Non-parametric Inference for Longitudinal and Repeated-Measures Neuroimaging Data with the Wild Bootstrap

Bryan Guillaume[1,2], Thomas E. Nichols[1] and the ADNI

[1]University of Warwick, Coventry, United Kingdom, [2]University of Liège, Liège, Belgium.

Introduction

Longitudinal neuroimaging data present significant modelling challenges due to subject dropout and within-subject correlation. Another complication is that non-parametric permutation inference [6,8] cannot be generally applied as the longitudinal correlation violates the assumption of null hypothesis exchangeability, and subject dropout can render different subjects incompatible for permutation. To address these limitations, we propose the use of an alternate non-parametric procedure, the Wild Bootstrap [4], and isolate which of its many variants are the most appropriate for longitudinal and repeated-measures neuroimaging data.

The Wild Bootstrap

Broadly speaking, the Wild Bootstrap (WB) generates samples of the data based on a fitted model, as follows: (1) the model's residuals are adjusted using a small sample bias correction, (2) the set of adjusted residuals for each subject is resampled by multiplying them by a single random value, and (3) the resampled residuals are added back to the predicted model. The random values are drawn independently from a resampling distribution, such as the Rademacher distribution (-1 or 1 with equal probability) [4] or the Mammen distribution $((1-\sqrt{5})/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$, $(1+\sqrt{5})/2$ otherwise) [5].

For the original data and each resampling, we use the Sandwich Estimator method [2] to construct a Wald statistic, creating a null distribution. Like permutation tests, this null distribution can be used to make inference on the original Wald statistic. In the context of neuroimaging, the Wild Bootstrap procedure can be used to estimate the null distribution of the maximum Wald statistic or the maximum cluster size, allowing a control of the voxel-wise or cluster-wise Family-Wise Error Rate (FWER), respectively.

Monte Carlo Simulations

We used Monte Carlo simulations to compare 80 WB variants, differing by the choice between (1) an unrestricted (not imposing the null hypothesis; U-WB) or a restricted (imposing the null hypothesis; R-WB) model, (2) five resampling distributions (i.e. the Rademacher distribution, the Mammen distribution, the Normal distribution $\mathcal{N}(0,1)$ and two distributions proposed in [7] referred to as Webb4 and Webb6 distributions), (3) four Sandwich Estimator versions, differing by whether they impose the null hypothesis (R-SwE) or not (U-SwE), and assume heterogeneous (S^{Het}) or per-group homogeneous (S^{Hom}) within-subject covariance matrices, and, (4), for both the bootstrap residuals and the Sandwich Estimator, the use of a small-sample bias adjustment (S_{C2}) or none (S_0). The settings considered corresponded to unbalanced designs with missing data (inspired by the real longitudinal neuroimaging ADNI dataset described in the top right section of this poster) under six different types of subject covariance structures: (1) Compound Symmetry (CS), (2) Toeplitz assuming a linear decay of the correlation over time, (3) heterogeneous variance over time, (4) heterogeneous variance between groups, (5) compound symmetric correlations & heterogeneous variance over time, and (6) Toeplitz correlations & heterogeneous variance over time. For each realisation (10,000 per setting), we used each of the 80 WB variants to test 24 contrasts at 5% level of significance. Fig. 1 summarises the obtained results (see caption for more details).
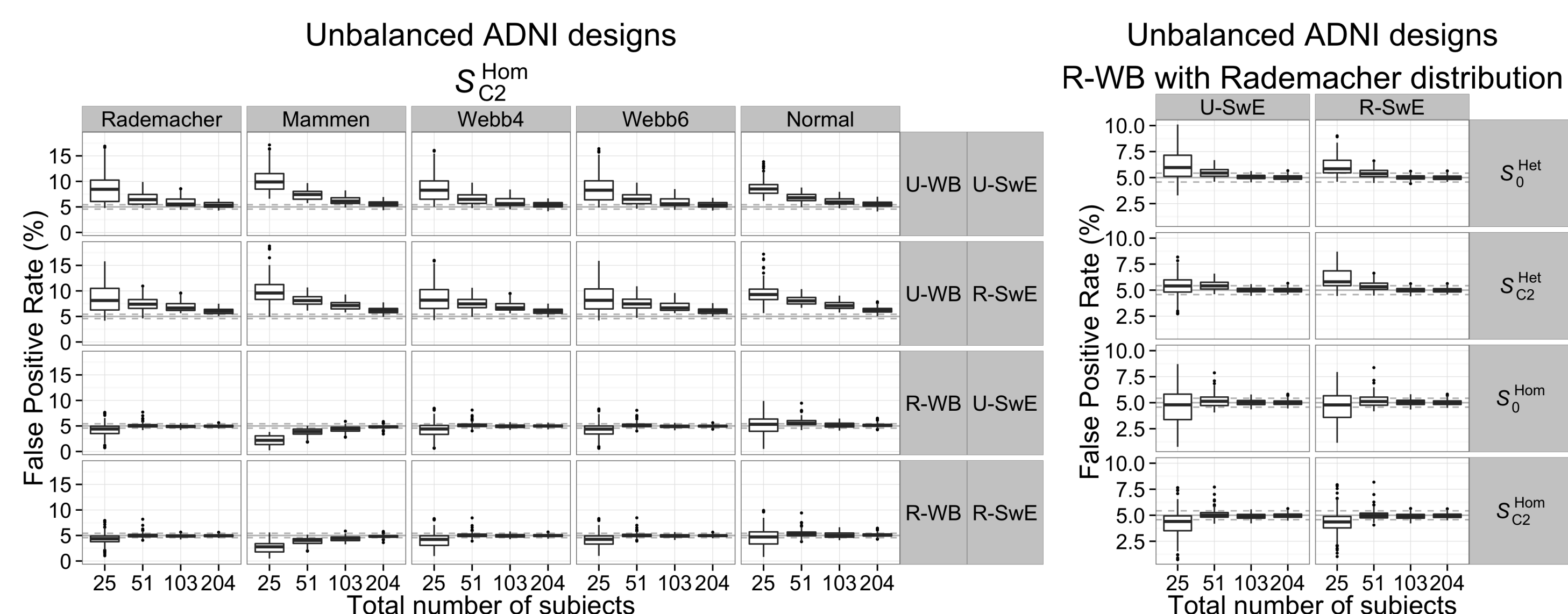

Fig. 1: Boxplots showing the False Positive Rate control of several WB variants as a function of the total number of subjects over 144 scenarios (consisting of the 24 contrasts tested and the 6 within-subject covariance structures considered). The best WB variants seem those considering a restricted model (R-WB) and the Rademacher distribution (which seems slightly better than the Webb4 and Webb6 distributions in some scenarios). No appreciable differences between the Sandwich Estimator versions and the use of a small-sample bias adjustment were observed, but it seems that the use of a small-sample bias adjustment and the Sandwich Estimator considering per-group homogeneous within-subject covariance matrices yielded a slightly better control of the False Positive Rate in some settings.

References
[1] Barch, D. M. et al. (2013), [2] Guillaume, B. et al. (2014), [3] Hua, X. et al. (2013), [4] Liu, R. Y. (1988), [5] Mammen, E. (1993) [6] Nichols, T. E. et al. (2002), [7] Webb, M. D. (2013), [8] Winkler, A. M. et al. (2014).

Electronic version of this poster
http://warwick.ac.uk/tenichols/ohbm

Real Data Longitudinal Analysis

We used the R-WB with the Rademacher distribution (with resampled residuals adjusted as in the R-SwE S^{Hom}_{C2}) and the R-SwE S^{Hom}_{C2} to analyse an unbalanced longitudinal Tensor-Based Morphometry dataset. Data are from the Alzheimer's Disease Neuroimaging Initiative (ADNI), consisting of 817 subjects divided in 3 groups (229 Normal, 400 MCI and 188 AD subjects) with a varying number of scans/subject (min = 1, max = 6, mean = 4.1 scans/subject) [2,3]. We tested for a greater atrophy effect in the AD vs. the Normal subjects by controlling for a voxel-wise FWER at 5%. For comparison, we also tested for the same effect using a permutation test on subject summary slopes extracted from subject-specific models; this model is not generally valid [2] but is commonly used. The results of the analysis are summarised in Fig. 2.
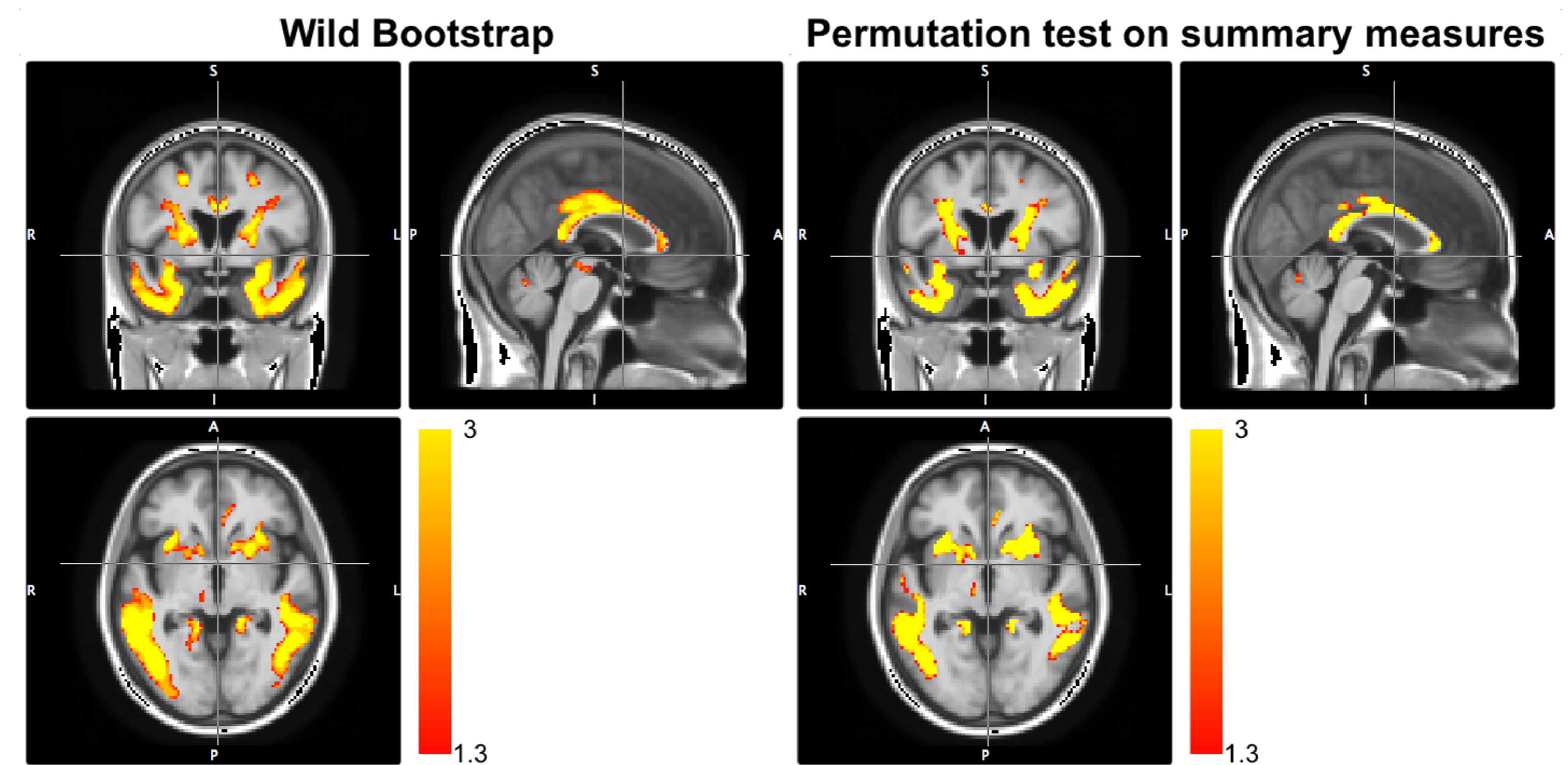

Fig. 2: ADNI data, test of AD volume decline vs. Normal volume decline with WB (left) and with a permutation test on summary measures (right). The voxel-wise 5% FWER-corrected thresholded images are $-\log_{10}(p_{FWER})$ images. The WB inference yielded more significant voxels (44,793 voxels) than the permutation test on summary measures (33,341 voxels).

Real Data Repeated Measures Analysis

We used the R-WB with the Rademacher distribution (with resampled residuals adjusted as in the R-SwE S^{Hom}_{C2}) and the U-SwE S^{Hom}_{C2} to analyse a repeated-measures fMRI dataset. Data are from the Human Connectome Project (HCP) [1], consisting of 2-back vs. 0-back memory task contrasts for four visual stimuli (body parts, faces, places and tools). We tested for positive & negative average effects across stimuli (two 1-degree of freedom tests) and for any difference between stimuli (one 3-degrees of freedom F-test) by controlling for a voxel-wise and a cluster-wise (primary threshold at $p = 0.001$) FWER at 5%.
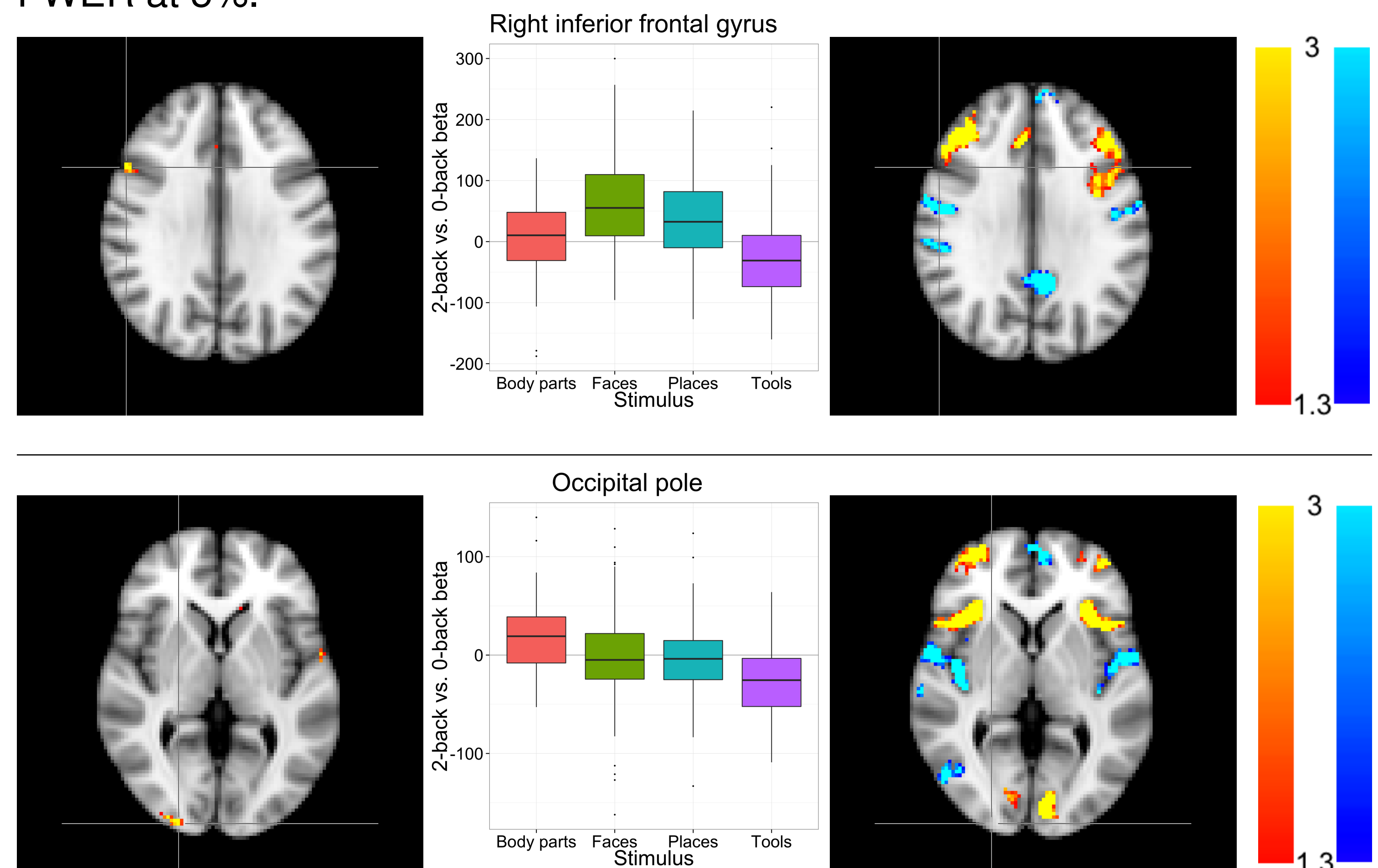

Fig. 3: Real data analysis of the HCP dataset with WB. Left: voxel-wise 5% FWER-corrected thresholded $-\log_{10}(p_{FWER})$ images on any differences of the 2-back vs. 0-back tasks between stimuli (406 voxels & 45 clusters survived the voxel-wise & cluster-wise thresholding, respectively). Middle: boxplots of the 2-back vs. 0-back beta for each stimuli at the selected voxel. Right: voxel-wise 5% FWER-corrected $-\log_{10}(p_{FWER})$ images on average positive (red/yellow colour; 13,749 voxels & 8 clusters survived the voxel-wise & cluster-wise thresholding, respectively) and negative (blue colour; 7,101 voxels & 11 clusters survived the voxel-wise & cluster-wise thresholding, respectively) effects of the 2-back vs. 0-back tasks across stimuli. The slices are located at $z = 30$mm (top) and $z = 2$mm (bottom) of the anterior commissure.

Discussion

In this work, we have proposed the use of the Wild Bootstrap to make non-parametric inferences for longitudinal and repeated-measures neuroimaging data. Using Monte Carlo stimulations we have isolated the best WB variants and we have illustrated its use on two real datasets. The WB has been implemented into the SwE toolbox for SPM (http://warwick.ac.uk/tenichols/SwE) and is currently under development for FSL.