

Comparison of Two Exploratory Data Analysis Methods for fMRI

Dr.–Ing. Anke Meyer–Bäse

Department of Electr. & Comp. Engineering, Florida State University

Outline

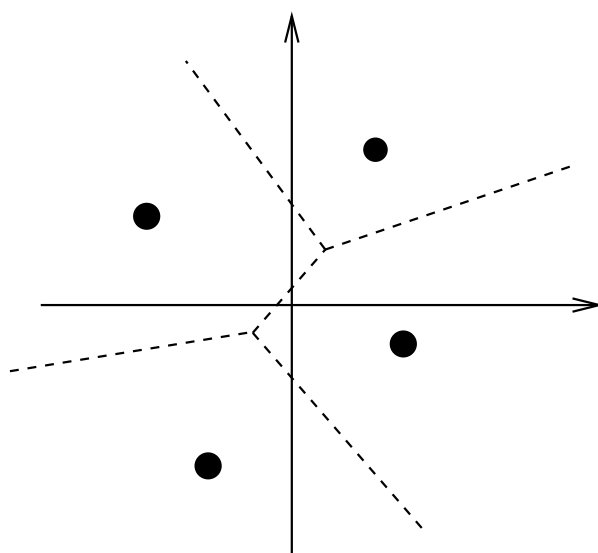
1. Introduction
2. Exploratory data analysis methods for fMRI
3. Vector quantization based on neural networks
4. Fuzzy cluster techniques
5. Estimation of the fuzzy model for fMRI
6. Characterization of task–related and artifactual effects
7. Conclusions

1 Introduction

- Application of ICA and clustering techniques to fMRI
- Introduction of a new technique in fMRI: Gath–Geva algorithm
- Clustering techniques: Kohonen map, ”neural gas” network, minimal free energy vector quantization
- ICA techniques: PCA, FastICA, Infomax, topographical ICA
- Characterization of task–related and artifactual effects
 1. Activity component maps and resulting time courses for the IC
 2. Differences between ICA and clustering techniques in the activation maps
 3. Motion artifacts and co–activation of the frontal eye fields

2 Vector Quantization Based on Neural Networks

- Vector quantization: fast clustering technique for feature vectors describing pixel time courses (fMRI or MRI)
- Goal: Determine cluster centers effectively



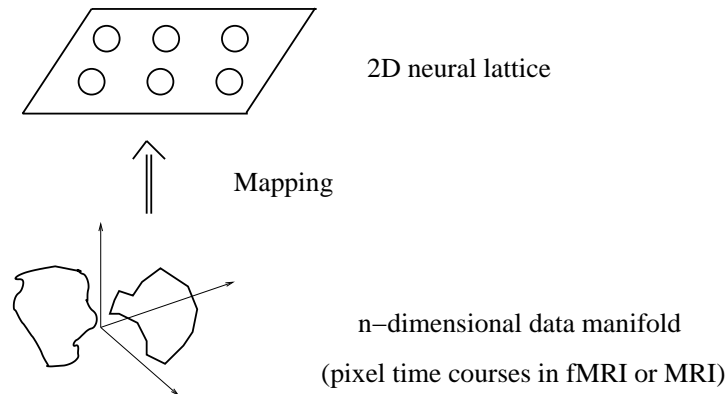
- Implementation of mapping from data space to space of code-words (cluster centers)
- Cluster centers adaptively updated:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \epsilon(t)a_i(\mathbf{x}(t), C(t), \kappa)(\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (0.1)$$

$\epsilon(t)$: learning parameter, a_i : codebook, $C(t)$: dependent cooperativity function, κ : cooperativity parameter, and \mathbf{x} : feature vector.

2.1 Kohonen's Map

- Prespecified lattice (1D or 2D)



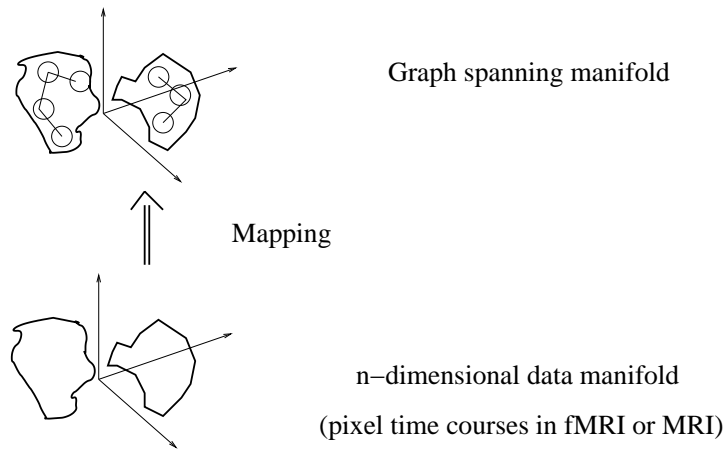
- Cluster update rule:

$$\mathbf{w}_i(t + 1) = \mathbf{w}_i(t) + \epsilon(t) \exp\left(-\frac{d_{ij}}{\sigma^2}\right)(\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (0.2)$$

- Kohonen map: mapping between data space V and prespecified graph G is not necessarily topology-preserving
- Failure of Kohonen algorithm: topological structure of V not *a priori* known or too complicated
- Problem: topological structure of preset graph G differs from that of V

2.2 "Neural Gas" Network

- Solution to topology constraints: induced Delaunay triangulation



- Neural gas network: topology preserving map
- Method: refreshing of connections in graph
- Neural gas algorithm: codewords are refreshed based on neighborhood-ranking
- Advantages: lower distortion rate and faster quantization than Kohonen map
- Update for cluster centers:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \epsilon(t) \exp(-k_i(\mathbf{x}, \mathbf{w}_i/\lambda))(\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (0.3)$$

2.3 Minimum Free–Energy Vector Quantization

- No prespecified lattice: one cluster at the beginning
- Learning rule: stochastic gradient descent
- Hierarchical and multiresolution aspect of data analysis
- Straightforward cluster splitting: control parameters (free energy, entropy)
- Learning rule:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \epsilon(t) \frac{\exp(-\|\mathbf{x}(t) - \mathbf{w}_i(t)\|^2/2\rho^2)}{\sum_i \exp(-\|\mathbf{x}(t) - \mathbf{w}_i(t)\|^2/2\rho^2)} (\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (0.4)$$

ρ : fuzzy range (mechan. stat. temperature $T = 2\rho^2$)

- Cooperativity function a_i :

$$a_i = \frac{\exp(-\|\mathbf{x}(t) - \mathbf{w}_i(t)\|^2/2\rho^2)}{\sum_i \exp(-\|\mathbf{x}(t) - \mathbf{w}_i(t)\|^2/2\rho^2)} \quad (0.5)$$

a_i is a softmax activation function

- Advantage over fuzzy c –means: no prespecified clusters

3 Fuzzy Clustering Algorithms

- Idea: pixel time course belongs to more than one cluster
- Degree of association to cluster specified by membership grade

3.1 Fuzzy n -Means Clustering

- Defined by a set of functions: $u_j : X \rightarrow A, j = 1, \dots, n$ with $A = [0, 1]$
- Cluster: point representative and fixed locations
- Goal: best partition and optimal cluster numbers
- Nonadaptive metric:

$$d^2(\mathbf{x}_j, \mathbf{L}_i) = (\mathbf{x}_j - \mathbf{L}_i)^T (\mathbf{x}_j - \mathbf{L}_i) \quad (0.6)$$

$d(\mathbf{x}_j, \mathbf{L}_i)$ is distance metric, \mathbf{x}_j is data point, \mathbf{L}_i is cluster prototype

- Limitations: predefined number of clusters
- Improvement: not specified number of clusters and variable locations (fuzzy n -means with unsupervised initialization)

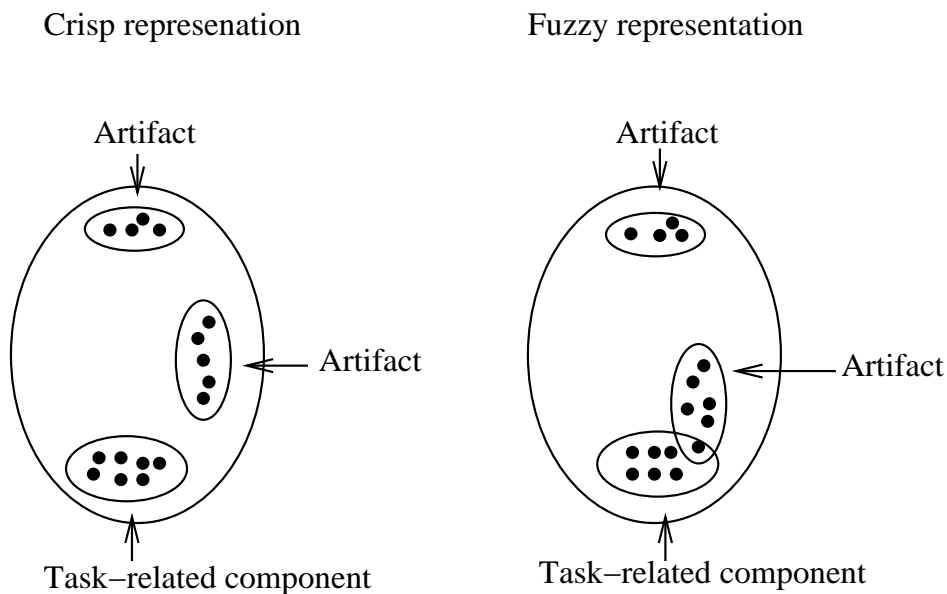
3.2 Gath–Geva Algorithm

- Idea: cluster has a shape, not point representative
- Cluster forms: spherical to ellipsoidal clusters
- Adaptive metric:

$$d^2(\mathbf{x}_j, \mathbf{L}_i) = (\mathbf{x}_j - \mathbf{L}_i)^T \mathbf{M}_i (\mathbf{x}_j - \mathbf{L}_i) \quad (0.7)$$

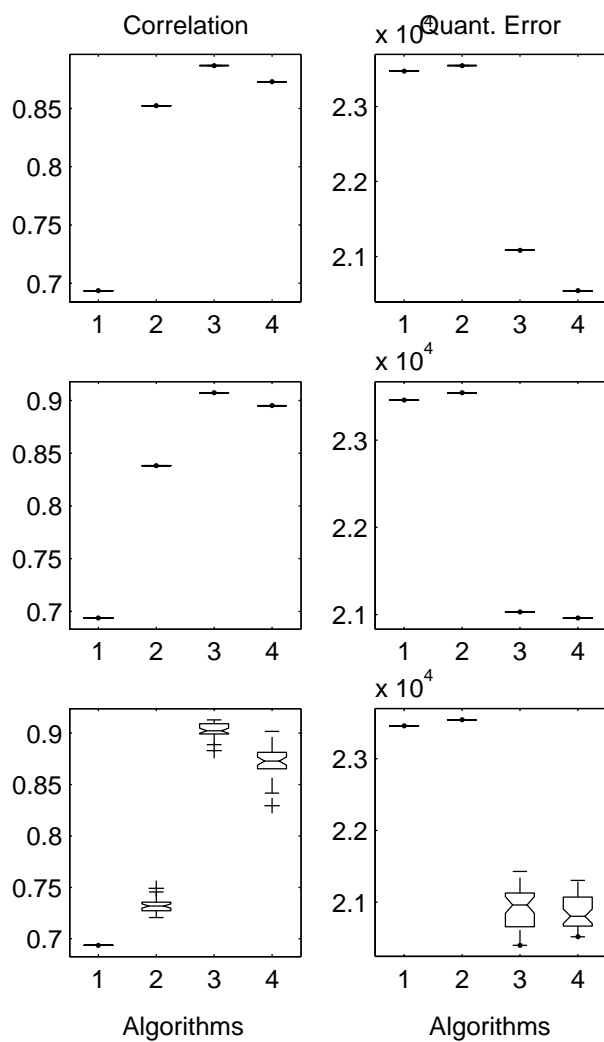
\mathbf{M}_i is the shape matrix

- Main benefit: good partition results for unequally variable features and densities
- Visualization of crisp and fuzzy clustering applied to fMRI data:

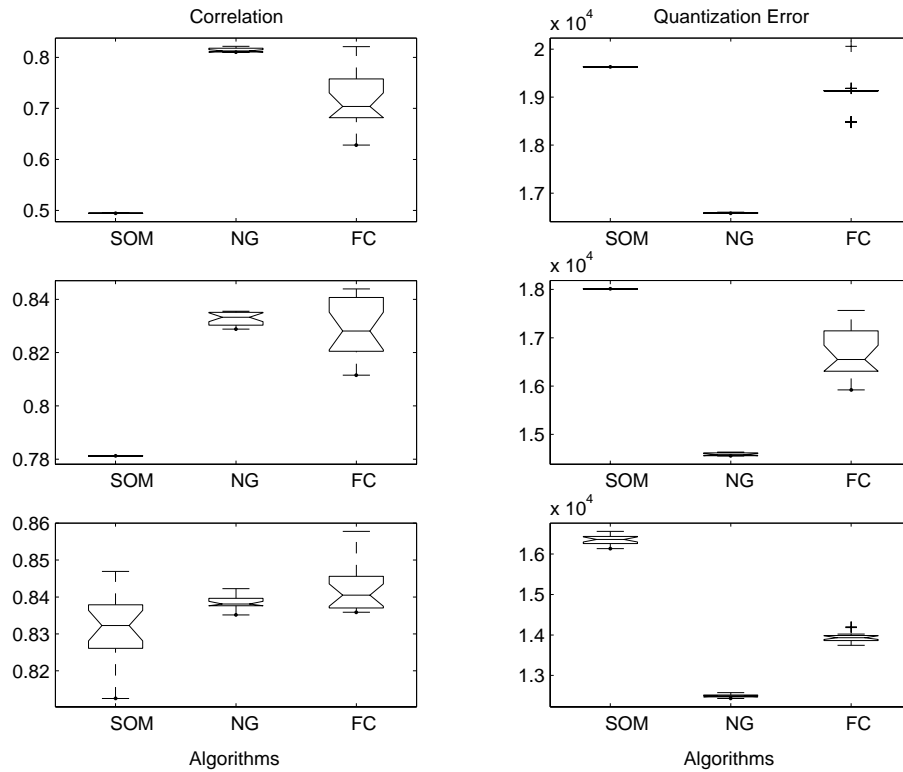


4 Estimation of the Optimal Numbers of ICs and Codebook Vectors

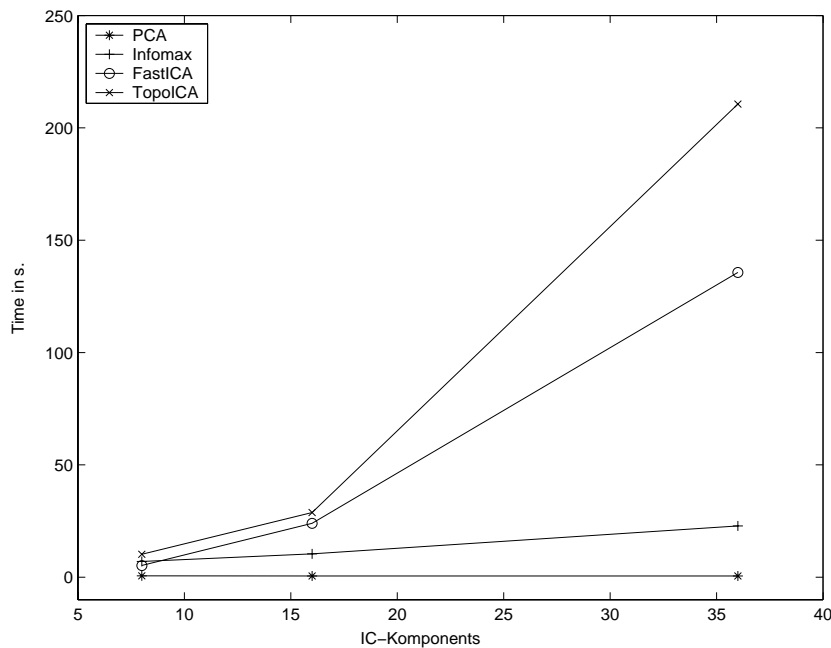
- Key issue: how many ICs or codebook vectors are necessary for a correct representation?
- Comparison between different ICA algorithms: 1 (PCA), 2 (Infomax), 3 (Fast ICA), and 4 (TopoICA) and for 8, 16, 36 independent components (top to bottom)



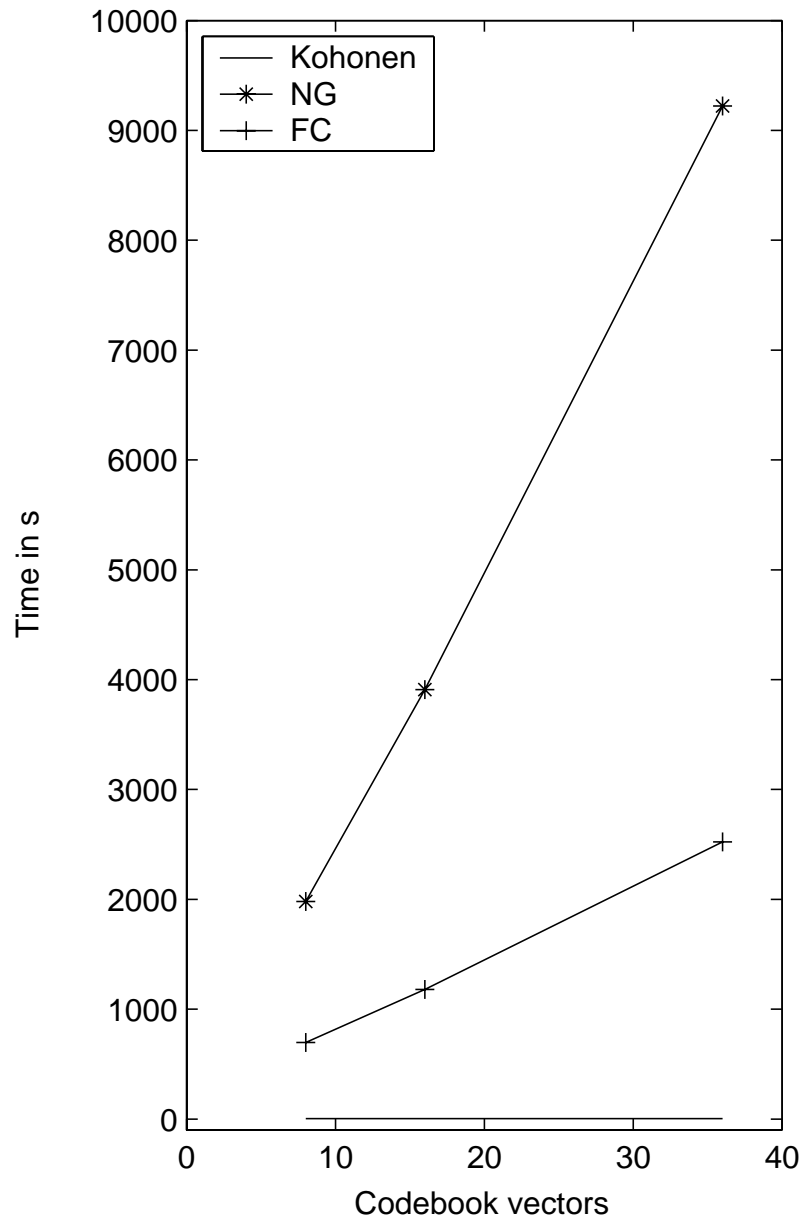
- Comparison between different clustering algorithms: 1 (SOM), 2 ("neural gas"), and 3 (minimal free-energy) and for 8, 16, 36 independent components (top to bottom)



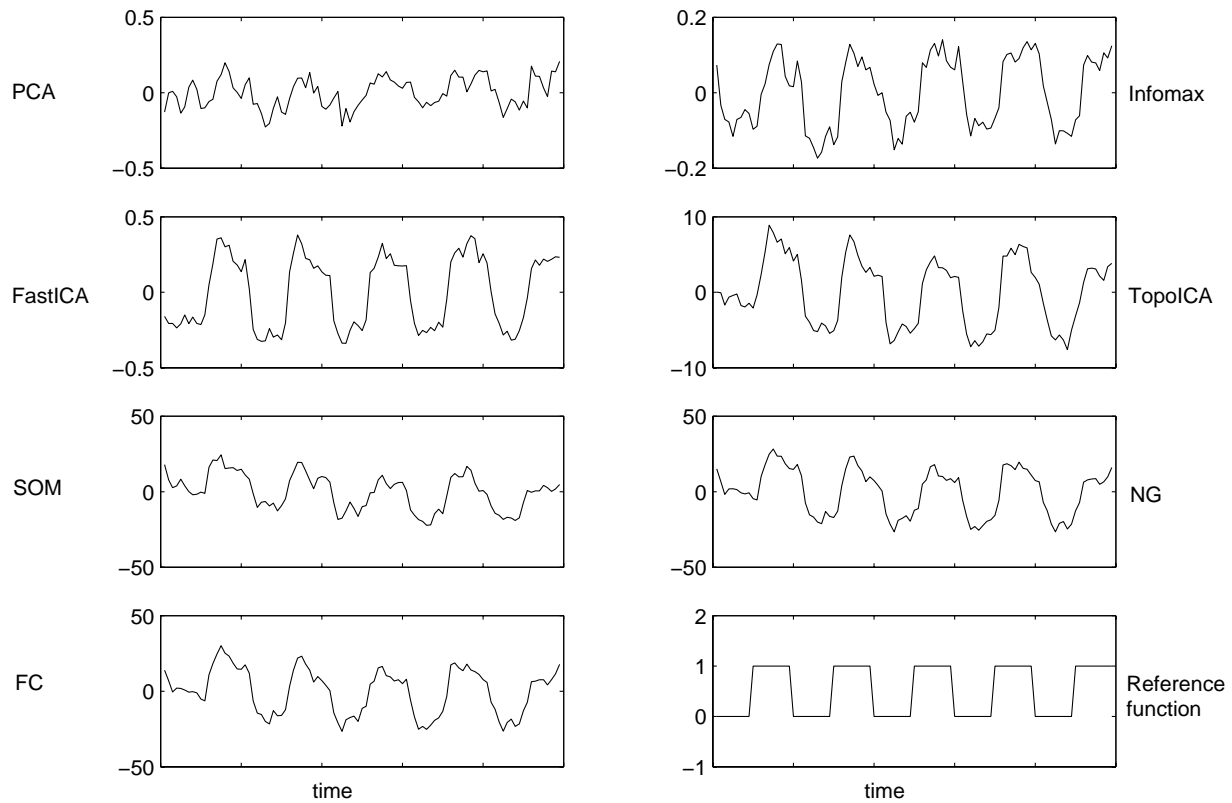
- Time considerations for ICA techniques:



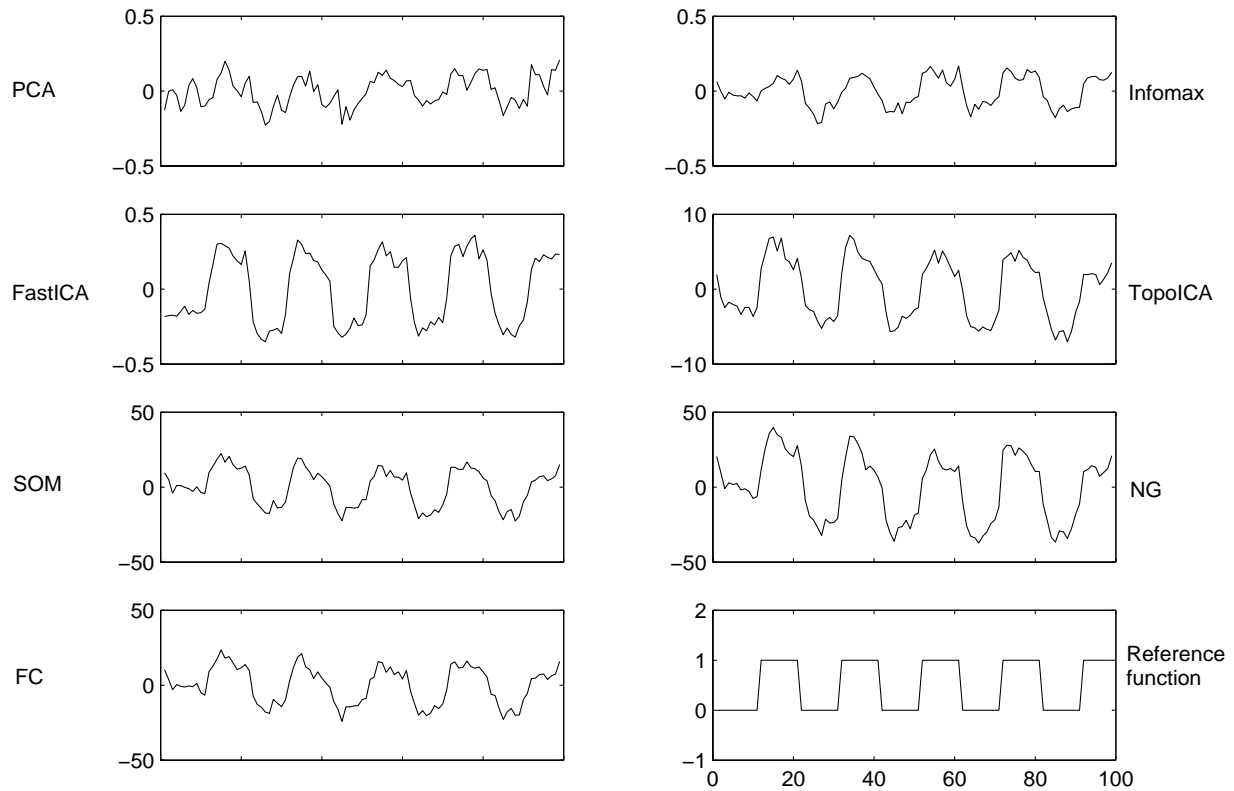
- Time considerations for clustering techniques:



- Normalization of data (mean=0, variance=1) on time axis, however spatial ICA!
- Reference functions: correlation coefficients $r = 0.69$ for PCA, $r = 0.85$ for Infomax, $r = 0.93$ for FastICA, $r = 0.86$ for topographic ICA, $r = 0.72$ for SOM, $r = 0.83$ for "neural gas" and $r = 0.82$ for minimal free-energy (8 ICs or codebook vectors)
- Increased signal sensitivity of topographical ICA versus others ICA techniques: range [-10:10]
- Highest signal sensitivity for clustering techniques

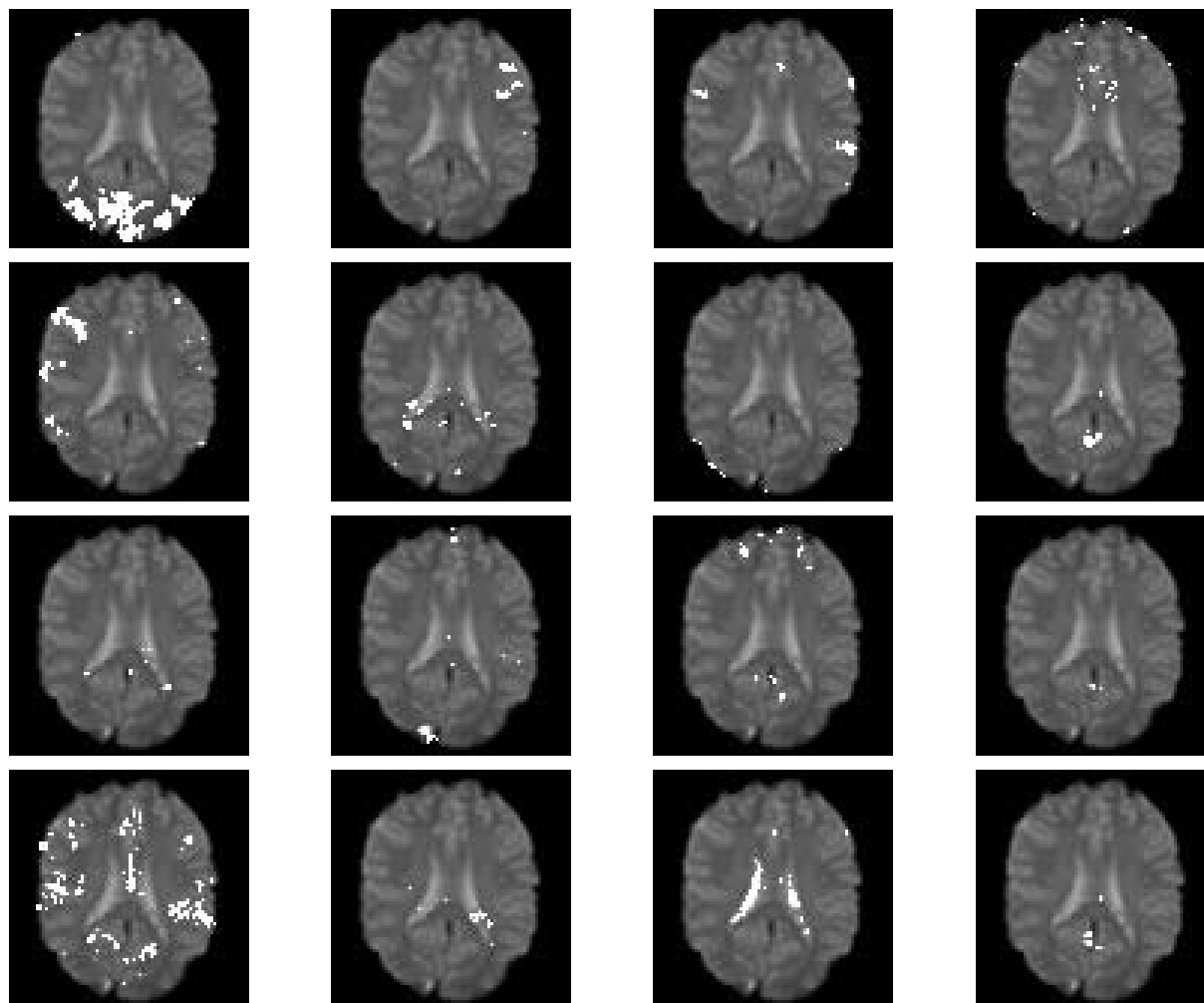


- Reference functions: correlation coefficients $r = 0.69$ for PCA, $r = 0.85$ for Infomax, $r = 0.93$ for FastICA, $r = 0.90$ for topographic ICA, $r = 0.84$ for SOM, $r = 0.84$ for "neural gas" and $r = 0.84$ for minimal free-energy (36 ICs or codebook vectors)

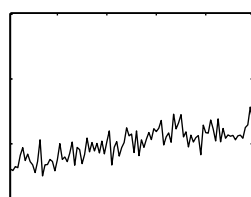
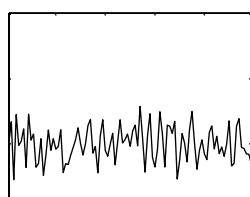
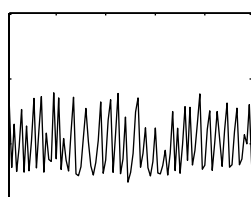
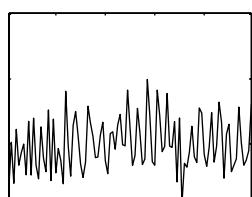
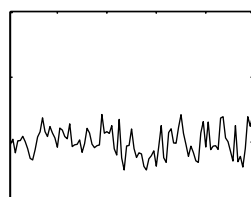
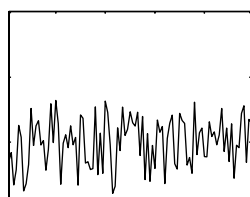
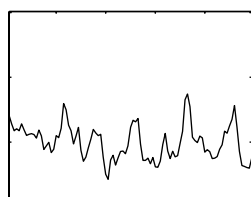
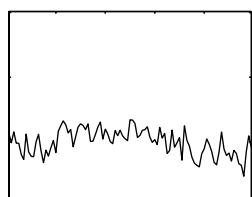
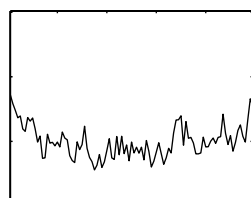
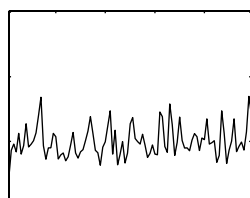
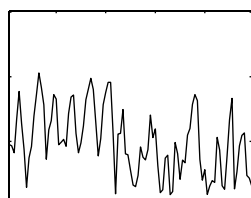
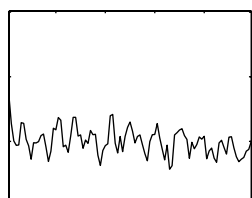
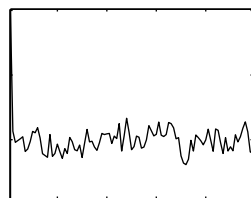
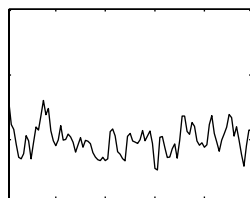
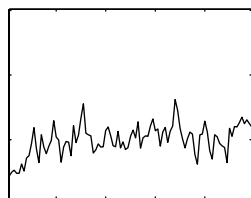
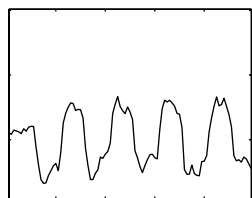


5 Characterization of Task-Related and Artifactual Effects

- FastICA cluster assignment maps for 16 components:

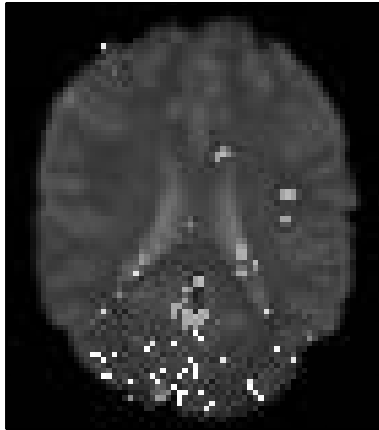


• Associated ICs:

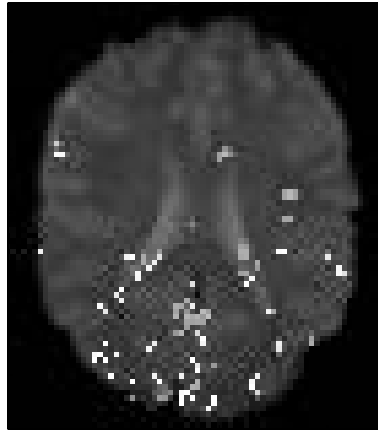


- Comparisons between ICA techniques based on pairwise pixel differences:

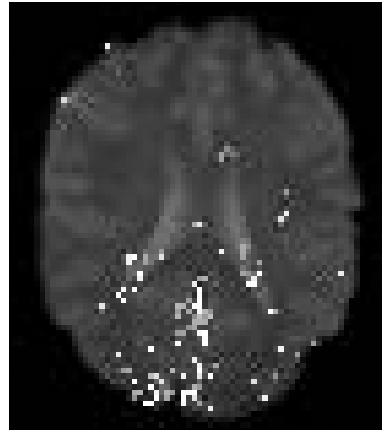
PCA versus Infomax



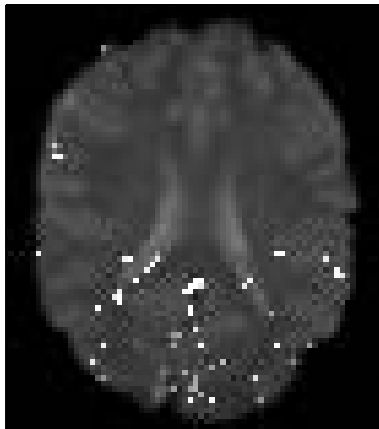
PCA versus FastICA



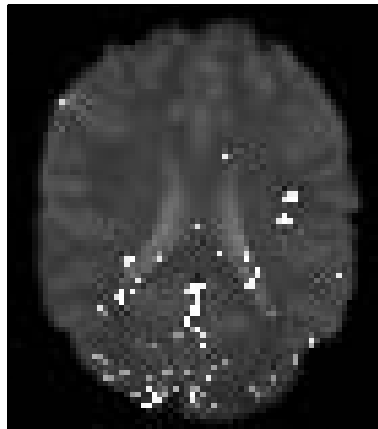
PCA versus TopoICA



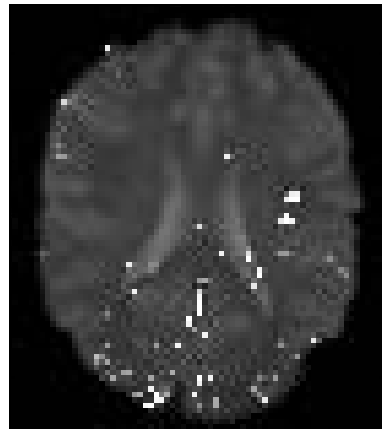
Infomax versus FastICA



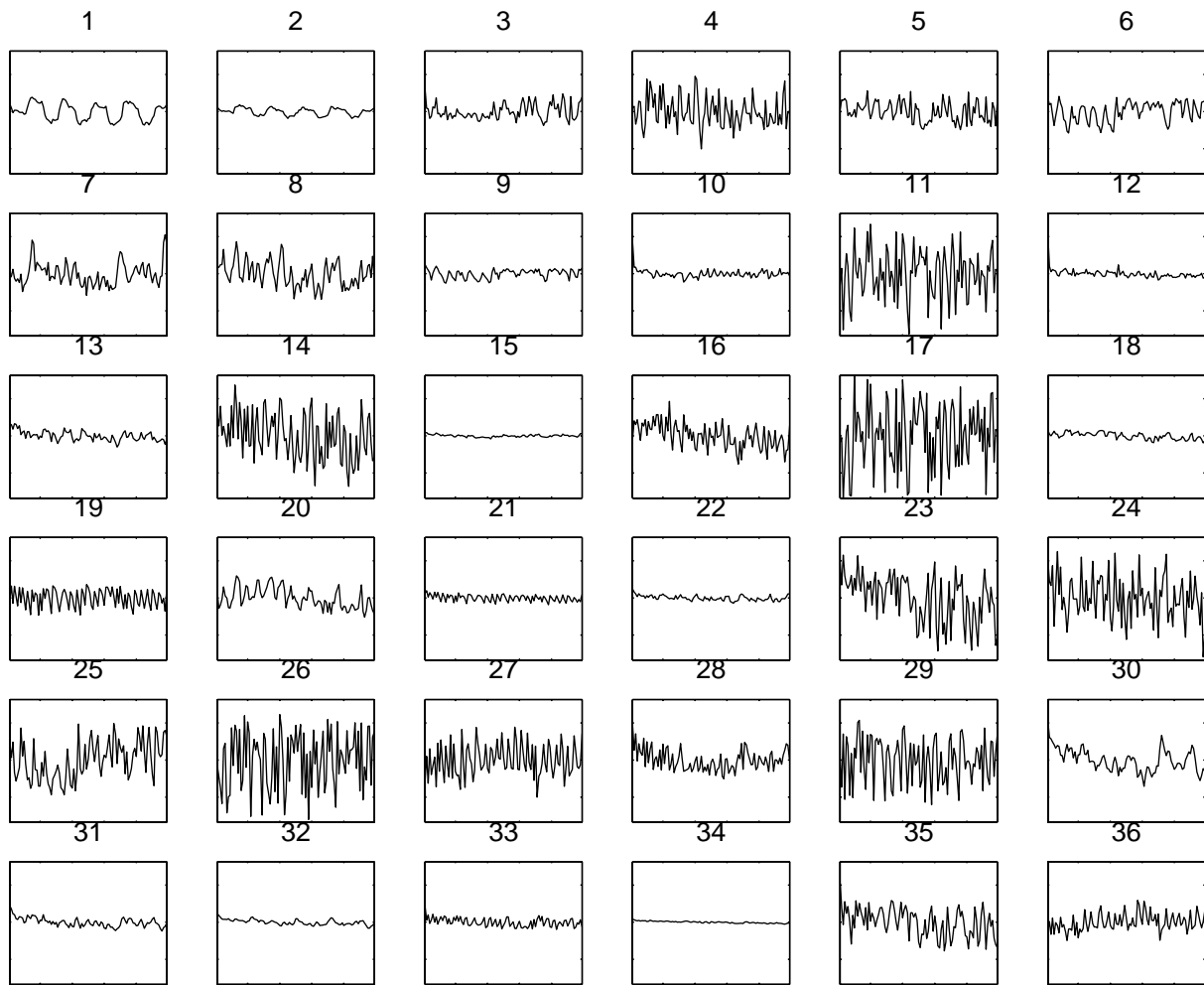
Infomax versus TopoICA



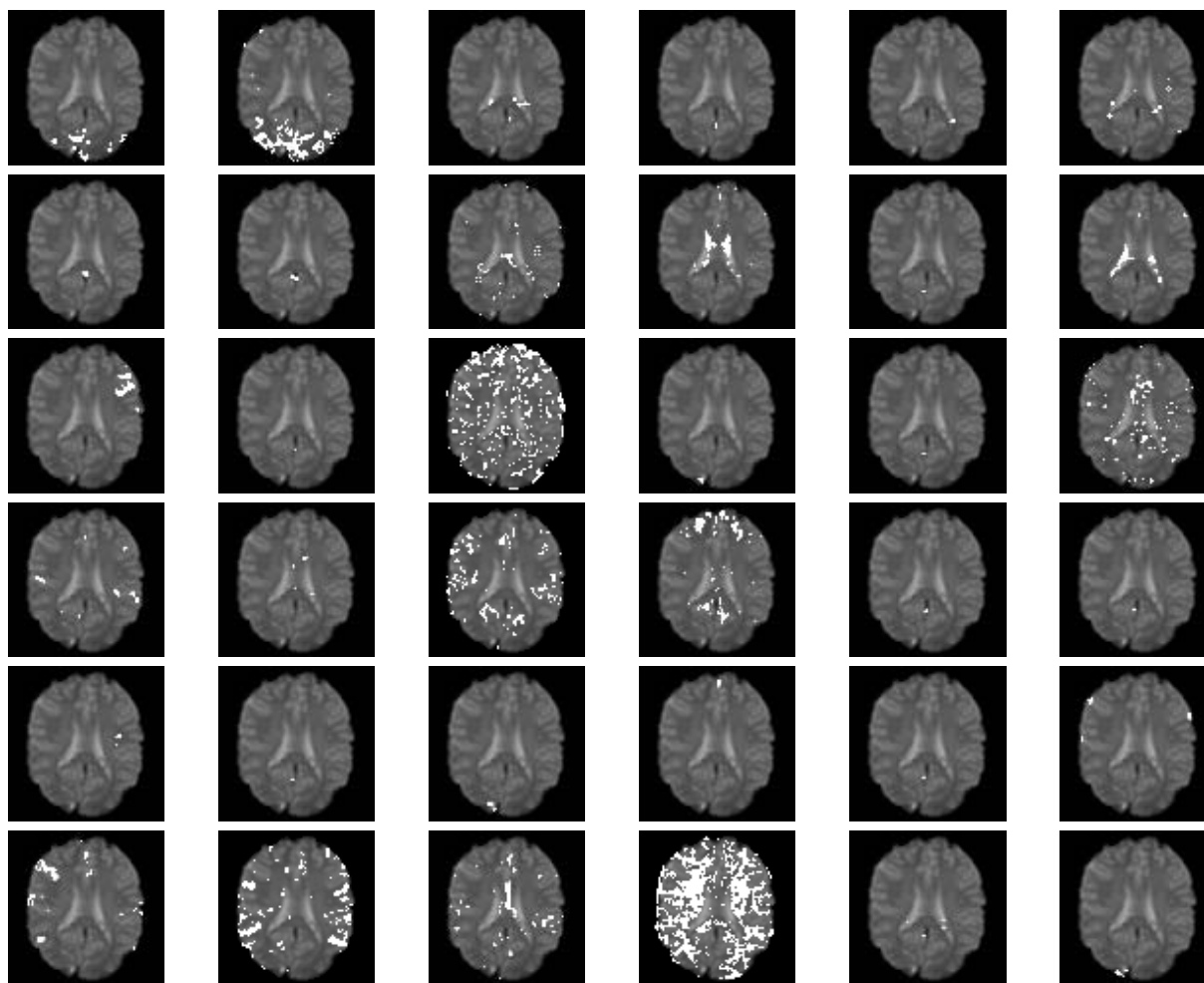
FastICA versus TopoICA



- "Neural gas" network cluster assignment maps for 36 components:

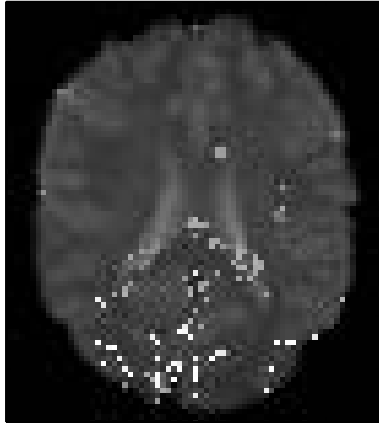


- Associated codebook vectors:

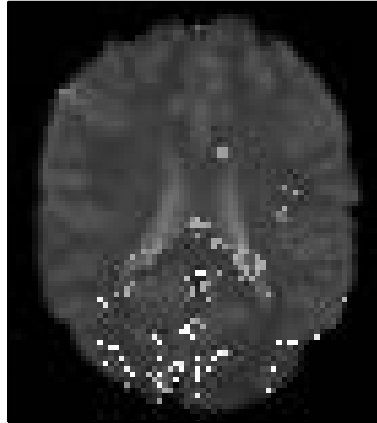


- Comparisons between cluster techniques based on pairwise pixel differences:

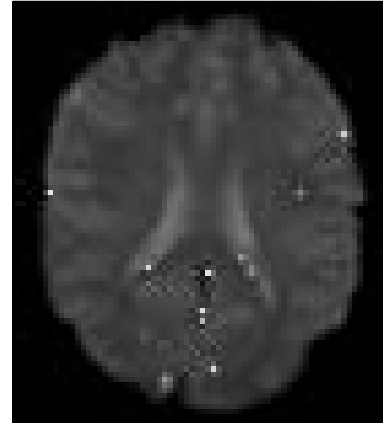
SOM versus NG



SOM versus FC

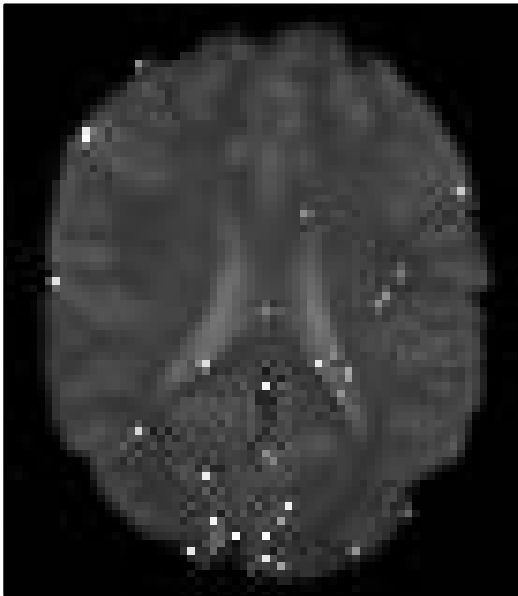


NG versus FC

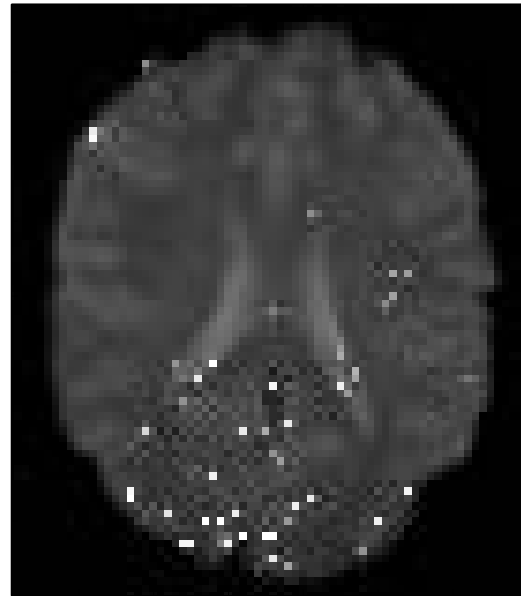


- Comparisons between ICA and cluster techniques based on pairwise pixel differences:

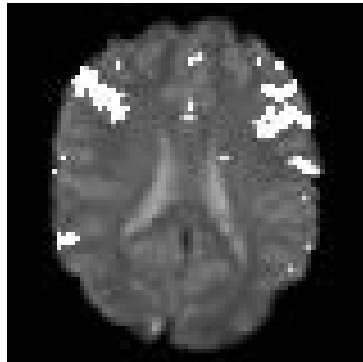
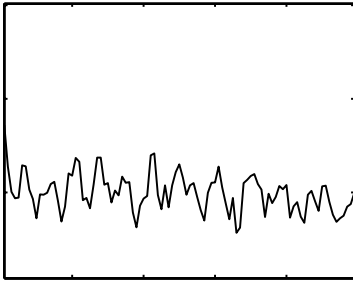
TopoICA versus Neural Gas



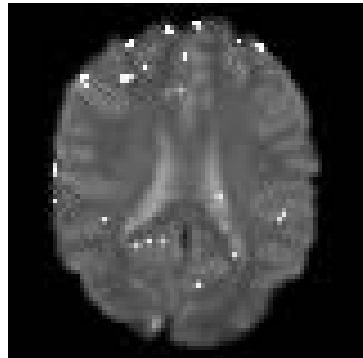
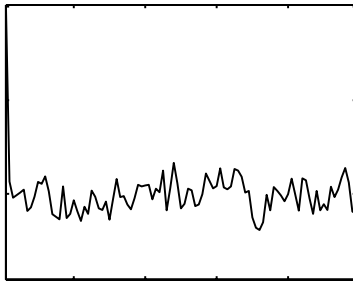
TopoICA versus FC



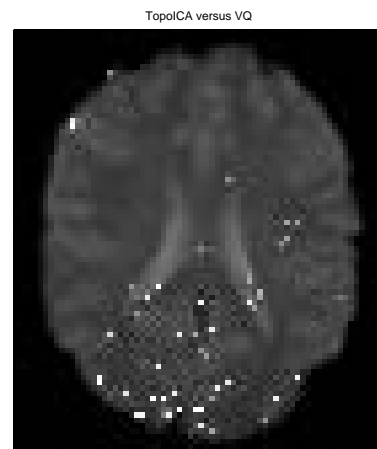
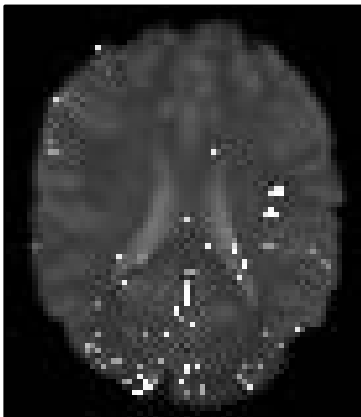
- Artifactual effect: co-activation of the frontal eye fields



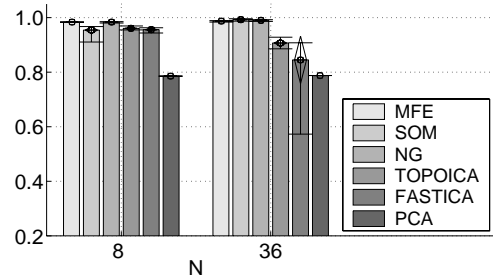
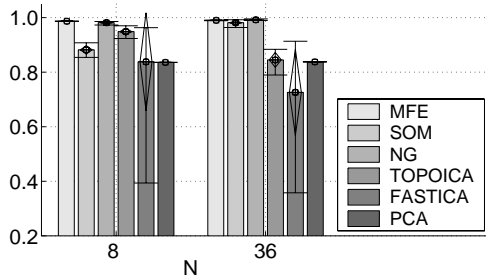
- Through-plane motion artifact:



- Comparisons between clustering and transformation-based methods in fMRI:

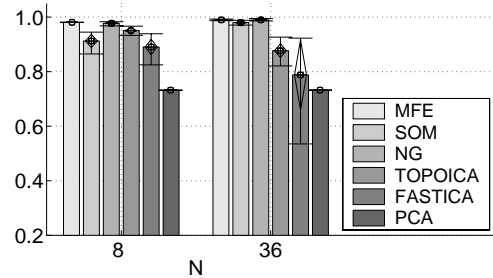
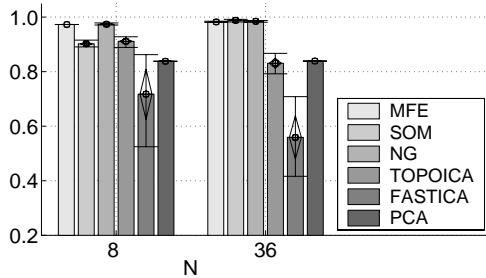


- ROC analysis based on correlation map (threshold of 0.4): 8 and 36 ICs or codebook vectors



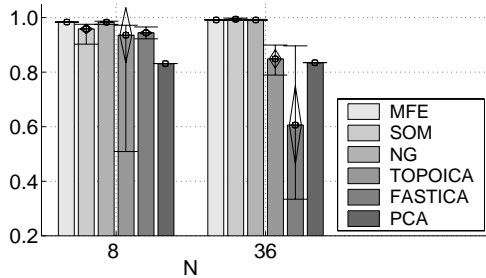
Subject 1

Subject 2



Subject 3

Subject 4

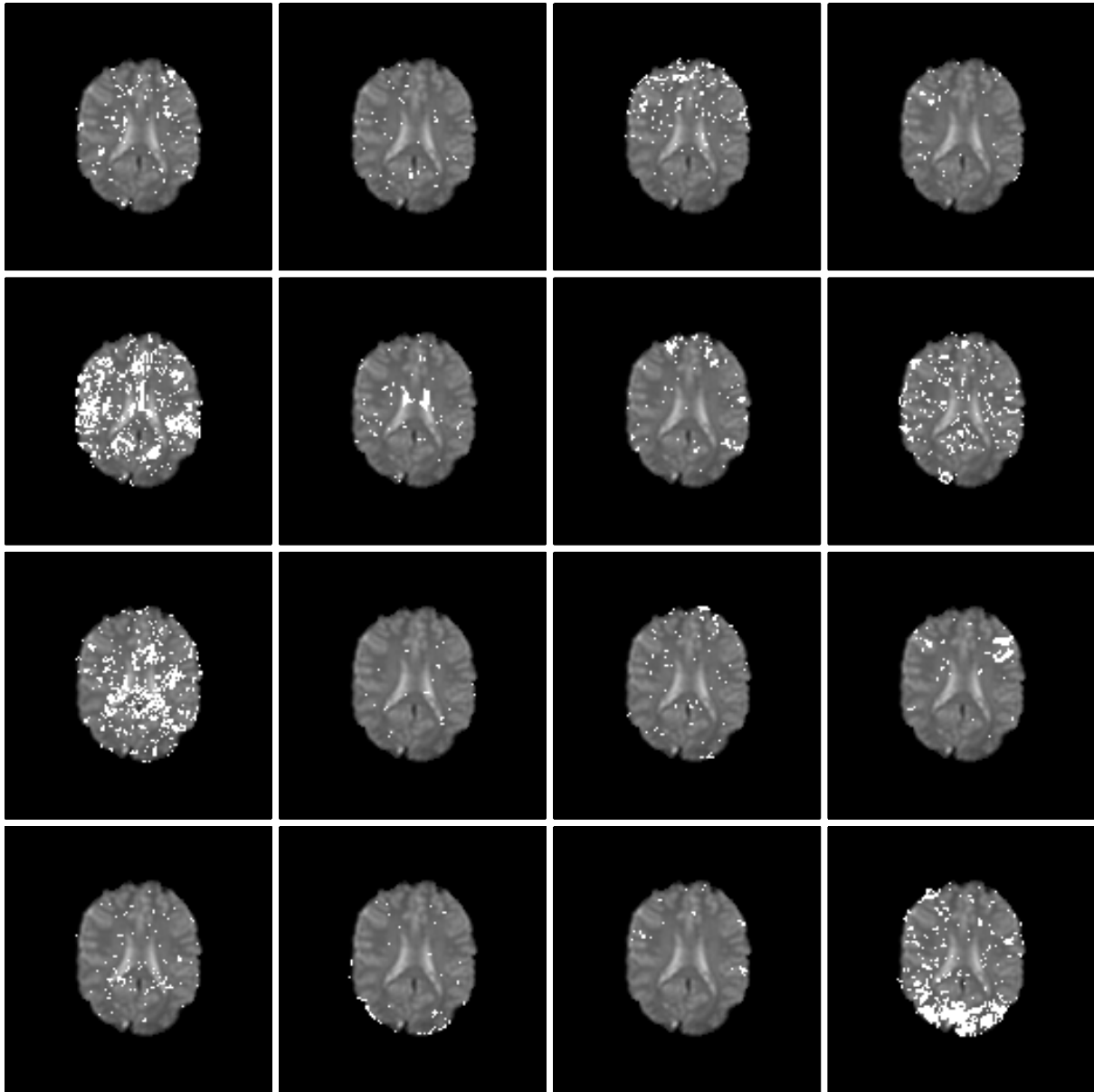


Subject 5

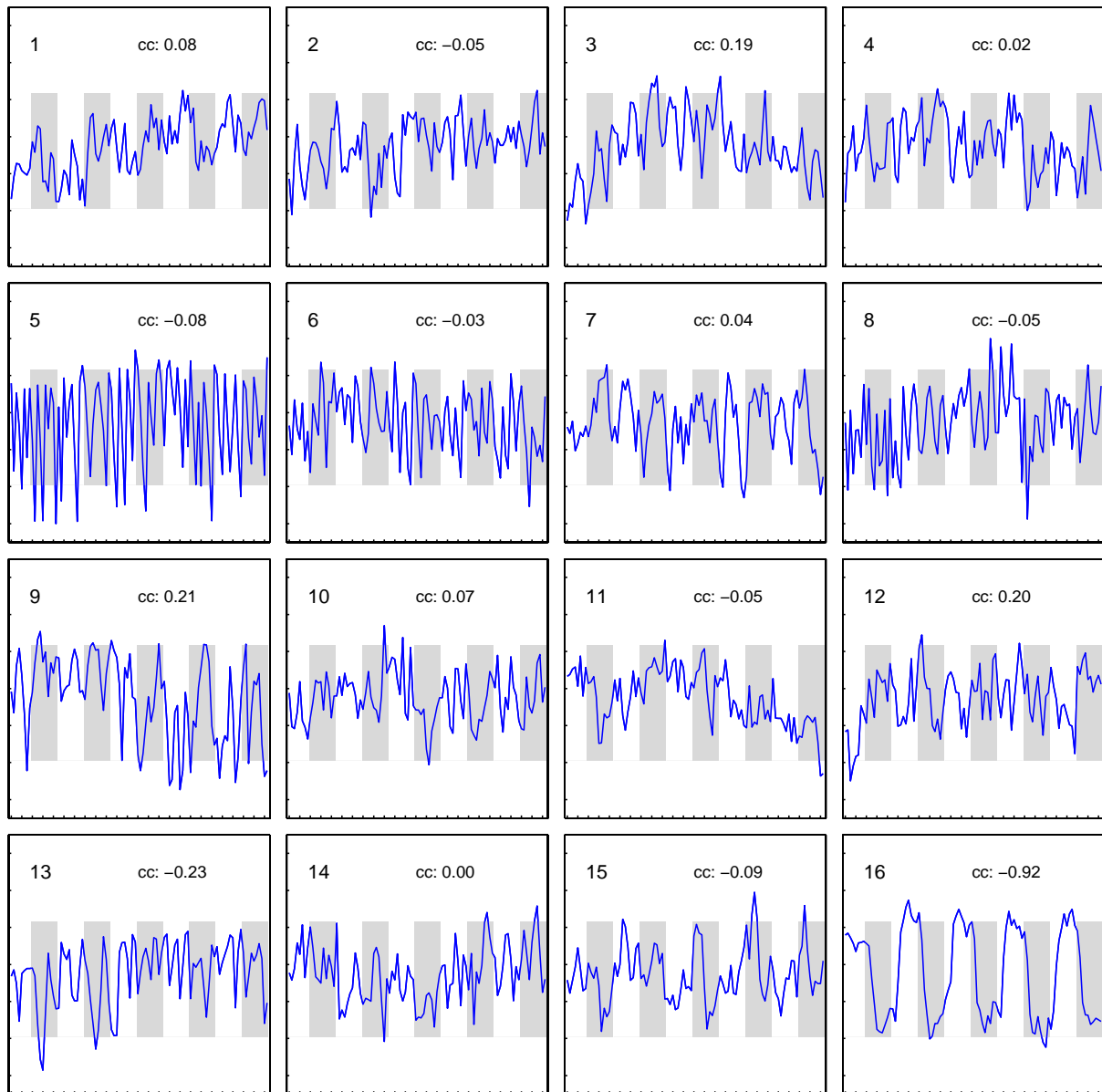
- "Neural gas" network and minimum free-energy network outperform all others
- Motivation: more detailed analysis of crisp and fuzzy clustering techniques

5.1 Advanced Clustering Techniques

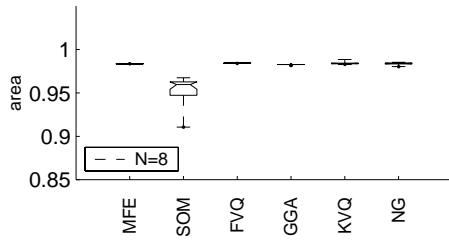
- Gath–Geva cluster assignment maps for 16 components:



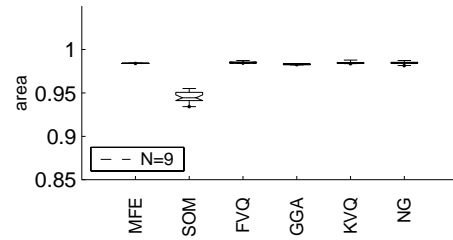
● Associated ICs:



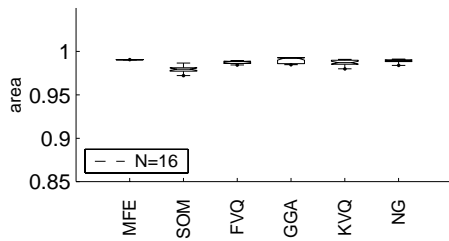
- Comparison between different algorithms: number of components $N=8,9$ and 16 based on ROC analysis and correlation map with a chosen threshold of 0.4



(a)

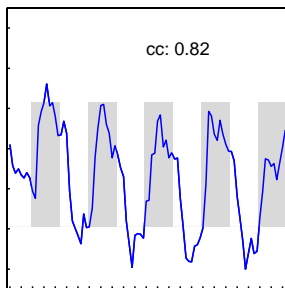


(b)

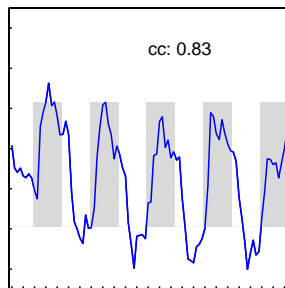


(c)

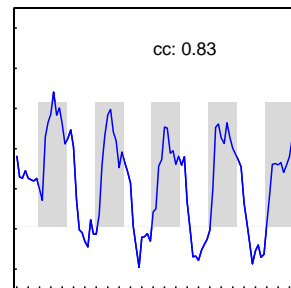
- Reference functions for Gath–Geva algorithm for $CV=8,9$ and 16



$CV=8$

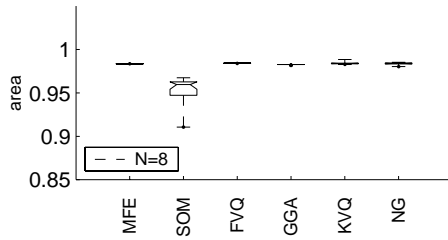


$CV=9$

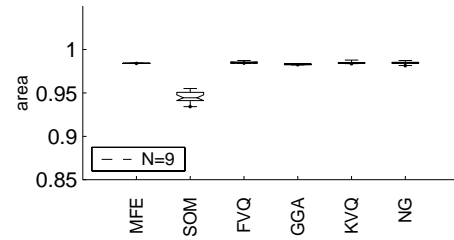


$CV=16$

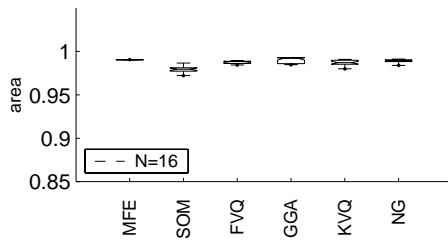
- Comparison between different algorithms: number of components $N=8,9$ and 16 based on ROC analysis and correlation map with a chosen threshold of 0.6



(a)



(b)



(c)

6 Conclusions

- Basic concepts of advanced vector quantization
- Basic concepts of fuzzy clustering
- Comparison between ICA and clustering techniques
- Characterization of task-related and artifactual effects
 1. Activity component maps and resulting time courses for the IC
 2. Differences between ICA and clustering techniques in the activation maps
 3. Motion artifacts and co-activation of the frontal eye fields
- Main result: advanced crisp or fuzzy techniques outperform ICA techniques