# A generalizable approach to quality assessment of voxel-based morphometry data

Albena Vassileva[1,2,3,6], Shahrzad Kharabian-Masouleh[1,3], Deepthi P Varikuti[1,3,4], Felix Hoffstaedter[1,3,4], Holger Schwender[5], Svenja Caspers[7,8,9], Christian Grefkes[2,6], Simon B Eickhoff[1,3,4]

1,2 Institute of Neuroscience and Medicine (INM-7[1], INM-3[2]), Research Center Jülich, Jülich, Germany; 3. Institute for Systems Neuroscience Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 4. Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany; 5. Mathematical Institute Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 6.Department of Neurology, University of Cologne, Cologne, Germany; 7. C. and O. Vogt Institute for Brain Research, Medical Faculty, Heinrich Heine University Düsseldorf, 40001 Düsseldorf, Germany; 8. Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany; 9. JARA-BRAIN, Jülich-Aachen Research Alliance, Jülich, Germany

## Introduction

We introduce and evaluate an automatized quality assessment approach based on low-rank representations via Orthonormal Non-negative Matrix Factorization (OPNMF)[1] of preprocessed structural MRI data. OPNMF produces:

- *W* matrix - sparse spatial distribution of the components
- *H* matrix - loading coefficients for each image

**Aim**: to provide a method for detection of aberrant images, which is:

1. **Generalizable**
2. **Data-driven** (i.e. no manual labels required)

The performance of the proposed approach is assessed as:

1. "Outlier detection", i.e. within-sample factorization
2. "Quality control", i.e. out-of-sample factorization - generalizability

## Methods

### Datasets

**I. Outlier detection paradigm**

1000BRAINS[2] dataset (n=1,324; age-range 18-85)

**II. Quality control paradigm**

multi-site clinical cohort (334 schizophrenia patients, 372 healthy controls; age-range 18-65; 13 sites)

### Preprocessing

- CAT12[3] with standard settings
- DARTEL-normalization and tissue segmentation
- Unsmooth, non-linear modulated only GM segments are used.
- Manual visual assessment by two raters of both datasets

### Analysis outline

**Flagging protocol**

(1) Calculation of Quality metrics (z-scored):
- **Component loadings** (**CL**; from *H* matric)
- **Whole-brain reconstruction error** (**RE**; from *W* and *H*)
- **ROI reconstruction error** (**ROI-RE**; maximal taken).

(2) Any image with at least two flags of |z-score|>2 is suspicious of artifacts.

**I. Outlier detection paradigm (within-sample factorization)**

(0) OPNMF is performed on 1000BRAINS

(1):(2) Perform Flagging protocol

**II. Quality control paradigm (out-of-sample factorization)**

(0) Factorization from 1000BRAINS is transferred to clinical sample.

(1):(2) Perform Flagging protocol

**III. Evaluate performance against manual visual inspection**

## Results

### Flagged images in outlier detection paradigm (1000BRAINS)

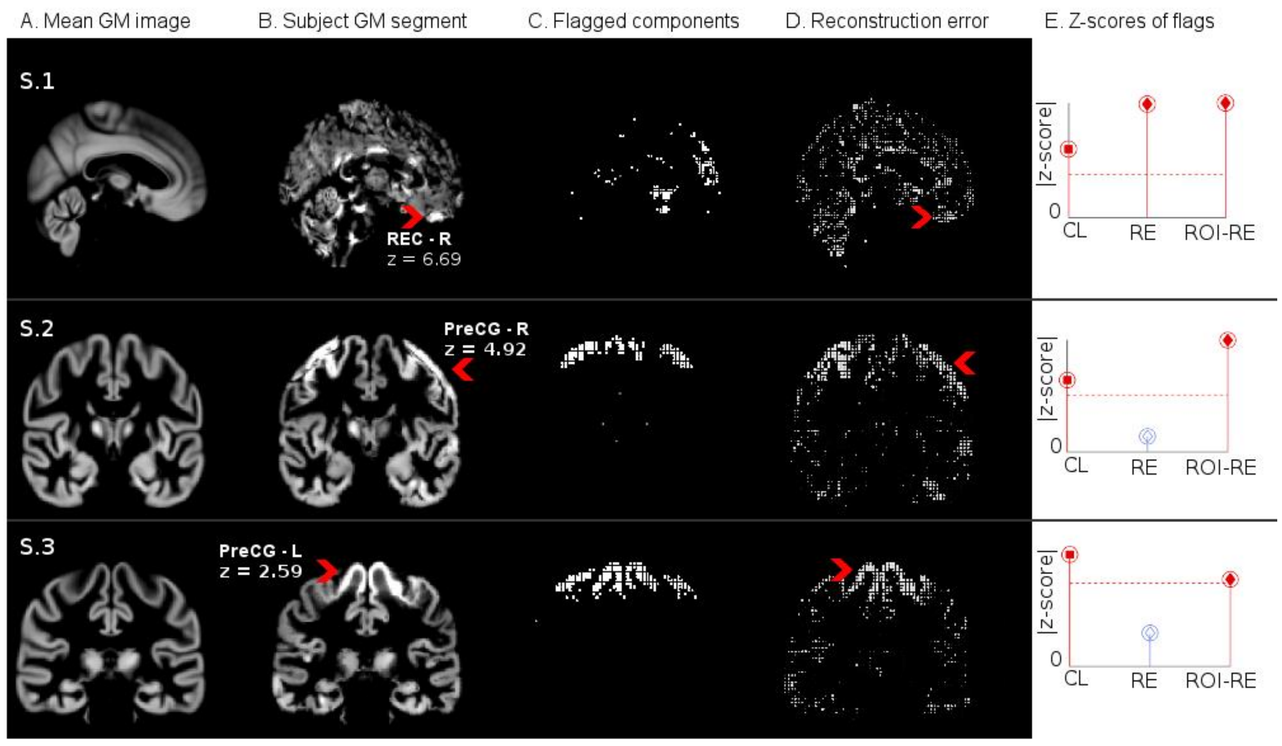
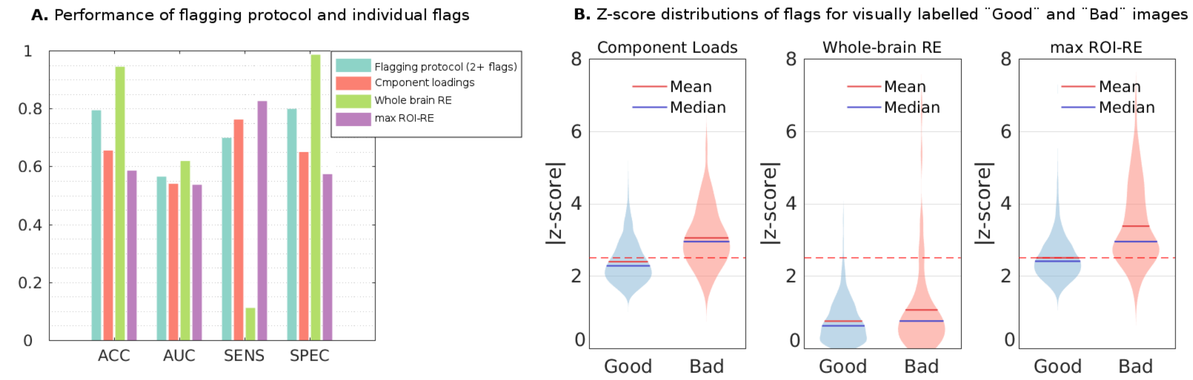



**Figure 1. Examples of images correctly flagged as "Bad".** The individual rows illustrate examples of images flagged as suspicious by at least two flags. Column A displays the mean GM image for the whole sample, for use as a visual reference. In column B there are the actual GM segments of interest. The red arrows indicate the ROIs with the largest |z-score|. Column C illustrates the voxels belonging to flagged components (i.e. components with |z-score| of at least 2.5); in column D is the reconstruction error for the voxels; and in column E the z-scores of each flag are plotted. Abbreviations: REC-R – Gyrus rectus (right hemisphere); PreCG-R/L – precentral gyrus (right/left)/

### I. Performance as outlier detection in healthy sample (1000BRAINS)




### II. Performance as quality control (in clinical sample)

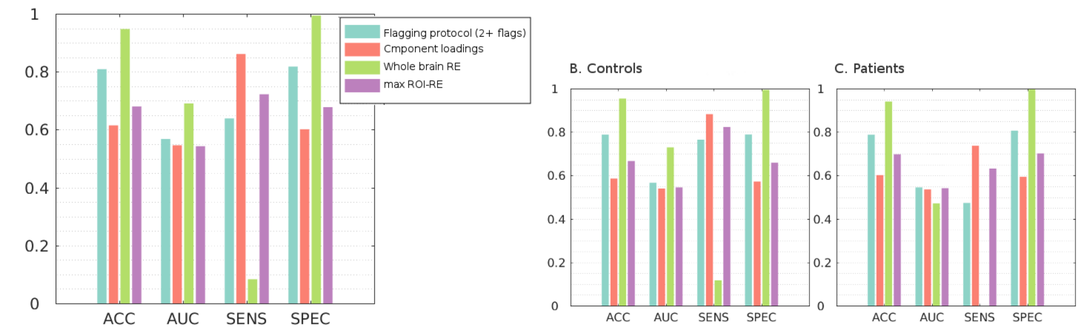



**Figure 2. Performance of the flagging protocol** when applied in the outlier detection paradigm **(I)** and when applied as quality control **(II)**. The violin plots on **I.B.** illustrate the separation between the images manually labelled as "Good" and "Bad" by the human raters, as measured by each of the individual flags – component loadings, whole-brain reconstruction error, and maximal ROI reconstruction error.
Figure **II.A.** shows the performance of the flagging protocol on the whole clinical sample (i.e. when applied to controls and patients pooled together). The performance of the protocol when applied to the two groups separately is shown on Figure **II.B.** and **C.**

### Images missed by the human raters

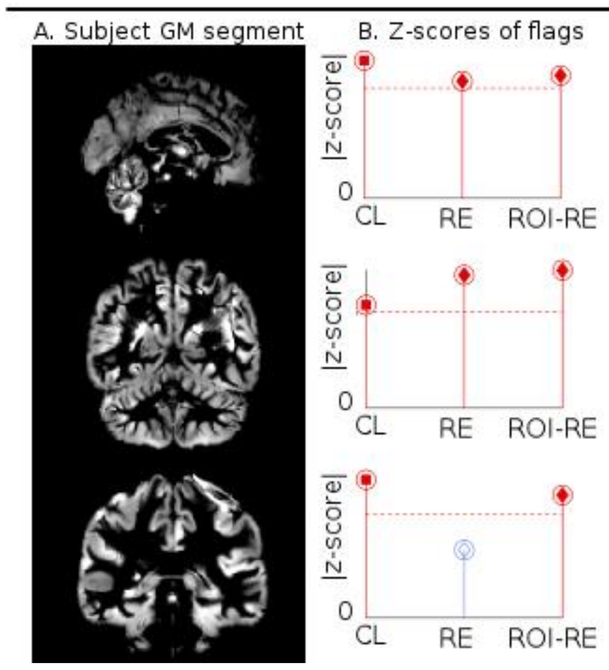



**Figure 3. "False positive" images containing artifacts, i.e. real true positives.** These are examples of three individual images detected by our automatic flagging protocol but missed by the human raters. The left column A shows GM slices at the coordinates of the voxel with the highest reconstruction error for each respective image. The scores of the individual flags are plotted in column B.

## Discussion

Our method is applicable to any large, already preprocessed VBM data at a low computational cost. The evaluation showed that:

- The component loadings and the ROI reconstruction errors can be used to flag aberrant images with good performance compared to manual ratings.
- Our method allows generalization to new datasets to identify aberrant images with good performance, especially on healthy subjects.
- Our method detects artifacts human raters can miss.

Data-driven techniques eliminate the dangers of human error and subjectivity. Unsupervised methods such as OPNMF have the advantage of not requiring training, yet are generalizable to new, unseen data.

**References:**

1. Sotiras et al.(2014): Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. Neuroimage 108:1–16. http://dx.doi.org/10.1016/j.neuroimage.2014.11.045.
2. Caspers S et al.(2014): Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. Front Aging Neurosci 6:1–14.
3. Gaser & Dahnke (2012): CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. HBM Conf 2012 32:7743.

**Acknowledgments**: The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, EI 816/11-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme "Supercomputing and Modeling for the Human Brain" and and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2).