Available online at www.sciencedirect.com
ScienceDirect
Magnetic Resonance Imaging 25 (2007) 1249–1262
MAGNETIC RESONANCE IMAGING

# Dense mode clustering in brain maps☆

Stephen José Hanson*, Donovan Rebbechi, Catherine Hanson, Yaroslav O. Halchenko
*Psychology Department, Rutgers University, Newark, NJ 07102, USA*
Received 22 October 2006; revised 2 March 2007; accepted 5 March 2007

**Abstract**

A mode-based clustering method is developed for identifying spatially dense clusters in brain maps. This type of clustering focuses on identifying clusters in brain maps independent of their shape or overall variance. This can be useful for both localization in terms of interpretation and for subsequent graphical analysis that might require more coherent or dense regions of interest as starting points. The method automatically does signal/noise sharpening through density mode seeking. We also discuss the problem of parameter selection with this method and propose a new method involving 2-parameter control surface, in which we show that the same cluster solution results from tradeoff of these 2 parameters (the local density $k$ and the radius $r$ of the spherical kernel). We benchmark the new dense mode clustering by using several artificially created data sets and brain imaging data sets from an event perception task by perturbing the data set with noise and measuring three kinds of deviation from the original cluster solution. We present benchmark results that demonstrate that the mode clustering method consistently outperforms the commonly used single-linkage clustering, $k$ means method (centroid method) and Ward's method (variance method).


*Keywords:* fMRI; Clustering; Brain maps; Localization; ROI analysis; Parcellation

## 1. Introduction

Neuroimaging analysis usually consists of two basic kinds of inferences; first, estimates of excursions from a logical baseline (whether in blocks or background event-related time points) collected within the same session, which are often normalized as a common statistic ($z$ or $t$ values) by voxel-wise time series variance; second, the locations or regions of interest (ROIs) that constitute the presumed "activated" spatial clusters due to the subject's response to the stimulus task. Many procedures have been proposed and used for inferring excursion signal and reliability, especially in linear frameworks (e.g., General Linear Model [GLM]), but less has been explored in identification of spatial structure, extent or shape of the ROIs. The majority of the clustering methods in neuroimaging rely on detecting the largest local maximum in the statistic map and then doing nearest-neighbor (often using Euclidean metrics) search to identify cluster membership. Clusters are typically thresholded with the longest link difference between single-link membership sets in a hierarchical cluster history, hence often referred to as "single linkage clustering" in the statistical taxonomy literature [1,2]. Other methods have been proposed and used with neuroimaging data, including $k$ means clustering [3], minimum variance methods (e.g., Ward's [4]), probabilistic kernel methods (so-called Fuzzy clustering [5]) as well as self-organizing elastic maps (SOMs) and neural networks (Kohonen [6]), but none of these methods have tended to enter into general usage, nor do they seem to provide a consensus on what should be considered normative localization in brain maps. Perhaps the closest method to what we will be proposing here and reviewing are the SOMs that Kohonen [7] and others introduced into the neural network literature during the mid 1980s. These methods are, in fact, mode-seeking algorithms as well and represent simply a different framework for introducing concepts such a kernel function, distance metric and group membership rules. The most common methods used in neuroimaging have been focused on the simplest and most computationally efficient algorithms but not necessarily ones that would be best for clustering 3D spatial densities. Consequently, this has led to methods such as single linkage or "connected components" (see below) to appear in most neuroimaging data analysis packages [AFNI

☆ This research was supported by NSF (EIA-0205178) and the James S. McDonnell Foundation.
* Corresponding author. Tel.: +1 973 353 5440x253.
*E-mail address:* jose@tractatus.rutgers.edu (S.J. Hanson).

0730-725X/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.mri.2007.03.013

[8], SPM, [9]; VoxBo (www.voxbo.org), FSL (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl) [10] and BV (www.brainvoyager.com)] as a default due to the fact that they are simple to implement and fast. Unfortunately, these methods are also well known in the statistical clustering literature to be fraught with most difficulties in terms of stability, classification error (in known targets) and density estimation. In this article, we will propose a new class of clustering methods which, in contrast to the commonly used "single-linkage" methods, explicitly searches for density in the spatial brain maps. These methods are based on the assumption that activation in brain maps is fundamentally based on spatial contiguity and spatial variance, and although this assumption is unlikely to be universally true, it must account for much of the neural activity that is expressed in the hemodynamics during mental activity. We are also assuming that spatial clustering should generally occur in the statistical parametric map where spatial structures may be emerging, which require cognitive or perceptual interpretation. This, of course, is a starting point that is arbitrary and could be substituted with any preprocessing that results in some initial signal/noise increase; this could include classifiers [11] or blind or semiblind methods (Independent Component Analysis [ICA], Principle Component Analysis [PCA], wavelet, etc.) that produce differential coefficients that can be associated with voxels. Still, another motivation for developing new spatial clustering methods is to create semiautomatic methods for detecting ROIs that might be candidates for graphical analysis and causal modeling approaches that have become popular in recent years [12,13]; without spatially dense ROIs, the resultant extracted time series are likely to be either incoherent or misspecified, representing neither a well-defined ROI or interactivity with other ROIs. In what follows, we provide a context for our multivariate clustering method as it relates to the general field of numerical taxonomy and how it relates to functional magnetic resonance imaging (fMRI) data in particular, in order to help frame the questions and results to follow in a broader context.

### 1.1. Brief overview of clustering

Statistical clustering originated in the early 1960s with the advent of convenient access to midsize computing environments [14,15] (see also Ref. [2] for more historical perspective), which encouraged implementations of new exploratory statistical analysis approaches. By the 1970s and 80s, a general framework appeared, which laid out the potentially vast landscape of clustering methods. Our class of methods falls into this framework and was first studied by Wishart [15]. Logically, there are two kinds of clustering, agglomerative (data merging) or divisive (data cutting), and both of these are hierarchical in that they follow a merge history linking clusters below or above as inclusive single-member sets. There are also nonhierarchical types of clustering, including seeded clustering, which specifies parametrically the number of clusters as input (e.g., $k$ means) or so-called overlapping clustering methods [16], in which members can be in more than one cluster. Within the most general kind of clustering method, there are basically two elements, which can be combined to produce a large array of clustering algorithms. The first is the similarity or dissimilarity measure that can be used between objects to be clustered. The second is the group membership rule, which describes the way in which new members are added to clusters, given two or more cluster membership candidates. In terms of similarity or distance measures, there are potentially an infinite number of such measures as long as they conform to the distance axioms ([17] identity, symmetry, triangle inequality) and what is often called the "ultrametric axiom," expressed in the next equation:

$$d\{x,y \leq \max[\mathrm{d}(\mathrm{x},\mathrm{z}),\mathrm{d}(\mathrm{y},\mathrm{z})]\}$$

All hierarchical clustering, in fact, results in ultrametric trees, often called "dendrograms," and allows the relative distance to be measured between all clusters and their members. Distance measures often used are typically part of the Minkowski family: Euclidean ($L=2$), city block ($L=1$) or L-Infinity. for binary features; hamming distance is often used. In terms of group membership rules, there are three common ones used [object $x$($y$ or $z$) distance to cluster $u$):

1. Single linkage $d(x,u)=\min[d\ (y,u),d(z,u)]$
2. Complete linkage $d(x,u)=\max[d\ (y,u),d(z,u)]$
3. Average linkage $d(x,u)=(1/2)*[d\ (y,u),d(z,u)]$

Combining distance metrics with group membership rules produces a large combinatory set of clustering algorithms. If we focus on group membership rules for the moment and fix the distance metric to be Euclidean, there are some common observations that have been made in the statistical clustering literature.

First, in average linkage clustering, bias can result when clusters are not multivariate gaussian but are, in fact, more complex cluster shapes (long spindles, bananas, etc.) or nonconvex. Nonetheless, average linkage is resistant to noise and can be very stable when cluster shape distributions approximate Rayleigh PDFs.

Second, complete linkage is the most conservative of the group membership rules and will tend to miss clusters structure and, in the worse case, create a large number of small fractionated islands ("archipelagos"). Finally, single linkage is the most liberal of group membership rules and will tend to create long spindling strings (often called "chaining" in that a single member ties to larger cohesive clusters together), with no modal central tendency. If there is no interesting density or underlying focal structures, single linkage is probably appropriate. However, single linkage is highly sensitive to noise, extremely unstable and cannot locate dense regions in cluster space.

In general, all the group membership rules discussed so far have no explicit way to detect density; rather, they are focused on spatial distribution without further noise

reduction. The motivation for density-based clustering methods is in resolving this conflict between noise filtering and flexibility. Density-based clustering methods work by identifying dense collections of points. The collections need not be spherical; they could have a wide diameter in a single dimension, which would probably be disallowed in a variance minimizing method. Yet, the fact that density is the criterion means that these methods are not sensitive to sparse noise. In the early 1980s, self-organizing maps were introduced, which are basically a special case of the general density mode estimators introduced first in the mid to late 1960s by Wishart [15], which provides the context and starting point for our new dense mode clustering methods.

### 1.2. Mode density clustering

One of the first mode density clusterers was proposed by Wishart [15] in a paper entitled "One Level Mode Analysis." In this seminal piece of work, Wishart asserted that "clustering methods should be able to detect and resolve distinct data modes independently of their shape and variance." To motivate this argument, we show in Fig. 1 his original example, where star clusters are plotted in a 2D temperature and luminosity plot showing that the cluster structure is modal and linear in shape. Worse, the two star groups ("dwarfs" and "giants") intersect causing most gaussian or minimum variance clustering to separate clusters at the midpoint of the spatial distribution. This type of data often is related to another distinction that is made in clustering between parametric and nonparametric clustering. In the former, the data are assumed to be generated from a known density such as multivariate gaussians, and hence, the task is to estimate the parameters of the mixture of these densities. In the nonparametric case, clusters are assumed to be associated with the modes of some unknown density. In Wishart's proposal, the goal of the nonparametric mode clusterer is to find the modes and assign each observation to the "domain of attraction" for each mode.

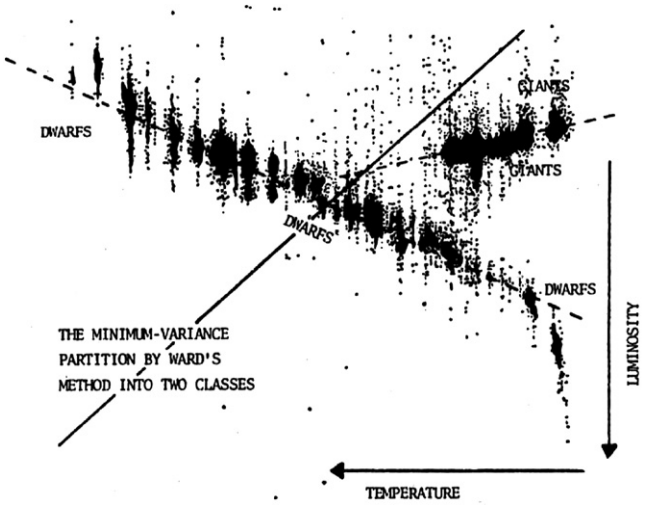


Fig. 1. Wishart's problem due to H.N. Russell, showing the distribution of star clusters (dwarfs, giants) as a function of temperature and luminosity.

In the 1D case, Wishart proposed fitting the data with a kernel density estimator and then successively thresholding the density, setting aside data in each "well-separated" mode until all levels had been identified. Once all modes were found, then high-density observations previously set aside could be clustered using single linkage, since the data were already conditioned on being from the same mode. This proposal had been reinvented a number of times in the past 30 years, including the ideas put forth under the term *sharpening* [18]. Still another density estimation approach recently proposed is often used for image segmentation and is called "mean shift" analysis [19]. Like the Tukey and Tukey method, "mean shift" is a standard density estimation approach, which iteratively locates gradient directions in 2D images in a given spatial scale that identifies modes; these are subsequently pruned using some threshold heuristics, and then clusters are identified, creating a segmentation of pixel intensities. Where there are sharp boundaries, this method can be very useful; however, if unit intensities are slowly graded, the method is unlikely to be very useful. Also, due to the gradient search method at the heart of "mean shift," it does not scale well in higher dimensions. This is in contrast with the present method, which is an exhaustive search but in a limited set of spatial scales. Note that in the present algorithm, we will not be choosing a scale or "steering" in scale space, but rather, we will rapidly search a limited range of spatial scales exhaustively to optimize a sparseness criterion. In terms of relevant neuroimaging methods, it is worth noting the work by Stanberry et al. [20], which explicitly used the Tukey and Tukey sharpening methods on the a resultant cluster tree (dendrogram) produced in hierarchical methods (e.g., single linkage). Their method within the same family of the present proposal, however, focused on temporal clustering as opposed to spatial; they unfortunately selected single linkage, which we will see introduces biases towards low-density clusters and did not provide a general control surface for their dendrogram-sharpening algorithm. In general, we believe that Stanberry et al. were on the right track in exploring the space of possible dense mode algorithms similar to the earlier proposal of Wishart [15], which we further explore with this new method. Wishart's methods did not gain wide acceptance at the time due to the choice of an arbitrary cut level, which was in the user's control. Wishart proposed a "hierarchical mode analysis" in order to identify potential cut levels for the modes; unfortunately, this involved a somewhat arbitrary and convoluted iterative merging process that attempted to estimate the cluster dendrogram of the kernel density estimate, with no guarantee that the algorithm would converge on the correct tree, except in some of the simplest cases. Nonetheless, we find the idea of mode density clustering in brain imaging data compelling and pursue a generalization of the method here. Basically, we estimate modes using a spherical kernel centered on each active voxel and use a local voting procedure to elect "dense

points." In order to avoid potentially arbitrary threshold problems, we exhaustively search, in a highly dense grid, for all possible kernel radii and thresholds producing an interpolated control surface. We further "sharpen" the modes by imposing a "sparseness" function on the control surface and maximize it. Surprisingly, this produces a single peaked function with effectively the same cluster solution at multiple radii/threshold combinations only along the maxima. What we provide next is the detailed mode clustering algorithm we call "dense mode clustering" (DMC), but before that, we provide a relevant side discussion of spatial vs. temporal clustering in order to further define the kind of method we have developed.

### 1.3. Spatial vs. temporal clustering

Logically, there are two possible ways to cluster a voxel position by scan time matrix, one in the spatial dimension (e.g., $x$,$y$,$z$) and the other in the temporal dimension by scan time for each voxel. For methods dependent on rank of the matrix, such as PCA, the shortest dimension in this case would be scan time, which is often an order of magnitude less than the number of total voxels, even if voxels are first thresholded as significant or "active." Over whole brain, this might still be in the range of 1000 to 10 of 1000 s of voxels, where scan times may be more typically in the order of 100 s. Moreover, spatial similarity may or may not be consistent with temporal similarity. If two voxels are modulated by the same regressor, this may occur in the same part of the brain or very different parts of the brain. It should be obvious that dense regions that often are taken to define functional regions are also likely have high temporal coherence. Any temporal similarity between voxels must be established by similarity over spatial voxel variation, while spatial similarity is dependent on scan time resolution variation. In any case, it is clear there must be some feature reduction in the spatial domain in order to establish reliable distance measures. This can be done by starting with a $z$ map or more sophisticated (but possibly less interpretable) feature selection process (e.g., ICA, PCA, etc.).

### 1.4. The mode clustering algorithm: DMC

The present DMC begins with any convenient-feature map; often, a liberally thresholded $z$-stat map will be an appropriate input (although arbitrary, every clustering method must start with some thresholded map; however, DMC could, in principle, use raw input). In Fig. 2, we show the various steps involved in the present clustering algorithm. First, we will describe the method in terms of filtering and density estimation, and then, we provide below the detailed pseudocode and more formal description of the method, which would allow replication of the method (code is also available online: www.rumba.rutgers.edu/soft/dmc). Starting with $r$ values near the minimum resolution of voxels in the map (~2–3 mm) to half the typical brain (20 mm) in steps of .5 mm, we do an exhaustive search for each $r$ over a sensitive range of $k$ values. This produces a ($k$,$r$) control surface,

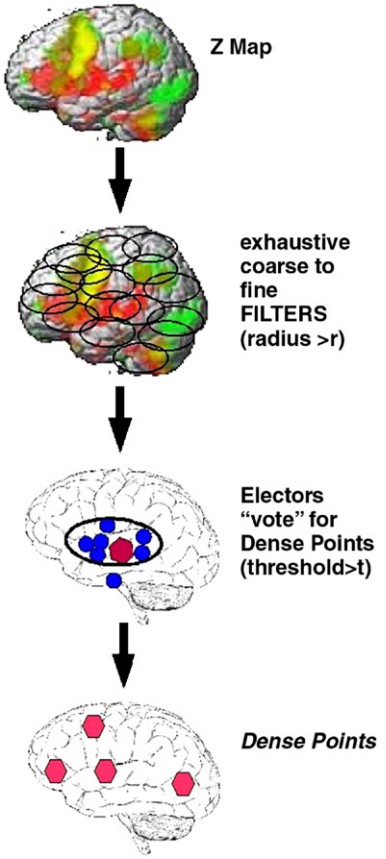


Fig. 2. Overview of DMC algorithm this process is repeated f or all ($k$,$r$) values.

which allows further optimization of density sensitive indices. The cluster method proper begins with one set of ($k$,$r$) values, which, in effect, places a spherical surface at every active voxel. The voxels coincident in each spherical region are considered to be "electors" for the central voxel, and if there are at least $k$s present, then the central voxel is proposed as a dense point. Voxels that are not dense points are eroded, leaving behind a minimal set of dense points that have mutually reinforced the estimated underlying nonparametric densities. What follows next is a more formal description of the method (we show the pseudocode in Appendix A) that could be used to reimplement our method. The clustering algorithm has two parameters: a threshold $k$ and a radius $r$. A point $p$ is said to be ($k$,$r$)-dense if there are at least $k$ other points of distance no greater than $R$ from $p$. The algorithm consists of the following two phases:

1. The introduction phase. For each dense point $p$, if there exists a cluster that contains at least one point of distance less than $R$ from $p$, then $p$ joins that cluster. In order to prevent any indeterminacy of cluster introduction, we further require that, if there are several such clusters ($c_1$, $c_2$,...$c_n$) that are within distance $R$ from $p$, then all clusters ($c_1$,$c_2$,...$c_n$) are merged to form a new cluster, which includes dense point $p$.

2. The merging phase. After the introduction phase, there are several clusters. We perform the following procedure until no merges are performed for each distinct pair of clusters $(c_i,c_j)$: let $p$ and $q$ be in $c_i,c_j$, such that

$$\forall p' \in c_i, q' \in c_j, d(p,q) \leq d(p',q')$$
$$a = 1, n_j \sum x \in c_i \; d(p,x) \quad \text{and let} \quad b = 1, n_j \sum y \in c_j \; d(q,y)$$

where $n_i$ and $n_j$ are the number of elements in the clusters $c_i$ and $c_j$. If $d(p,q)<(a+b)/2$, then the clusters are merged — we call this the "Romeo and Juliet" rule.

### 1.5. Romeo and Juliet merging rule

In DMC, we introduce a new cluster merging rule that is a variant on complete linkage. In what we call the Romeo and Juliet rule, clusters are merged when their nearest members are closer than the average distance between all members of each of the clusters (shown in Fig. 3). This type of rule ensures that clusters are only merged, which tend to increase the local density of each cluster overall ensuring density maximization. Next, we provide a new approach to cluster identification using the $(k,r)$ control surface.

### 1.6. The control surface

#### 1.6.1. Identifying the cluster solution

One of the dilemmas with a parametrized clustering method is resolving the question of selection of parameters. For an $n$ parameter method, there is an $n$-dimensional family of clustering solutions. So, to justify a particular choice of parameter value, one needs to either argue that the choice is arbitrary (that other parameter choices would give a similar answer) or that the choice is optimal. To argue that a choice is optimal, it is useful to have an objective function to appraise different solutions. For mode clustering, the objective function we will use is the pseudo-$F$ ratio of the total mean squared, nearest-neighbor distance between clusters, compared to the total variance within clusters. What makes this a pseudo-$F$ ratio is the choice of nearest-neighbor distance over centroid distance. The reason to prefer nearest-neighbor distance to centroid distance is the numerator should reflect the degree of separation between clusters. If some clusters are very large, the variance of the centroids will not reflect this. We can plot this function of the radius $r$ and the threshold $k$. The plot is called the "control surface." One of the first observations is that there is not a single optimal solution. Instead, there is a ridge with an

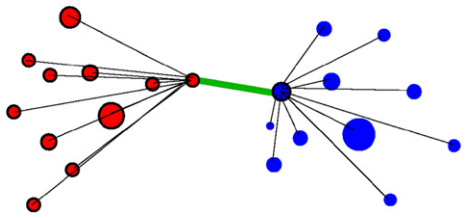

Fig. 3. the Romeo and Juliet cluster merge rule. Here, the green link represents the link between nearest neighbors of the red and blue groups.

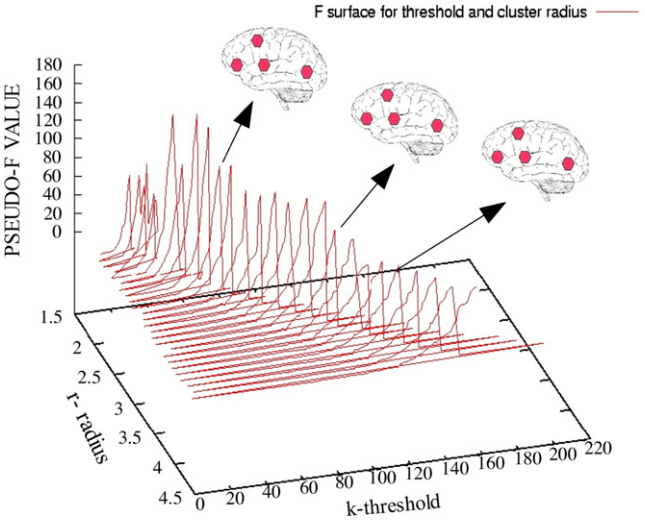


Fig. 4. Control surface for typical exhaustive clustering for $r$ and $k$; note the maximum $F$ value centered in the control surface. Each point represents a cluster solution volume, which, along the maximum point $(k,r)$, is highly similar in location, shape and size.

optimal threshold for each value of $r$. The control surface degenerates at the extremes, as $r$ gets smaller than the bold resolution or large enough for clustering to become too coarse. In this family of solutions, the threshold increases linearly with $r^3$, as one might expect (if we view $k/r^3$ as a density threshold, each solution on the ridge has the same density). Another observation is that reducing $r$ tends to increase the ratio without obvious improvement in the quality of the solution. In fact, any family of solutions that run parallel to the ridge are qualitatively very similar in appearance (see Fig. 4). Consequently, use of the control surface reduces the parameter selection issue to a 1D problem — choosing an appropriate value of $r$. An appropriate value of $r$ should reflect the Blood Oxygenation Level Dependency (BOLD) resolution. It needs to be in the range of nondegenerate solutions, and in the interests of good localization, a good choice of $r$ would be the lowest value, which permits stable control surface behavior. Empirically, allowing the 26 nearest neighbors works quite well. For a 3.375×3.375×4-mm volume, the resulting radius is 7.2 mm. When multiple subjects are used, allowing a small margin for differences in normalization, a radius of about 8.2 mm seems to be effective. Since the choice of $r$ is independent of the outcome of different clustering methods, it is only necessary to run a variety of thresholds for a single fixed radius. This greatly reduces the amount of computing resources required by the method.

An example of the signal/noise sharpening of DMC is shown in the Fig. 5, which on left is a single-subject thresholded ($P<.01$ corrected) $z$ map of bilateral finger tapping. Note the high level of noise. On the right-hand side is the output of DMC on the original $z$ map. Significance values of DMC maps can be assigned using standard cluster-size statistics that are commonly used in most statistical packages [21–23] (but see discussion below concerning cluster size/intensity based statistical inference). Finally, further

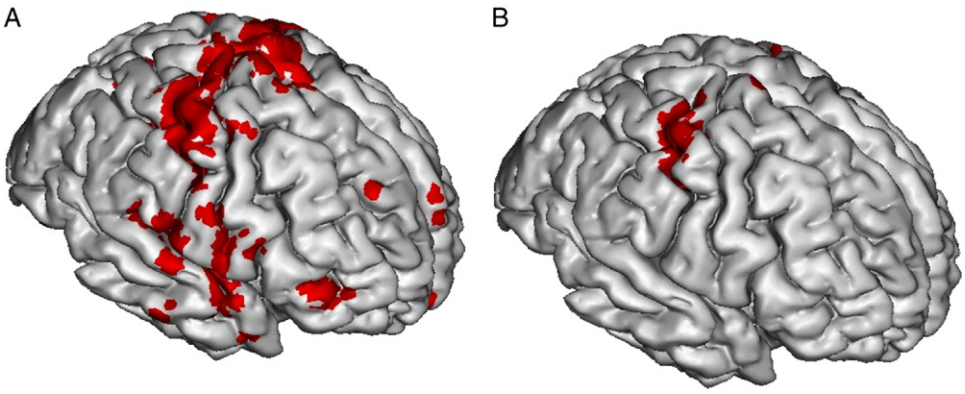


Fig. 5. Example of DMC application to bilateral finger tapping. The figure on the left (A) shows thresholded ( $P<.01$ corrected) $z$ values for a single subject. The figure on the right (B) shows the application of DMC to the original $z$ map.

thresholding will not, in general, produce results similar to DMC unless the local density is homogeneous throughout the brain map, which, for brain imaging data is highly unlikely.

## 2. Methods

### 2.1. Example artificial data

We consider next the response of the dense mode clustering on two kinds of artificial data, which are common fragments one might see in brain imaging data. Both cases are constructed from nongaussian random spatial processes [24,25] and possess nonisotropic variance commonly seen in fMRI data. The first case represents long thin closely contiguous clusters separated by a gap similar to the width of the smallest variance. The second case is similar to the first, except that the source noise is smaller per cluster, creating tighter strings of data points and a weakly joined connective noise cluster about twice the width of either source cluster (Fig. 6). We apply each hierarchical clustering methods (Ward's, single linkage) to each of the benchmarks and show the results in a series of graphs in the next results section, and we also apply $k$ means by seeding the method with 2 clusters and finally show the results of DMC in the final panels later in Fig. 8 (below).

### 2.2. Example experimental case

#### 2.2.1. The event perception task

In order to provide a diverse set of ROIs, we will use a cognitive comprehension task that requires subjects to "parse" familiar action sequences into action clusters bounded by event change points. Six subjects were shown an animated movie consisting of a geometric shape moving randomly among other geometric structures. Subjects were asked to press a button when they identified an event change.

Scanning was performed with a 3T Siemens Allegra head-only MRI scanner (Erlangen, Germany). We used a 3D magnetization-prepared rapid acquisition gradient-recalled echo (MP-RAGE) T1-weighted scanning sequence with 2-mm isotropic resolution to acquire structural images for each participant. A T2*-weighted asymmetric spin-echo, echo planar sequence with flip angle of 90° and a 30-ms time to echo were used to acquire functional data. There were 32 4-mm slices, with each slice consisting of $3.75\times3.75$–mm cells in $64\times64$ grid acquired in whole volumes. The time to repetition for each volume was 2 s. All data analysis was done using FSL [10]. MP-RAGE for all subjects were registered to standard atlas using FSL's FLIRT subprogram to the Montreal Neurological Institute (MNI) atlas template. Data were realigned and detrended also using the standard FSL FLIRT tool. All localization was done in the Talairach and Torneax atlas [26] using appropriate affine transformations from MNI-registered T1 and BOLD images. Subjects responses in this task are highly correlated and with strong between-subject

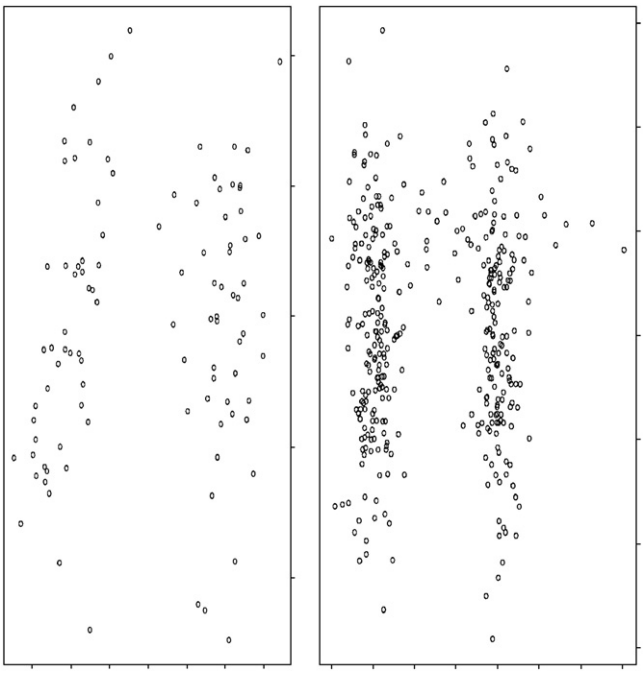

Fig. 6. Two artificial data sets consisting of nongaussian clusters and nonisotropic variance; in the left top panel are two thin clusters with unequal variance; in the top right panel, two weakly connected clusters with higher density central modes.

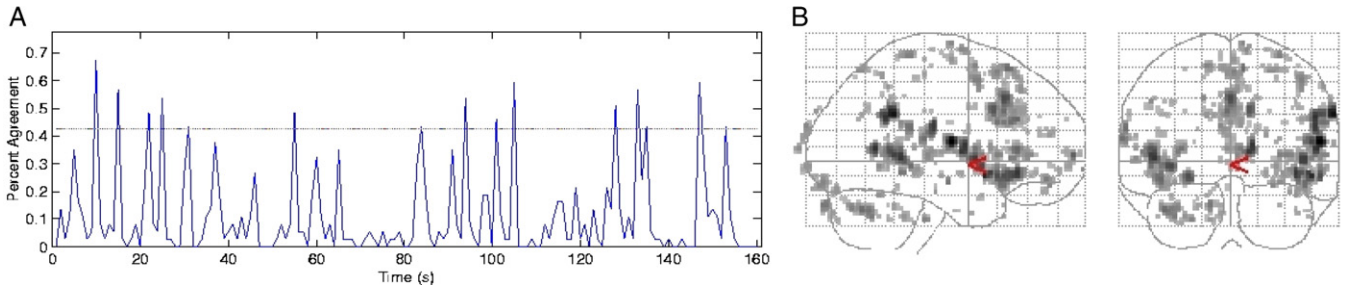


Fig. 7. Panel (A) shows a typical groupwise TRD (temporal response density) showing the response of 22 subjects indicating a "significant event change" in the familiar action sequence. The dashed line shows significant excursion above a $z$ value with $P<.01$ indicating approximately 10 event change points. In panel (B), we show the GLM using the TRD regressor and the resultant brain activity at $P<.05$ (corrected) shown in two views.

agreement on change points as well as within-subject parsing rates. The subject responses were used as regressors to perform GLM analysis. Analysis was performed out using FMRI Expert Analysis Tool Version 5.1, part of FSL. The following preprocessing was applied: high-pass temporal filtering (gaussian-weighted straight line fitting, with sigma=12.5 s). Time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model [27]). Registration to high resolution and standard images was carried out using FLIRT [28,29] using 3- and 9-parameter models, respectively.

### 2.3. Experimental design

First, a group response function [temporal response density (TRD)] was formed by forming the histogram of all subject responses, using 1-s bins. So, the group response function would have a value of 3 at $t=10$ if three subjects pressed the button between 10 and 11s after the commencement of the movie. The subjects' own response TRD (which is binary) was weighted by the group response function producing a mixture of group/individual agreement. This group/individual weighted response was further convolved with the Hemodynamic Response Function (HRF) and used as a regressor for the GLM analysis, thus indicating selective tissue per subject's perceptual event change points. In Fig. 7, we show an example of a group TRD (based on 22 subjects), which was used in the construction of the regressor discussed before and a corresponding SPM glass brain plot at $P<.05$ (corrected).

### 2.4. Adding noise to data

A uniform random number generators was used to add 100, 500, and 1000 random points in space to the masks. We have also used gaussian noise as well as more point centered noise such as Laplacian, which produces similar results; however, uniform noise produced the least bias toward any particular clustering algorithm. The noise bench mark performs clustering on each of these data sets and looks at two different indices of noise that measure the quality of the clustering:

1. The number of imposters — voxels that come from the noise set and not the original data [this is identical to a false alarm rate with highest hit rate — thus a maximal point on a receiver operating characteristic (ROC) curve]. Note that even a perfect clusterer will include some imposters, because by chance, several of the imposters will fall in or near the original clusters. But generally, fewer imposters is better.
2. *Volatility*: we compare the centroids and the numerical count of the overlaps and symmetric differences of clusters in the methods. The less these variables change with the addition of noise, the better. This is also a stability test, which allows one to assess the perturbation of the true clusters to noise.

These noise levels are chosen to significantly stress the clustering method and reveal its inherent stability and type I/type II error profile (or ROC measure) rather than what might be considered "typical" noise levels.

### 2.5. Parameter selection

Because DMC automatically erodes or removes voxels that do not pass the dense point threshold, we allow all other cluster algorithms to also erode (based on rank order similarity) voxels at a similar rate to DMC to make proper inferential comparisons. For the $k$ means clusterer, we use an erosion with an initial value of $k=20$, since for the event perception data, DMC tends to identify approximately 20 clusters (although, based on contiguity, there are about 5 or 6). The erosion constant is also set to 20%, which is similar to DMC's overall erosion loss in this same data set. For the Ward's based clusterer, in order to reduce computational overhead, we begin with an initial $k$ means clustering with $k=1000$ and merge until there are 20 clusters again similar for the DMC solution. For the DMC, we use a value of $r=7.2$ and select the value of $k$ that maximizes the pseudo-$F$ ratio, as discussed above.

## 3. Benchmarking

### 3.1. Imposter voxels

For each cluster, we track the number of voxels that come from the added noise. This is done as follows — the implementation of all of the clustering methods we have used allow the addition of additional information to the data points.

So we tag each point with a "subject ID" field, while a separate subject ID is assigned to the noise data. We then compare the number of voxels from this "noise subject" across different methods, for different numbers of noise voxels.

### 3.2. Symmetric difference

We can compute set-symmetric differences as follows: first, it is possible to make a correspondence between clusters by assigning each cluster in the no-noise condition to a cluster in the noise condition, by using nearest centroid. This is not necessarily a one-to-one correspondence, but in the case of low volatility, it is. Then we examine the intersection of the two corresponding clusters. We would expect that the difference should be small, and consist mostly of noise voxels.

### 3.3. Comparing to specific clustering methods

Erosion was done for all benchmark clustering comparisons ($k$ means and Wards), which uses an erosion process

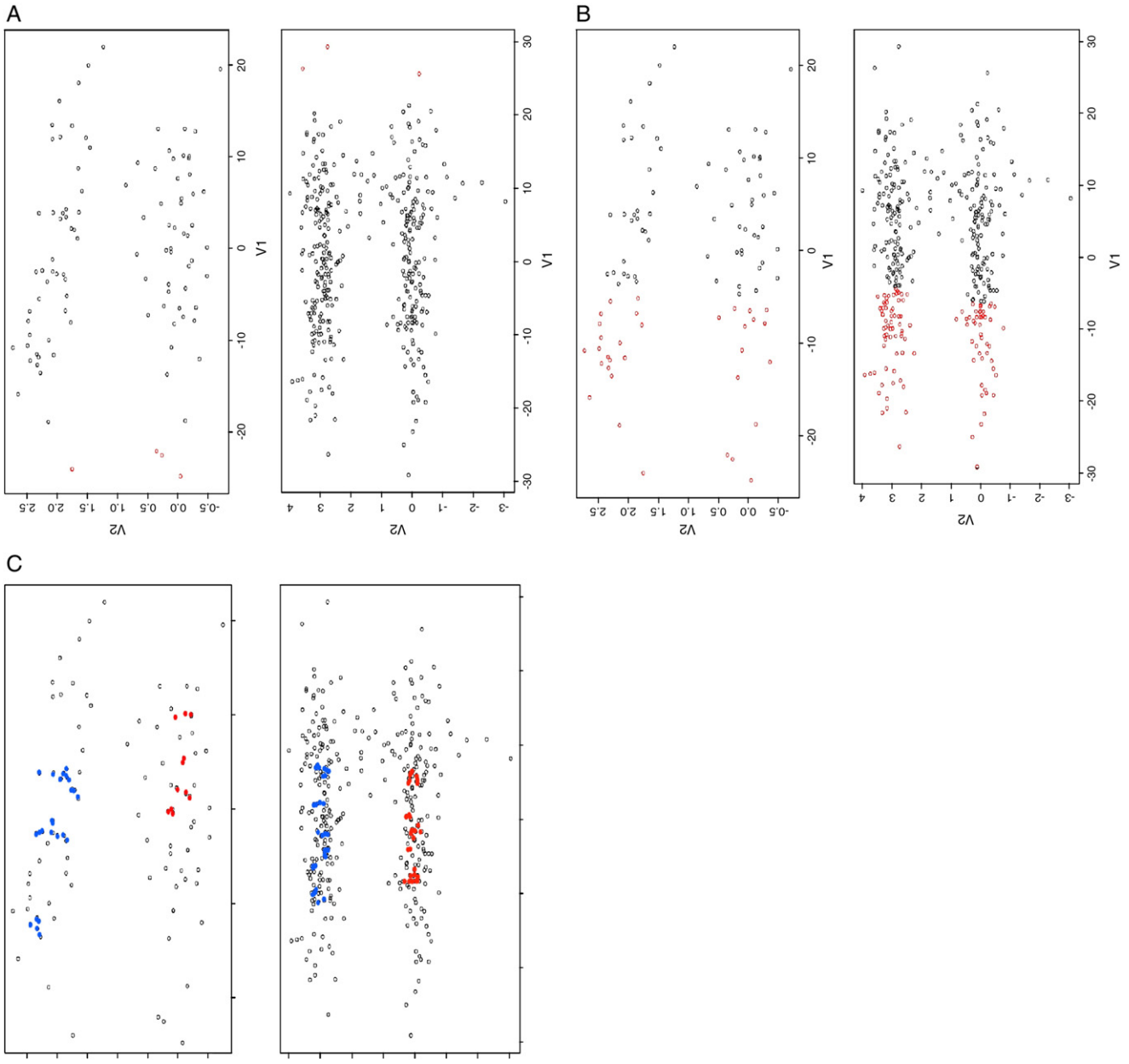


Fig. 8. Benchmark tests of two artificial data sets consisting of nongaussian clusters and nonisotropic variance shown above in Fig. 6. In panel (A), we show the results for single linkage clustering in which the tendency for this method to "chain" produces one large cluster in both cases, with a small outlier cluster. In panel (B), we show the results for Ward's clustering, which with $k=2$ in $k$ means will produce roughly the same outcome and, in this case, splits the distribution in two parts on the dimension with the largest variance while missing the main modes of the data. Finally, in panel (C), we show the typical DMC output on these two benchmarks, with erosion parameters set to 40% of the data set; as erosion is set to 0%, DMC detects both modes and misses less than 3% of the two clusters. Note that DMC finds local density within each group.

that operates after sorting the points by distance from centroid of the cluster that they have been assigned. Only voxels in the top 20% are retained. This parameter is based on DMC, which empirically erodes about 80% of the voxels near a dense point. Erosion will vary with the original noise in the data set and the level of thresholding used as input for the clustering method. In the present data, we started with a very liberally thresholded ($P<.1$) $z$ map, hence an associated high level of erosion. Single linkage technically has no centroid and typically produces a single-cluster; erosion, therefore is not applicable. Besides single-linkage clustering, we will also compare to one of the more popular minimum variance methods, that of Ward's method. This method is very similar to $k$ means — the difference is that $k$ means is a seeded method with a fixed number of clusters, whereas Ward's is agglomerative, with a variable number of clusters. Indeed, these are somewhat related, and we will use the $k$ means method (with a large $k$, e.g., $k=1000$) to initialize Ward's method, as Ward's is prohibitively slow for very large data sets. Another simple approach is an agglomerative single-linkage method that merges the two closest clusters in nearest-neighbor distance. Similar to this is the simple approach of simply taking connected components [30]. The advantage of the two variance-minimizing approaches is that they are very effective in noisy data. They are good at identifying regions that contain a tightly grouped set of points and are not prone to chaining. The disadvantage of these methods is that they are somewhat biased towards finding spherical clusters. On the other hand, single-linkage clustering is very permissive in allowing oddly shaped clusters. This is useful in circumstances where the clusters are likely to be nonspherical, e.g., contiguous regions of suprathresholded voxels. However, it is easily affected by noise and very prone to chaining as we have previously discussed. DMC has the potential of a compromise between flexible cluster finding and noise reduction. In what follows, we provide benchmark comparisons between aforementioned clustering algorithms (single-linkage, Ward's, $k$ means) and DMC.

## 4. Results

Next, we provide comparisons between DMC and the other clustering methods when noise is added, and their effects are measured by the benchmark measures discussed before (symmetric difference, centroid deviation and imposters).

### 4.1. Qualitative results

Following the benchmark in Fig. 6, we show in Fig. 8A the results for both cases using the single-linkage merging rule. In both cases, as expected, single linkage found one large cluster and then one small (3–4 points) and, in neither case, detected the actual modes. In panel B, we show the result of Ward's analysis on the two benchmarks; in this case, the dimension with the largest variance provides a split point for the clustering. This of course masks the actual modes in data set. Starting k means with 2 clusters produces a similar outcome to Ward's. Finally, we show the DMC solution which does appear to detect both the modes and shape of the underlying clusters. Without erosion, DMC

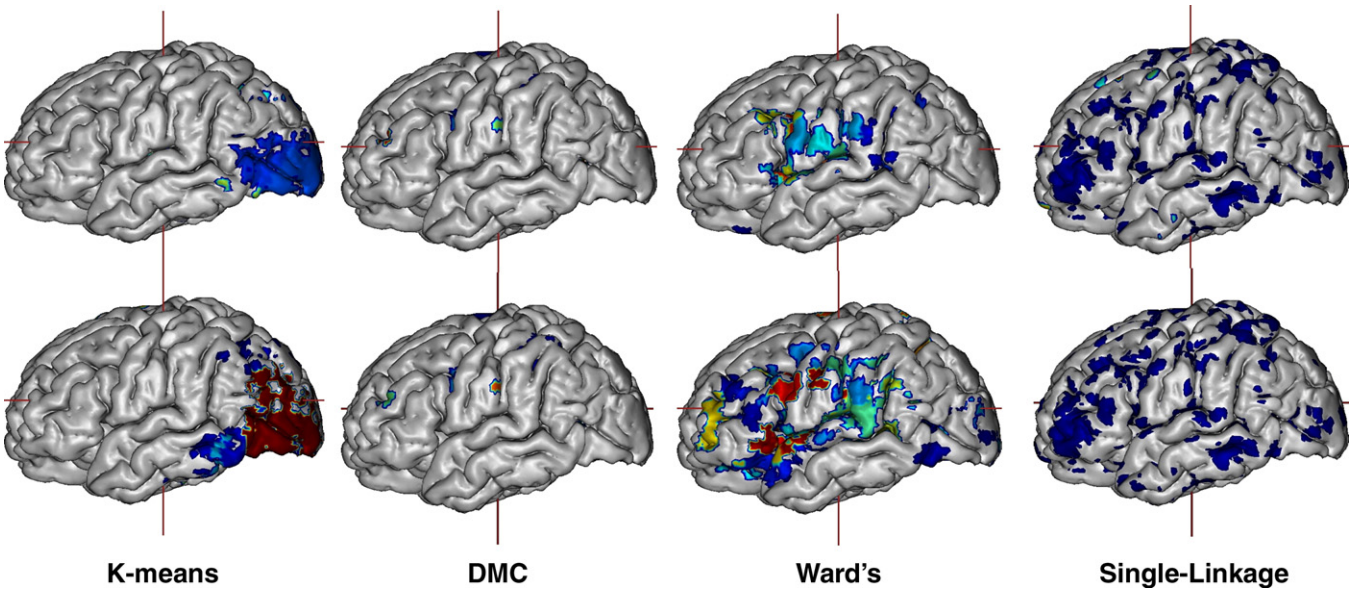


Fig. 9. Benchmark cases: shown above are the results for no-noise clustering (first row) of the benchmark data set and the extreme noise condition when 1000 voxels are added uniformly to the volume (second row). Note the wide diversity in clustering solutions of the same data set. Note that the orientation is maximized visually to show the largest clusters for that given type of clustering and is fixed for before the noise was added and after the noise was added. Note that single-linkage cluster distribution corresponds to the glass brain distribution in Fig. 7.

actually detects more than 97% of the modes in the two thin cluster cases and splits the thin noise island near the midpoint. In the next test, we run the clustering methods on data typical of a integrative cognitive/perceptual task.

### 4.2. Qualitative comparison

The first thing to note is the wide diversity across clustering methods with the same data set. The first row of Fig. 9 shows, from left to right, the cluster solution for *k* means's DMC, Ward's and single-linkage. For Ward's, DMC and *k* means, all six subjects were clustered concurrently, and results show the common spatial clusters across subjects. Single-linkage does not scale with voxel number well and, consequently, cannot handle more than one subject at a time; shown therefore in the last column is a typical outcome for one subject, which corresponds to the glass brain (due to single-linkage's tendency to cluster all voxels active in the map).

In the second row, we show the result of adding noise (1000 voxels), showing *k* means, DMC, Ward's and single-linkage with a typical subject. Qualitatively, one can observe the following:

- The DMC clustering solutions look very similar and produce highly select areas on gyri or along specific sulci (e.g., see red area in DMC clustering in Fig. 9). The noise resistance is primarily due to high-density clusters found in fMRI data sets.
- The two Ward's method solutions are similar with the noise condition producing more clusters distributed more widely over the same area, but the method tends to drop clusters.
- The *k* means solutions tend to repartition — find new clusters that are probably noise and, therefore, must drop originally good clusters. Dynamically then, noise for *k* means can be disastrous; as in this case, it repartitions on noise clusters that focus primarily on visual cortex as compared with the existing regressor, which doesn't even appear in the original *z* map. Single linkage in all subjects tends to find one large cluster throughout most of the brain, and hence, adding noise does little to disturb that particular degenerate solution. Applying erosion to this set (relative to the global centroid) would randomly remove 80% of the voxels in solution neither producing a plausible set of clusters or any further density in the map.

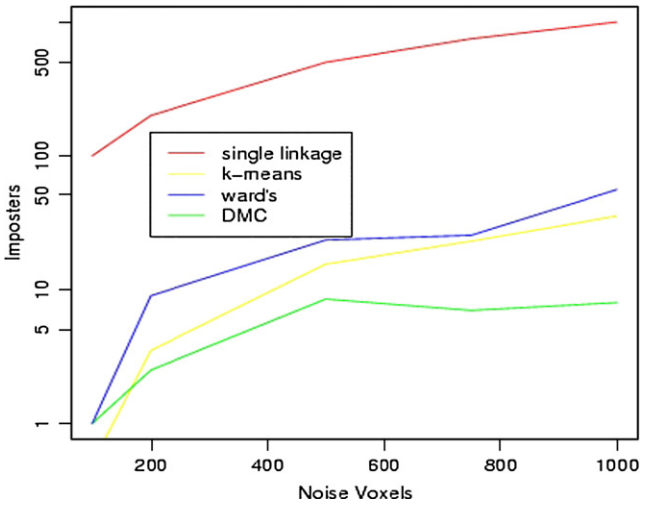


Fig. 10. Imposters as function of noise perturbation for all benchmark clustering methods. Note DMC resists false alarms, while single linkage has greatest sensitivity; *k* means and Ward's are similar in response.

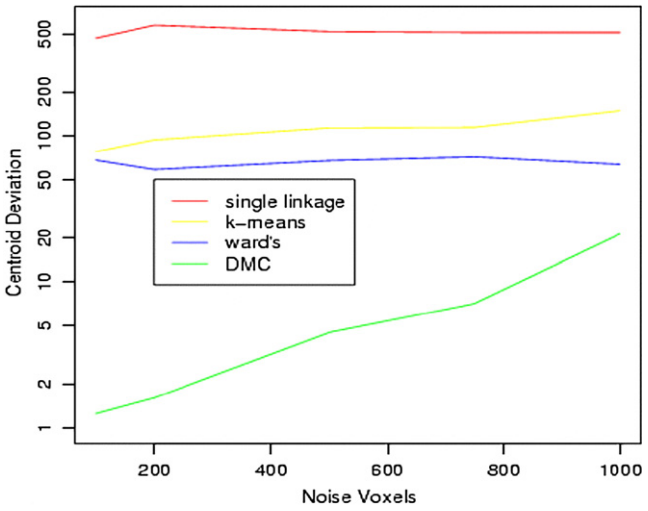


Fig. 11. Centroid deviation is a measure of stability of the cluster solution as measured by the original centroids of the clusters and their movement under noise perturbation. Note that most methods under even low-noise stress, produce high volatility; DMC shows small but systematic changes.

### 4.3. Imposters

This benchmark is a measure of the extent to which the set of voxels selected by the method changes. It does not penalize subdivision or merging of clusters (Fig. 10). For example, if adding noise produced a clustering solution that was identical, except for the fact that two of the original clusters were merged and one of the original clusters was split into two, this benchmark would not penalize this. Hence, a partitioning method such as *k* means would be expected to do well on this benchmark, as global reconfiguration is not penalized. The results do indeed show the *k* means methods doing reasonably well, although the mode clustering method is substantially stronger. Ward's method performs poorly on this benchmark. Single linkage as expected is particularly poor in that it consistently chains noise voxels. Moreover, as shown in Fig. 9, one large cluster is produced; hence, noise can only add imposters to the overall solution.

### 4.4. Centroids

This benchmark is a measure of similarity of two cluster solutions, as measured by how close the centroids of the two solutions are (Fig. 11). The mode clustering method

performs substantially better than other methods on this benchmark. This is an important result — one might expect that this benchmark should favor minimum variance methods because the mode clustering method is prone to allowing imposter voxels to "chain" on the boundary of the cluster, which should have a greater impact on the centroid. The potential weakness of $k$ means methods is that adding noise forces a global repartitioning, so one cluster can (and, in this data, does) split into two clusters. In this case, single linkage produces one large blob for a cluster and hence has a degenerate form of stability. Given the size of the supercluster, it possesses no centroid in any local sense.

### 4.5. Symmetric difference

This is a more stringent version of the first benchmark — it measures not only the global agreement between the voxels selected but also the agreement in partitioning (Fig. 12). So one would expect it to produce similar results to the first benchmark but punish partitioning methods. This does indeed happen (partitioning methods consistently produce set-symmetric differences of $>0.48$, higher than mode cluster in the high noise condition), although Ward's method performs surprisingly poorly compared to DMC. Single linkage clustering again, produces a degenerate solution that is resistant to noise in as much it produces only one cluster.

## 5. Discussion

Mode seeking proved to be effective in detecting spatial density in noisy data. Ward's method performed poorly on benchmarks, proving volatile under noise conditions, despite producing clustering solutions that subjectively looked good. Of these benchmarks, the symmetric difference is the most stringent — any deviation from the noise-free solution is punished with this benchmark. The symmetric difference results showed that even adding a very small amount of noise will produce solutions that have little overlap with the original solution for all methods except the mode-clustering method. In fact, the mismatch in the high noise condition, with noise voxels added for DMC, was 10%, while for $k$ means and Ward's, it was over 40% and 50%, respectively. In the low noise condition, the mode cluster measure had a mismatch of less than 1%. Methods for choosing "good" clusters from thresholded statistical maps have received relatively little attention in the literature. For the most part, the emphasis is on choosing "statistically valid" thresholds without addressing whether or not the partitioning into clusters is appropriate. This is problematic because cluster-level statistics work by trading off spatial localization for increased statistical power, so if identified clusters are too large or too sparse, one loses meaningful spatial localization of activity. Single-linkage clustering, which often identifies large or sparse clusters, is the defacto standard method for postprocessing activation maps supported in popular softwares such as FSL, AFNI and SPM. There are some software packages that offer at least some flexibility. For example, AFNI allows different forms of connectivity (26 or 18 nearest-neighbor, etc.), permits one to specify an actual search radius and also allows matching on nearby parameter values and not just coordinates. However, the state of the art is largely to put the onus on the user to choose a threshold which provides appropriate spatial localization and density. Dimitriadou et al. [31] benchmarked a number of different clustering methods, using simulated data and real data. This article compared clustering in the time domain and focused on correctly detecting activations in data sets. A replication of the same experiment with the mode clustering method would be an interesting way to test the method for time-domain clustering. The constraints in this domain are somewhat different to spatial clustering, so there is no reason to expect the same results. For example, $k$ means and other methods that are biased towards spherical clusters may be more appropriate in this domain (indeed, Ward's performed better in the present benchmarks). Partitioning methods may be more appropriate when there is a single "signal class" than our case, where we had an unknown number of sources.

### 5.1. Intersubject clustering

It is possible to use the DMC method or, for that matter, any clustering method across subjects; this is not unique to DMC. However, we encourage this particular type of analysis since it focuses the analysis on a tradeoff between individual differences and aggregate spatial location. Basically, rather than cluster spatial points in averaged coordinates taken over subjects, one can first register each subject's data to a common space (MNI) and then create a concatentation of all coordinates of different subjects as if

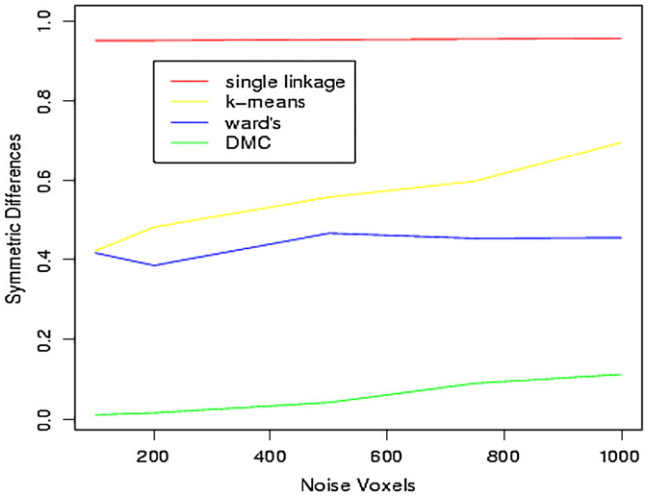


Fig. 12. Symmetric difference: is sensitive to both centroid movement and original cluster partitions. In this case, note that DMC shows a highly stable solution in the face of noise relative to all other methods.

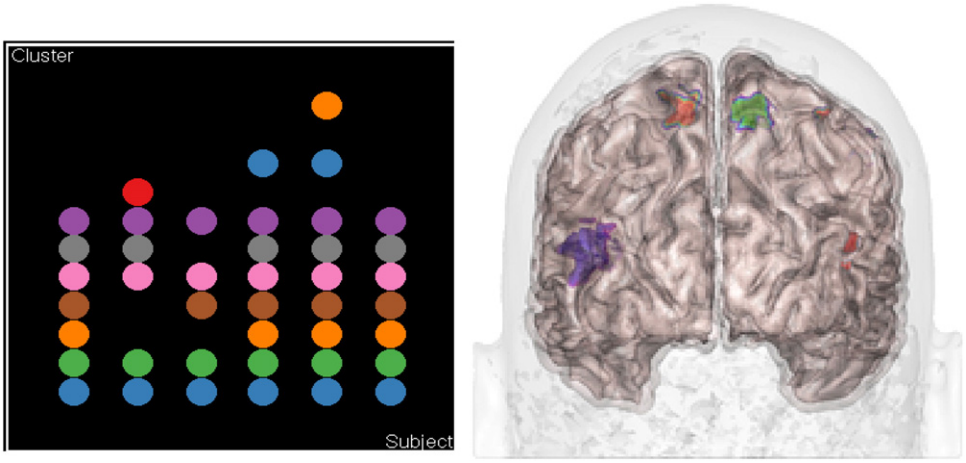


Fig. 13. Intersubject clustering using DMC. Note color coding of clusters indicating common location agreement. Clustering is done by combining all subjects voxels in one input file using an voxel subject index, allowing the clustering to be scored for common spatial location through subject agreement.

they were taken from the same subject. Specially, the input for the clustering algorithm consists of the concatenation of the lists of coordinates obtained from all subjects with an appropriate subject index which allows scoring (using the total subject agreement on that voxel as a figure of merit; see Fig. 13). Note that this method allows the clustering mechanism to discover agreement between subjects, instead of resorting to averages (which may misrepresent the data, especially if some subjects have a greater degree of activity than others; see also Ref. [32]). Compared to other clustering methods, DMC would tend to discover a higher subject agreement as compared due to the mode seeking aspect of the method. A future version of intersubject clustering would include a likelihood function that trades off a function of subject agreement with spatial location variables (intensity, local density etc.).

### 5.2. ROI selection

Since most experimenters have a hypothesis about active brain regions, one typically wants to use anatomical criteria to guide ROI selection. Because of prior understanding of the coupling of function and anatomy, one will typically use anatomical information to postprocess cluster information. For example, a spatially contiguous cluster that spans more than one anatomical region is a good candidate for subdivision. The advantage of selecting via functional activity is that the selection is finer — all voxels selected demonstrate substantial condition-dependent change and, hence, are good candidates for downstream analysis. So, ROI selection guided by functional criteria is likely to produce ROIs that are less noisy in the sense that they conform well to the investigators hypotheses.

### 5.3. Inferential Statistics for cluster size and set

Inferential statistics for cluster size inferences have been covered in depth in prior work [9,30,33–35]. Early work defined cluster size statistics based on the assumption that there is some functional relationship between activation intensity and number of contiguously activated clusters. This work first characterized the cluster size distribution as a spatial point process, assuming a Poisson process, while later work using Random Field Theory (RFT), introduced more complex alternatives which can trade off intensity and spatial extent in order to do inference. For the present algorithm (DMC), since the clusters are spatially contiguous, cluster size-based statistics using either RFT or permutation methods will still be valid. Nonetheless, DMC clusters would tend to produce cluster size distributions more peaked with lighter tails then densities based on independent spatial point processes that might arise in single-linkage algorithms. Consequently, it is probably the case that the standard cluster size inference framework so far developed in neuroimaging analysis is not independent of specific clustering algorithms. A more serious issue with simple cluster size approaches is the assumption that size of the cluster is a measure of significance; clearly, familywise error at the cluster level is minimized in larger clusters (although with a resultant size bias); it depends on an unspecified assumption of density of the clusters themselves, a factor highlighted by the present algorithm. Bivariate cluster size approaches that incorporate both cluster size and intensity have a similar bias, since intensity will tend to co-occur with larger clusters. Nonetheless, it is possible to imagine a small but highly dense area of the brain that is more significant in terms of intensity magnitude but, at the same time, one of the smallest clusters in a field of clusters (e.g., consider how the “FFA” is defined). In general, $P$ values can be assigned at least in a similar way to the present clusters results as in the standard single-linkage case, using permutation methods suggested by Hawasaki et al. It is clear from this discussion, however, that the relationship between cluster size, density and cluster algorithms will require a closer look at the underlying probability assignment even in the standard case [36–39].

## Appendix A. Dense mode clustering algorithm

```
Pseudocode(main function)
#cluster_distance returns the distance from a point to the
  nearest cluster in that point function cluster_distance
  (cluster,x)
 return min([distance(x,y) for y in clusters])
# nearest_cluster(clusters,x) returns the shortest distance
  from a point x to any point in any cluster in clusters.
  function nearest_cluster(clusters,x)
 precondition(not empty(clusters))
 d=distance(clusters.first(),x)
 for c in clusters
  d<_min(cluster_distance(x,c),d)
  return d
#pairs returns all distinct pairs (but not including permuta-
  tions of the same pair, e.g., pairs (1,2,3 )=(1,2),(1,3),(2,3)
functionpairs(L1,L2)
  result=0.
 for x in L1
 for y in L2
  append(result, (x,y))
  return result
function pairs (L1)
  result=0.
 for i in 1..length(L)
 for j in i+1..length(L)
  append(result,L[i],L[j])
  return result
# Find nearest point to given cluster function nearest_point
  (c1,c2)
 for x,y in pairs(c1,c2)
  if distance(x,y) < d
   p1 <_x
   p2 <_y
   return p1,p2
#w merge with Romeo and Juliet rule starting with smallest
  pairwise distance. function rj_merge (clusters)
 for c1, c2 in pairs (clusters)
  p1,p2 <_nearest_point (c1, c2)
  d1 <_mean ([distance(p1,x) for x in c1])
  d2 <_mean ([distance(p2,x) for x in c2])
  if d(p1,p2) < (d1+d2)/2
  merge_clusters(c1,c2)
  return TRUE
  return FALSE
  dense_points <_empty_list()
 for x in points
  count <_0
 for y in points
  if distance(x,y) < r
  count <_count+1
 if count >=THRESHOLD
  dense_points <_append(dense_points,x)
  clusters <_empty_list()
 for x in dense_points
  if not empty (clusters)
  c=nearest_cluster (clusters,x)
  d=cluster_distance (c,x)
 if d < R append(c,x) else
   append(clusters,[x])
 else
  append(clusters,[x])
# keep merging until we can't do it any more
 try_merge<_1
 while(try_merge)
 try_merge<_rj_merge(clusters)
```

## References


[1] Hartigan J. Clustering algorithms. New York: Wiley; 1975.
[2] Everitt B. Cluster analysis. John Wiley & Sons, Inc; 1993.
[3] Baumgartner R, Rayner L, Richter W, Summers R, Jarmasz M, Somorjai R. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis — statistical approaches to human brain mapping by functional magnetic resonance imaging. Magn Reson Imaging 2000;18, 1(6):89–94.
[4] Goutte C, Hansen LK, Liptrot MG, Rostrup E. Feature-space clustering for fMRI meta-analysis. Hum Brain Mapp 2001;13(3):165–83.
[5] Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn Reson Med 1998;(2):249–60.
[6] Chuang KH, Chiu MJ, Lin CC, Chen JH. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means. IEEE Trans Med Imaging 1999;18(12):1117–28.
[7] Kohonen T. Self-Organizing Maps, Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 1995, 1997, 2001.
[8] Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29:162–73.
[9] Friston KJ. Analyzing brain images: principles and overview. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, editors. Human brain function. New York: Academic Press; 1997. p. 25–41.
[10] Smith S, Jenkinson M, Woolrich M, Beckmann CF, Behrens TEJ, Johansen−Berg H, et al. New York: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004;23(Sup 1):208–19.
[11] Hanson SJ, Matsuka T, Haxby JV. Combinational codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage 2004;23(1):156–66.
[12] Friston KJ, Harrison L, Penny W. Dynamic causal modeling. Neuroimage 2003;19(4):1273–302.
[13] McIntosh AR. Towards a network theory of cognition. Neural Netw 2001;13:861–76.
[14] Johnson SC. Hierarchical clustering schemes. Psychometrika 1967;2: 241–54.
[15] Wishart D. Mode analysis: a generalization of nearest neighbor which reduces chaining effects. In: Cole AJ, editor. Numerical taxonomy. New York: Academic Press; 1969. p. 282–311.
[16] Corter JE, Tversky A. Extended similarity trees. Psychometrika 1986; 51:429–51.
[17] Krantz D, Luce RD, Suppes A, Tversky A. Foundations of measurement: volume 1. New York: Academic Press; 1971.
[18] Tukey PA, Tukey JW. Preparation: prechosen sequences of views. In interpreting multivariate data. In: Barnett V, editor. Chichester: Wiley; 1981. p. 189–213.
[19] Comaniciu D, Meer P. Distribution free decomposition of multivariate data. Pattern Anal Appl 1999;2:22–30.
[20] Stanberry L, Nandy R, Cordes D. Cluster analysis of fMRI data using dendrogram sharpening. Hum Brain Mapp 2003;20:201–19.

[21] Hayasaka S, Nichols TE. Validating cluster size inference: random field and permutation methods. Neuroimage 2003;(4):2343–56.
[22] Forman SD, Cohen JD, Fitzgerald JD, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 1995;33:636–47.
[23] Poline JB, Worsley KJ, Evans AC, Friston KJ. Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage 1997;5(2):83–96.
[24] Hanson SJ, Bly BM. The distribution of BOLD susceptibility effects in the brain is non-Gaussian. Neuroreport 2001;12(9):1971–7.
[25] Chen CC, Tyler CW, Baseler HA. Statistical properties of BOLD magnetic resonance activity in the human brain. Neuroimage 2003; 20(2):1096–109.
[26] Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain: 3-dimensional propotional system. Stuttgart: Thieme; 1988.
[27] Woolrich MW, Ripley BD, Brady JM, Smith SM. Temporal autocorrelation in univariate linear modeling of fMRI data. Neuro-Image 2001;14(6):1370–86.
[28] Jenkinson M, Smith SM. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001;5(2):143–56.
[29] Jenkinson M, Bannister P, Brady M, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17(2):825–41.
[30] Friston K, Worsley K, Frackowiak R, Mazziotta J, Evans A. Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1994;1:214–20.
[31] Dimitriadou E, Barth M, Windischberger C, Hornik K, Moser E. A quantitative comparison of functional MRI cluster analysis. Artif Intell Med 2004;31(1):57–71.
[32] Thirion B, Flandin G, Pinel P, Roche A, Ciuciu P, Poline J-B. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets In Press, Hum Brain Mapp.
[33] Poline J-B, Mazoyer B. Cluster analysis in individual functional brain images: some new techniques to enhance the sensitivity of activation detection methods. Hum Brain Mapp 1994;2:103–11.
[34] Hayasaka S, Nichols TE. Combining voxel intensity and cluster extent with permutation test framework. Neuroimage 2004;(1):54–63.
[35] Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE. Nonstationary cluster-size inference with random field and permutation methods. Neuroimage 2004;22(2):676–87.
[36] Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 1996;4(3 Pt 1):223–35.
[37] Hanson S, Rebbechi D, Matsuka T, Hanson C, Zaimi A. Top down and bottom up processing of everyday visual action. Submitted.
[38] Murase K, Kikuchi K, Miki H, Shimizu T, Ikezoe J. Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast enhanced MR imaging. J Magn Reson Imaging 2001;13: 797–806.
[39] Shiffman S, Ng YR, Brosnan TJ, Eliez S, Links JM, Kelkar UV, Reiss AL. Interactive specification of regions of interest on brain surfaces. Neuroimage 2003;20(3):1811–86.