Robust estimation of fractal measures for characterizing the structural complexity of the human brain: optimization and reproducibility

Joaquín Goñi1,*, Olaf Sporns1, Hu Cheng1, Maite Aznárez-Sanado2, Yang Wang3, Santiago Josa2, Gonzalo Arrondo2, Vincent P Mathews3, Tom A Hummer4, William G Kronenberger4, Andrea Avena-Koenigsberger1, Andrew J. Saykin3, and María A. Pastor2

1Department of Psychological and Brain Sciences. Indiana University, Bloomington, IN, USA
2Functional Neuroimaging Laboratory. Center for Applied Medical Research. University of Navarra, Pamplona, Spain
3Center for Neuroimaging. Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
4Indiana Alzheimer’s Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA
5Department of Psychiatry. Indiana University, Indianapolis, IN, USA

Abstract

High-resolution isotropic three-dimensional reconstructions of human brain gray and white matter structures can be characterized to quantify aspects of their shape, volume and topological complexity. In particular, methods based on fractal analysis have been applied in neuroimaging studies to quantify the structural complexity of the brain in both healthy and impaired conditions. The usefulness of such measures for characterizing individual differences in brain structure critically depends on their within-subject reproducibility in order to allow the robust detection of between-subject differences. This study analyzes key analytic parameters of three fractal-based methods that rely on the box-counting algorithm with the aim to maximize within-subject reproducibility of the fractal characterizations of different brain objects, including the pial surface, the cortical ribbon volume, the white matter volume and the grey matter/white matter boundary. Two separate datasets originating from different imaging centers were analyzed, comprising, 50 subjects with three and 24 subjects with four successive scanning sessions per subject, respectively. The reproducibility of fractal measures was statistically assessed by computing their intra-class correlations. Results reveal differences between different fractal estimators and allow the identification of several parameters that are critical for high reproducibility. Highest reproducibility with intra-class correlations in the range of 0.9–0.95 is achieved with the correlation dimension. Further analyses of the fractal dimensions of parcellated cortical and subcortical gray matter regions suggest robustly estimated and region-specific patterns of individual variability. These results are valuable for defining appropriate parameter configurations.
Introduction

Fractal properties of the structure of an object capture its self-similarity in shape over a range of spatial scales or resolutions (Mandelbrot, 1967; Mandelbrot, 1983). Fractal measures (Hentschel and Procaccia, 1983, Henry et al., 2004) have been used extensively as estimators of topological complexity in nature, e.g. in river basins (Cieplac et al., 1998), plant development (Corbit and Garbary, 1995), dendritic arborization of spinal cord neurons (Milosevic et al., 2006) or the morphology and classification of α ganglion cells in the rat retina (Jelinec et al., 2011). A number of studies have used fractal measures to characterize the complexity of grey matter (GM) or white matter (WM) structures of the human brain (Bullmore et al., 1994; Henry et al., 2004; Lopes and Betrouni, 2009). Several recent neuroimaging studies have explored fractal measures, for instance, in relation to cognitive changes and age (Mustafa et al., 2012), in relation to diffuse WM damage (Esteban et al., 2007) and GM damage (Esteban et al., 2009) in multiple sclerosis, or in relation to GM neurodegeneration in mild Alzheimer’s disease (King et al., 2010). Fractal measures complement more standard quantitative analyses of brain structure based on grey matter volume, cortical thickness (Fischl et al. 2000) or voxel-based morphometry (Ashburner et al., 2000). Fractal measures describe topological characteristics based on scaling properties that can provide an estimate of the structural complexity of the object under study, and changes in structural complexity may occur independent of changes in volume. Phenomena related to morphological changes in the brain such as neural reorganization, plasticity or neuronal death may be amenable to fractal quantifications. However, little is known about the reproducibility of fractal measurements and their sensitivity to variations in key analysis parameters.

This study analyzes three fractal-based measures (Hentschel and Procaccia, 1983), the Kolmogorov capacity dimension, the information dimension and the correlation dimension. All rely on the box-counting algorithm, and key parameters of this algorithm that impact the reproducibility of these measures are explored. Fractal measures were computed for the pial surface, the cortical ribbon volume, the white matter volume, and the grey matter/white matter boundary. Two independently collected datasets were analyzed comprising 50 subjects with three separate scanning sessions and 24 subjects with four scanning sessions per subject. The reproducibility of fractal measures was assessed by computing intra-class correlations (Shrout and Fleiss, 1979), derived for volumes and surfaces covering whole hemispheres, as well as for parcellated brain regions.

Materials and Methods

Dataset A

Dataset A was collected at Indiana University School of Medicine (Indianapolis, IN, USA) and consisted of 150 T1-weighted scans (NIfTI format) from 50 healthy controls (all males, 24±3.2 years). For each subject, images were acquired during 3 successive visits scheduled 3 weeks apart on a 3.0 Tesla (T) MR scanner (Siemens TRIO, Germany) using a 12-channel head coil. Anatomical images were acquired using a T1-weighted MPRAGE sequence: TR/TE = 2300/2.91ms, inversion time (TI) = 900ms, FOV = 256×240×160 mm³, flip angle = 9°, 160 slices, voxel size =1.00×1.00×1.20mm³. The study was approved by the local Institutional Review Board, and written informed consent was obtained from subjects prior to scanning.
to any study procedures. Visual inspection of images did not reveal noticeable undesired effects such as ringing motion or ghost images.

**Dataset B**

Dataset B was collected at the Center for Applied Medical Research, University of Navarra Hospital (Pamplona, Spain) and consisted of 96 T1-weighted scans (NIFTI format) from 24 healthy controls (16 males, 57±8.6 years). For each subject, images were acquired during 4 visits scheduled 45 days apart on a 3.0 Tesla (T) MR scanner (Siemens TRIO, Germany) using a 12-channel head coil. Anatomical images were acquired using a T1-weighted MPRAGE sequence: TR/TE = 1620/3.09 ms, inversion time (TI) = 950 ms, FOV = 250×187.5×160 mm$^3$, flip angle = 15°, 160 slices, voxel size =0.98×0.98×1.00 mm$^3$.

Subjects participated in the study approved by the Ethics Research Committee of the University of Navarra, after signing a written informed consent. Visual inspection of images did not reveal noticeable undesired effects such as ringing motion or ghost images.

**Preprocessing and Segmentation**

The general workflow for the image and fractal analysis was performed as follows. In order to approximate an orthogonal orientation, all images were manually reoriented to according to the Montreal Neurological Institute (MNI) reference brain. This step aims to minimize the impact of different and non orthogonal orientations on fractal measures (Foroutan-pour et al. 1999). Preprocessing and segmentation of images was performed using the “recon-all” standard procedure available in the neuroimaging suite FreeSurfer V5.1.0 (http://surfer.nmr.mgh.harvard.edu) for each session of each subject. Briefly, this process includes the following steps. First, several preprocessing steps were taken including motion correction, intensity normalization, removal of extra-cerebral voxels and normalization of head position along the commissural axis. Second, a segmentation procedure applied to GM and WM allowed the definition and extraction of several brain objects (Figure 1). These objects are the cortical ribbon (left and right) and its outer pial surface (left and right), as well as the white matter and gray/white matter boundary (left and right). The cortical ribbon and white matter are three-dimensional volumes (1 mm isotropic voxel size) whereas the pial and gray/white surfaces are expressed as sets of faces and vertices and show sub-voxel accuracy. Details of the automated preprocessing and segmentation steps of FreeSurfer have been reported in a number of articles (Dale et al., 1999; Fischl et al., 2001; Fischl et al., 2002; Fischl et al., 1999; Fischl et al., 2004; Han et al., 2006; Jovicich et al., 2006; Segonne et al., 2007).

**Gray Matter Parcellation**

The cerebral cortex was divided into sixty-eight regions according to the gyral-based atlas proposed by Desikan et al. (2006) available in FreeSurfer (denominated as ‘aparc’). Furthermore, subcortical GM structures were divided, for each hemisphere, into thalamus, caudate, putamen, pallidum, accumbens, hippocampus and amygdala. Overall, eighty-two GM regions were analyzed, corresponding to the GM ‘aparc+aseg’ partition available in FreeSurfer.

**Voxelization of Inner and Outer Surfaces of the Cortical Ribbon**

As opposed to the rest of the brain objects and to the parcellated divisions, all of them represented by volumes, the outer (pial) and inner (gray/white) surfaces of the GM cortical ribbon were extracted by FreeSurfer as sets of faces and vertices. Since box-counting methods require volumes instead of mesh structures, a “voxelization” step was required for both the outer (pial) and inner (gray/white) surfaces. This process did not involve the rest of brain structures and brain regions analyzed. To perform this step, we used the x2v (surface to
volume) function available at the iso2mesh toolbox (Fang and Boas, 2009) for Matlab (freely available at website http://iso2mesh.sourceforge.net/cgi-bin/index.cgi?Home). The input parameter \( \text{div} \) defines the number of voxels that will be used in the smallest Euclidean dimension of the object. In order to avoid undesired volume effects, the parameter was set, in each individual data set (subject and session), to the range of the smallest dimension occupied by the mesh formed by faces and vertices. Similar approaches of voxelization or cube-tiled volumes have also been proposed in the literature (Jiang et al., 2008, King et al., 2010) for this purpose. Assuming that different approaches might alter the resulting volumes and thus the fractal measurements, the results referring to these structures should be taken with caution when looking at them without contrasting other volumetric approaches.

**Fractal Measures**

All fractal measures explored in this study are based on the box-counting method (Russel et al., 1980) that is generally applied to sets of points or volume elements. In this study, a set corresponds to a three dimensional (3D) object \( O \) formed by voxels. Hence, box- or cube-counting refers to quantifying, for different 3D grids made of regular cubes of size \( r \), the number of cubes \( N_c(r) \) necessary to fully cover the object. Let us also define \( N_v(r, i) \) as the number of voxels of the object contained at the \( i \)-th non-empty cube \( (1 \leq i \leq N_c(r)) \) of size \( r \) (note that empty cubes were not counted). Hence the number of voxels (i.e. volume) of the object \( |O| \) can be expressed for a given \( r \) as

\[
|O| = \sum_{i=1}^{N_c(r)} N_v(r, i).
\]

Theoretically, the Kolmogorov capacity dimension \( D_0 \) (widely referred to as the box-counting dimension, or simply the fractal dimension) is expressed as (Henry et al., 2004)

\[
D_0 = \lim_{r \to 0} \frac{\log(N_c(r))}{\log(r)}
\]

This measure was later placed into a generalized dimension spectrum (Hentschel and Procaccia, 1983) where the \( q \)-th dimension, denoted as \( D_q \) is expressed as

\[
D_q = \lim_{r \to 0} \frac{1}{q-1} \frac{\log(\sum_{i=1}^{N_c(r)} p_i^q)}{\log(r)}, \quad q = 0, 1, 2, \ldots,
\]

with \( p_i \) giving the probability of finding a voxel of the object within cube \( i \) of size \( r \), i.e.

\[
p_i = \frac{N_v(r, i)}{|O|}.
\]

As the exponent \( q \) increases, fractal measurements take into account higher-order correlations. In particular, the information dimension \( (D_1) \) is expressed as

\[
D_1 = \lim_{r \to 0} \frac{\sum_{i=1}^{N_c(r)} p_i \log p_i}{\log(r)}
\]

and the correlation dimension \( (D_2) \) is expressed as
For a given object, it is always the case that $D_0 \geq D_1 \geq D_2$ (Henry et al., 2004). Applications of measures of fractal dimensions to real systems in general and to the brain in particular have generally been limited to $D_0$, while an evaluation of $D_1$ and $D_2$ has been infrequent.

In practical applications, when studying objects that are not pure fractals, the limit $\lim_{r \to 0}$ is not attainable (Henry et al., 2004). Instead, $N_c(r)$ and $N_v(r, i)$ are measured for a limited range of $r$ values corresponding to intermediate resolutions of the object under study. Then, estimation of $D_0$ corresponds to the slope obtained by fitting a linear regression (least squares) on the scatter plot of $\log(r)$ versus $\log(N_c(r))$. Analogous linear fits are performed for $D_1$ (scatter plot of $\log(r)$ versus $\sum_{i=1}^{N_c(r)} p_i \log(p_i \log(p_i))$) and for $D_2$ (scatter plot of $\log(r)$ versus $\log(\sum_{i=1}^{N_c(r)} p_i^2)$). The linear fits used for these three measurements rely on their scale invariance, ideally allowing that evaluations at different resolutions or after object re-scaling do not alter the fractal measurements, at least in the case of pure fractals.

Due to individual differences in brain geometry across subjects, brain regions and surfaces have different sizes that could lead to volume-effect biases. Hence a grid based on cubes of side $r$ does not necessarily represent the same spatial resolution across different subjects. For this reason, $r$ values were obtained as a function of the smallest dimension of the object under study. Selecting the appropriate resolutions to be evaluated is very important for the computation of FD measures based on box-counting (Foroutan-pour et al., 1999). The minimum and maximum resolutions used for grids were set to 5% and 40% of the smallest Euclidean dimension of the object, respectively. Note that we are expressing a percent and not an absolute magnitude. This implies that, in order to represent truly equivalent resolutions, slightly different cube-sizes might be used among subjects. Using very low or very high resolutions in box-counting methods applied to real objects has been found to produce inaccurate results (Foroutan-pour et al., 1999).

Under a box-counting approach, there are a number of parameters that can affect the computation and accuracy of $D_0$, $D_1$ and $D_2$:

- **GridScales**: This parameter corresponds to the number of different cube sizes used, and thus defines the number of levels of scale that contribute to estimating the slopes and thus the fractal values. Values of GridScales ranged from 3 to 9 (in increments of 1). For a given value of GridScales, the smallest and largest cube sizes were set to 5% and 40% of the smallest Euclidean dimension of the object, respectively. Intermediate cube sizes were chosen to be equidistant on the log-scale. Three different scales for an object can be seen in Figure 2A, B and C.

- **GridOffsets**: For each cube size, a grid is set up in order to compute $N_v(r, i)$ and $N_c(r)$. The placement of the grid with respect to the object along the three Euclidean dimensions is known to affect the computation of fractal measures based on the box-counting method (Barnsley 1988; Feder 1988; Falconer 1990). In order to avoid systematic bias, a number of randomly offset grid placements were chosen. The offset coordinates is formed by a set of 3 components $\{i, j, k\}$ indicating the starting point of the object with respect to the grid on each Euclidean dimension. For a given cube of side $r$, which determines the grid, each offset component can be from 1 up to $r$. Offsets were computed independently on each component.
generating random integer numbers uniformly distributed in the range \([1,2,\ldots, r]\). The number of randomly placed offsets is given by the parameter \(\text{GridOffsets}\) and ranges from 1 to 20 in this study. Higher numbers of \(\text{GridOffsets}\) up to a 100 were explored in some data sets but did not provide additional information. Three different offsets for an object and a fixed scale can be seen in Figure 2D, E and F.

- **GridMode**: Fractal theory suggests that as multiple randomly placed offsets are sampled, the fractal dimension should correspond to the slope obtained when including the smallest \(N_c(r)\) values found for each cube side \(r\) (Barnsley 1988; Feder 1988; Falconer 1990). Because finding the smallest value for the fractal dimension of a biological object may sensitively depend on the data acquisition and other parameters (e.g. the number of \(\text{GridOffsets}\) values), all \(N_c(r)\) values were retained and three different \(\text{GridModes}\) were pursued, i.e. keeping the smallest \(N_c(r)\), the mean \(N_c(r)\) and the largest \(N_c(r)\). Respectively, these correspond to \(\text{GridMode} = \text{min}, \text{avg}, \text{max}\).

The influence of these three parameters in quantifying \(D_0\) is shown in Table 1. These parameters were studied for the cortical ribbon, pial surface and gray/white matter boundary both at left and right hemispheres. The most robust configuration of parameters was later applied in a post-hoc analysis of structural complexity and ICC for cortical and subcortical GM structures. All computations of fractal measures and their statistical analyses were performed in Matlab R2012a (Mathworks, Natick, MA, USA).

In order to test our computational procedure, we evaluated a 3D pure fractal, the Menger sponge (with 5 iterations). Theoretical results of \(D_0\) are \(\log(20) \log(3) \approx 2.7268\) whereas our empirical results were, on average after 50 runs, as accurate as \(\langle D_0 \rangle = 2.7078\) when using \(\text{GridMode} = \text{min}, \text{GridOffsets} = 100\) and \(\text{GridScales} = 5\) (in particular cube sides were 4, 8, 16, 32 and 64). We also observed that the accuracy monotonically increased with \(\text{GridOffsets}\) within the range studied (from 1 to 100).

### Intra-Class-Correlation of Fractal Measurements

Intra-class-correlation (ICC; Shrout and Fleiss, 1979) is a descriptive statistic widely used to assess the consistency or reproducibility of quantitative measurements made by different observers when measuring the same quantity. In our study, those quantitative measurements are, for each dataset, the values of \(D_0, D_1\) and \(D_2\) for different brain regions and surfaces, and the “observers” are the scan sessions (three sessions per subject for dataset A and four sessions per subject for dataset B). The ICC represents a composite of intra-observer and inter-observer variability. On the one hand, FD descriptors obtained from different scan sessions of the same subject in a short period of time are desired to be as similar as possible, in order to show low intra-subject variability across sessions. On the other hand, FD descriptors are also desired to be specific to each individual, and thus different between individuals, i.e. are desired to show high inter-subject variability (with respect to the intra-subject variability). The ICC modality used in this study for both datasets was the two-way random single measure, also referred to as ICC(2,1) (Rajaratnam, N. 1960, Shrout and Fleiss, 1979). For the sake of simplicity, from now on, we will refer to it simply as ICC.

Let’s define \(x_{ij}\) as one measurement provided by observer \(i\) when evaluating a subject \(j\). The linear model for the design corresponds to a two-way random effects model:

\[
x_{ij} = \mu + a_i + b_j + (ab)_{ij} + e_{ij},
\]

This shows that an observation is composed of the grand mean and four deviation terms from that mean. Parameter \(a\) stands for the observers deviations (scanner sessions of the
same subject), parameter \( b \) stands for the target deviations (subjects), \( ab \) stands for the interaction of the former two, and \( e \) stands for error. The estimated sources of variation associated to this model are defined in Table 2, and permit to estimate the ICC coefficient as follows:

\[
ICC(2, 1) = \frac{BMS - EMS}{BMS + (k-1)EMS + k(JMS - EMS)/n}
\]

were all \( k \) observers evaluate each of all the \( n \) targets (subjects). Sources of variation in the formula are between targets (BMS) and within target divided into between judges (JMS) and residual (EMS). See Table 2 for details. It ranges from 0 (no agreement among observers), to 1 (perfect agreement among observers). \( ICC(2, 1) \) is an estimation of the covariance ratio between the between targets variability and all four kinds of variability described above, namely the population value \( \rho \):

\[
\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_j^2 + \sigma_e^2}
\]

The ICC of \( D_0 \), \( D_1 \) and \( D_2 \) was measured for each GM-object for different configurations of parameters. Those configurations were determined by different settings of GridScales, GridOffsets and GridModes. All calculations were performed for both datasets A and B, in particular to seven different objects: left and right cortical ribbon, left and right pial surface, as well as the white matter and left and right gray/white matter boundaries. The parameter configurations with the highest ICC were later applied in a post-hoc analysis to estimate FD measures from the overall eighty-two cortical and subcortical GM structures. In all cases, fractal and ICC measurements shown correspond not to a single realization but to the average obtained after 16 realizations.

Statistical significance for correlations between measurements was set to \( p < 0.05 \).

Results

For each scanner session of the two datasets, computation of Kolmogorov capacity dimension (\( D_0 \)), the information dimension (\( D_1 \)) and the correlation dimension (\( D_2 \)) was performed for seven objects, namely left and right pial surfaces, left and right ribbon, left and right white surfaces and white matter. The parameters evaluated were GridModes (min, avg and max), GridOffsets (from 1 to 20) and GridScales (from 3 to 9). The ICC was computed for all objects and all parameter configurations on each dataset.

Figure 3A shows representative examples of one fractal measure, the Kolmogorov capacity dimension \( D_0 \), and its ICC for the left-hemispheric pial surface of dataset A. Several relationships between key algorithmic parameters and the ICC are evident. First, the three different settings of the GridMode parameter result in sharp differences in the reproducibility of estimating \( D_0 \), i.e. the ICC. Using only either the maximal or minimal value for \( D_0 \) resulted in large variations in \( D_0 \) estimates across scanning sessions from the same subject. Instead, stable and consistent estimates were achieved by averaging \( D_0 \) estimates from all grid offsets, with ICC values achieving very high consistency at around GridOffsets =15. This strong effect suggests that \( D_0 \) cannot be stably estimated if only extreme values are sampled. Instead, averaging over different instantiations of the grid provide higher ICC. Second, the number of GridScales had little effect on ICC values. Even
a modest number of GridScales provided stable ICC. These results hold across all objects in both hemispheres (data not shown).

Figure 3B shows two representative examples of $D_0$ estimates for the left pial surface of all subjects and all scanning sessions derived from the middle panel in Figure 3A. The plot at the top illustrates $D_0$ estimates with low ICC (ICC = 0.20) calculated using suboptimal parameter settings ($\text{GridOffsets} = 1$, $\text{GridScales} = 5$). The plot at the bottom shows $D_0$ estimates with high ICC (ICC = 0.89; $\text{GridOffsets} = 20$, $\text{GridScales} = 8$).

Because of the superior reproducibility of averaging over all grid offsets, this setting ($\text{GridMode} = \text{avg}$) is used for all further analyses shown in this paper. Figure 4 presents a comparison of the three fractal measures $D_0$, $D_1$ and $D_2$ for all structures (cortical ribbon, pial surface, gray/white matter boundary and white matter) obtained from dataset A. Cortical ribbon, pial surface and gray/white matter boundary are shown for both left and right brain hemispheres. All plots are heat maps visualizing ICC across variations of the two parameters $\text{GridOffsets}$ and $\text{GridScales}$. Consistent with data shown in Figure 3, ICC values plateau at around 15 $\text{GridOffsets}$ and the number of $\text{GridScales}$ has relatively little effect. High ICC values of at least 0.8 are achieved for all objects under study, with highest ICC values obtained for the gray/white matter boundary and white matter structures. Across measures and for all objects and both data sets, ICC levels for $D_0$ tended to be lower than for measures $D_1$ and $D_2$. Corresponding plots for dataset B are shown in Figure 5.

Table 3 summarizes the values for all three fractal measures $D_0$, $D_1$ and $D_2$, for all objects under study, for parameter settings $\text{GridMode} = \text{avg}$, $\text{GridOffsets} = 20$, $\text{GridScales} = 9$. In addition, the table provides values for each structure’s regional volume. Consistent with fractal theory (Henry et al., 2004), we observe that $D_2 > D_1 > D_0$ for all structures and both datasets. Furthermore, with the exception of white matter, fractal dimensions are significantly greater for dataset A compared to dataset B.

Building on these results obtained for whole-brain or whole-hemisphere structures, we performed a post hoc analysis to assess fractal properties of individual cortical and subcortical GM regions, using parameter settings $\text{GridMode} = \text{avg}$, $\text{GridOffsets} = 20$, and $\text{GridScales} = 5$. We adopted a lower setting for $\text{GridScales}$ because of the smaller sizes of individual regions. We chose to perform this regional analysis using the correlation dimension $D_2$ which consistently resulted in the highest ICC for the parameter settings mentioned above.

The resulting distributions of $D_2$ and ICC revealed regional variations of fractal dimension as well as its corresponding ICC. Importantly, regional variations in $D_2$ and ICC showed consistent patterns across the two datasets under study. Figure 6A shows a comparison of the regional estimates of $D_2$, and Figure 6B the corresponding regional ICC values, averaged across subjects and sessions from dataset A (abscissa) and dataset B (ordinate). Both, $D$ and ICC are highly correlated across the two datasets ($r = 0.97$, $r = 0.74$, respectively, both $p < 0.001$). We note that fractal estimates from dataset B are consistently lower than those from dataset A (see also Table 3), an effect we attribute to differences in the mean age of the two subject cohorts (see Discussion).

Fractal measures of the human cortical surface have been found to correlate, to some extent, with cortical thickness (Im et al., 2006) among other features. We examined the relation of regional estimates of $D_2$ and the corresponding regional ICC with each other, as well as with regional estimates of volume and (for cortical regions) thickness. Regional $D_2$ and ICC were only weakly correlated with each other ($r = -0.35$ significant correlation in dataset A, and not significantly correlated in dataset B). $D_2$ was not significantly correlated with regional
volume in any of the two datasets, whereas ICC was significantly correlated with regional volume for both dataset A ($r = 0.27$) and dataset B ($r = 0.34$). Correlations between the two measures and cortical thickness were significant in all cases, but lower in dataset A ($r = 0.25$, $r = -0.31$) for $D_2$ and ICC, respectively than in dataset B ($r = 0.44$, $r = -0.43$). The weak correlation with regional volume suggests that $D_2$ estimates are not strongly dependent on the size of the region from which they are computed. The strengthening of the relation of both measures with cortical thickness in dataset B relative to dataset A may be due to age differences in the two subject cohorts (see Discussion).

The high consistency of $D_2$ and ICC across the two datasets (Figure 6) allowed us to aggregate these data across both datasets under study and to generate maps of their anatomical distributions ($D_2$ and ICC shown in Figure 7 and Figure 8, respectively). Measures are computed by first averaging over all realizations per session, sessions per subjects, then over subjects and finally over datasets A and B. Neither $D_2$ nor ICC exhibit any obvious anterior-posterior, medial-lateral or dorsal-rostral gradients that could be attributed to acquisition artifacts. Subcortical regions exhibited higher structural complexity, but with lower ICC. Both $D_2$ and ICC exhibit pronounced symmetry across the two hemispheres, both for subcortical and cortical regions. Cortical regions with high $D_2$ included the superior frontal cortex, rostral middle frontal cortex, the precuneus, inferior parietal cortex and the banks of the superior temporal sulcus. All of these regions also exhibited high ICC. Regions with low $D_2$ included the parahippocampal gyrus, posterior cingulate cortex, and the cuneus. Only a few regions exhibited low ICC, most prominently the pallidum (see Figure 9A) and the insula.

**Discussion**

Fractal measures have been used in a number of studies to characterize the complexity of human brain structures. The approach may have some promises in identifying differences across individual subjects that index differences in behavioral or cognitive capacities, developmental stages, neurodegeneration or other disease-related processes. In order to be operationally useful, fractal estimates need to exhibit robustness and reproducibility, with high consistency between multiple acquisitions from the same subject, and significant differences between subjects. To establish these characteristics for fractal measures requires test-retest studies across subject cohorts. In neuroimaging, test-retest studies have been performed in the context of fMRI activation (Raemaekers et al. 2012), pulsed arterial spin labeling perfusion (Wang et al. 2011b) in graph theoretical analyses of structural (Bassett et al., 2011) and resting-state functional networks (Wang et al. 2011), and in metrics of morphological atrophy (Duchesne et al. 2012). Robustness of fractal measures has been examined in non-neural applications, e.g. in studies of regional blood flow in rectal cancer (Sanghera et al. 2012), but methodological analysis applied to brain imaging is lacking so far. In this study, we report findings about the reproducibility and robustness, specifically the intra-class-correlation (ICC), of whole-brain and regional estimates of several measures of fractal dimension applied to structural MRI images. Applying our results in two separately acquired datasets we identify regional patterns of structural complexity and ICC across cortex and subcortical structures.

Estimates of fractal dimension derived from biological objects such as cortical surfaces or brain regions are necessarily approximations, as only idealized mathematical or “pure” fractals exhibit a single true fractal dimension manifesting at all scales. Real biological objects have not only finite size but, in the case of neuroimaging, noisy acquisition and limited resolution due to voxel size impose further constraints on fractal estimates. The constraints must be taken into account if valid estimates of fractal dimension are to be derived from realistic brain imaging data. In the absence of a single true fractal dimension
for the brain, an accurate estimate may be expected to show high consistency across multiple observations acquired over short time periods, as well as to express differences between structures coming from individual subjects whose brain structure varies. This reasoning is the main rationale for aiming to derive fractal estimates with high ICC, thus maximizing test-retest consistency as well as between-subject differences.

Our results show that it is possible to derive fractal estimates with high ICC over a wide range of algorithmic parameters (Figure 4 and Figure 5). Since all measures are based on box-counting approaches, their computation sensitively depends on the placement of the grid that determines the orientation and spacing of the boxes. As opposed to what is considered good practice in pure fractals (Barnsley 1988; Feder 1988; Falconer 1990), our results suggest that in brain objects high ICC is difficult to achieve when only minimal (or maximal) coverage (GridMode = {min, max}) is taken into account. This effect is most likely due to the resolution limits that underlie the neuroimaging acquisition. Instead, averaging over multiple grid placements gave robust ICC values, for modest numbers of GridOffsets and across small numbers of GridScales. This allows reproducible estimates for fractal dimensions at reasonable computational cost. In both datasets studied here, high ICC values could be achieved for all whole-brain or whole-hemisphere brain structures (ICC>0.8), as well as for the majority of cortical and subcortical regions (ICC>0.6).

Our results show not only that it is possible to achieve high ICC within one subject cohort, but we demonstrated high consistency in both $D_2$ and ICC across two datasets acquired independently (Figure 6). This suggests that regional differences in estimates for the correlation dimension $D_2$ reflect true differences in structural complexity across the brain. Hence, it may be possible to create a map of structural complexity and relate this map to underlying neurobiological processes.

Our post-hoc analysis revealed that some cortical and subcortical regions show higher ICC than others. Importantly, ICCs were highly consistent across two different subject cohorts acquired at different imaging centers. This may suggest that regional differences in ICC reveal some real patterns in the way brain structures vary across individuals. ICC as a measure expresses a trade-off between intra-subject variability (across sessions) and inter-subject variability (across individuals). In our study, high ICC (ICC $\approx 1$) may be due to significant inter-subject variability that outweighs inconsistencies in estimates across short time periods, while near-zero ICC implies that intra- and inter-subject variability cannot be distinguished. The latter effect may be due to low inter-subject variability or high intra-subject variability (e.g. due to acquisition noise or registration errors). An exploratory analysis of patterns of variability in low-ICC and high-ICC regions (Figure 9) suggests that intra-subject variability is relatively constant, but that lower ICC in some structures may be due to lower inter-subject variability. Hence, it appears that the level of inter-subject variability is driving the level of ICC.

ICC and fractal estimates could be influenced by the volumes or spatial location of regions. In dataset A, linear regression of ICC and $D_2$ on measures of regional volume and cortical thickness explained less than 10% of the variance. In dataset B, cortical thickness was found to be moderately correlated with ICC and $D_2$ but still accounting only for less than 20% of the variance. The difference in relation to cortical thickness between these two datasets may be due to the age difference (expressed in both mean and dispersion) in the two subject cohorts. Although the visual inspection of anatomical images did not reveal noticeable artifacts, we cannot totally exclude the possibility that these aging effects are related to small motion artifacts (Stockman et al., 2012). This could also explain a generally lower ICC in brain structures of Dataset B with respect to Dataset A (see Figures 4 and 5). The anatomical distribution of ICC and $D_2$ across the cortical surface does not indicate a spatial pattern that
may be due to linear gradients in signal strength/accuracy or other geometric effects. Overall, \(D_2\) and ICC estimates appear to express structural characteristics that are fairly independent from more classical measures such as volume and thickness, and independent of the spatial location of each region within the brain.

The high consistency of \(D_2\) and ICC across datasets suggests that fractal measures can capture aspects of regional structural organization that are stable and reproducible across studies and represent true structural features. The anatomical distribution of \(D_2\) across cortical and subcortical regions did reveal some interesting patterns. First, we found strong hemispheric symmetry across all regions. Second, subcortical regions tended to exhibit higher structural complexity, an effect that may be due to their specific geometry. Third, regions with high \(D_2\) tended to be regions identified in other studies as structurally highly connected and highly central brain “hubs” (e.g. Hagmann et al., 2008’ van den Heuvel and Sporns, 2011). Finally, the regional distribution of ICC may give an indication of the degree to which structural complexity varies across individuals, an idea for which the data shown in Figure 9 provide some support. An exploratory analysis suggests that lower ICC is not due to greater within-subject variability but instead is associated with brain regions whose structural complexity is more constant across individuals.

The current study was intended to explore parameters for the robust estimation of the structural complexity of whole-brain objects and individual brain regions. We have identified parameter settings that should allow capturing inter-subject differences even from data derived from single scanning sessions. Our results now open the possibility to measure regionally specific differences in structural complexity across subjects, and thus add a potentially valuable tool for characterizing changes in brain structure associated with ageing, neurodegenerative processes, substance abuse, or pharmacological intervention.

**Acknowledgments**

JG acknowledges Spanish Government grant, contract number E-28-2012-0504681. JG and OS acknowledge support from the James S. McDonnell Foundation. AS acknowledges support by the National Institutes of Health (P30 AG10133, R01 AG19771 and R01 LM011360) and the National Science Foundation (IIS-1117335).

**References**


*Neuroimage. Author manuscript; available in PMC 2014 December 01.*


Highlights

- The study quantifies the reliability of fractal measures within the human brain.
- Parameter settings that ensure high reproducibility are identified.
- All results provided were validated in two separate subject cohorts.
- Estimates of fractal dimension can be derived with high intra-class correlation.
- We found robust and reproducible regional differences in fractal dimension.
Figure 1.
Examples of brain objects under study. Panels showing the brain objects: cortical ribbon (left and right hemisphere), pial surface (left and right hemisphere), gray/white matter boundary (left and right hemisphere) and white matter (whole brain). The insert at the bottom right shows details of the cubes forming part of the white matter object.
Figure 2.
Examples of different scales and offsets for an object. The parameter GridScales is defined by the number of different cubeSides used to create grids and the parameter GridOffsets is defined by the number of positionings (offsets) of the object within the grid. A, B and C show three different grids corresponding to cubeSides $r=16$, $r=24$, $r=32$ respectively. D, E and F show 3 different offsets (indicated by a green solid circle) for the same grid (cubeSide $r=32$).
Figure 3.
Representative examples of $D_0$ and ICC in relation to parameter settings. The examples shown correspond to calculations of the Kolmogorov capacity dimension $D_0$ of the left pial surface of dataset A. $D_0$ and ICC heat maps (top to bottom) are generated with $\text{GridMode} = \text{min}$, $\text{GridMode} = \text{avg}$, $\text{GridMode} = \text{max}$, respectively. B. The plots show fractal estimates of $D_0$ from three scanning sessions for each of the individual subjects (dataset A). In each plot, subjects are arranged by their mean $D_0$ values in increasing order (left to right). The plot at the left was obtained with settings of $\text{GridOffsets} = 1$ and $\text{GridScales} = 5$, resulting in ICC = 0.20. The plot at the right was obtained with settings of $\text{GridOffsets} = 20$ and $\text{GridScales} = 8$, resulting in ICC = 0.89.
Figure 4.
ICC results for Dataset A, with gridMode set to avg. The plots show data for left and right hemisphere structures (pial surface, cortical ribbon, gray/white matter boundary) and white matter. Rows of heat-maps represent $D_0$, $D_1$ and $D_2$ respectively.
Figure 5.
ICC results for Dataset B, with gridMode set to avg. The plots show data for left and right hemisphere structures (pial surface, cortical ribbon, gray/white matter boundary) and white matter. Rows of heat-maps represent $D_0$, $D_1$ and $D_2$ respectively.
Figure 6.
A. Post-hoc analysis of the reproducibility of correlation dimension ($D_2$). All data are for parcellated cortical regions (derived from the cortical ribbon) and subcortical regions, with parameters set to GridMode = avg, GridOffsets = 20, GridScales = 5. The correlation between the two datasets was very high ($r=0.97, p<0.001$) while most of the regions had a smaller correlation dimension $D_2$ in dataset B. B. Most of the regions showed a high ICC (>0.6) in both datasets. Furthermore, there was a significant correlation between the ICC ($r=0.74, p<0.001$) of the regions from each dataset, which indicates a similar level of reproducibility of $D_2$ for each region regardless of the dataset studied.
Figure 7.
Characterizing the structural complexity of parcellated cortical and subcortical regions. Data shown are values of the correlation dimension ($D_2$), averaged across the two data sets A and B. The bar graph at the left shows values for both hemispheres and all (cortical and subcortical) regions under study. The surface plots on the right show the regional distribution of $D_2$ across the cortical surface (subcortical regions not shown).
Figure 8.
Characterizing the reproducibility of the parcellated complexity of different cortical and subcortical regions. Data shown are values of the ICC of the correlation dimension (ICC-D2), averaged across the two data sets A and B. The bar graph at the left shows values for both hemispheres and all (cortical and subcortical) regions under study. The surface plots on the right show the regional distribution of ICC-D2 across the cortical surface (subcortical regions not shown).
Figure 9.
Examples of different ICC values obtained in dataset A for three different regions. In each plot, subjects are arranged by their mean $D_2$ values in increasing order (left to right). A. The pallidum (right) gave rise to a low ICC (0.19). B. The cuneus (right) gave rise to a medium-high ICC (0.63). C. The caudal anterior cingulate (left) gave rise to a high ICC (0.84). Insets on each panel show the region as identified by the segmentation procedure in one representative subject of dataset A.
Example of how computation of $D_0$ depends on parameters $\text{GridScales}$, $\text{GridOffsets}$ and $\text{GridMode}$. The object studied is the left pial surface as shown in Figure 2. Furthermore, the three offset coordinates used for $r = 32$ also correspond to those shown in Figure 2D, E and F, respectively.

### Table 1

<table>
<thead>
<tr>
<th>GridScales=4</th>
<th>GridOffsets=3</th>
<th>$N_c (r)$ for GridMode</th>
</tr>
</thead>
<tbody>
<tr>
<td>cubeSide $r$</td>
<td>offset 1</td>
<td>$N_c (r)$</td>
</tr>
<tr>
<td>$r = 8$</td>
<td>(7,8,2)</td>
<td>1303</td>
</tr>
<tr>
<td>$r = 16$</td>
<td>(11,13,12)</td>
<td>249</td>
</tr>
<tr>
<td>$r = 24$</td>
<td>(2,3,20)</td>
<td>93</td>
</tr>
<tr>
<td>$r = 32$</td>
<td>(2,20,8)</td>
<td>53</td>
</tr>
</tbody>
</table>

$D_0$ for GridMode

<table>
<thead>
<tr>
<th>min</th>
<th>avg</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.334</td>
<td>2.308</td>
<td>2.271</td>
</tr>
</tbody>
</table>
**Table 2**

ICC sources of variation

Two-way ANOVA with random effects for \( n \) subjects and \( k \) observers (also known as judges or raters). Symbol \( \sigma \) denotes standard deviation, and sub-indexes \( S, O, I, E \) refer to subjects, observers, interaction and error respectively.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>term</th>
<th>degrees of freedom</th>
<th>Expected mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td>BMS</td>
<td>( n-1 )</td>
<td>( k\sigma^2_S + \sigma^2_I + \sigma^2_E )</td>
</tr>
<tr>
<td>Within targets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- between observers</td>
<td>JMS</td>
<td>( k-1 )</td>
<td>( n\sigma^2_O + \sigma^2_I + \sigma^2_E )</td>
</tr>
<tr>
<td>- residual error</td>
<td>EMS</td>
<td>( (k-1)(n-1) )</td>
<td>( \sigma^2_I + \sigma^2_E )</td>
</tr>
</tbody>
</table>
Fractal measures of structural complexity of the human brain. These values represent mean (above) and standard deviation (below) of the distribution of fractal measures \((\text{GridMode} = \text{avg})\) computed for each brain object across all subjects in the two data sets. The contribution of each subject is computed as the mean of the subject’s scan sessions. Parameters are \(\text{GridOffsets} = 20\) and \(\text{GridScales} = 9\). Term \(|O|\) represents the number of isotropic voxels \((1\ mm^3)\) used in the different structures, except in pial and gray/white surfaces, were units are adimensional.

<table>
<thead>
<tr>
<th>Dataset A</th>
<th>Pial surface</th>
<th>Cortical ribbon</th>
<th>Gray/white boundary</th>
<th>White matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>hemisphere</td>
<td>left</td>
<td>right</td>
<td>left</td>
<td>right</td>
</tr>
<tr>
<td>(D_0)</td>
<td>2.3446 (0.0109)</td>
<td>2.3504 (0.0105)</td>
<td>2.4398 (0.0099)</td>
<td>2.4423 (0.0101)</td>
</tr>
<tr>
<td>(D_1)</td>
<td>2.4129 (0.0135)</td>
<td>2.4176 (0.0126)</td>
<td>2.5603 (0.0115)</td>
<td>2.5600 (0.0116)</td>
</tr>
<tr>
<td>(D_2)</td>
<td>2.4245 (0.0143)</td>
<td>2.4294 (0.0136)</td>
<td>2.5874 (0.0125)</td>
<td>2.5871 (0.0127)</td>
</tr>
<tr>
<td>(</td>
<td>O</td>
<td>)</td>
<td>147208 (11846)</td>
<td>148947 (11841)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset B</th>
<th>Pial surface</th>
<th>Cortical ribbon</th>
<th>Gray/white boundary</th>
<th>White matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>hemisphere</td>
<td>left</td>
<td>right</td>
<td>left</td>
<td>right</td>
</tr>
<tr>
<td>(D_0)</td>
<td>2.3323 (0.0068)</td>
<td>2.3357 (0.0076)</td>
<td>2.4244 (0.0077)</td>
<td>2.4271 (0.0094)</td>
</tr>
<tr>
<td>(D_1)</td>
<td>2.4006 (0.0084)</td>
<td>2.4031 (0.0097)</td>
<td>2.5530 (0.0103)</td>
<td>2.5367 (0.0126)</td>
</tr>
<tr>
<td>(D_2)</td>
<td>2.4123 (0.0093)</td>
<td>2.4146 (0.0108)</td>
<td>2.5586 (0.0124)</td>
<td>2.5600 (0.0141)</td>
</tr>
<tr>
<td>(</td>
<td>O</td>
<td>)</td>
<td>135457 (7932)</td>
<td>135711 (7694)</td>
</tr>
</tbody>
</table>