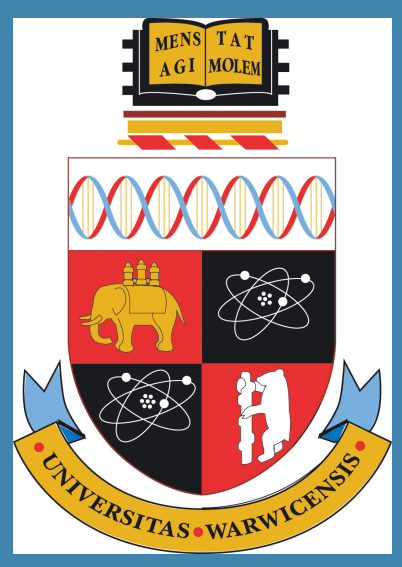# Threshold–Free Cluster Enhancement revisited: Increasing Power and Spatial specificity of the TFCE Statistic

Habib Ganjgahi[1], Steve Smith[2], and Thomas Nichols[1]

[1]University of Warwick [2]University of Oxford

Warwick Statistics

Introduction

Voxel-wise, cluster-wise and Threshold-Free Cluster Enhancement (TFCE) are 3 commonly used inference tools in Neuroimaging data analysis. Voxel-wise inference is generally the least powerful but is the most accurate spatially. Cluster-wise inference is more sensitive, but depends on a cluster forming threshold, and at low thresholds has poor spatial specificity and may lead to misinterpretation of findings (Woo 2014). TFCE was developed (Smith 2009) to obtain the power of cluster-wise inference without the dependence on an arbitrary cluster forming threshold. The TFCE statistic summarises the cluster-wise evidence at each voxel, considering a fine grid of cluster-forming thresholds; it introduces two new parameters that are fixed to values justified by theoretical and empirical results. While TFCE was found to be as or more powerful than cluster size inference (Smith 2009), significant TFCE voxels may get support from extremely large clusters, degrading spatial specificity. Here we introduce variants of TFCE statistic to improve its spatial specificity while maintaining or improving its power.

Methods

As originally defined, the TFCE statistic as voxel p is:

$$TFCE(p) = \int_{h=0}^{h_p} e(h)^E h^H dh \qquad (1)$$

where h_p is the statistic value at voxel p, h is a cluster-forming threshold, and $e(h)$ is cluster extent found at voxel p with cluster-forming threshold h. We consider 3 variants:

$$TFCE_1(p) = \int_{h=0}^{h_p} \min(e(h), e_{\max})^E h^H dh \qquad (2)$$

$$TFCE_2(p) = \max_h \left(e(h)^E h^H \right) \qquad (3)$$

$$TFCE_3(p) = \int_{h=h0}^{h_p} e(h)^E h^H dh \qquad (4)$$

In $TFCE_1$ the score at voxel p is not allowed to get support from clusters bigger than a pre-specified cluster size $e_{\max}$ where it corresponds to a cluster extend when the statistic image is thresholded at 5% level. $TFCE_2$ corresponds to searching for an optimal h, as measured by $e(h)^E h^H$. Finally $TFCE_3$ ignores all clusters with h less than h_0; we set $h_0 = 1.64$, the 5% uncorrected critical Z value. To compare these variants we define different Free-Response Receiver Operating Characteristic (ROC) curves: Each ROC has familywise error on x-axis (1-specificity), while the y-axis variable has one of various sensitivity and specificity measurements that are defined in Figure 1. Our comparison metric is Area Under the Curve (AUC), measured for Familywise Error (FWE) < 0.05 (scaled by 1/0.05, so $0 \geq AUC \leq 1$). When the y-axis has a True Positive (TP) rate, AUC reflects sensitivity, and when the y-axis has a False Positive (FP) rate, it measures (lack of) specificity; while the latter case is unusual, this type of AUC usefully measures (lack of) specificity when FWE is controlled at 0.05. We desire a method with a high AUC value for TP and a low AUC for FP.

Measuring Spatial Specificity

We define a comprehensive set of 4 sensitivity and 4 (lack of) specificity measures. We consider a cluster to be a "true positive" if it touches *any* true positive voxels, and "false positive" if it comprised entirely of null voxels. In the following, W is the total number of voxels in the image, N total number of detected clusters:

	Sensitivity		1 – Specificity
VP:	Voxel-wise Power	VFP:	Voxel-wise False Positive Rate
VTD:	Voxel-wise True Discovery Rate	VFD:	Voxel-wise False Discovery Rate
CTD:	Cluster-wise True Discovery Rate	CFD:	Cluster-wise False Discovery Rate
CVTD:	Cluster Voxel-wise True Discovery Rate	CVFD:	Cluster Voxel-wise False Discovery Rate

Sensitivity

$$TP = \frac{R}{R+G}$$
$$VTD = \frac{R}{R+B+V}$$
$$CTD = \frac{T}{N}$$
$$CVTD = \frac{R}{N(R+B)}$$

1-Specificity

$$VFP = \frac{B+V}{W-(G+R)}$$
$$VFD = \frac{B+V}{W-(R+B+V)}$$
$$CFD = \frac{N-T}{N}$$
$$CVFD = \frac{1}{N}\left(\frac{B}{R+B} + \frac{V}{V}\right)$$
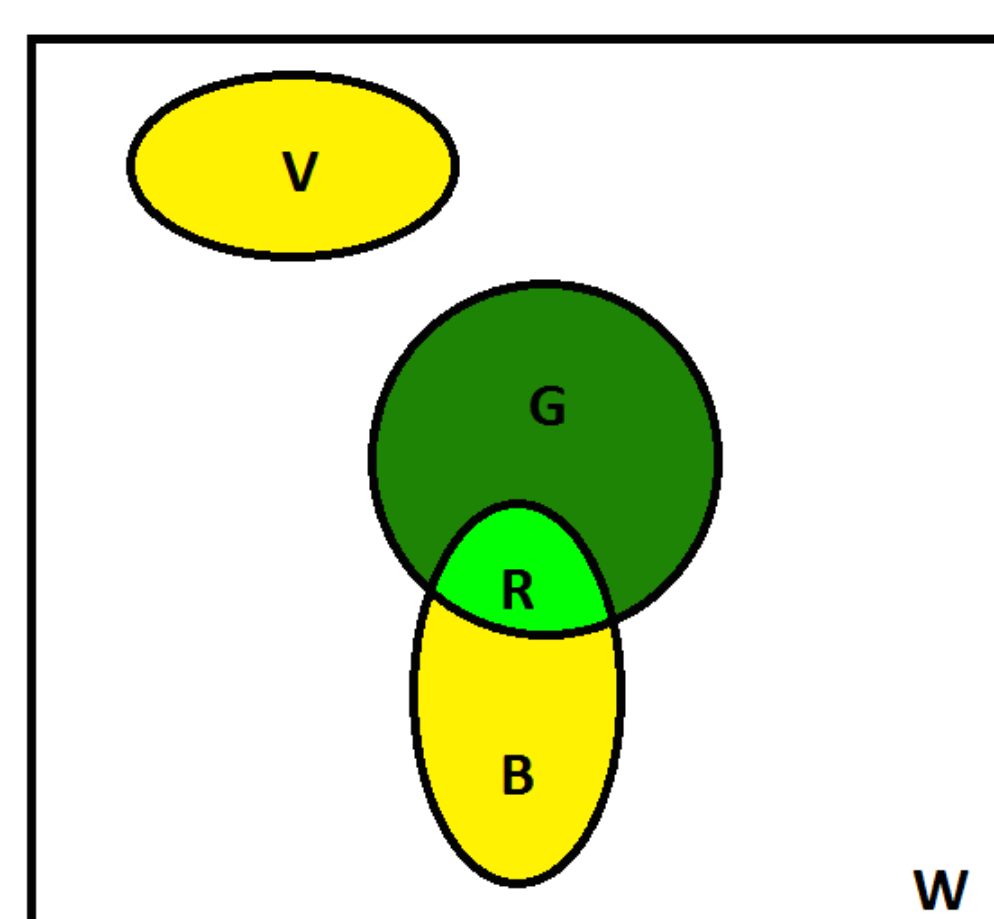

Figure 1

Evaluation with Simulated and Real Data

Simulations are conducted similar to (Smith2009) using a wide range of true smooth signals, in 2D on 32x32 images and 1000 realisations: noise only realisations are used to define FWE rates. We also use HCP working memory task fMRI data from 80 unrelated subjects to evaluate the methods. Ground truth signal is generated from a one-sample t-test on a random pool of 40 "held out" subjects, where the Z image is thresholded at 4 and truncated at 10. The signal+nosie data and noise only data were simulated in two ways. In the first approach, 1000 synthetic noise images are added to the scaled true signal to construct signal+noise data. In the 2nd approach, from the pool of 40 "held in" subjects, a one-sample t-test on 15 randomly chosen subjects is considered as signal + noise data; noise-only data is obtained by random sign-flipping of subject's data. This procedure is repeated 1000 times to derive analogous TP and FP measures as with simulated data. Finally HCP motor task fMRI data is used for demonstration, producing FWE inferences for each TFCE variant (3000 sign-flips).

Results: Simulated Data Evaluation

- True signals

Sphere 3 Spheres Bar Bell Row Of Spheres 3 Penny Real Signal
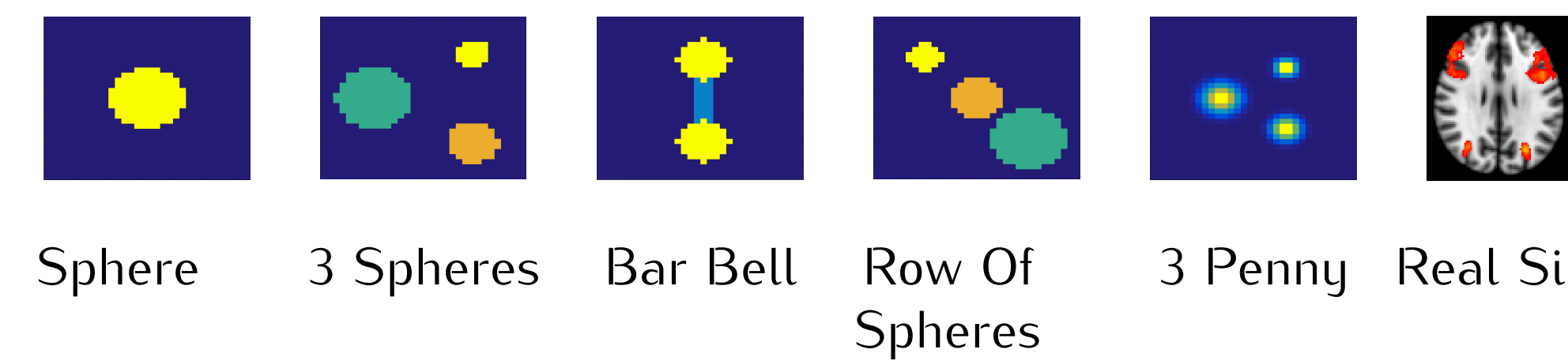
- Sensitivity and spatial specificity on simulated data, smoothing FWHM 2 voxels. $TFCE_1$ and $TFCE_3$ are roughly as powerful, but have better specificity (smaller FP AUC). $TFCE_2$ has very poor specificity.
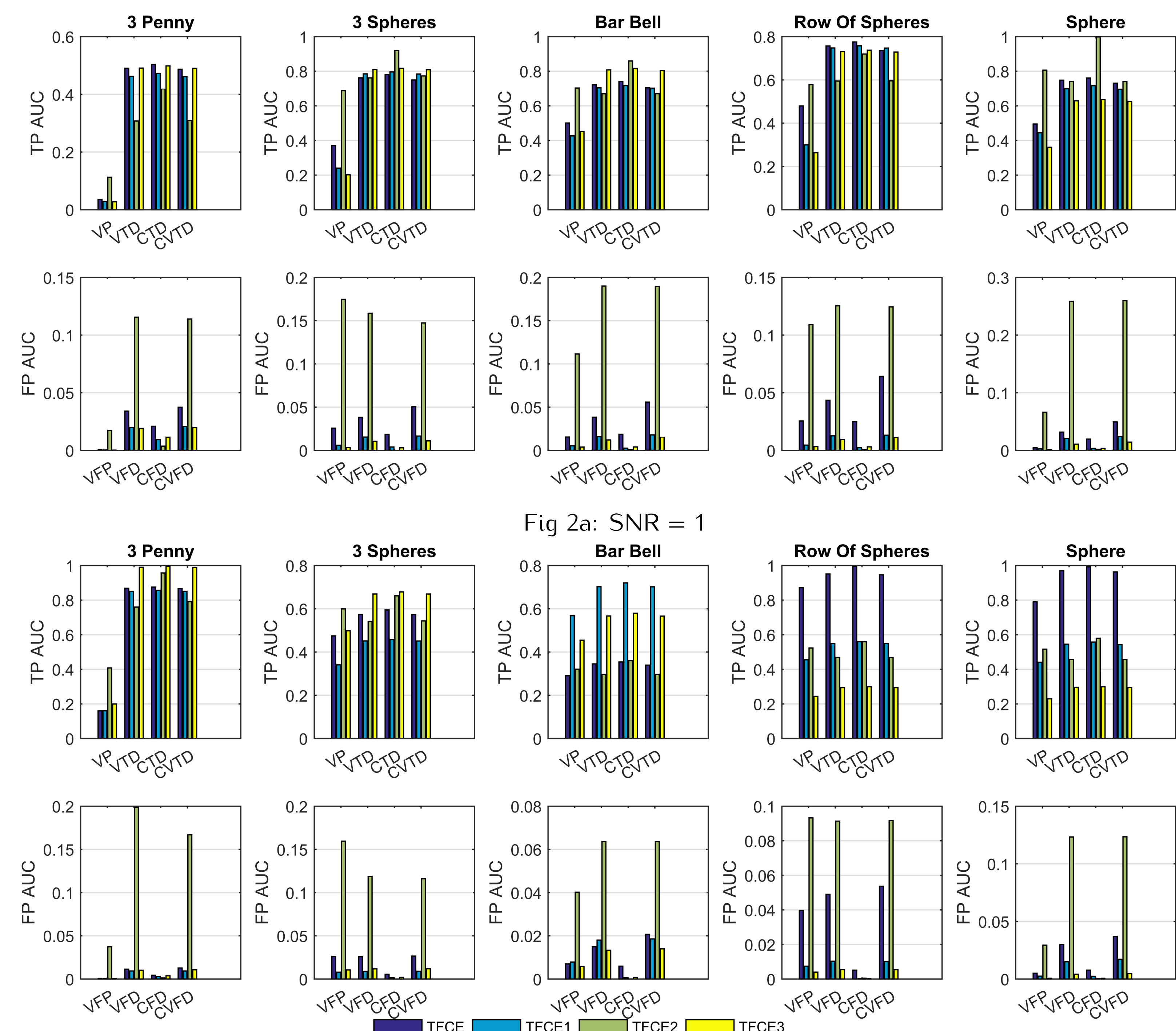

Fig 2a: SNR = 1

Fig 2a: SNR = 2

Results: Real Data Evaluation

- Evaluations using real data, TP measures (Fig 3a) and FP measures (Fig 3b). For comparison with above, we show the real data signal with synthetic noise (Fig 3a left & 3b left); both TFCE1 & 3 have improved power and specificity.
- With empirical signal and noise (Fig 3a right, 3b right), TFCE1 and 3 are comparable though have sightly worse power and specificity.
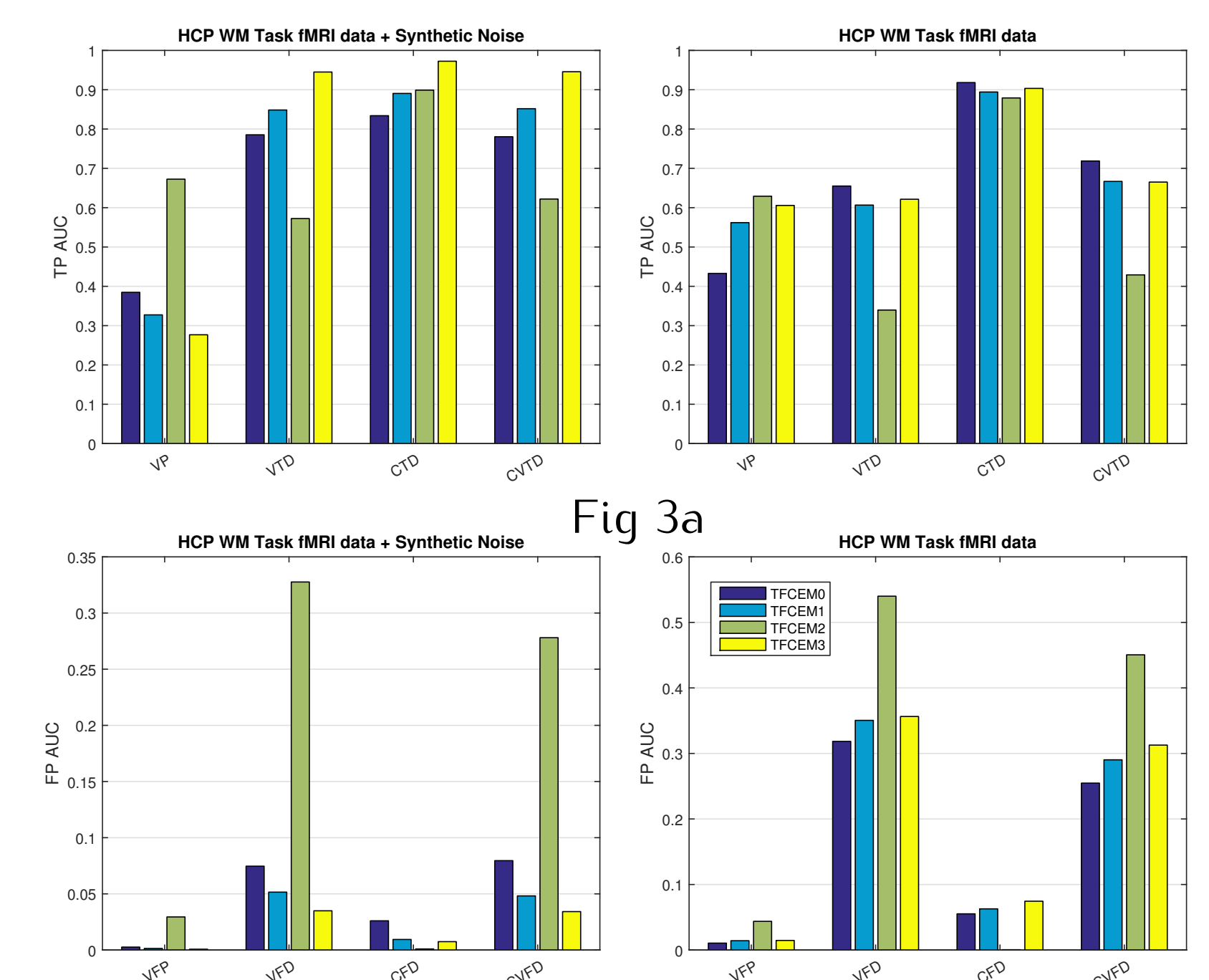

Fig 3a

Fig 3b

Results: Real Data Application

- One-sample t-test on HCP motor fMRI data, 5% FWE level. While $TFCE_1$ and $TFCE_3$ are similar and similar to the original TFCE, $TFCE_2$ has evidence of bleeding (Fig 4).
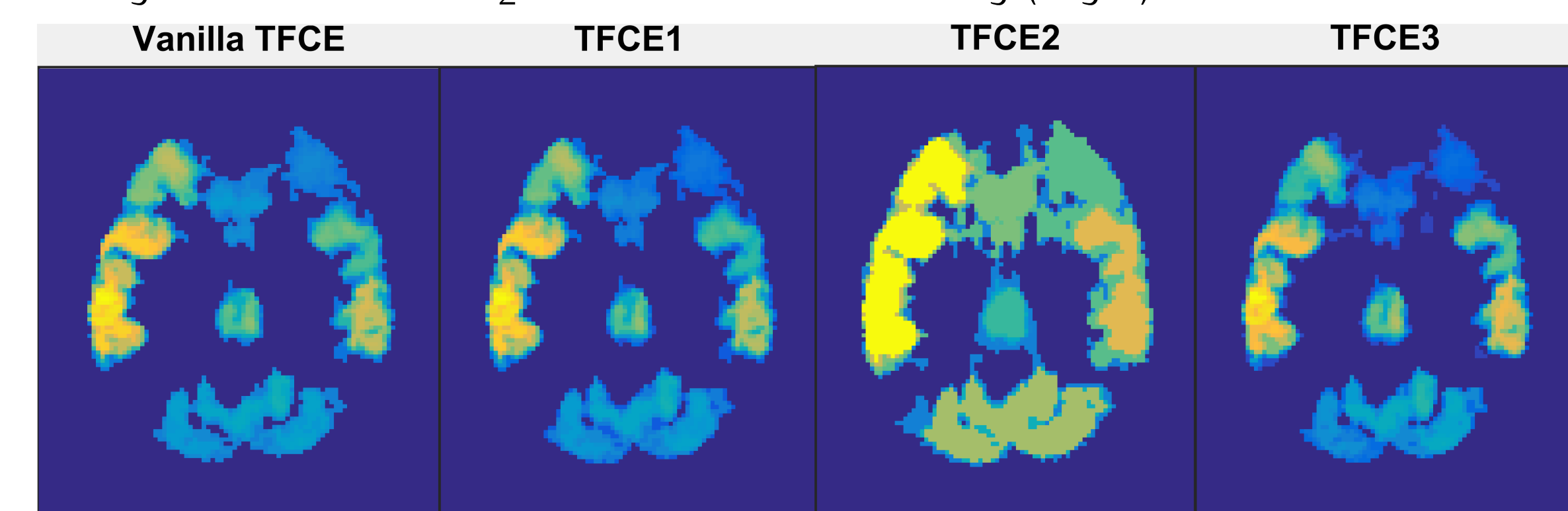

Vanilla TFCE   TFCE1   TFCE2   TFCE3

Fig 4 Percentage of significant voxels are 27%, 24%, 31% and 24% respectively.

Conclusions

$TFCE_1$ and $TFCE_3$, variants that explicitly limit the spatial support of the TFCE statistic, were found to perform about as well as the original, suggesting they can be used when 'spatial bleeding' is a concern without sacrifice of power.

References

S. Smith. et al., Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage, 44(1):83-98, 2009.
C.W. Woo, et al., Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. NeuroImage, 91:412-9, 2014.
Electronic copy of this poster: http://warwick.ac.uk/tenichols/ohbm