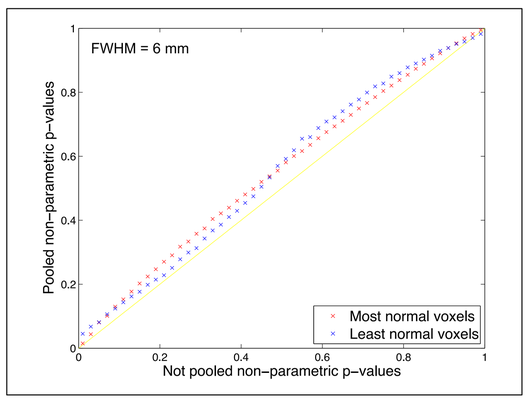


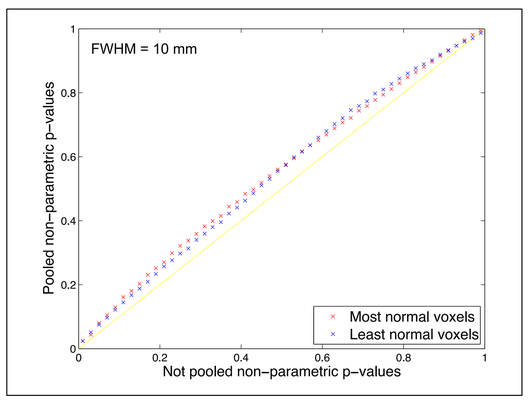


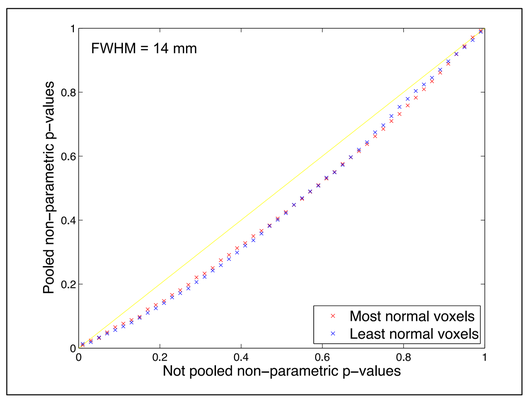
The not-pooled-vs-pooled nonparametric P-values show a similar pattern as the parametric-vs-nonparametric plots, with an additional skew. This skew is more pronounced at higher smoothing levels, and can be in the direction of conservativeness or anticonservativeness by chance, depending only on the set of 10 permutations that are realized.

In other words, even the most normally distributed voxels diverge from the identity, and in an unpredictable way, suggesting that the pooling approach is not stable [Figure 4].

Fig 4: Pooled non-parametric vs. not-pooled non-parametric.







Discussion

We have replicated others' findings that VBM data exhibits non-normality. Unlike other work, we present a straightforward solution to the problem: nonparametric permutation, as available for major software packages (e.g. SnPM for SPM and randomise for FSL).

We also show that pooling over space for inference at the voxel-level is not a stable approach, resulting in erratic and unpredictable control over the error rate.

References

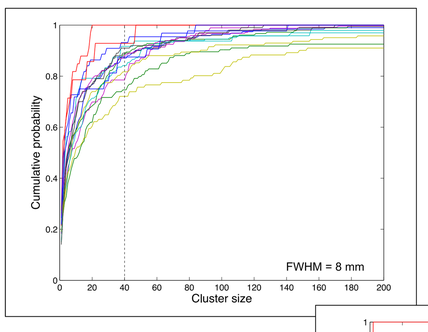
[1] Viviani R et al. NeuroImage, 35:121-30, 2007;
[2] Salmond CH et al. NeuroImage, 17:1027-30, 2002;
[3] Holmes AP et al. J Cereb Blood Flow Metab, 16:7-22, 1996;
[4] Nichols TE and Holmes AP. Hum Brain Mapp, 15:1-25, 2001;
[5] Bullmore ET et al. IEEE Trans Med Imag, 18:32-42, 1999;
[6] Kochunov P et al. Hum Brain Mapp, 27:957-62, 2006.

A look at the cluster level...

The same exercise repeated also with a null, but larger dataset (147 subjects, divided at random into two groups), found similar findings. We also used this larger dataset to examine cluster-level inference.

Uncorrected cluster-level inferences show substantial variability when pooling the permutation distribution over the brain. When using just 10 permutations to build an empirical distribution, we found highly unstable uncorrected cluster P-values. For example, at 8 mm smoothing we observed that a cluster of size 40 voxels can have a pooled uncorrected P-value as low as 0.02 (or even closer to zero) or as high as 0.28, depending only on what set of random permutations were realized (Figure 5, upper), with even greater variability with higher smoothing (e.g. for 12 mm, P≈0 to P=0.4, Figure 5, lower).

We also observed that the distribution cluster size varies with the average smoothness in each cluster (measured in terms of FWHM per voxel). The smallest clusters can arise in both rough and smooth regions and, while there is substantial variability, the largest clusters are more likely to be found in smooth regions. This may result in conservative uncorrected P-values in rough regions and liberal p-values in smooth regions. No extrinsic smoothing could completely remove this trend [Figure 6].



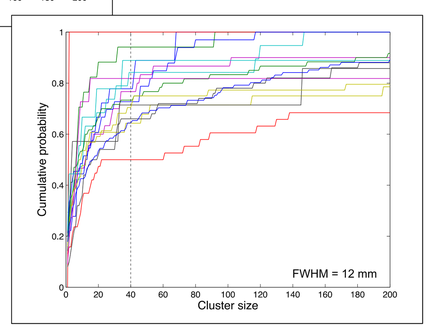
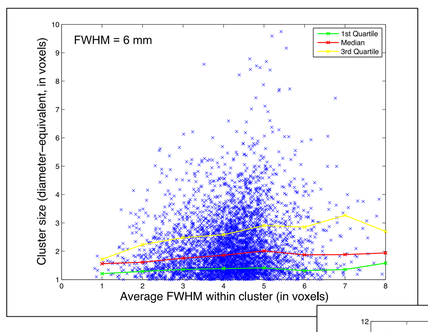


Fig 5: Cluster size CDFs of 15 realizations, each pooling 10 permutations chosen at random.



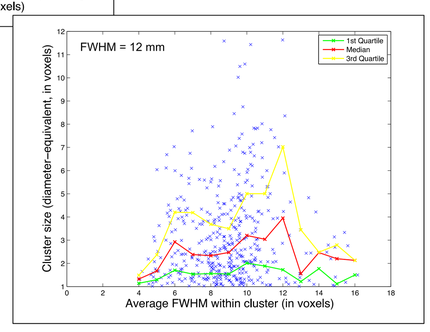


Fig 6: Average FWHM per voxel, within each cluster.

WE MAKE LIVES BETTER
UT HEALTH SCIENCE CENTER™
SAN ANTONIO

Clinical Imaging Centre
GlaxoSmithKline