Consistent Activation Across Trials and Field Strengths by ROC-reproducibility Thresholding

Tynan Reid Stevens[1], David B Clarke[2], Ryan CN D'Arcy[1], and Steven D Beyea[1]
[1]IBD-Atlantic, NRC, Halifax, Nova Scotia, Canada, [2]QEII Health Science Centre, Halifax, Nova Scotia, Canada

Background: Setting activation thresholds remains a significant challenge in fMRI. Traditional approaches that control error rates at a fixed level (e.g. FDR or Bonferroni) do not account for individual variability, differences in task demands, or changes in scanning hardware [1]. These deficiencies are particularly problematic in individual level analyses, as illustrated by comparisons with cortical stimulation mapping [2]. Reproducibility-based, data-driven methods have surfaced as a possible alternative [3]. We present a test-retest ROC method for detecting reliable activation patterns in individual fMRI. The ROC-reproducibility (ROC-r) thresholds are demonstrated for multiple subjects, tasks, and field strengths, and are shown to produce consistent activation maps across replications of a task, and even between field strengths.

Methods: 8 healthy volunteers (4 male, 24.4 +/- 3.5 y) were scanned twice at 4 T (TE=15 ms) and twice at 1.5 T (TE=40 ms) using a spiral out trajectory (TR=2 s, α=90°, 64 x 64 matrix, 22 slices, and 3.75 x 3.75 x 5 mm voxels, 0.5 mm gap). Each participant completed a motor and a cognitive task at each session. The motor task consisted of blocks of paced finger-to-thumb tapping with each hand (1 Hz). The cognitive task consisted of language and math blocks. Language blocks contained four simple English sentences that were congruent or incongruent ("He stroked her face with a feather" / "I drank some ability from California"). Math blocks contained four consecutive single digit addition/subtraction operations that were correct or incorrect (4+2=6 / 9-4=4). Participants were asked to respond to sentence or math stimuli by pressing one of two buttons for correct or incorrect respectively. Functional MRI images were analyzed with AFNI. The fMRI data were first motion corrected, and registered to a high-resolution T1 MRI, then spatially smoothed (FWHM 6mm). The generalized linear model was used to produce t-statistic maps for each condition contrasted to rest.

Test-retest ROC curves were created by using each of the two replications of a task (i.e. collected from the same subject, at the same field strength) as a template for the other. The threshold on the template was allowed to vary, and the area under the curve (AUC) of the ROC was computed for each template threshold. The AUC was plotted against the threshold templates used, producing AUC profiles that typically rise sharply at low thresholds, and plateau as the threshold approaches an optimal value. To quantitatively determine optimal template thresholds, the first maximum negative curvature was used. The ROC-r optimized template thresholds were then used to identify ROC-r optimized retest thresholds for each image, as the threshold that produced the best combination of retest sensitivity and specificity according to the ROC curve associated with the optimal template. Each image is used as template or retest, and both optimized template images or both retest images are overlaid for display. The ROC-r template and ROC-r retest thresholds were then compared with FDR and Bonferroni thresholds using q=0.01 and p=0.01 respectively.

Results: An example of the ROC-r AUC profiles used to select template thresholds is shown in figure 1, along with the resulting ROC-r optimized template and retest images. The average ROC-r template thresholds (5.27 +/- 0.22) were similar to the Bonferroni levels (5.29 +/- 0.01) for these datasets. The ROC-r template thresholds were slightly higher for the image collected first (5.55 +/- 0.27 vs. 4.98 +/- 0.34), and were greater at 4 T than 1.5 T (5.77 +/- 0.36 vs. 4.77 +/- 0.23). The ROC-r retest thresholds (3.71 +/- 0.18) were similar to the average FDR threshold (3.94 +/- 0.02). Like the ROC-r template thresholds, the ROC-r retest thresholds were slightly higher for the first of the two images (3.79 +/- 0.18 vs. 3.64 +/- 0.30), as well as for 4 T images than 1.5 T (4.29 +/- 0.31 vs. 3.14 +/- 0.12). The ROC-r template and retest thresholds both produced more consistent activation extent between sessions, as well as field strengths.

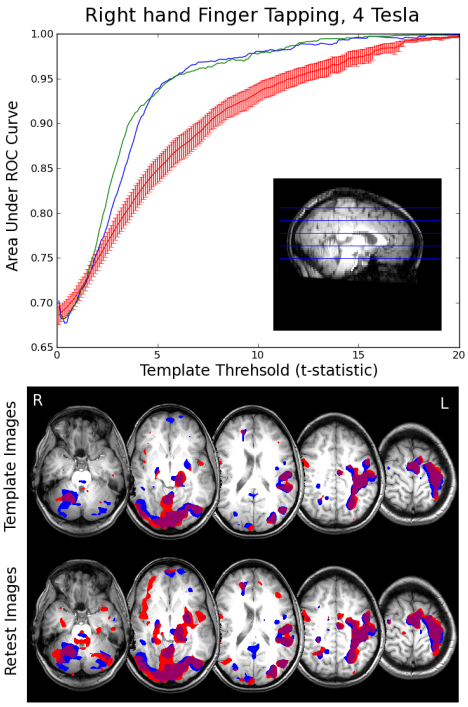

Figure 1: Example AUC profile for the first (blue) and second (green) image of the test-retest pair as template. The average AUC profile (red) for this task and scanner is shown as well. The ROC-r template (image 1, red, t= 4.0; image 2, blue, t=3.4), and retest thresholds (image 1, t= 2.9; image 2, t=2.9) were determined, and the resulting template and retest image pairs are shown below.

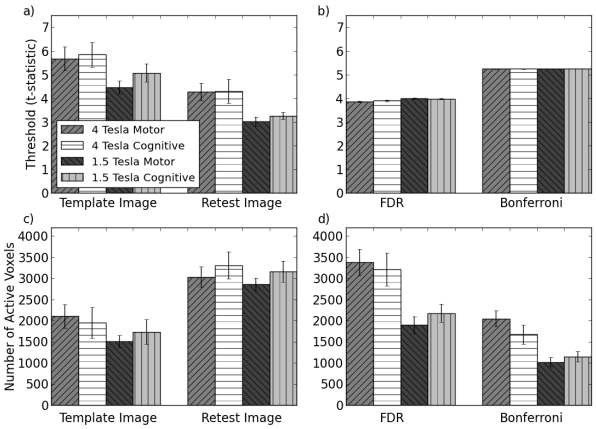

Figure 2: Average threshold (a,b) and number of active voxels (c,d) for the ROC-r methods (a,c) and the FDR/Bonferroni methods (b,d), by field and by task. The ROC-r method uses higher thresholds for 4 T data than 1.5 T, producing the same activation extent at both field strengths. The FDR and Bonferroni use the same thresholds across tasks and field, resulting in more active voxels at 4 T.

Discussion & Conclusion: The ROC-r threshold algorithm produces two optimized threshold levels, which provide average FPR control equivalent to the Bonferroni (template) or FDR (retest) methods. The ROC-r thresholds produce more consistent activation extent across replications, whereas the FDR and Bonferroni methods produce less activation in the second session, likely because of habituation to testing conditions. Additionally, the ROC-r thresholds eliminate the difference in number of active voxels observed between 1.5 T and 4 T, by applying higher thresholds to the 4 T data. This suggests that there are reliable activation patterns of similar extent at the two field strengths, which makes sense given that the fMRI task and therefore neural processes producing these maps are presumably the same at both field strengths. The FDR and Bonferroni methods produce more active voxels at 4 T than 1.5 T because of increased signal to noise. However, the same activated regions are likely present at low field, albeit at lower thresholds.

References: [1] Logan BR, Rowe DB, 2004. NeuroImage, 22, 95-108. [2] Rutten GJM, Ramsey NF, van Rijen PC, Noordmans HJ, van Veelen CWM, 2002. Ann Neurol, 51, 350-360. [3] Genovese CR, Noll DC, Eddy WF, 1997. Mag Reson Med, 38, 497-507.