MethodsX 3 (2016) 477–482

Contents lists available at ScienceDirect

MethodsX

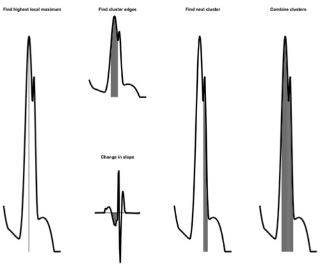
journal homepage: www.elsevier.com/locate/mex

# A landscape-based cluster analysis using recursive search instead of a threshold parameter

Thomas E. Gladwin[a,b,*], Matthijs Vink[b,c], Roger B. Mars[d,e]

[a] Military Mental Health Research Centre, Ministry of Defense, P.O. Box 90.000, 3509AA Utrecht, The Netherlands
[b] Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands
[c] Departments of Experimental and Developmental Psychology, Faculty of Social Sciences, Utrecht University, Utrecht, The Netherlands
[d] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands
[e] Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, United Kingdom

G R A P H I C A L   A B S T R A C T




A B S T R A C T

Cluster-based analysis methods in neuroimaging provide control of whole-brain false positive rates without the need to conservatively correct for the number of voxels and the associated false negative results. The current method defines clusters based purely on shapes in the landscape of activation, instead of requiring the choice of a statistical threshold that may strongly affect results. Statistical significance is determined using permutation testing, combining both size and height of activation. A method is proposed for dealing with relatively small local peaks. Simulations confirm the method controls the false positive rate and correctly identifies regions of activation. The method is also illustrated using real data.

* Corresponding author at: Ministry of Defense, Military Mental Health Care Research Group, PO Box 90.000, 3509AA Utrecht, The Netherlands.

*E-mail address:* thomas.gladwin@gmail.com (T.E. Gladwin).

http://dx.doi.org/10.1016/j.mex.2016.06.002

- A landscape-based method to define clusters in neuroimaging data avoids the need to pre-specify a threshold to define clusters.
- The implementation of the method works as expected, based on simulated and real data.
- The recursive method used for defining clusters, the method used for combining clusters, and the definition of the "value" of a cluster may be of interest for future variations.




ARTICLE INFO
*Method name:* Recursive clustering
*Keywords:* fMRI, Cluster analysis, Threshold-free, Permutation, Recursive, Derivative
*Article history:* Received 31 March 2016; Accepted 30 June 2016; Available online 7 July 2016

## Method details

The method involves three steps: (1) defining clusters using a recursive search function aimed at detecting an upwards change in the differential of the activation, moving away from a local maximum; (2) defining a condition when to combine adjacent clusters; and (3) permutation tests for the whole-brain maximum of a score per cluster that combines size and activation level. Functions from the SPM toolbox (www.fil.ion.ucl.ac.uk/spm) in Matlab [8] were used for reading and saving files and smoothing data.

Clusters are defined as follows, for a given statistical activation map of $-\log(p)$ values derived from a T-map or F-map. First, the voxel with the highest value is selected. Then, a recursive function is used to iteratively visit neighboring voxels, then their neighbors, and so on. Voxels are only visited if they are further away from the peak voxel, to avoid doubling back. New voxels are added until the slope in the value from the previous voxel to the new one is more positive than the previous slope. This procedure thus selects the edges of clusters, which start at the peak and at some point must increase their derivative as they drop in the activation landscape. After a cluster is defined, the voxels in that cluster are excluded from further processing and the cluster surrounding the next highest peak in the image is calculated, until no local maxima remain. Local maxima were defined as any voxel for which all eight neighboring voxels had a lower value.

Since local maxima within clusters may occur, depending on the smoothness of the data, the following criterion was used to combine adjacent clusters into a single cluster. If no activation threshold is used at all (which is unnecessary with the current method, although for purposes of speed a liberal threshold of $p = 0.05$ could be used), but some form of cluster-combination is used, this step is particularly important. With a too-liberal combination criterion, the whole "floor" of the activation landscape will be combined into a single very extensive cluster, which may acquire large values during permutation testing under the null hypothesis. In our method, for each cluster the proportion of the edge voxels that border on a different cluster (ProportionConnected) is determined. If this proportion is above zero (that is, if there is any adjacent cluster), it is determined whether to combine the clusters. Two additional values are used for this: The difference between the peak values of the two clusters (PeaksDifference), and the difference between the peak value of the cluster with the lower peak and the mean activation level at the edge-voxels adjacent to the neighboring cluster (SmallerPeakToConnectingEdge). The clusters are combined under the following condition:

$$\text{PeaksDifference}/(\text{PeaksDifference} + \text{SmallerPeakToConnectingEdge}) \geq 1 - \text{ProportionConnected}$$

That is: As the amount of connection increases, the more likely the clusters will be combined. In the extreme case, a fully surrounded cluster will always be incorporated into the surrounding cluster. Further, combination is more likely as the lower cluster is less well separated: If the lower cluster's peak is not much higher than the connecting flank with the higher cluster, it will be combined. The criterion thus differentiates the case of two clearly separated peaks, versus a bump lying on the flank of a larger hill in the activation landscape.