# Sampling stability and processing parameter-dependent characteristics of the 3D fractal dimension as a marker of structural brain complexity in magnetic resonance images

Stephan Krohn[1], Martijn Froeling[2], Alexander Leemans[3], Dirk Ostwald[1,4], Jesús Jiménez[5], Pablo Villoslada[6,7], and Francisco J. Esteban[8]

[1]Computational Cognitive Neuroscience Laboratory, Freie Universität Berlin, Germany

[2]Department of Radiology, University Medical Center Utrecht, Netherlands

[3]Image Sciences Institute, University Medical Center Utrecht, Netherlands

[4]Center for Adaptive Rationality, Max-Planck Institute for Human Development, Berlin, Germany

[5]Department of Computer Science, Universidad de Jaén, Spain

[6]Center of Neuroimmunology, Institut d'Investigacions Biomediques August Pi Sunyer (IDIBAPS), Barcelona, Spain

[7]University of California, San Francisco, USA

[8]Systems Biology Unit, Department of Experimental Biology, Universidad de Jaén, Spain

Authors for correspondence:

| Stephan Krohn | Francisco J. Esteban |
|---|---|
| Habelschwerdter Allee 45 | Campus Las Lagunillas s/n |
| 14195 Berlin, Germany | E-23071 Jaén, Spain |
| stephan.krohn@fu-berlin.de | festeban@ujaen.es |

Keywords: fractal analysis, structural brain complexity, MRI biomarker, computational image analysis

## Abstract

Fractal analysis, i.e. the estimation of an object's fractal dimension (FD) as a marker of its morphometric complexity, has attracted increasing interest as a versatile tool for the analysis of structural neuroimaging data in both health and disease. However, a number of important methodological questions regarding fractal analysis in magnetic resonance images have so far remained unaddressed. This includes the stability of the FD over repeated within-subject measurements, i.e. the susceptibility of fractal analysis to noise, a formal assessment of its sampling distribution, and the impact of image acquisition and processing parameters. Importantly, fractal analysis has not yet been explored in detail in T2 contrast images. To address these issues, we analyzed structural images from the recently published MASSIVE data set (Multiple Acquisitions for Standardization of Structural Imaging Validation and Evaluation). We conduct a fine-grained stratification of image parameters, leading to 32 distinct analysis groups as a combination of image contrast, spatial resolution, segmentation procedures, tissue type, and image complexity. We estimate 3D tissue models based on the thus obtained input volumes and compute the FDs as the box-counting regression on these models. Furthermore, we present a detailed deviation analysis including resampling methods, composite normality assessment, outlier detection, and multivariate comparisons to establish the susceptibility of the FD to noise. We find that in both T1 and T2 contrasts, the FD of gray matter (GM) segmentations was generally higher than in white matter volumes (WM). FDs in both image contrasts were sampled in comparable range and showed similar responses to processing parameters, e.g. as regards the effects of binary vs. partial volume segmentation and a decrease in FD by image skeletization. Lower spatial resolution invariably resulted in decreased FDs in unskeletized images, while the response depended on the segmentation procedure in image skeletons. Furthermore, in multiple measurements, the FD can be assumed to be sampled from an underlying normal distribution. We tested different options for a sensible within-group deviation criterion and found that outlier detection by Grubbs testing and a 2 standard-deviation interval around the sample mean performed very well in this regard. Even with the more conservative threshold, the overall robustness of the FD to noise was well above 90 %. Most deviations were found in T1-weighted images, and binarized image skeletons were most susceptible to deviations. Importantly, our analysis was able to detect sample-wise deviation clusters, and we identify image registration as a source of noise in fractal analysis. Interestingly, registration-induced deviations were limited to T1-weighted images, lending even further support for the usefulness of T2 contrast in fractal analysis. In conclusion, we provide detailed evidence for the stability of the FD as a marker of structural brain complexity and its parameter-dependent characteristics in magnetic resonance images and thus contribute to the development of fractal analysis as a scientifically and clinically useful neuroimaging tool.

# 1 Introduction

Fractal analysis has attracted increasing interest as a versatile tool for the analysis of structural brain data on a cellular as well as a macroscopic scale and in both health and disease [1, 2]. The concept of fractality was predominantly developed by Benoît Mandelbrot [3] and essentially corresponds to the insight that most natural objects do not adhere to the smooth whole-integer dimensions of Euclidean geometry. Structurally complex objects are indeed better described by the fractal dimension (FD), which is not limited to integers and can be regarded as a measure of morphometric complexity [4]. Natural objects are of course not perfect fractals in the mathematical sense because they cannot be scaled up or down infinitely and their self-similarity is statistical (repeating patterns) rather than structural (repeating composition). That said, real-world fractals abound in nature: from the inanimate (e.g. coastlines, clouds, lightning) to the cellular (e.g. protein surfaces, viral receptor molecules, cellular shapes) and higher-order organisms (e.g. human bronchial and vascular ramifications) [1–4]. In the field of biomedical neuroimaging, the FD has been successfully applied in the anatomical description of cortical geometry [5] and as a biomarker in the detection of brain tissue alterations in early multiple sclerosis [6,7], interuterine growth restriction [8], the cortical features in Alzheimer's disease [9], cerebral tumors [10] as well as age-related brain atrophy [11].

However, in exploring fractal analysis as a tool for both basic research and clinical investigations, a number of methodological questions have not been addressed so far. In terms of validity, it is crucial to investigate the stability of the FD over repeated within-subject measurements without any clinical alterations, i.e. to capture the susceptibility of the FD to measurement noise, a fundamental aspect if we are to transfer fractal analysis into diagnostics. Formally, it is also interesting to see if the FD can be assumed to follow a normal distribution, a central assumption to many diagnostic markers. Furthermore, the impact of image processing parameters such as segmentation procedures, tissue type, image complexity and interpolation effects needs detailed study. Moreover, the effect of spatial resolution on the FD has not yet been addressed. Importantly, fractal analysis has not yet been explored in detail in T2 contrast images, which are substantial to both neuroimaging studies and clinical neuroradiological assessment.

To address these issues, we analyzed structural T1 and T2 images in both high and low spatial resolution from the recently published MASSIVE database (Multiple Acquisitions for Standardization of Structural Imaging Validation and Evaluation) [12]. The MASSIVE data set is uniquely suited to investigate the above questions because it provides high-quality magnetic resonance images acquired in the same subject over repeated scans in a highly controlled setting. We hypothesized that in a single healthy subject (25-year-old female) scanned multiple times over a short interval of time (2 weeks), it

is reasonable to assume that structural brain complexity did not change at all. Therefore, if the FD is indeed an empirically valid measure of structural brain complexity, it should show high stability across these short-interval measurements. In more formal terms, we thus assumed that the true but unknown underlying distribution of the sampled fractal dimensions was always the same and that any deviations in the estimated FDs were essentially due to noise.

Our methodological approach to the points raised above rests on an image processing procedure differentiating between image contrast, resolution, segmentation method, tissue type, and image complexity. We then calculated the FD on the 3-dimensional tissue models (therefore 3DFD) obtained from the thus processed data. As detailed below, this leads to 32 distinct analysis groups as a combination of image category and processing parameters. We then first carry out comparative analyses of the mean FDs across analysis groups to investigate the effect of image processing parameters, establish a detailed description of the fractal dimension in T2 images and to compare high and low spatial resolution. Furthermore, we derive a detailed deviation analysis algorithm including resampling methods, composite normality assessment, outlier detection, and multivariate comparisons to investigate whether the individual FDs within a particular analysis group differed from each other in order to capture parameter-dependent repeated-measurements stability. We examine different options for a sensible within-group deviation criterion and apply them to the whole data set to establish the susceptibility of the FD to noise. Finally, we explore interpolation effects as a potential source of noise and discuss our findings in light of both technical validation and the clinical potential of fractal analysis in brain imaging.

## 2 Methods

### 2.1 Image acquisition

Structural T1 and T2 data analyzed here are part of the recently published MASSIVE dataset, openly available from `www.massive-data.org`. The data was acquired on a clinical 3 T system (Philips Achieva) with an eight-channel head coil. In brief, the T1 and T2 data were acquired with a FOV of 240 x 180 x 140 mm$^3$ and an acquisition matrix of 240 x 90 x 140 and reconstructed with a 1 x 1 x 1 mm$^3$ isotropic resolution. In total, ten T1 and ten T2 datasets were acquired in five sessions within a two week period (on days 0, 3, 7, 9, and 14). Each session consisted of two scans with a two hour pause in between. All datasets were registered to a common space using a rigid registration algorithm (www.elastix.org, see [13]) with the first T1 volume as the registration target. Furthermore, the data for both contrasts were resampled to a lower resolution resulting in two image sets, one with a 1 x 1 x 1 mm$^3$ and one

4

with a 2.5 x 2.5 x 2.5 mm$^3$ isotropic resolution. As a result, we obtained the four image categories T1 high resolution, T1 low resolution, T2 high resolution, and T2 low resolution for further processing. For additional details on the acquisition parameters, please refer to [12].

## 2.2 Image processing & fractal analysis

An FSL-based analysis pipeline [14–16] was used to preprocess the MR images for subsequent fractal analysis. Figure 1 displays a schematic representation of the according processing steps.
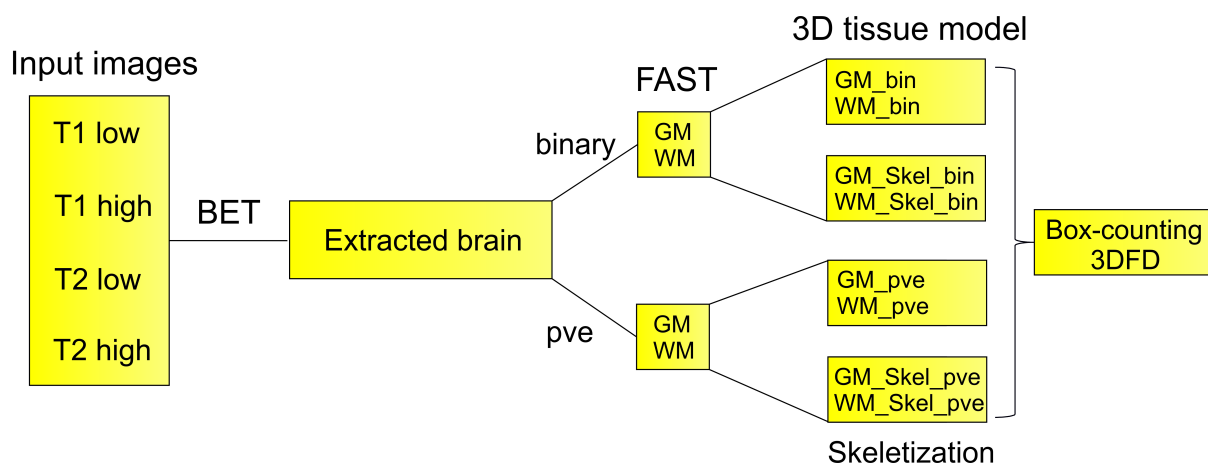


**Figure 1: Image processing pipeline.** The schematic displays the major steps applied in the processing of the structural images from the MASSIVE database. BET refers to the brain extraction routine and FAST to the tissue segmentation procedure. Skeletization means the estimation of a topological image skeleton (see main text for details). There were ten volumes in each of the four input categories: T1 and T2 sequences in low (2.5 mm$^3$) and high (1mm$^3$) spatial resolution. Based on the different analysis parameters, there were 8 processed output images for every input volume, corresponding to the segmentation procedure (partial volume estimates (pve) vs. binary segmentation (bin)), tissue type (gray matter (GM) vs. white matter (WM)) and image complexity reduction (skeletized (Skel) vs. unskeletized images). In total, we thus obtained 320 fractal dimension values in the data set (4 image categories x 10 input volumes each x 8 processed volumes for every input volume), calculated by evaluation of the box-counting dimension on a 3D tissue model estimated from the images (3DFD).

Specifically, the brain extraction routine (BET) was applied to all individual 3D volumes with default fractional intensity threshold [17]. The brain-extracted images were passed on to the FAST routine for tissue segmentation into gray matter (GM), white matter (WM) and cerebrospinal fluid classes with default analysis parameters [18]. We estimated partial volume maps for each of the three tissue classes, of which the GM and WM estimates entered the fractal analysis. For qualitative comparison, we also included a forced-decision binary classification ("hard" segmentation), in which voxels are labeled as 0 or 1 for a specific tissue class. The resulting images were grouped according to analysis class and transferred to a server for further processing based on fast cloud computing. From the segmented data,
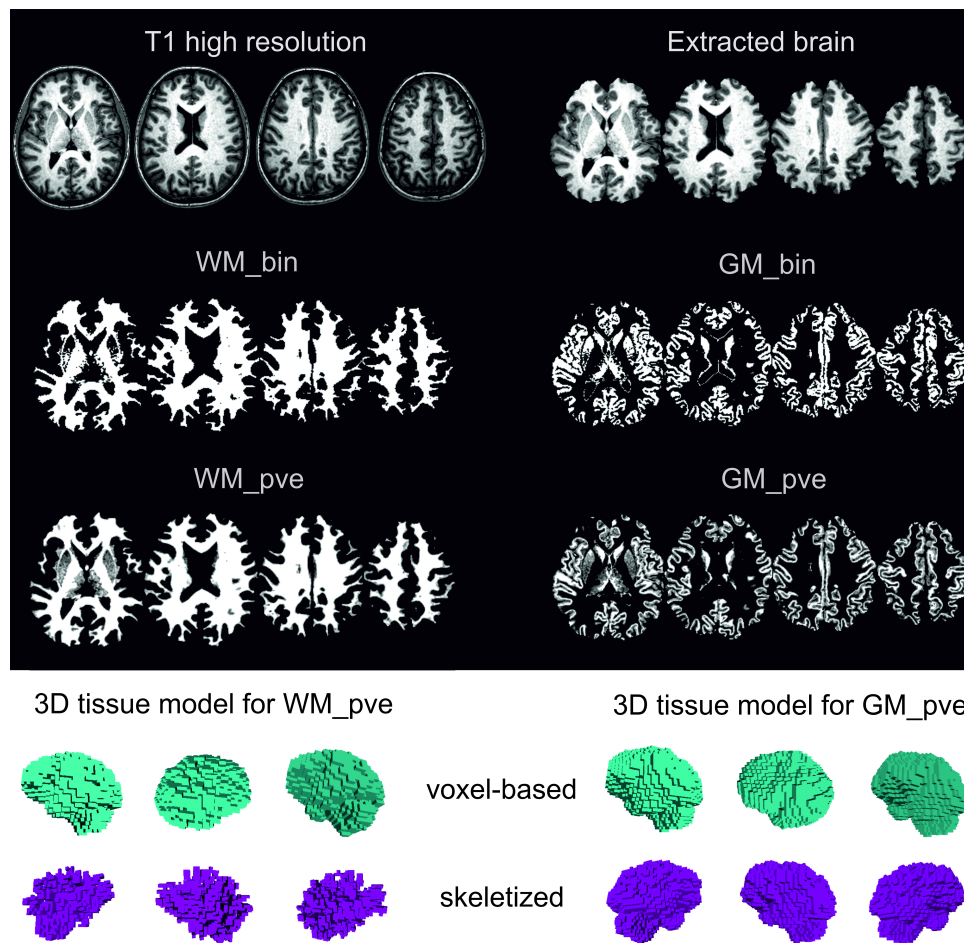
**Figure 2: Implementation of image processing.** Here the image processing steps in Fig. 1 are visualized for the first volume of the T1 high resolution images. Note the absence of gray voxels in the binary forced-decision segmentations (bin) as compared to the partial volume estimates (pve). Based on the segmented images, 3D tissue models for the respective volume are estimated. For each input volume, we computed a voxel-based model as well as a topological skeleton, corresponding to a minimum complexity representation of the tissue model. Some snapshots of the output corresponding to the pve-segmented gray and white matter volumes are displayed here to illustrate the 3-dimensional nature of the models. Here the upper row corresponds to the tissue models for WM_pve and GM_pve and the lower row to WM_Skel_pve and GM_Skel_pve according to our taxonomy. The 3D fractal dimension was then computed as the box-counting regression slope on these tissue models. bin: binary segmentations; GM: gray matter; pve: partial volume estimates; WM: white matter.

3D tissue models were computed as a function of the voxel intensities, and curve skeleton topological representations were estimated for each input volume [19]. Image skeletons are the result of an iterative reduction algorithm that computes a minimum complexity version of the 3D model by testing how global image properties change by removing a particular voxel [19]. The image skeletons thus aim at capturing the "essence" of an image and have been shown to be more sensitive to pathological changes in some cases [6–8, 19], which is why we included them in the present study. For every input volume, we thus obtained eight models as a combination of gray matter vs. white matter, standard partial volume vs.

6

binary segmentation, and skeletization vs. non-skeletization.

The resulting models then provided the input for the cloud-based calculation of the 3DFD, see [19] for details. In the empirical sciences, the FD is commonly estimated by the box-counting dimension $D_{bc}$ which is given by

$$D_{bc} = \lim_{x \to 0} \frac{logN(x)}{log(\frac{1}{x})} \tag{1}$$

where $x$ is the box edge length and $N(x)$ the minimum number of boxes needed to cover the object under scrutiny (cf. [1]). Since the zero limit does not apply to natural objects, the $D_{bc}$ is in practice calculated as the slope of the linear regression line over an interval of $x$. In terms of structural MR images, these intervals correspond to the range of the voxel edge sizes over which the box-counting dimension is computed. For the current study, we used box edge sizes of 4-16 voxel units for all GM images and 5-20 for all WM images, which was previously established to yield the best correlation results in terms of the box-counting regression, see e.g. [7, 19]. Fractal dimensions were calculated with an intensity threshold of 70 for all images (which most closely resembles typical clinical images [8]) and considering black and gray voxel types (which optimally considers the border between tissue types [7]). The resulting fractal dimension values were written into csv format and entered statistical analysis as detailed below based on custom-written Matlab code (The MathWorks, Inc., Natick, MA, United States). For the reader wishing to retrace our analysis, the FSL preprocessing scripts, the fractal analysis results files as well as the Matlab code for data analysis are available from the Open Science Framework (`http://osf.io/3mtqx`) and the corresponding authors. In summary, we obtain a total of 32 analysis groups as a combination of image processing parameters, on which we base the taxonomy applied throughout the manuscript: image contrast (T1 vs. T2) and resolution (low vs. high), segmentation procedure (common partial volume estimates (pve) vs. binary segmentation (bin)), tissue type (gray matter (GM) vs. white matter (WM)), and image complexity reduction (skeletized vs. non-skeletized images, where the former is abbreviated by "Skel").

## 2.3 Data analysis

With the procedure detailed above, we obtained 10 FD values within each of the 32 analysis groups. Without any a priori assumptions about the data, we first assessed the processing parameter-dependent characteristics of the FD across analysis groups. To this end, we compared the mean fractal dimensions between analysis groups by first computing an analysis of variance (ANOVA), which invariably yielded significant differences in the mean FDs across groups. Subsequently, we performed a conservative post-hoc Tukey-Kramer test [20] to investigate significant FD differences between analysis groups in pair-wise comparisons. In order to compare the mean FDs in the original and reregistered data (section 3.4), we

7

carried out a series of Welch's t-tests (i.e. without the assumption of equal variances), with Bonferroni-Holm correction for multiple comparisons. Similarly, the average image FD values were compared to the original and registered data, respectively, by means of a one-sample t-test. Effect sizes are given as $r_{equivalent}$, following Rosenthal & Rubin [21]. For all statistical tests employed herein, we defined a minimum significance level of 0.05. Where we report significance levels in intervals, the reader interested in the exact p-values is kindly referred to our Matlab code which generates all results including the figures in detail.

### 2.3.1 Deviation analysis

In order to qualitatively asses the data within each analysis group, we first applied a combination of random and systematic resampling procedures. Specifically, we performed a bootstrapping procedure in order to randomly sample the mean and the 99 % normal approximation confidence interval (CI) of the FD over 2000 resampling iterations. Bootstrapping provided an objective way of qualitative data assessment in terms of the tightness of the confidence interval, which served as an indicator for the deviations within the analysis group, and the presence or absence of a skew in the clusters of the resampled means, indicative of important singular deviations in the original FDs. Moreover, the bootstrapped CI was subsequently assessed as one of several criteria to identify meaningful deviations in the sampled FDs within each analysis group. We then applied a jackknife procedure, where we systematically resampled the means by iteratively omitting each of the 10 scans in order to see if the variance changed significantly as assessed by the Bartlett test. We then made the explicit assumption that the FDs obtained within each analysis group were sampled from a true but unknown normal distribution. To test this assumption, we fitted a Gaussian distribution to the sampled FDs and assessed the coherence to a corresponding theoretical distribution by means of a quantile-quantile plot. In order to obtain a formal criterion of whether the sampled data was reasonably assumed to follow a normal distribution, we furthermore computed the Shapiro-Wilk test [22], which is well suited to assess composite normality for smaller sample sizes.

As an example of the above, Figure 3 visualizes these analysis steps for the exemplary analysis group of binarized and skeletized WM images in the T2 low resolution category (T2 low WM_Skel_bin). The same analysis steps were applied to all 32 analysis groups. In doing so, we sought to define a sensible criterion of when to "flag" an FD value due to a meaningful deviation within an analysis group. To this end, we applied and compared various measures to find an optimal trade-off between detection ability and conservativeness. First, we assessed whether a single FD value was inside or outside the bootstrapped confidence interval. As a second method, we assessed whether a particular value was within one or respectively two standard deviations (SD) of the sample mean. Third, we assessed whether the variances
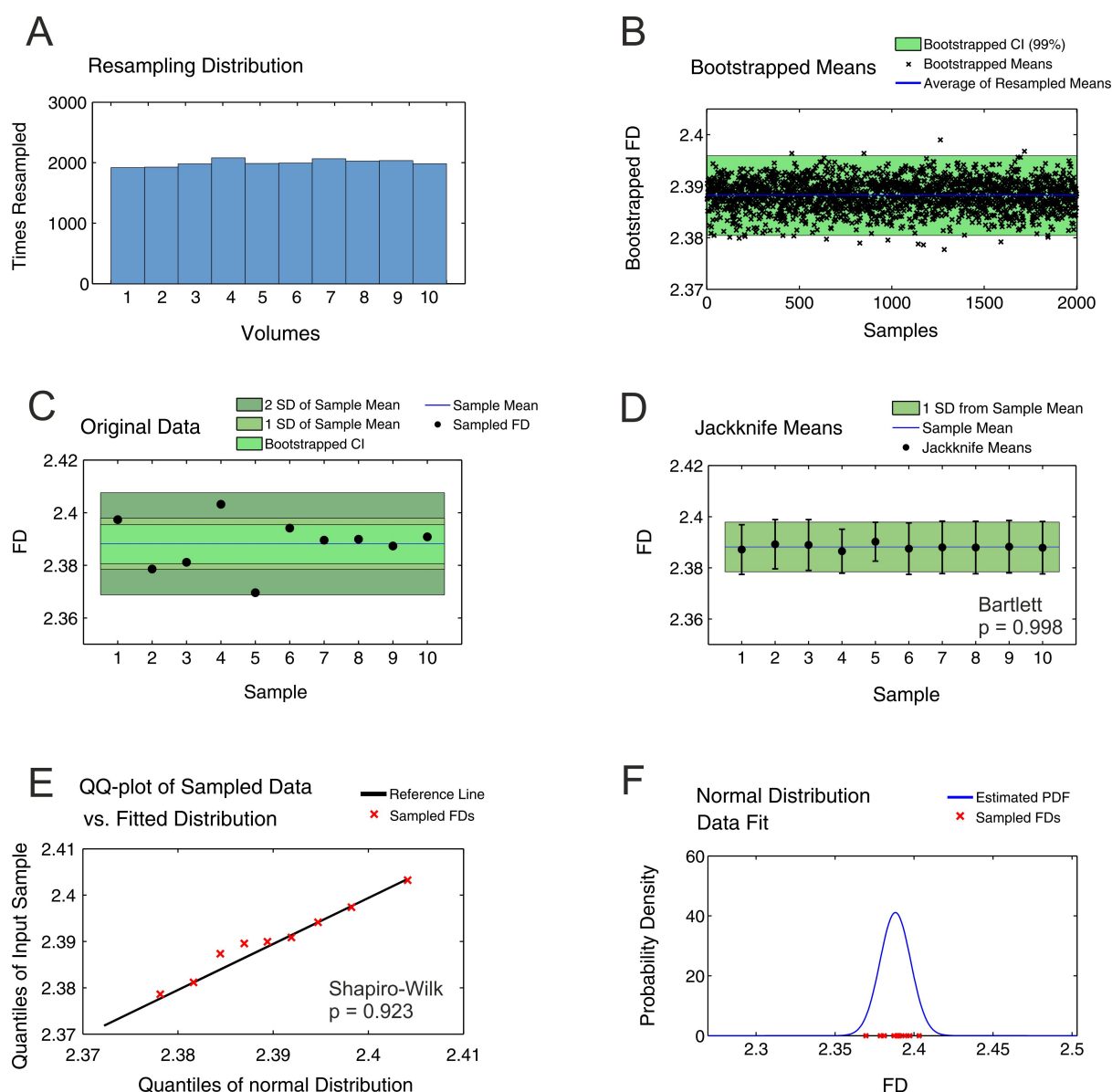
**Figure 3: Main steps of within-group deviation analysis.** The figure displays the deviation analysis for the exemplary analysis group of binarized and skeletized WM T2 images in the low resolution category. Panel A shows a near-uniform resampling distribution for bootstrapping, indicating that no *a priori* weights were used. Panel B displays the bootstrapped mean fractal dimensions as well as the resulting 99 % confidence interval and average over all bootstrapped means. Panel C plots the original estimated FDs for the 10 volumes within each analysis group and their mean, together with the bootstrapped confidence interval and the intervals spanning one and two standard deviations, respectively. Panel D represents the jackknife means, i.e. a systematic resampling, where each of the 10 original samples was iteratively omitted to compute the mean over the remaining nine samples. The Bartlett test to see if the variances of the thus obtained means significantly differed from one another was insignificant. Panel E shows a quantile-quantile plot for the original data vs. a fitted normal distribution, where a theoretical Gaussian would precisely follow the reference line. The values of the current analysis group reasonably adhere to this reference, and the Shapiro-Wilk test suggested that the data can be confidently assumed to follow an underlying normal distribution. Panel F shows the corresponding estimated normal distribution together with the cluster of the sampled FDs. We also estimated the kernel density of the sampled FDs, which invariably peaked together with the estimated normal distribution, even for analysis groups with deviations. The same analysis algorithm was applied to all 32 analysis groups. CI: confidence interval; FD: fractal dimension; PDF: probability density function; SD: standard deviation.

of the jackknife means significantly differed from one another by computing the Bartlett test. Finally, we computed the Grubbs test to detect outliers within a given analysis group [23]. The different methods were then assessed in terms of the original data and the effect that removing a flagged value had on the analysis in fig. 3. Specifically, we checked the flags against whether or not they occurred in those groups in which the assumption of composite normality was first violated when considering all 10 original FDs and whether the removal of the flagged volume changed this. Across the whole data set, 22 of all 32 analysis groups adhered to normality without any further assumptions. As a quality check of the different deviation criteria, we thus examined if a deviation criterion could identify those 10 analysis groups in which composite normality was first violated. Based on this approach, the first method was deemed too conservative because the CI was tighter than even the one standard deviation interval of the sample mean and because it is sensitive to arbitrary choices regarding the type of computation (normal approximation vs. percentile-based, studentized or not, etc.). Systematic resampling nicely showed the qualitative effect that a single volume had on the overall mean and its variance but only resulted in one flag over all 32 analysis groups (as given by a significant Bartlett test), which was considered too liberal for our purposes given that 10 out of 32 analysis groups did not adhere to the criterion of composite normality. When the 1 SD interval around the sample mean was considered, volumes were more selectively flagged. However, this criterion does not account for the range of the data scatter, which was generally very small in the data set. See for example fig. 3, where the data were sampled in the subdecimal scatter range of around 0.03. As a result, all ten scanning sessions were flagged at least once, indicating relatively low selectivity. Choosing a 2 SD interval, in contrast, increased selectivity and closely identified the 10 analysis groups that first violated composite normality. Even more selective, the Grubbs test procedure uniquely flagged volumes in those analysis groups, and all but one adhered to composite normality after removal of the flag (T1 high WM_Skel_pve being the exception, p = 0.043; see also table 1 in section 3.4). Therefore, we deemed this method as the most appropriate deviation criterion with the more conservative 2 SD method as a cross check. For an exemplary identification of a flag see section 3.3 where we also analyze the occurrence of flags by scanning session, image group, analysis parameters and examine the overall susceptibility to deviations in the data set.

## 3   Results

### 3.1   Fractal dimension in T1 and T2 contrast image groups

In the following, we first report the mean fractal dimensions by image groups to evaluate the effect of the various image processing parameters. Figure 4 displays the FD analysis results for the T1 sequences
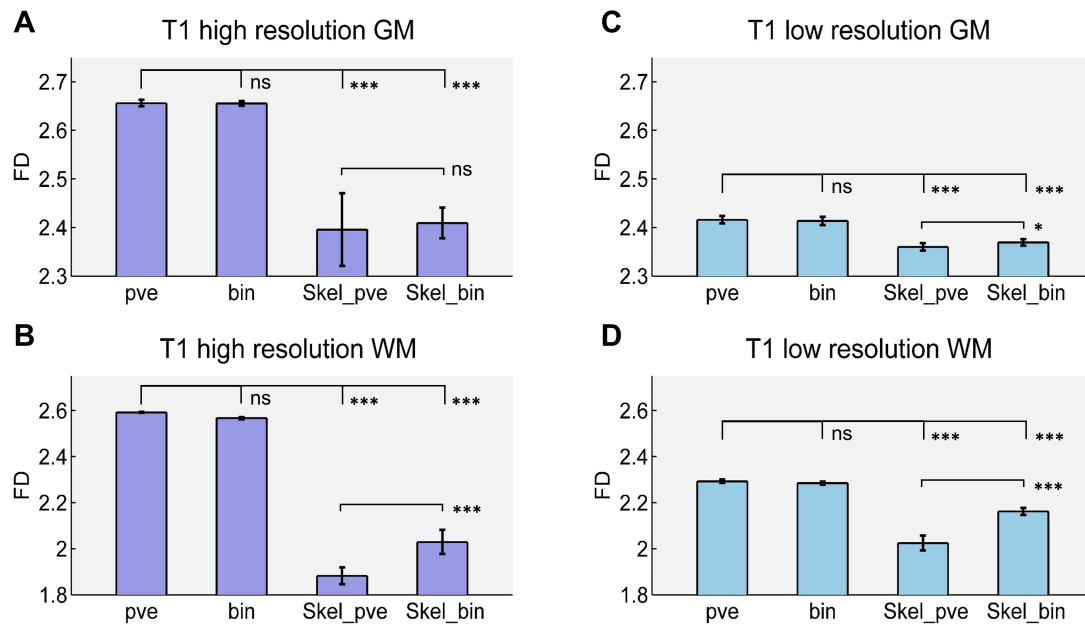
**Figure 4: Comparative analysis of the fractal dimension in T1 images.** Panels A and B display the mean fractal dimensions for the high resolution condition in GM and WM images, respectively. Similarly, panels C and D represent the fractal dimensions for the low resolution across analysis parameters. The horizontal bars reflect pair-wise significance levels in the Tukey-Kramer test. The upper bar refers to the MR images. The pair-wise comparisons for MR_bin images invariably yielded the same significance levels so they were omitted here for visual coherence. Note that for visual comparison, the FD intervals of the plots are the same for GM (panels A and C) and WM (panels B and D), respectively. ns: not significant; *: $p < 0.05$; ***: $p < 0.001$; bin: binary tissue segmentation; pve: partial volume estimates; Skel: skeletized tissue model; Skel_bin: skeleton model of binarized image.

and figure 5 for the T2 contrast images, considering both WM and GM images in the high as well as the low resolution category. The T1 FDs were in the expected range, compatible with previous reports. Of note, the FD values for the T2 contrast images were sampled in a comparable range and indeed showed a qualitative behavior very similar to the T1 images. Note, for instance, that in both T1 and T2 contrast, binary segmentation of GM as well as WM images did not affect the mean FD values while skeletization resulted in a significant decrease of the FD for both high and low resolution images. Interestingly, the impact of skeletization on the FD varied with the segmentation procedure. Specifically, the reduction of the FD by skeletization was less pronounced in binarily-segmented images (Skel_bin) than in partial volume estimate segmentation (Skel_pve), and especially so in white matter images in both T1 and T2 contrast and high as well as low spatial resolution. Furthermore, considering that the input sample for each image group consisted of only 10 volumes, note the relatively narrow standard deviations throughout the low resolution category and the unskeletized images of the high resolution condition for T1 images (although no prior assumptions about the data were made at this stage); notably, standard deviations were even more constrained across the whole T2 contrast group (cf. also section 3.3). Another interesting
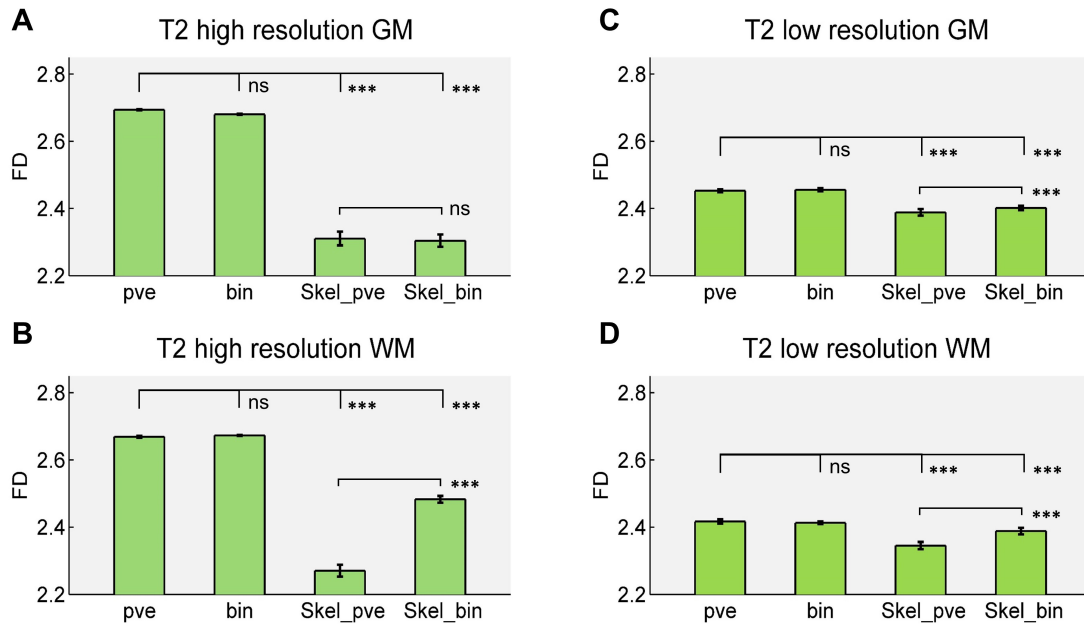
11

**Figure 5: Comparative analysis of the fractal dimension in T2 images.** Similar to fig. 4, panels A (GM) and B (WM) display the mean fractal dimensions for the high resolution images and panels C (GM) and D (WM) for the low resolution images. The horizontal bars reflect pair-wise significance levels. Note the plausible range of the fractal dimensions across the T2 images and the similar qualitative behavior compared to T1, while showing even more limited deviations from the mean. ns: not significant; *: $p < 0.05$; ***: $p < 0.001$; bin: binary tissue segmentation; pve: partial volume estimates; Skel: skeletized tissue model; Skel_bin: skeleton model of binarized image.

general finding was that all fractal dimension values (except for high-resolution T2_MR_bin images) in GM were significantly larger than their corresponding WM values (e.g., GM_MR vs. WM_MR), both for T1 and for T2 images and in both high and low resolution categories.

## 3.2 High vs. low spatial resolution

Figure 6 represents the differences in the mean FDs between high ($1\ mm^3$) and low ($2.5\ mm^3$) voxel resolution, analyzed by image group and analysis parameters.

As a general finding, the spatial resolution had a highly significant impact on the absolute mean FD across the data set. Specifically, in both pve-segmented and binary-segmentation unskeletized images, the lower resolution invariably resulted in decreased fractal dimensions, regardless of contrast and tissue type. For skeletized images, the effect of lower resolution was more complex: In skeletized T1 GM images, the resolution had no significant effect on the FD. Similar to the results for unskeletized images, the mean FD was decreased in low resolution solely for binary-segmentation skeletized T2 images (T2 WM Skel_bin, cf. fig. 5, panels B and D). Strikingly, however, for all other comparisons, the mean FD was in fact *increased* in the lower resolution images. Also note that while the absolute FD values of T1
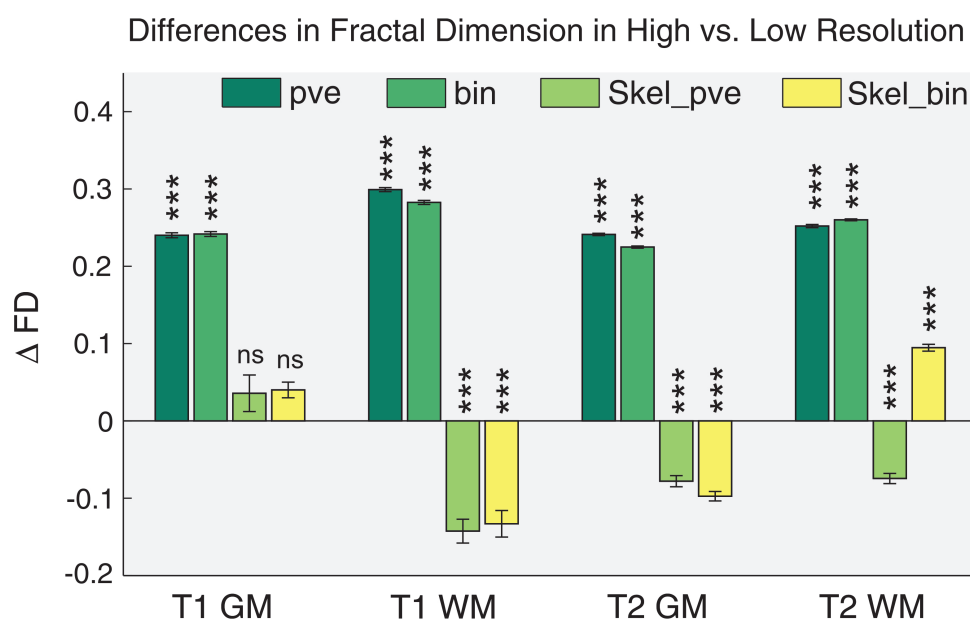
12

**Figure 6: Fractal dimension difference in high vs. low spatial resolution.** Bars represent the respective difference in the mean fractal dimensions of high and low resolution images. Deviations correspond to the difference's sampling distribution error. Changing the spatial voxel resolution from 1 mm$^3$ to 2.5 mm$^3$ had a highly significant impact on the value of the respective fractal dimension. This effect was most pronounced in unskeletized partial volume estimates and binary segmentations, where the lower spatial resolution invariably resulted in decreased FDs. Note that the FDs in T1 and T2 images are affected by spatial resolution in a similar way. $\Delta$ FD : difference in mean fractal dimension in high vs. low contrast condition. ns: not significant; ***: $p < 0.001$; bin: binary tissue segmentation; pve: partial volume estimates; Skel: skeletized tissue model; Skel_bin: skeleton model of binarized image.

and T2 contrast images were different (cf. figs. 4 and 5), the impact of the different voxel resolutions on the mean FDs were surprisingly similar in unskeletized images (FD reduction of around 0.25-0.3) regardless of tissue type.

## 3.3 Susceptibility to measurement noise

We now turn to the within-group analyses and first illustrate the application of our deviation analysis presented above. To this end, consider fig. 7 which exemplifies the identification of a flag in the high-resolution T1 GM_pve images. Here, the Grubbs test flags the FD that corresponds to the first scanning session (note that the more conservative 2 SD criterion equivalently identifies this flag). Systematic resampling shows that omitting the flagged value causes an upward shift of the mean and reduces its variance but this does not reach significance level in multivariate testing. Moreover, the flagged FD causes the assumption of composite normality to be invalid although the remaining samples tightly follow the reference for normality. Omitting the flag, in turn, restores normality and clearly "tightens" the Gaussian (cf. panel D). However, the comparison between these two distributions was invariably insignificant, here
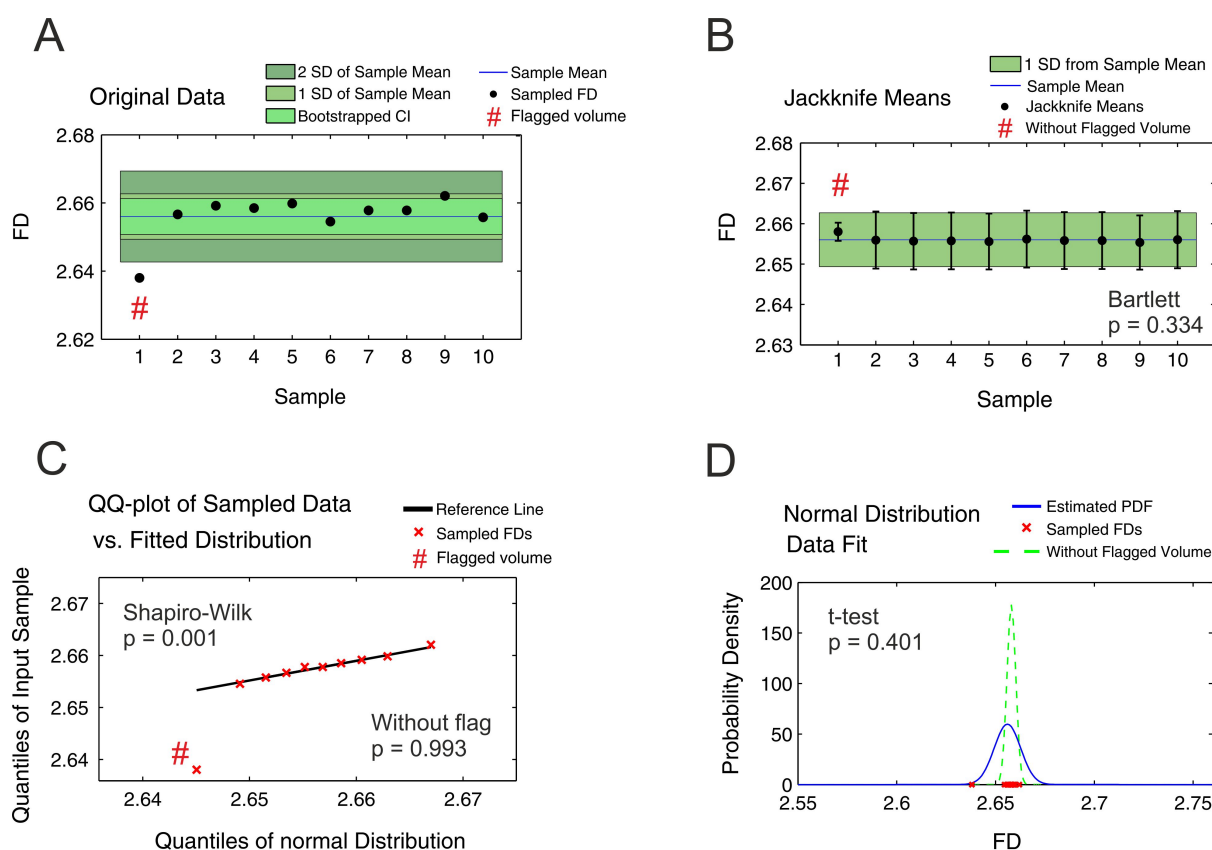
**Figure 7: Exemplary identification of a flag.** The data presented here belongs to the T1 GM_pve images in the high resolution category. If the fractal dimension of an image was identified to deviate from the remaining analysis group according to the chosen deviation criterion, the corresponding volume was flagged (indicated here by #). In this case, the FD value belonging to the first scan was flagged, and its deviation from the remaining sample population is visible from panel A. Note that in panel B, the variance of the jackknife mean without this flagged volume is notably smaller, although this did not reach significance level by multivariate variance comparison. Panel C shows the corresponding quantile-quantile plot. Although the flagged FD only deviates by about 0.02 from the other sampled FDs, the estimate of normality suggests that assuming an underlying Gaussian distribution is not recommendable. Clearly, however, the remaining samples tightly follow the normality reference and discarding the flagged FD indeed restores the assumption of normality. Furthermore, comparing the fitted normal distributions with and without the flagged volume invariably yielded insignificant results in the data set, exemplified here in panel D. CI: confidence interval; PDF: probability density function; SD: standard deviation.

and across all 10 analysis groups containing flagged volumes.

This procedure was applied to all 32 analysis groups, the result of which is shown in Figure 8 (by sample and image group in panel A and by processing parameters in panel B). Here, it is first crucial to note that the overall robustness of the FD against deviations was very high across the entire data set. Of the 320 FDs estimated in the present study, only 11 were flagged based on the chosen deviation criterion. Accordingly, 96.6 % of all FD values were not flagged, which remained well above 90 % even when we applied the more conservative standard-deviation-based criterion. Importantly, a sample-based analysis uniquely identified a single scanning session that was responsible for the majority of the deviations in the

14

data set, in this case sample number 1 (figure 8, panel A). Notably, this remained true when compared with the more conservative deviation criterion (which also identified a deviation in session 9 and in session 3), and it was those flagged volumes that caused the a priori violation of composite normality in the corresponding analysis groups. In terms of image condition, all but two flags were found in high resolution images. Interestingly, flagged volumes were almost exclusively limited to T1 images. In terms of analysis parameters, most flags were found in binary segmentation skeletons, and white matter images were more susceptible to deviations than gray matter images (figure 8, panel B).
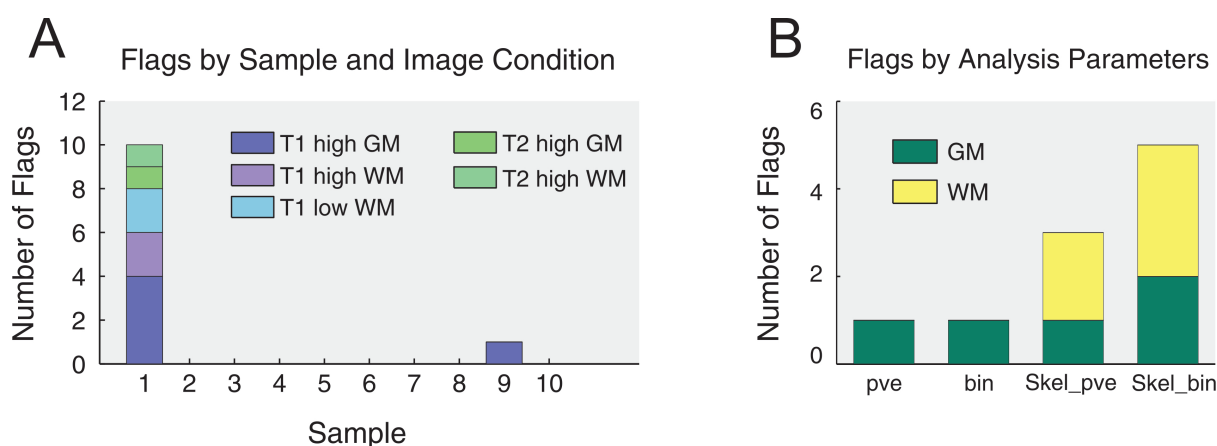


**Figure 8: Susceptibility of fractal dimension to deviations.** Panel A visualizes the number of flagged measurements by sample (i.e. scanning session) and analysis group, which uniquely identifies the first scan as the main source of noise in the data set. Note furthermore that flags were mostly found in T1 contrast and in high resolution images. Panel B displays the number of flagged measurements by analysis parameters. Most flags were found in the binary-segmentation image skeletons, and white matter images were more affected than gray matter images. The number of total flags was very low, however, with well over 90 % unflagged volumes across the whole data set. GM: gray matter; bin: binary tissue segmentation; pve: partial volume estimates; Skel: skeletized tissue model; Skel_bin: skeleton model of binarized image; WM: white matter.

## 3.4 Impact of image registration on fractal dimension profile

Finally, we investigated the effects of image registration and the ensuing interpolation on the fractal dimension profile across the 32 analysis groups. The motivation for this was that our deviation analysis identified a cluster of deviations in the first scanning session, which mostly affected T1-weighted images. Since the first volume in the T1 high-resolution category was the target for image registration, we explored if and how a different registration target would alter our results. To this end, we modified the data set in two distinct ways: First, we reregistered the original images to the mean of the FLAIR sequences in the MASSIVE data base and repeated the above analysis in the reregistered data set. Furthermore, for each of the four image categories (T1 high resolution, T1 low resolution, T2 high resolution, and T2 low resolution), we computed an average image over the 10 scanning sessions, resulting in a single FD

value for every analysis group. Table 1 summarizes the corresponding fractal analysis results. The 32 analysis groups showed a differential response to image reregistration. In terms of the deviation analysis, reregistration decreased the number of total flags even further from 11 (3.4 %) to 4 (1.3 %) flagged volumes. Interestingly, this effect was exclusively seen in T1 images. Here, reregistration greatly reduced the number of flags by abolishing the first scanning session as a source of noise while it also induced two distinct singular deviations in previously unflagged volumes. In T2 images, the deviation analysis identified exactly the same flags in the original and reregistered data sets. Furthermore, reregistration also entailed a differential change in the absolute FD values. The most pronounced effect of reregistration was seen in WM images in the low-resolution T2 contrast group. The impact on skeletized images was less often significant. In this regard, it is noteworthy that the standard deviations in high-resolution skeletized images are about one order of magnitude higher (in the second decimal place) than in unskeletized images, both in the original and the reregistered data set. The data scatter in unskeletized images was generally very limited so that even relatively subtle FD differences resulted in a significant test outcome (e.g. see T2 high-resolution WM_pve). Importantly, the effect of processing parameters within the image groups was virtually unaltered by reregistration (e.g. little effect of binary segmentation on unskeletized images, significant FD reduction by skeletization and in lower resolution, etc., cf. figs. 4 and 5). Finally, we compared the original and reregistered data set to the average images. Across all analysis groups, the average image FDs were more often similar to the corresponding mean FD in the reregistered data set (indicated by 23 out of 32 confirmed null hypotheses), especially in T2 images. The least difference to the original data was observed in low-resolution T1 WM and T2 GM, while the most pronounced difference was found in low-resolution T2 WM (as judged by magnitude of effect sizes). Notably, in high-resolution T1 images, image averaging caused a marked drop in the FDs over image skeletons with respect to both the original and the reregistered data and especially in WM segmentations.

| Analysis group | Original Images | | Reregistered Images | | | | Avg. Image | vs. original | | vs. reregistered | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean FD ± SD | Flag | Mean FD ± SD | Flag | $p$ | $r_{eq}$ | FD | $p$ | $r_{eq}$ | $p$ | $r_{eq}$ |
| **T1 high** | | | | | | | | | | | |
| GM_pve | 2.6560 ± 0.0067 | 1 | 2.6732 ± 0.0024 | - | 1.4e-4 | 0.92 | 2.6744 | 1.9e-4 | 0.95 | $ns$ | 0.45 |
| GM_bin | 2.6554 ± 0.0047 | 1 | 2.6703 ± 0.0024 | - | 9.3e-6 | 0.93 | 2.6713 | 3.9e-5 | 0.96 | $ns$ | 0.41 |
| GM_Skel_pve | 2.3959 ± 0.0745 | 1,9 | 2.4315 ± 0.0172 | 2 | $ns$ | 0.42 | 2.3886 | $ns$ | 0.10 | 7.1e-4 | 0.93 |
| GM_Skel_bin | 2.4094 ± 0.0313 | 1 | 2.4143 ± 0.0098 | - | $ns$ | 0.14 | 2.3530 | 0.004 | 0.88 | 3.1e-7 | 0.99 |
| WM_pve | 2.5921 ± 0.0026 | - | 2.6053 ± 0.0020 | - | 1.2e-8 | 0.95 | 2.6035 | 5.4e-6 | 0.98 | $ns$ | 0.68 |
| WM_bin | 2.5672 ± 0.0048 | - | 2.5808 ± 0.0033 | - | 2.7e-5 | 0.88 | 2.5760 | 0.004 | 0.89 | 0.032 | 0.84 |
| WM_Skel_pve | 1.8821 ± 0.0371 | 1 | 1.8776 ± 0.0137 | - | $ns$ | 0.10 | 1.8325 | 0.027 | 0.82 | 7.6e-5 | 0.96 |
| WM_Skel_bin | 2.0293 ± 0.0522 | 1 | 2.0112 ± 0.0141 | - | $ns$ | 0.31 | 1.9417 | 0.006 | 0.87 | 2.4e-6 | 0.98 |
| **T1 low** | | | | | | | | | | | |
| GM_pve | 2.4159 ± 0.0077 | - | 2.4470 ± 0.0064 | - | 3.5e-7 | 0.92 | 2.4497 | 5.2e-6 | 0.98 | $ns$ | 0.42 |
| GM_bin | 2.4137 ± 0.0087 | - | 2.4477 ± 0.0097 | 8 | 3.5e-6 | 0.89 | 2.4484 | 1.1e-5 | 0.97 | $ns$ | 0.07 |
| GM_Skel_pve | 2.3601 ± 0.0077 | - | 2.4075 ± 0.0073 | - | 1.0e-9 | 0.96 | 2.3970 | 2.6e-6 | 0.98 | 0.034 | 0.83 |
| GM_Skel_bin | 2.3694 ± 0.0065 | - | 2.4298 ± 0.0055 | - | 6.3e-13 | 0.98 | 2.4311 | 6.9e-9 | 0.99 | $ns$ | 0.25 |
| WM_pve | 2.2930 ± 0.0076 | - | 2.3061 ± 0.0082 | - | 0.025 | 0.66 | 2.2982 | $ns$ | 0.58 | $ns$ | 0.71 |
| WM_bin | 2.2847 ± 0.0068 | - | 2.3057 ± 0.0063 | - | 2.2e-5 | 0.86 | 2.2911 | $ns$ | 0.71 | 0.001 | 0.92 |
| WM_Skel_pve | 2.0246 ± 0.0317 | 1 | 2.0481 ± 0.0258 | - | $ns$ | 0.40 | 2.0491 | $ns$ | 0.63 | $ns$ | 0.04 |
| WM_Skel_bin | 2.1623 ± 0.0150 | 1 | 2.1684 ± 0.0143 | - | $ns$ | 0.21 | 2.1589 | $ns$ | 0.24 | $ns$ | 0.57 |
| **T2 high** | | | | | | | | | | | |
| GM_pve | 2.6939 ± 0.0014 | - | 2.7012 ± 0.0017 | - | 1.8e-7 | 0.93 | 2.7007 | 2.5e-6 | 0.98 | $ns$ | 0.26 |
| GM_bin | 2.6804 ± 0.0017 | - | 2.6881 ± 0.0016 | - | 1.7e-7 | 0.92 | 2.6852 | 1.7e-4 | 0.95 | 0.011 | 0.88 |
| GM_Skel_pve | 2.3105 ± 0.0204 | - | 2.3115 ± 0.0197 | - | $ns$ | 0.03 | 2.2984 | $ns$ | 0.53 | $ns$ | 0.57 |
| GM_Skel_bin | 2.3041 ± 0.0183 | 1 | 2.2979 ± 0.0142 | 1 | $ns$ | 0.20 | 2.2810 | 0.034 | 0.80 | $ns$ | 0.78 |
| WM_pve | 2.6688 ± 0.0029 | - | 2.6638 ± 0.0026 | - | 0.007 | 0.71 | 2.6611 | 1.1e-4 | 0.95 | $ns$ | 0.74 |
| WM_bin | 2.6731 ± 0.0015 | - | 2.6647 ± 0.0009 | - | 6.8e-9 | 0.97 | 2.6648 | 9.3e-7 | 0.98 | $ns$ | 0.16 |
| WM_Skel_pve | 2.2706 ± 0.0177 | - | 2.2688 ± 0.0172 | - | $ns$ | 0.05 | 2.2498 | 0.048 | 0.78 | $ns$ | 0.76 |
| WM_Skel_bin | 2.4829 ± 0.0100 | 1 | 2.4714 ± 0.0117 | 1 | $ns$ | 0.49 | 2.4605 | 9.3e-4 | 0.92 | $ns$ | 0.70 |
| **T2 low** | | | | | | | | | | | |
| GM_pve | 2.4528 ± 0.0046 | - | 2.4501 ± 0.0035 | - | $ns$ | 0.33 | 2.4487 | $ns$ | 0.68 | $ns$ | 0.38 |
| GM_bin | 2.4557 ± 0.0043 | - | 2.4504 ± 0.0035 | - | $ns$ | 0.58 | 2.4537 | $ns$ | 0.44 | $ns$ | 0.70 |
| GM_Skel_pve | 2.3884 ± 0.0100 | - | 2.3931 ± 0.0081 | - | $ns$ | 0.27 | 2.3958 | $ns$ | 0.61 | $ns$ | 0.33 |
| GM_Skel_bin | 2.4015 ± 0.0062 | - | 2.4075 ± 0.0091 | - | $ns$ | 0.40 | 2.4219 | 4.9e-5 | 0.96 | 0.019 | 0.86 |
| WM_pve | 2.4169 ± 0.0062 | - | 2.3103 ± 0.0056 | - | 2.0e-17 | 0.99 | 2.3095 | 3.7e-11 | 0.99 | $ns$ | 0.15 |
| WM_bin | 2.4131 ± 0.0037 | - | 2.3140 ± 0.0051 | - | 5.6e-18 | 0.99 | 2.3206 | 1.5e-12 | 0.99 | $ns$ | 0.81 |
| WM_Skel_pve | 2.3451 ± 0.0110 | - | 2.2635 ± 0.0112 | - | 7.4e-11 | 0.97 | 2.2603 | 4.4e-8 | 0.99 | $ns$ | 0.29 |
| WM_Skel_bin | 2.3882 ± 0.0097 | - | 2.3083 ± 0.0090 | - | 7.1e-12 | 0.98 | 2.3122 | 4.0e-8 | 0.99 | $ns$ | 0.42 |

**Table 1: Impact of image registration and interpolation on fractal dimension profile**. The table presents the fractal dimension values by image group (T1 and T2 contrast in high and low spatial resolution, respectively) for the original data set (corresponding to the values presented in detail above) and the images when registered to the mean of the FLAIR images in the MASSIVE data set. Entries in the flag columns correspond to within-group deviations as determined by the deviation analysis presented above (numbers indicate the corresponding scanning session). The respective group means were compared by Welch's t-test (i.e. without the assumption of equal variances). Furthermore, we computed the average images over all 10 respective scanning sessions for each image category to compare the mean FDs with the corresponding FD of the averaged image. This comparison was carried out by a one-sample t-test with respect to the original images and the reregistered images. All p-values are Bonferroni-Holm-corrected for multiple comparisons. Effect sizes are indicated by $r_{equivalent}$ after Rosenthal & Rubin [21]. Abbreviations: avg: average; bin: binary segmentation; FD: fractal dimension; GM: gray matter; ns: not significant; pve: partial volume estimates; $r_{eq}$: effect size indicator $r_{equivalent}$; SD: standard deviation; Skel: image skeleton; Skel_bin: skeleton model of binarized image; WM: white matter.

# 4    Discussion

In the present study, we investigate the 3D fractal dimension as a marker of structural brain complexity in magnetic resonance images from the highly standardized MASSIVE data set (Multiple Acquisitions for Standardization of Structural Imaging Validation and Evaluation) [12], featuring high-quality images obtained from a single healthy subject scanned 10 times over an interval of 2 weeks. We hypothesized that the structural brain complexity in such a subject should not change at all and that the FD should therefore show high stability across the dataset. To assess this hypothesis, we provide a detailed analysis procedure, featuring a stratification of various image characteristics, resulting in 32 analysis groups as a combination of image contrast, resolution, segmentation procedure, tissue type, and image complexity. We first assessed the processing parameter-dependent characteristics of the FD and then examined the within-group FD values with a combination of random and systematic resampling methods, which proved useful in the qualitative assessment of the data (confidence interval bounds, skewness of bootstrapped means) and for hypothesis generation without prior assumptions about the data. We investigated various methods of identifying meaningful deviations of FD values within a sample group, and found that a 2 standard deviation interval around the sample means and outlier detection by Grubbs testing performed very well in this regard. The overall robustness of the 3DFD to measurement noise was very good ($>90$ %). Moreover, based on the results obtained herein, it is reasonable to assume that the fractal dimension values sampled repeatedly from the same subject (without change in structural brain complexity) adhere to a normal distribution.

Furthermore, to our knowledge, this is the first study to investigate fractal analysis in detail for T2 sequences. The T2 contrast images yielded surprisingly robust results, both in comparison to T1 images and in terms of stability over repeated measurements. T2 images showed a sample range comparable to T1 images and were equally affected by image resolution and analysis parameters but in fact showed less susceptibility to deviations and image registration. In terms of resolution, it is interesting to note that the lower voxel resolution invariably resulted in lower FD values in the unskeletized images. Intuitively, a measure of structural brain complexity could be lower in coarser resolution because structural complexity is blunted by partial volume effects. Low resolution images showed a comparable qualitative behavior and were furthermore less susceptible to measurement noise, which is a promising finding for further research. For skeletized images, the resolution difference was more complex. Here, lower spatial resolution even yielded significantly higher FD values in some cases. One possible explanation of this finding is that lower-resolution images convey less structural information to begin with and are therefore less affected by skeletization while high-resolution images are severely impacted by complexity reduction. In this context, it is also noteworthy that skeletized images invariably showed significantly lower FDs than their

unskeletized counterparts. Since the skeletization procedure estimates a minimum complexity model of the input image, the FD as a marker of tissue complexity should indeed be reduced by skeletization. In this context, we found that binary segmentation did not affect the unskeletized images, while it did have a significant impact on the skeletized images. Furthermore, binary segmentation skeletized images were most prone to measurement deviations. In this context, a sample-based analysis approach proved useful in identifying deviation clusters over scanning sessions responsible for most of the measurement noise in the FD values across the whole dataset. Related to this, we show that image registration and the ensuing interpolation constitute a source of noise. Reregistration decreased deviations even further to under 2 % across the entire data set. Intriguingly, registration-induced changes in within-group deviations were limited to T1-weighted images, lending even further support to the usefulness of T2 contrast images in fractal analysis. Moreover, image registration had a differential effect on the absolute FD values across all analysis groups, inducing significant changes predominantly in unskeletized high-resolution images and low-resolution T1 gray matter and T2 white matter volumes. We therefore suggest that a particular FD value must be interpreted in light of the applied image processing parameters in general and registration in particular as well as in the context of associated FD values derived from the same image (its "fractal profile"), which we elaborate on below.

## 4.1 Future directions

While our results confirm the stability of the 3D fractal dimension as a useful marker of structural brain complexity, there are some questions that remain unaddressed but deserve mentioning. First, the data evaluated here were limited to a single subject. While within-subject multiple sampling stability is precisely what we wanted to address in the current work, it of course raises the question of second-level analysis, i.e. if potential reference values for a single subject are specific to that subject or if they are at least in principle generalizable to a population level. Tightly connected to this issue is the question of technical variance. The MASSIVE data set provides highly standardized images, while this may not always be the case in empirical reality. Motion artefacts, for instance, can be expected to obscure the utility of fractal analysis. Similarly, just as the reference values for blood analysis differ depending on the laboratory where they are measured, the fractal dimension may be influenced by the type of scanning equipment or the software used to preprocess the images. While we used a well-established preprocessing pipeline in the current study, clearly more work needs to be done in this regard to validate the FD on a population level. In a clinical context, it will be interesting to see whether we can eventually define diagnostic cut-offs, i.e. FD values that suggest a pathology rather than a measurement deviation of a single scan with a reasonable amount of confidence. In this regard, our work shows that some

procedural standardization may eventually be necessary to facilitate comparability across studies and centers. For instance, image registration both acts as a source of noise in repeated sampling and also exerts an influence on the absolute FD values. Therefore, further work is warranted to examine the impact of registration on between-subject variability. We here explore one possible way to establish reference values, namely by fractal analysis on average images. This yielded comparable results with respect to the FLAIR-reregistered FDs, especially in T2-weighted images and low image resolution. As we detail above, however, this may be cumbersome in skeletized images as it seems that in some cases (here, predominantly in high-resolution T1 contrast) the minimum complexity of the average image may not be the same as the average minimum complexity of the individual images, i.e. that image complexity is altered by averaging.

Finally, we briefly elaborate on the notion of a fractal profile. Naturally, we want an empirically useful measure of structural brain complexity to remain stable throughout a healthy subject (at least within a short amount of time), while it should be sensitive to deviations in order to rapidly detect any structural changes (as has indeed been shown by the early alterations of the FD in multiple sclerosis [6,7]). Most studies to date have focused on a single (mean) FD value for a specific tissue type to carry out group-wise comparisons between study populations. In the current study, we obtain 8 FD values for every input image due to the stratification of image processing parameters (tissue type, segmentation procedure, image complexity). We thus compute a profile of 32 FD values for the 4 image inputs (T1/T2 contrast in low/high resolution, cf. table 1). Since the different analysis groups show differential responses to deviations, considering the fractal profile of the same input image could perhaps ameliorate the sensitivity-specificity trade-off. For instance, image skeletons have been shown to be more sensitive to pathological changes in some cases [6,7], and it will be interesting to see if the binarized image skeletons that were even more susceptible to deviations in the present work could in fact be useful to detect early pathological changes. Unskeletized images, in contrast, generally showed less susceptibility to measurement deviations. Accordingly, it may be worthwhile to examine if a simultaneous change of both deviation-susceptible and relatively deviation-robust analysis groups can perhaps help distinguish noise from real tissue alterations. On a similar note, it will be interesting to see if additional information can be gained by computing the fractal dimension profiles not on global tissue segmentations but on brain parcellations, especially given the increasingly sophisticated estimation methods of the latter [24].

## 5  Conclusion

In the current study, we provide evidence for the stability of the FD as a marker of structural brain complexity and evaluate its processing parameter-dependent characteristics in detail. Essentially, fractal

analysis in the brain amounts to the attempt of a maximum dimensionality reduction, where we try to map the complexity of brain tissue onto a single scalar number. In order to maximize the potential of this promising new field of study, we have to apply both high-quality image acquisition protocols and rigorous analytical methods. With the present work, we aim to combine both and make progress towards the development of fractal analysis as a scientifically and clinically useful neuroimaging tool.

# 6    Acknowledgments

# References

[1] A. Di Ieva, F. Grizzi, H. Jelinek, A. J. Pellionisz, and G. A. Losa, "Fractals in the neurosciences, Part I: general principles and basic neurosciences," *The Neuroscientist*, vol. 20, no. 4, pp. 403–417, 2014.

[2] A. Di Ieva, F. J. Esteban, F. Grizzi, W. Klonowski, and M. Martín-Landrove, "Fractals in the neurosciences, Part II: clinical applications and future perspectives," *The Neuroscientist*, vol. 21, no. 1, pp. 30–43, 2015.

[3] B. B. Mandelbrot, *The fractal geometry of nature*, vol. 173. Macmillan, 1983.

[4] B. B. Mandelbrot, "How long is the coast of Britain? Statistical self-similarity and fractional dimension.," *Science*, vol. 156, no. 3775, pp. 636–638, 1967.

[5] V. G. Kiselev, K. R. Hahn, and D. P. Auer, "Is the brain cortex a fractal?," *Neuroimage*, vol. 20, no. 3, pp. 1765–1774, 2003.

[6] F. J. Esteban, J. Sepulcre, N. V. de Mendizábal, J. Goñi, J. Navas, J. R. de Miras, B. Bejarano, J. C. Masdeu, and P. Villoslada, "Fractal dimension and white matter changes in multiple sclerosis," *Neuroimage*, vol. 36, no. 3, pp. 543–549, 2007.

[7] F. J. Esteban, J. Sepulcre, J. R. de Miras, J. Navas, N. V. de Mendizábal, J. Goñi, J. M. Quesada, B. Bejarano, and P. Villoslada, "Fractal dimension analysis of grey matter in multiple sclerosis," *Journal of the neurological sciences*, vol. 282, no. 1, pp. 67–71, 2009.

[8] F. J. Esteban, N. Padilla, M. Sanz-Cortés, J. R. de Miras, N. Bargalló, P. Villoslada, and E. Gratacós, "Fractal-dimension analysis detects cerebral changes in preterm infants with and without intrauterine growth restriction," *Neuroimage*, vol. 53, no. 4, pp. 1225–1232, 2010.

[9] R. D. King, B. Brown, M. Hwang, T. Jeon, A. T. George, A. D. N. Initiative, *et al.*, "Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease," *Neuroimage*, vol. 53, no. 2, pp. 471–479, 2010.

[10] K. M. Iftekharuddin, J. Zheng, M. A. Islam, and R. J. Ogg, "Fractal-based brain tumor detection in multimodal MRI," *Applied Mathematics and Computation*, vol. 207, no. 1, pp. 23–41, 2009.

[11] C. R. Madan and E. A. Kensinger, "Cortical complexity as a measure of age-related brain atrophy," *NeuroImage*, vol. 134, pp. 617–629, 2016.

[12] M. Froeling, C. M. Tax, S. B. Vos, P. R. Luijten, and A. Leemans, "MASSIVE brain dataset: Multiple acquisitions for standardization of structural imaging validation and evaluation," *Magnetic resonance in medicine*, 2016.

[13] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[14] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, pp. S208–S219, 2004.

[15] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in FSL," *Neuroimage*, vol. 45, no. 1, pp. S173–S186, 2009.

[16] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[17] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[18] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[19] J. Jiménez, A. López, J. Cruz, F. J. Esteban, J. Navas, P. Villoslada, and J. R. de Miras, "A web platform for the interactive visualization and analysis of the 3D fractal dimension of MRI data," *Journal of biomedical informatics*, vol. 51, pp. 176–190, 2014.

[20] A. J. Hayter, "A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative," *The Annals of Statistics*, pp. 61–75, 1984.

[21] D. B. Rubin and R. Rosenthal, "r (equivalent): A simple effect size indicator," 2003.

[22] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[23] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[24] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, 2016.