**Supplementary references:**


[Banerjee et al. 2005] Banerjee A., Dhillon I. S., Ghosh J. and Sra S.: Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. Journal of Machine Learning Research, 6: 1-39 (2005)

[Baurngartner et al. 1997] Baurngartner R., Scarth G., Somorjai R. and Moser E.: Fuzzy Clustering of Gradient-Echo Functional MRI in the Human Visual Cortex.Part I: Reproducibility. Magnetic Resonance Imaging, 7(6): 1094–1101 (1997)

[Baune et al. 1999] Baune A., Sommer F. T., Erb M., Wildgruber D., Kardatzki B., Palm G. and Grodd W.: Dynamical Cluster Analysis of Cortical fMRI Activation. NeuroImage 9: 477-489 (1999)

[Bezdek et al. 1984] Bezdek J. C., Ehrlich R and Full W.: FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3): 191–203 (1984)

[Cohen et al. 2008] Cohen A.L., Fair D.A., Dosenbach N.U.F., Miezin F.M., Dierker D., Van Essen D.C., Schlaggar B.L. and Petersen S.E.: Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage, 41: 45–57 (2008)

[Cordes et al. 2002] Cordes D., Haughton V., Carew J. D., Arfanakis K. and Maravilla K.: Hierarchical clustering to measure connectivity in fMRI resting-state data. Magnetic Resonance Imaging, 20: 305-317 (2002)

[Duda et al. 2000] Duda R. O., Hart P. E. and Stork D. G.: Pattern Classification. 2nd edition, Wiley Interscience (2000)

[Filzmoser et al. 1999] Filzmoser P., Baumgartner R. and Moser E.: A hierarchical Clustering Method for analyzing functional MR images. Magnetic Resonacne Imaging 17(6): 817-826 (1999)

[Frey and Dueck 2007] Frey B. J. and Dueck D.: Clustering by Passing Messages Between Data Points Brendan. Science, 315: 927–951 (2007)

[Fischer et al. 1999] Fischer H and Hennig J.: Neural Network-Based Analysis of MR Time Series. Magnetic Resonance in Medicine, 41: 124–131 (1999)

[Golland et al. 2008] Golland P., Lashkari D. and Venkataraman A.: Spatial Patterns and Functional Profiles for Discovering Structure in fMRI Data. 42nd Asilomar Conference on Signals, Systems and Computers, 1402 –1409 (2008)

[Golland et al. 2009] Golland Y., Golland P., Bentin S. and Malach R.: Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. Neuropsychologia, 46(2):540–553 (2008)

[Goutte et al. 1999] Goutte C., Toft P., Rostrup E., Nielsen F. Å. and Hansen L. K.: On Clustering fMRI Time Series. NeuroImage, 9: 298–310

[van den Heuvel et al. 2008] van den Heuvel M., Mandl R. and Hulshoff Pol, H.: Normalized cut group clustering of resting-state FMRI data. PLoS One, 3(4): e2001 (2008)

[Langfelder et al. 2007] Langfelder P. and Horvath S.: Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol, 1(54) (2007)

[Liao et al. 2009] Liao W., Chen H., Yang Q. and Lei X.: Analysis of fMRI Data Using Improved Self-Organizing Mapping and Spatio-Temporal Metric Hierarchical Clustering. IEEE Transactions on Medical Imaging, 27(10): 1472–1483 (2009)

[McKeown et al. 1998] McKeown M.J. and Sejnowski T.J.: Independent component analysis of fMRI data: examining the assumptions. Hum Brain Mapping, 6(5-6):368-72 (1998)

[Mezer et al. 2009] Mezer A., Yovel Y., Pasternak O., Gorfine T. and Assaf Y.: Cluster analysis of resting-state fMRI time series. NeuroImage, 45: 1117–1125 (2009)

[Raichle et al. 2001] Raichle M.E., MacLeod A.M., Snyder A.Z., Powers W.J., Gusnard D.A., Shulman G.L.: A default mode of brain function. Proc Natl Acad Sci U S A. 16;98(2):676–82 (2001)

[Salvador et al. 2005] Salvador R., Suckling J., Coleman M. R., Pickard J. D., Menon D. and Bullmore E.: Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain. Cerebral Cortex, 15(9): 1332–1342 (2005)

[Windischberger et al. 2003] Windischberger C., Barth M., Lamm C., Schroeder L., Bauer H., Gur R.C. and Moser E.: Fuzzy cluster analysis of high-field functional MRI data. Artificial Intelligence in Medicine, 29: 203–223 (2003)

**Previous approaches to rs-fMRI parcellation, a very brief overview (supplementary material)**

A wide range of different clustering approaches have been applied to fMRI data in the past. Methods include: the 'normalised CUTS' (NCUTS) algorithm (Shi et al. 2001), which was used for fMRI analysis in (van den Heuvel et al. 2008), (Craddock et al. 2011) and (Shen et al. 2010); a spectral clustering approach based on network modularity optimisation (Newmann 2006) used for fMRI in (Shen et al. 2010); the 'infomap' algorithm of (Rosvall and Bergstrom 2008) used for fMRI parcellation in (Power et al. 2011); a range of hierarchical clustering approaches based around several linkage rules (with the Ward method (Ward 1963) being the most popular), which have been used for fMRI in (Cordes et al. 2002), (Goutte et al. 1999), (Salvador et al. 2005) and (Liao et al. 2009), whilst a hierarchical approach based on dynamic tree cutting (Langfelder et al. 2007) was used for fMRI in (Mumford et al. 2010). Other popular approaches include independent component analysis (ICA) (McKeown et al. 1998), (Beckmann and Smith 2004) or k-means clustering (Duda et al. 2000) and its fuzzy alternative (Bezdek et al. 1984), which have been used for fMRI in (Baune et al. 1999), (Filzmoser et al. 1999), (Goutte et al. 1999), (Baurngartner et al. 1997), (Golland et al. 2008), (Golland et al. 2009), (Mezer et al. 2009), (Bellec et al. 2010), (Kim et al. 2010), (Windischberger et al. 2003) and (Fischer et al. 1999). More complex statistical models have also been proposed, such as the approach based on a von Mises Fisher mixture model and expectation maximisation inference (Banerjee et al. 2005) which was used for fMRI analysis in (Lashkari et al. 2010). An alternative recent approach is based on the gradient of the connectivity profiles and edge detection, which has been used together with a watershed segmentation in (Cohen et al. 2008). The affinity propagation algorithm of (Frey and Dueck 2007) was used for fMRI in (Zhang el at. 2011), whilst self organising maps have been explored for fMRI in (Fischer et al. 1999) and (Liao et al. 2009). Finally, region growing approaches have been used previously in (Lu et al. 2006), (Bellec et al. 2006) and (Heller et al. 2006).

When using many of the aforementioned algorithms, there is often an additional flexibility in the measure used to quantify the functional homogeneity within a region. Popular measures include Pearson correlation, Euclidean distance and coherence. These measures are often used to directly measure the similarity between the timeseries of two voxels. An alternative approach to measure voxel similarity first constructs similarity maps for each voxel (also called connectivity fingerprints/profiles or functional neighbourhoods). These are typically computed by comparing a voxel's timeseries to the timeseries of other voxels, thus producing voxelwise feature vectors. Two of these feature vectors can then be compared using either correlation (possibly after a Fisher r to z transform) or Euclidean distance, whilst a similar "$\eta^2$" coefficient was used in (Cohen et al. 2008).

Furthermore, a wide range of additional variations on these measures has been used, including the use of spatial neighbourhood constraints (Craddock et al. 2011), temporal low-pass filtering (to reduce the influence of high-frequency artefacts and to improve SNR), thresholding (to reduce spurious small correlations

or to sparsify the distance matrix) and Fisher r to z transformation of correlation values (to increase their normality).

**Evaluating different aspects and variations of the approach (supplementary material)**

The clustering approach outlined here uses several computational steps and each of these steps can potentially be replaced with several alternative methods to achieve the same aim. During the design of the approach, several alternative approaches were tried and some of the results of these experiments are given here. We also study the importance of some of the steps in terms of the performance of the approach.

*Evaluating the importance of reliable seed identification*

How important is it to select "good" seed locations, and how important is it for the reproducibility of the clusters to be able to identify similar seeds? To answer this question, we used two sets of 60 minutes of rs-fMRI data taken from the same subject (data-set 1) and run the following four variations to the method. (1) using different seeds, each set of seeds estimated from one of the scans, (2) using the same seeds (estimated from one of the two scans) for both data-sets, (3) using different randomly located seeds for each of the scans and (4) using the same set of randomly selected seed locations for both of the scans. After seed selection, processing proceeded in the same way. Seeds were grown into initial regions as described above and these regions were further clustered using a hierarchical clustering approach based on Ward's linkage (no locality constraint was used here and a threshold was used to define a cluster resolution to analyse). Before the final analysis, spatially disjoint clusters were split into spatially homogeneous clusters.

The cluster reproducibility for each of the four settings is given in table 1, where we give the average Dice similarity and its standard deviation as well as the average number of seeds (and standard deviation). It is clear from this table that the reliable identification of seed locations is important for the reliable estimation of reproducible parcellations. It is also clear how our stability map based approach leads to better performance than random seed selection. The fact that random seed selection can still lead to fairly reproducible parcellations indicates that the final parcellation only partially depends on the initial seed locations and that the initial region growing approach is likely to be able to compensate to some extent for imperfect seed locations.

*Choosing representative region time-courses*

After the initial region growing step, representative time-courses have to be defined for each region, which can then be fed into the next stage of hierarchical clustering. We here compared several approaches, (1) taking the time-course of the original seed, (2) taking the mean time-course from an ROI centred around the seed, (3) taking the mean time-course from the entire region and (4) taking the principal time-course from the entire region. All other parcellation steps were identical to the standard approach developed here.

| | | before joining | | after joining | |
|---|---|---|---|---|---|
| | | Dice | # clusters | Dice | # clusters |
| stable seeds | same | 0.73 (0.018) | 611 (15) | 0.9 (0.006) | 468 (14) |
| | different | 0.55 (0.021) | 611 (14) | 0.76 (0.015) | 423(14) |
| random seeds | same | 0.69 (0.011) | 755 (15) | 0.88 (0.004) | 545 (14) |
| | different | 0.41 (0.011) | 755 (18) | 0.6 (0.011) | 490 (20) |

Table 1: Comparison between parcellation reproducibility based on seeds selected as stability map minima and randomly selected seed locations. We furthermore compare the use of the same seeds against the use of different seeds for the two parcellations. The left columns show the Dice similarity and number of clusters calculated before an additional parcel joining step whilst the right columns show the results after this joining.

The results are shown in table 2, where we show the mean Dice similarity of parcels in each of the parcellations both before and after parcel joining for a range of parcellation resolutions. Standard deviations are given in parenthesis.

It appears that, due to our choice of stable seeds, the seed time-course itself or the average over a small neighbourhood seems to be better than to use a time-course derived from the entire region. This might be due to the fact that the stable seed's time-course is more representative than many of the other region's time-courses.

*Comparing different hierarchical clustering approaches*

Finally, we compare the influence of different choices of final hierarchical clustering methods on parcellation reproducibility. We here compare five approaches, two locally constrained methods and three unconstrained approaches.

The three unconstrained approaches used a hierarchical clustering method with Ward's linkage rule and three different similarity measures. (1) the euclidean distance between connectivity fingerprints, (2) correlation of the Fisher transformed connectivity fingerprints and (3) Pearson correlation. The spatially constrained hierarchical clustering also used Ward's linkage and used (1) Pearson correlation and (2) correlation of the Fisher transformed connectivity fingerprints as similarity measures. Connectivity fingerprints were the vectors of Pearson correlation coefficients between the characteristic time-series of a cluster and a pre-specified set of 1000 randomly chosen vertex time-series[7].

[7]We used a random selection of vertices to reduce computational burden. We found that this random selection did not degrade performance and the similarity based on connectivity fingerprints based on different randomly chosen target vertices did not differ much.

| Number of regions | | 200 | 400 | 600 | 800 | 1000 |
|---|---|---|---|---|---|---|
| seed timecourse | before | .69(.21) | .70(.21) | .75(.2) | .77(.19) | .78(.18) |
| | after | .78(.17) | .79(.15) | .83(.13) | .84(.11) | .84(.11) |
| ROI timecourse | before | .69(.2) | .70(.22) | .75(.2) | .78(.18) | .78(.17) |
| | after | .79(.15) | .80(.15) | .83(.12) | .84(.11) | .84(.11) |
| mean timecourse | before | .71(.18) | .71(.20) | .72(.21) | .74(.19) | .74(.20) |
| | after | .77(.16) | .80(.14) | .82(.13) | .83(.12) | .89(.14) |
| PCA timecourse | before | .72(0.19) | .71(.21) | .74(.19) | .75(.20) | .76(.18) |
| | after | .80(.13) | .80(.14) | .84(.11) | .83(.12) | .83(.13) |

Table 2: Comparison of parcellation reproducibility using four different ways to calculate representative time-courses for each of the initial clusters. Results are shown for parcellations with 200, 400, 600, 800 and 1000 parcels. Results are given *before* and *after* joining clusters that are matched to a single cluster in the other parcellation.

For the unconstrained approaches, clusters were split into spatially contiguous clusters after clustering and the cluster numbers reported below are those found after this splitting.

The results are compared in table 3 where Dice similarity and cluster sizes are shown for both parcellations. Again, results are given for the initial parcellations (before) as well as for parcellations found after joining of clusters. We selected the results so that the final parcellations all had comparable numbers of clusters.

Non-local methods return far larger numbers of spatially homogeneous clusters than the locally constrained methods for a fixed parcellation resolution. To compare Dice, we here match the cluster sizes. Note that a greater number of small clusters will tend to have higher Dice. Thus, the fact that we achieve much higher Dice with fewer clusters using the local method speaks in favour of the local hierarchical approach.

| | | # of clusters | Dice |
|---|---|---|---|
| Euclidean connectivity fingerprint | before | 1201/1262 | .60(.24) |
| | after | 834/858 | .76(.16) |
| Correlation of connectivity fingerprint | before | 1267/1331 | .62(.23) |
| | after | 904/921 | .77(.15) |
| Timeseries Correlation | before | 1121/1181 | .59(.25) |
| | after | 766/791 | .75(.16) |
| Local Timeseries Correlation | before | 1000/1000 | .76(.18) |
| | after | 860/860 | .83(.13) |
| Local correlation of connectivity fingerprint | before | 1000/1000 | .75(.19) |
| | after | 836/830 | .83(.12) |

Table 3: Comparison of parcellation reproducibility for a single subject using five different hierarchical clustering methods - two locally constrained approaches and three unconstrained methods. For the unconstrained approaches, results were calculated after splitting the parcels into spatially homogeneous clusters. From top to bottom, we used the euclidean distance between connectivity fingerprints, correlation of the Fisher transformed connectivity fingerprints, Pearson correlation, Pearson correlation using a locally constrained hierarchical approach and a locally constrained hierarchical approach based on correlation of the Fisher transformed connectivity fingerprints. Connectivity fingerprints were calculated as the vector of Pearson correlation coefficients between the characteristic time-series and a pre-specified set of 1000 randomly chosen vertex time-series. Results are given *before* and *after* joining clusters that are matched to a single cluster in the other parcellation.

**Supplementary figures**

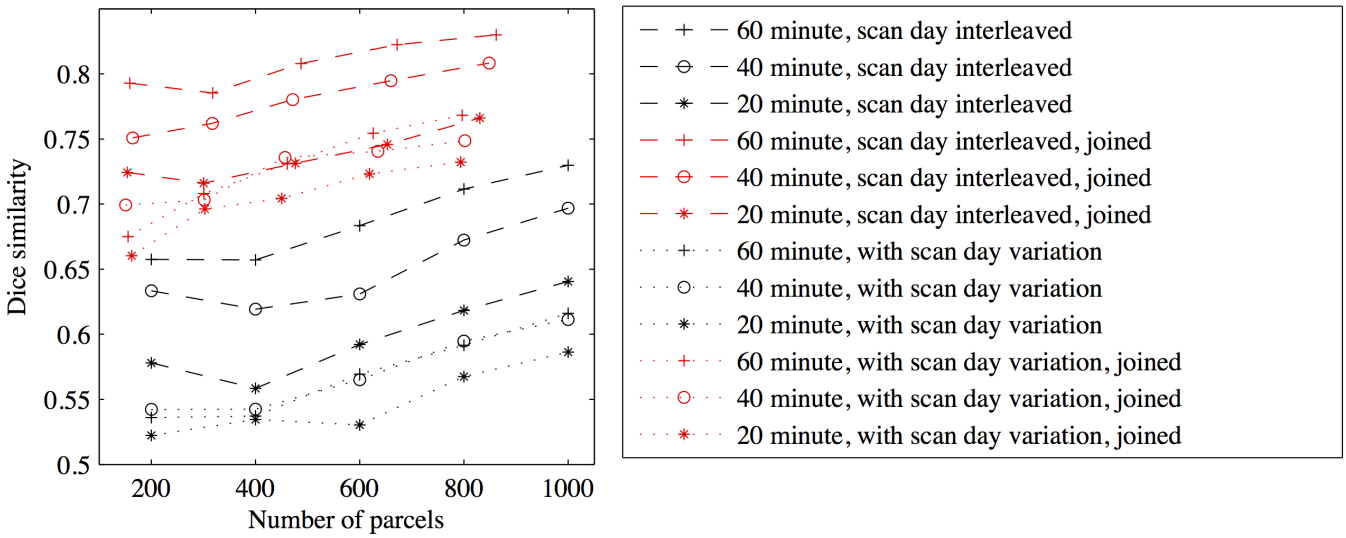


Figure 10: Average Dice similarity between matched parcellations vs. parcellation level, calculated for parcellations from different datasets of the same subject before (black) and after (red) joining of split clusters. Dashed lines show results calculated from a split of the data in which each of the two subsets contained an equal numbers of 10 minute scans from the two scan days, whilst dotted lines show results calculated from a split of the data in which each subset only contained scans from the same day, so that these results also reflect variations due to scan day effects. For each experiment condition, subsets contained data of 60 minutes ($6 \times 10$ min., (+)), 40 minutes ($4 \times 10$ min., (o)) and 20 minutes ($2 \times 10$ min., (*)).

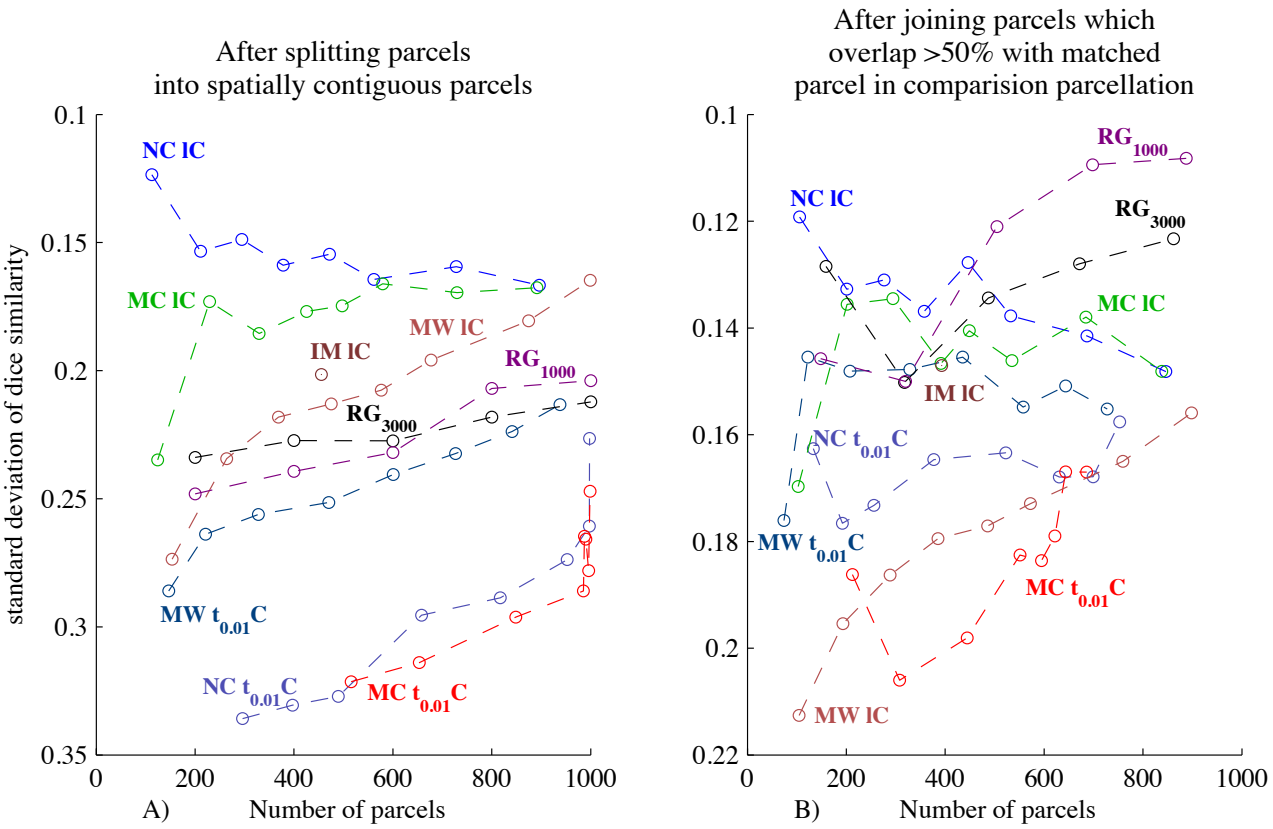


Figure 11: Standard deviation of dice similarity between matched parcellations vs. parcellation level, calculated for parcellations from different datasets of the same subject before (A) and after (B) joining split clusters. Results shown for the data-set one subject which was scanned twice for 60 minutes. The number of parcels is the total across both hemispheres. Region growing approach with ∼1000 ($RG_{1000}$) and ∼3000 ($RG_{3000}$) seeds outperforms all other tested approaches over a range of parcellation resolutions, especially after parcel joining (B). Also shown is a small subset of the results obtained with other methods (including the next best performing method, NCUTS (NC lC) (Shi et al. 2001) used in (Craddock et al. 2011), a spectral clustering approach to optimise network modularity (MC lC, black) (Newmann 2006), a hierarchical clustering approach using Ward's linkage rule (HW lC) (Ward 1963) and the infomap algorithm (IM IC) (Rosvall and Bergstrom 2008) as used in (Power et al. 2011), all with the same locally restricted correlation similarity measure. For comparison, the results obtained by the same approaches with a sparser similarity matrix (a correlation matrix in which values were thresholded so that only 1% of the entries in each column/row were non-zero (Power et al. 2011)) are also shown ({NC, MC, HW} $t_{0.01}$C).)

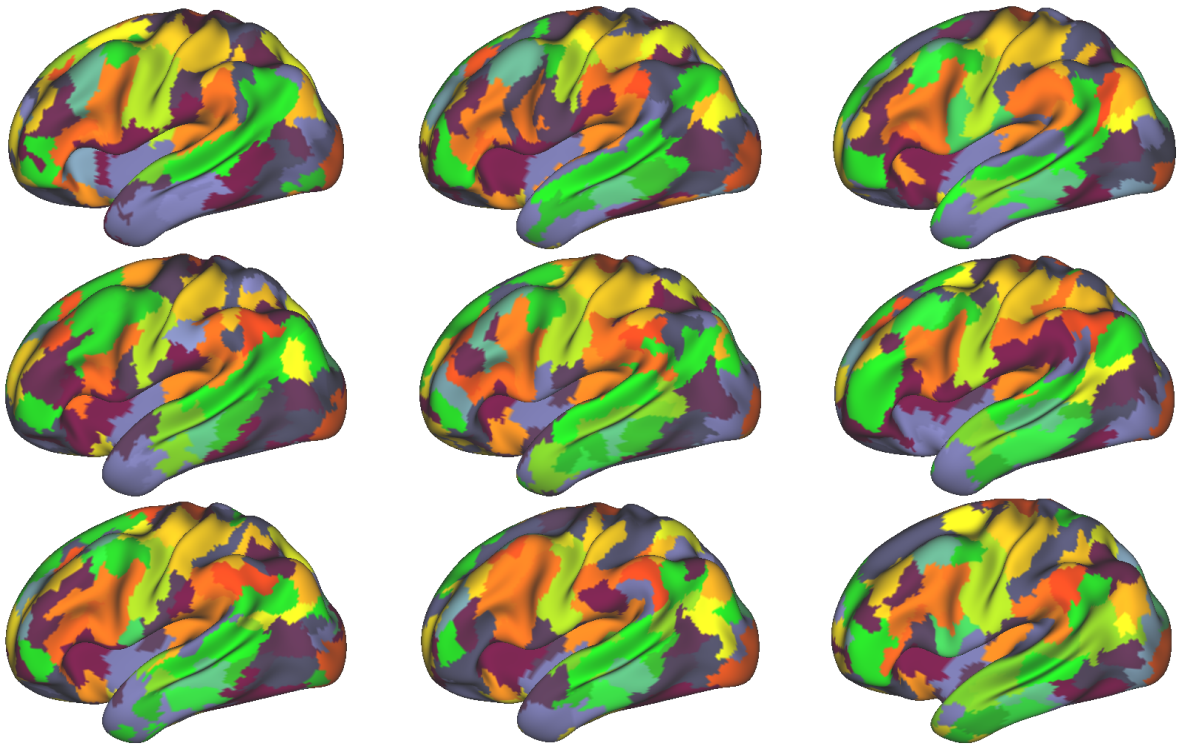

Figure 12: Comparison of region growing parcellations for 9 different subjects from data-set 2 at a resolution of 200 parcels. Parcel colours have been matched for ease of visualisation.

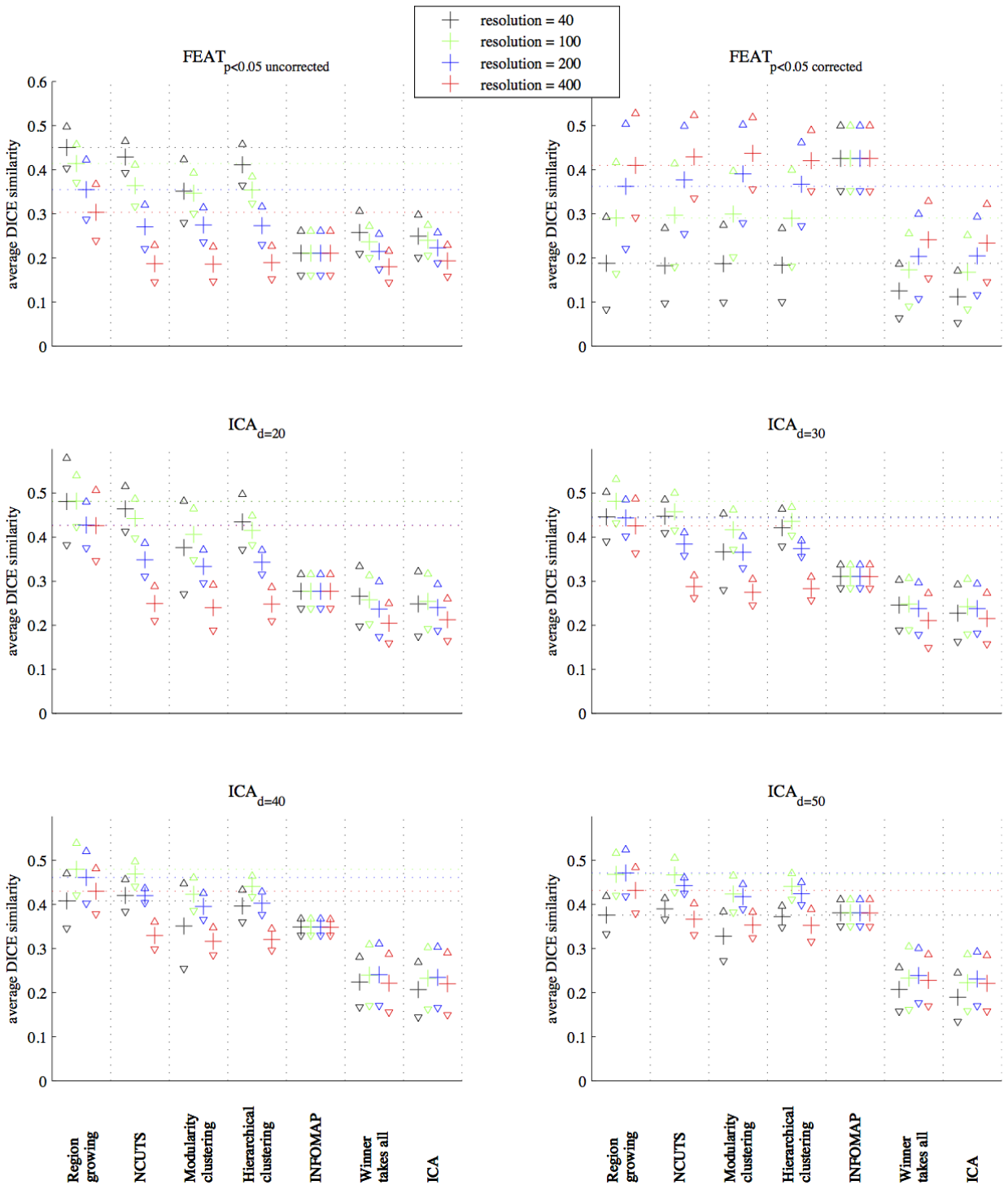


Figure 13: Comparison of parcellations from data-set 2, derived with different methods, to task fMRI derived clusters. Shown are the mean Dice similarity (+) and the ± 3 standard deviations (△ and ▽), where the mean is over the Dice measure between each task cluster and the best matching parcel from a parcellation. Results are shown for different parcellation methods and different parcellation resolutions. Task fMRI clusters were derived using general linear model analysis with thresholds at $p < 0.05$ (both, uncorrected and Bonferroni corrected results are shown) as well as ICA analysis using the FSL melodic tool (Beckmann and Smith 2004) with the number of components set to 20, 30, 40 and 50 respectively. Parcellations were derived for different resolutions producing 40, 100, 200 and 400 clusters, apart form the infomap results, where the resolution is determined automatically.

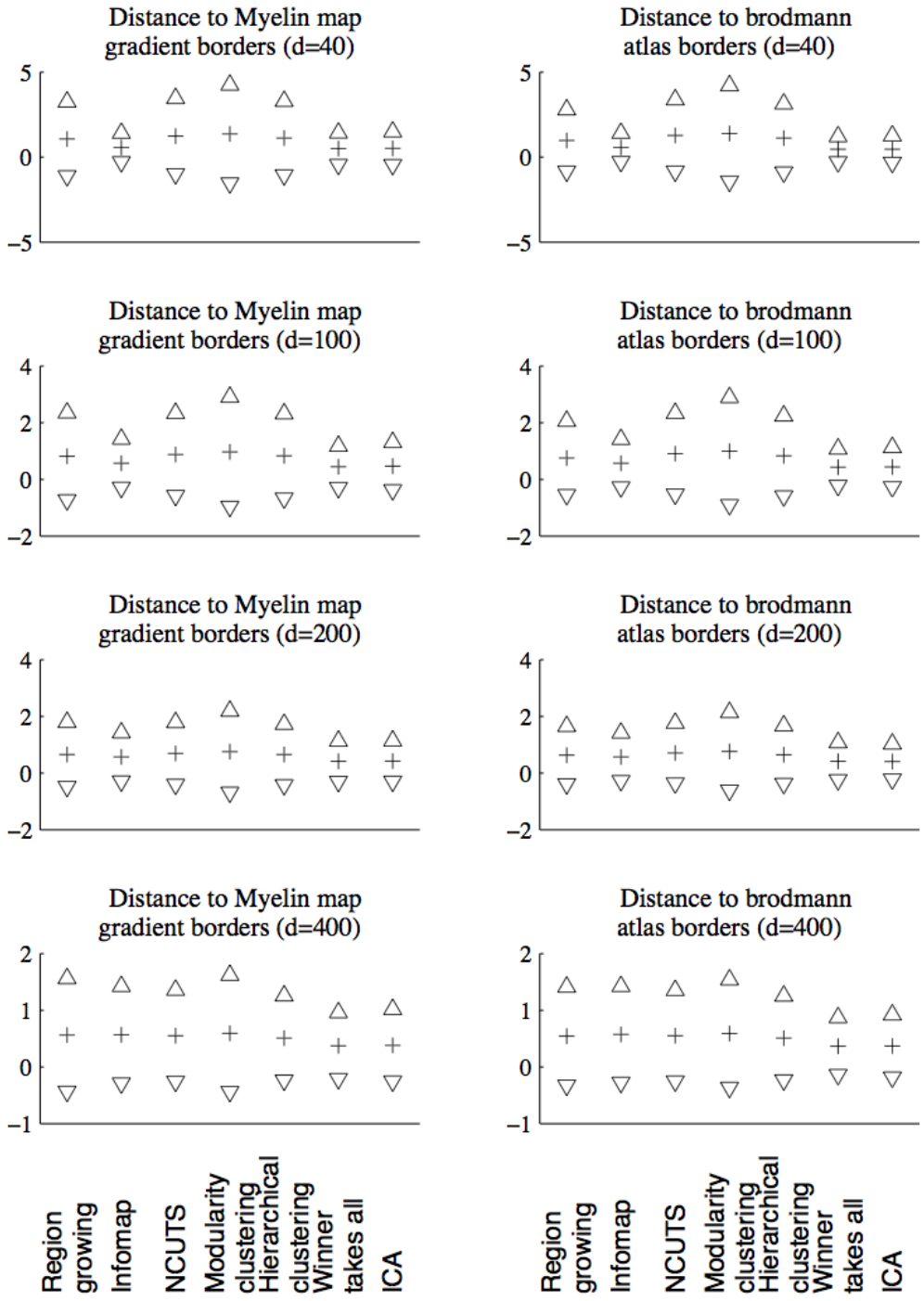


Figure 14: Comparison of parcellations from data-set 2, derived with different methods, to Brodmann atlas borders and Myelin map derived borders (defined as maximal gradients from the Myelin atlas (Glasser et al. 2011a)). Shown are the mean distances (+) and the ± 3 standard deviations (△ and ▽), between each point on the cluster borders to the closest point on the borders derived from the Brodmann atlas (right) or the borders derived from the Myelin atlas (left).

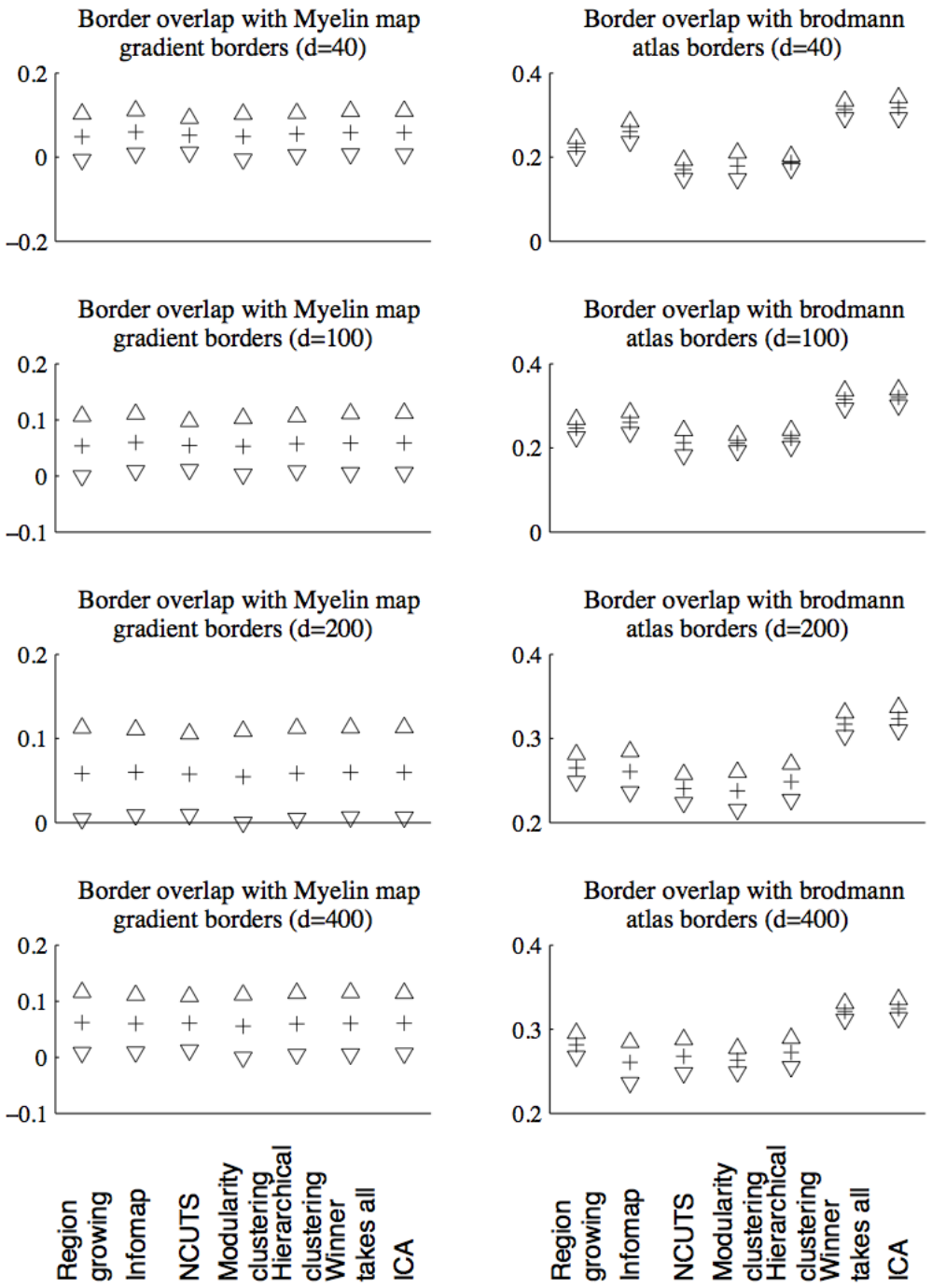


Figure 15: Comparison of parcellations from data-set 2, derived with different methods, to Brodmann atlas borders and Myelin map derived borders (defined as maximal gradients from the Myelin atlas (Glasser et al. 2011a)). Shown are the mean Dice border overlap (+) and the ± 3 standard deviations (△ and ▽), between the cluster borders and the borders derived from the Brodmann atlas (right) or the borders derived from the Myelin atlas (left).