Optimizing Complexity Measures for fMRI Data: Algorithm, Artifact, and Sensitivity

Denis Rubin[1,2], Tomer Fekete[2], Lilianne R. Mujica-Parodi[2]*

1 Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York, United States of America, **2** Department of Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, New York, United States of America

Abstract

Introduction: Complexity in the brain has been well-documented at both neuronal and hemodynamic scales, with increasing evidence supporting its use in sensitively differentiating between mental states and disorders. However, application of complexity measures to fMRI time-series, which are short, sparse, and have low signal/noise, requires careful modality-specific optimization.

Methods: Here we use both simulated and real data to address two fundamental issues: choice of algorithm and degree/type of signal processing. Methods were evaluated with regard to resilience to acquisition artifacts common to fMRI as well as detection sensitivity. Detection sensitivity was quantified in terms of grey-white matter contrast and overlap with activation. We additionally investigated the variation of complexity with activation and emotional content, optimal task length, and the degree to which results scaled with scanner using the same paradigm with two 3T magnets made by different manufacturers. Methods for evaluating complexity were: *power spectrum*, *structure function*, *wavelet decomposition*, *second derivative*, *rescaled range*, *Higuchi's estimate of fractal dimension*, *aggregated variance,* and *detrended fluctuation analysis*. To permit direct comparison across methods, all results were normalized to Hurst exponents.

Results: Power-spectrum, Higuchi's fractal dimension, and generalized Hurst exponent based estimates were most successful by all criteria; the poorest-performing measures were wavelet, detrended fluctuation analysis, aggregated variance, and rescaled range.

Conclusions: Functional MRI data have artifacts that interact with complexity calculations in nontrivially distinct ways compared to other physiological data (such as EKG, EEG) for which these measures are typically used. Our results clearly demonstrate that decisions regarding choice of algorithm, signal processing, time-series length, and scanner have a significant impact on the reliability and sensitivity of complexity estimates.

Introduction

The prevalence of power-law or scale-free behavior in natural processes is well-established [1], with theoretical justification for the complexity found in neurobiological and physiological systems [2,3]. These include considerations of robustness, in which redundancies maximize system integrity in the event of damage, as well as adaptability: operating on the edge of chaos, complex systems position themselves for optimal responsivity to inputs, as well as ability to maintain homeostatic regulation. The complexity of brain activity has been observed and modeled on many levels, from neuronal spiking [4,5] and local field potentials [6] to EEG [7], suggesting that scale-free behavior may be fundamental to neural information processing. Recent fMRI experiments demonstrate that brain signal fluctuations differ from other scale-free natural processes, suggestive of different underlying mechanisms [8], and that fMRI time series exhibit complexity that is functionally significant [9]. Moreover, there is evidence that

deviations from chaotic behavior can be used diagnostically in identifying disease, using ECG [10–12], EEG [13], MEG [14], NIRS [15], and fMRI [8,16–18].

FMRI applications of complexity have dealt with both between-voxel and between-subject differences. For example, active and inactive voxels in the human visual cortex show markedly different power-law and Hurst exponents during a simple rotating checkerboard fMRI paradigm [19]. Shimizu et al. [20] applied a multifractal version of the Hurst exponent (the Hölder exponent) to fMRI data to show that active voxels are clearly distinct from both non-active ones and white matter. More recent results illustrate that the power law exponent varies across networks of voxels (e.g. attention, default, motor, saliency, and visual) [8] and is affected by cognitive load [21]; for review, see [22] and [3]. Between-group differences in time series complexity either focused on a particular disorder, such as Alzheimer's disease [17], autism [16], and schizophrenia [18], or individual variability across psychophysiological variables such as trait anxiety, heart rate

variability [23], age, cholinergic effects, and cognitive performance [24,25].

Complexity measures in physiology have historically been applied to time-series that are long, rich, and have strong signal/noise, as with 24-hour ECG. In contrast, fMRI time-series tend to be short (5–10 minutes), sparse (TR = 1–2.5 s), and subject to sufficient scanner and physiological artifacts that significant pre-processing is the norm. Despite the application of complexity methods to fMRI data, and the challenges associated with it, no systematic analyses of the optimal way to compute such complexity measures specifically for fMRI data have been published. To this end, here we evaluate the most common methods for computing time series complexity, investigate the effects of data preprocessing, activation and scanner differences, with eye towards optimizing the balance between detection sensitivity and resilience to artifact. For the sake of permitting direct comparison between methods, each of which is scaled differently, all values will be normalized to the Hurst exponent.

Methods

2.1 Overview

This paper references a range of fMRI datasets, with all data collected on a single Siemens Magnetom Trio 3T scanner, unless specified otherwise. The tasks include a resting state design (RST; N = 11), a "guided-rest" design modified to standardize content across subjects by presentation of a movie, the pilot episode of the television series "Lost" (GRST-LOST; N = 11), event-related presentation of emotionally valent faces (ER-FACES; N = 22), block presentation of faces (BL-FACES; N = 22), and a block design auditory anticipation task (BL-ANT; N = 22). The GRST-LOST task was implemented mainly to provide a realistic design-free stimulus that would ensure that participants do not fall asleep during the lengthy scan session designed to provide long time series. The particular choice of show was made based on the general public consensus that the show is highly rated (IMDB rating of 9.3) and therefore would be engaging. The other tasks (BL-ANT, ER-FACES, BL-FACES) were chosen for the purpose of task design comparison, as all three sets of stimuli were presented to the same 22 participants.

2.2 fMRI Acquisition Parameters

For the three sets of stimuli presented to the same 22 participants (BL-ANT, ER-FACES, BL-FACES), we acquired fMRI data using a 3 Tesla Siemens Trio whole body scanner equipped with an eight-channel SENSE head coil. We acquired T2*-weighted whole-brain volumes with an EPI sequence with the following parameters: TR = 2500 ms, TE = 22 ms, flip angle = 83°, matrix dimensions = 96×96, FOV = 224 mm×224 mm, slices = 36, slice thickness = 3.5 mm, gap = 0. For RST and GRST-LOST the TR was changed to 2100 ms and the number of slices to 37. Standard preprocessing procedures were performed in SPM8, including image realignment corrections for head movements, slice timing corrections for acquisition order, normalization to standard 2×2×2 mm Montreal Neurological Institute space, and spatial smoothing with a 6-mm full width at half maximum Gaussian kernel. The BL-ANT task was preprocessed in SPM5, but the procedures used produce numerically identical results in SPM5 and SPM8. As for the other tasks, the standard preprocessing in SPM8 included realignment, normalization, and spatial smoothing with the same parameters as above. The number of acquired volumes in these tasks was as follows: BL-ANT = 232, ER-FACES = 226, BL-FACES = 264, RST = 143, GRST-LOST = 1190.

2.3. Task Description

The GRST-LOST data were collected during an uninterrupted scanning session, with a five-minute pure rest condition (RST) always preceding guided-rest condition (GRST-LOST). During the rest condition, participants were instructed to relax and let their minds wander while looking at a fixation cross and waiting for the show to start. Immediately after the rest condition, participants viewed the entire first part (42 minutes) of the pilot of ABC's TV-series "Lost" (GRST-LOST). Using simultaneous eye-tracking during scans, we verified that all participants stayed awake during the entire course of the experiment. Immediately after watching the episode, the participants rated the content of the show on a nine-point scale in terms of arousal ("excitement") and valence ("pleasantness"). The ratings were provided for each of the two hundred 10–15 second scenes into which the episode was split. The participants were instructed to rate their emotional responses rather than the content. The sequences of clip ratings were interpolated to match the length of the show and averaged across subjects. The valence scores were inverted for ease of interpretation.

In the BL-ANT task, the participants anticipated an unpleasant or a neutral stimulus during a 16-second countdown preceded by a cue informing them of the upcoming stimulus. The stimuli were aversive (loud) or benign (soft) bursts of white noise. A complete description of the anticipation task and its findings using standard (GLM) analyses were previously published [26].

For the BL-FACES task, the Ekman faces [27] were presented in 20-second blocks of nine faces, each block containing either angry, fearful, neutral, or happy faces. The sequence of blocks was counterbalanced for order across participants. This design was equivalent to one used in our previous studies [18,23]. The ER-FACES task required more stimuli for a comparable length of time, so we used a subset of Karolinska faces [28], with fearful and happy emotional expressions (26 individuals, 13 female). The images were cropped along the hairline and converted to grayscale. The design was generated using OptSeq [29] using 2–10 second inter-stimulus intervals.

2.4. Complexity Estimation and Normalization to Hurst Exponent

The generalized Hurst exponent was computed for time series $x(t)$ for $q = 1$ and $q = 2$ according to the structure function: $S(q,\tau) = \langle |x(t+\tau) - x(t)|^q \rangle \propto \tau^{qH(q,\tau)}$ [30]. For time series of length L, we used ten logarithmically spaced lags τ ranging between 1 and $log_{10}(L/50)$. The estimates of $H(q)$, denoted by H_{Q1a} and H_{Q2a}, were estimated by the slope of line fitting $\log(S(q,\tau))/q$ vs. $\log(\tau)$.

The generalized Hurst exponent was also estimated by a different implementation [31] of the same relation in accordance to [32–34]. Again, the Hurst parameter was estimated for $q = 1$ and $q = 2$ (H_{Q1b}/H_{Q2b}), with the default maximum lag of 19.

An alternative method for estimating Hurst exponent is the aggregated variance technique [35], available through Matlab File Exchange [36,37]. A series of length N is divided into N/m blocks of length m, and sample variance of each block $X^{(m)}$ is computed. For series of finite variance, the sample variance $var(X^{(m)})$ will be asymptotically proportional to m^{2H-2} for large N/m and m. The Hurst exponent H_{AV} is estimated by computing the slope $2H_{AV}$-2 of the graph of sample variance vs. m.

Another closely related quantity is Higuchi fractal dimension [38], which estimates the fractal dimension D of a time series curve from segments $L(k)$ constructed from k samples, with $L(k) \propto k^{-D}$, where D is related to Hurst estimate H_{HFD} through $D = 2-HFD$

[39]. We used $k = 5$ and $k = 10$ samples, denoted by H_{HFD-5} and H_{HFD-10}, respectively. A Matlab implementation of the algorithm [40] can be found through Matlab File Exchange [41].

The Matlab Wavelet Toolbox [42] also provides several estimates of the Hurst parameter. The first two estimates are based on the discrete second-order derivative, using an FIR and a wavelet filter, denoted H_{DD} and H_{DDW}, respectively [43]. The third estimate is obtained from the slope of the variance of local detail coefficients vs. scale [39]. For the last estimate, in addition to the default Haar wavelet, we used higher order Daubechies wavelets of scales 2, 4, 8, and 16, denoted by H_{db1}, H_{db2}, H_{db4}, H_{db8}, and H_{db16}.

The Hurst parameter can, under certain assumptions [44], be computed through the relation $H = (\beta - 1)/2$, where β is the negative slope of the power spectral density of the time series $PSD(x)$ vs. frequency f on a log-log scale ($PSD(x) \propto f^{-\beta}$). We used two methods to estimate the power spectrum; one employing the Matlab fast Fourier transform ($PSD(x) = |FFT(x)|^2$) and the other used Welch's periodogram with eight windows of 50% overlap. The corresponding values of the Hurst exponent calculated from β estimates are denoted H_{FFT} and H_{pWelch}, respectively.

The estimation of the Hurst parameter via Detrended Fluctuation Analysis (H_{DFA}) was performed in accordance with [11]. First, each N-point long time series $x(t)$, $t = 1..N$, is integrated $y = \sum_1^N (x(t) - \overline{x})$ and a set of window – or box – lengths is defined. We chose box sizes on a log-scale between 4 and $N/4$ points such that the ratio between successive box sizes was $2^{1/8}$. For each box size, time series segments of that many points are created by a sliding window method. A linear trend is fit and subtracted from each segment and the root-mean-square of the residuals is computed. More formally, for each window length n and set of k segments $y(k)$, the root-mean-square fluctuation is $F(n) = \sqrt{\frac{1}{n} \sum_1^n [y(k) - y_{fit}(k)]^2}$, where $y_{fit}(k)$ is the linear fit to the segment $y(k)$. This calculation is repeated for each box size n and the slope of the log-log scale fit of $F(n)$ vs. box size yields the scaling exponent α, where $H = \alpha - 1$ [2]. Since the DFA estimate is affected by the range of box sizes, we also computed DFA over small and large box sizes (H_{DFA-S}/H_{DFA-L}) by fitting the upper and lower parts of $F(n)$ vs. box size separately. For example, for our longest 1190-point time series, small boxes ranged 4–40 points and large ranged 40–300 points).

H via Rescaled Range was estimated using the relationship $E[R(n)/S(n)] \propto Cn^H$, where $S(n)$ is the standard deviation and $R(n)$ is the range of cumulative sums of zero-meaned n-size segments of time series X_i, $i = 1..L$. The recipe for computing is as follows. (1) Define a range of scales: $L, L/2, L/4, ... L/k$ such that $L/k >= 8$. (2) For each scale, define n-length segments of time series (e.g. one segment for scale L, two for $L/2$, etc.). (3) For each segment of length n, subtract the mean and compute cumulative sum $Y_m = \sum_1^m (X_i - \overline{X})$, with $m = 1..n$. (4) Calculate $R(n) = \max(Y_1..Y_n) - \min(Y_1..Y_n)$ over standard deviation $S(n)$. (5) Find $E[R/S]$ by averaging R/S over all segments for a particular scale. (6) Estimate Hurst exponent H_{RS} as the slope of $\log_2(E[R/S])$ vs. $\log_2$(scale) (e.g. $|\log_2(L)|$, $|\log_2(L/2)|$, etc.).

In order to test the consistency of measures, we generated 1000 time series with the *wfbm* function for wavelet-based generation of fractional Brownian motion with $H \in [0,1]$ (Matlab's Wavelet Toolbox) and compared the estimated values of H with the generated ones. Robustness of each measure was assessed through introducing spikes by changing values of a random subset of points within each of 1000 time series generated with $H = 0.5$. The magnitude of spikes ranged from zero to six standard deviations of the simulated time series. Spikes were both positive and negative in

magnitude. The number of spikes varied from zero to five percent of the time series length. The susceptibility to spikes was estimated as a t-statistic describing the change in estimated value of the Hurst exponent as compared to the estimate of H from time series without spikes.

2.5. Correlation between Measures

Since all of the methods attempt to estimate the same quantity, we assessed their consistency by applying the measures to the same time series. Measure precision – the consistency of estimating a particular Hurst exponent – was estimated by correlating Hurst estimates for different measures applied to the time series with $H = 0.5$ ($N = 1000$). Measure accuracy – the ability to estimate a range of Hurst exponents accurately – was assessed by correlating Hurst estimates for different measures across 1000 time series with Hurst exponent ranging from zero to one.

2.6. Error Drop-off

The error of the estimates σ decreases as the length of the time series L increases, which can be expressed by the relationship $\sigma \propto L^{-\alpha}$, where α is the error drop-off rate. α was computed for "brown noise" time series of ten lengths ranging from 100 to 10000 points over 100 instances.

2.7. Signal Processing

Various signal processing techniques are applied to fMRI time series in order to improve signal to noise ratio. For example, it is commonly accepted that high-pass filtering and inclusion of motion parameters in the regression model is appropriate for most GLM analyses. Other techniques involve inclusion of higher order motion terms to account for nonlinearity of magnetic field distortion. In resting state connectivity literature, band-pass filtering is often applied to exclude fluctuations beyond a certain frequency, since they are believed to be non-physiological. Inter-subject correlation analyses suggest removal of global signal to avoid inflation of correlation values [45]. However, to the best of our knowledge, no comprehensive comparison of effects of signal processing on complexity of fMRI time series has been performed to date.

The types of corrections applied to the time series were detrending (denoted by 'd'), regressing out motion parameters (denoted by 'm') and their squares (denoted by '2'), high-pass filtering at 0.01 Hz (denoted by 'fhi'), band-pass filtering in the 0.01–0.1 Hz range (denoted by 'fbp'), and regression of the global signal out of the time series (denoted by 'bcw'), which consisted of three regressors: mean time series of the whole brain, cerebrospinal fluid, and white matter. The motion parameters were derived from SPM's realignment procedure, the filtering was performed using 10th order Butterworth filter, and global regressors were obtained using canonical masks for brain, white matter, and cerebrospinal fluid included in the SPM8. We considered all reasonable combinations of corrections, excluding higher order motion correction '2' without the linear term and redundancy of 'fhi-fbp' combination, which resulted in 36 combinations that included the original untouched time series (denoted by 'o').

The success of the correction was assessed in two ways: by comparison of values Hurst estimates within grey and white matter voxels and by strength of relation to activation measured as the area under the curve (AUC) of ROC curves constructed using SPM's t-map of activation in response to *task>rest* contrast. The true positives were voxels with top 1% of the *t*-values – a criterion ensuring a constant ratio of true to false positives across subjects, regardless of strength of the individual response to the task. As an alternative measure of activation to the task we also computed

cross-subject, or inter-subject, correlations (ISC) that showed which time series co-varied across subjects throughout the task. The cross-subject correlations were computed via pairwise correlations between subjects' time series (for each voxel, after detrending, regressing out the motion and global signal, and high-pass filtering), converting correlations to z-scores using Fisher's transformation, and averaging across subjects [46]. The ISC, designed to pick up activation to GRST-LOST, allowed us to compare this task to BL-FACES and ER-FACES.

Results

3.1. Measure Error and Susceptibility to Spikes (Simulated Data)

As **Figure 1** illustrates, most measures return the same values of H that were used to generate the time series. The few exceptions are H_{AV}, H_{RS}, and power spectrum based measures, H_{FFT} and H_{pWelch}. For H_{AV} and H_{RS}, a derivative needs to be applied to transform fractional Brownian motion (fBm) to fractional Gaussian noise (fGn). H_{FFT} and H_{pWelch} are not intrinsically bounded at 0.5 since there is no natural upper limit on the slope of the log-power-spectrum: rather the upper bound of the spectrum exists due to the way wavelet-based fBm time series are generated. We therefore

used the derivative offset by one to obtain H estimates for simulated data.

The Hurst estimates of the integral and the derivative of the time series show that the majority of the measures are bounded from above and below, which means that H-estimates are artificially squeezed within the theoretically valid range for the Hurst exponent – a potential loss of sensitivity in case experimental data exhibits unusually high or low complexity. The unbounded measures – H_{FFT}, H_{pWelch}, H_{db*}(except H_{db1}), H_{DD}, H_{DDW} – show an approximately constant offset as a result of time series integration and differentiation. The average error on the H-estimates was on the order of ~0.05 and consistent over the [0,1] range, with H_{DFA-L} and H_{db*} showing the largest errors.

Figure 2 illustrates the susceptibility of the measures to spikes as a function of their number and magnitude. As expected, all measures gravitated toward "white noise" values as the number of spikes and their magnitude increases. H_{Q2a}, the second derivative estimates (H_{DD}, H_{DDW}) and FFT-based estimates (H_{FFT}, H_{pWelch}) were most sensitive to small number of spikes, reaching an upper bound quicker than other measures. H_{Q*b} and H_{RS} showed an almost uniform increase regardless of spike amplitude. H_{db*} estimates were least susceptible to spikes, but low-order H_{db*} showed slight decreases in H as the number and magnitude of spikes increased.
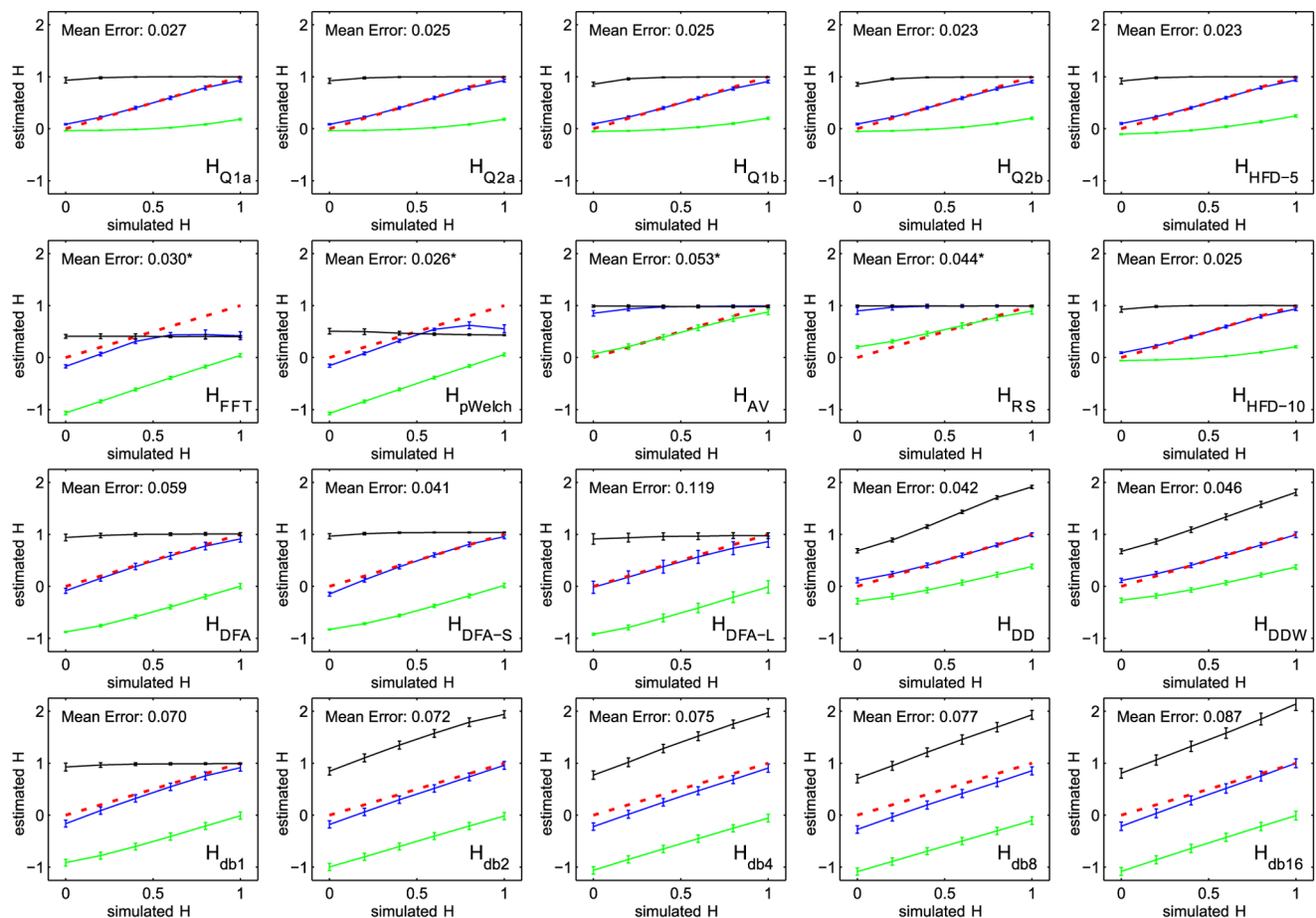
Figure 1. Estimates of the Hurst exponent are consistent and monotone for simulated wavelet-based fractional Brownian motion time series with Hurst exponent ranging from zero to one. Along with estimates for the simulated time series (blue), estimates for the time series integral (black) and derivative (green) are also shown. The red dotted line represents the theoretical value. Mean error is calculated as the average of the errors at each point; starred errors were computed for the derivative.
doi:10.1371/journal.pone.0063448.g001
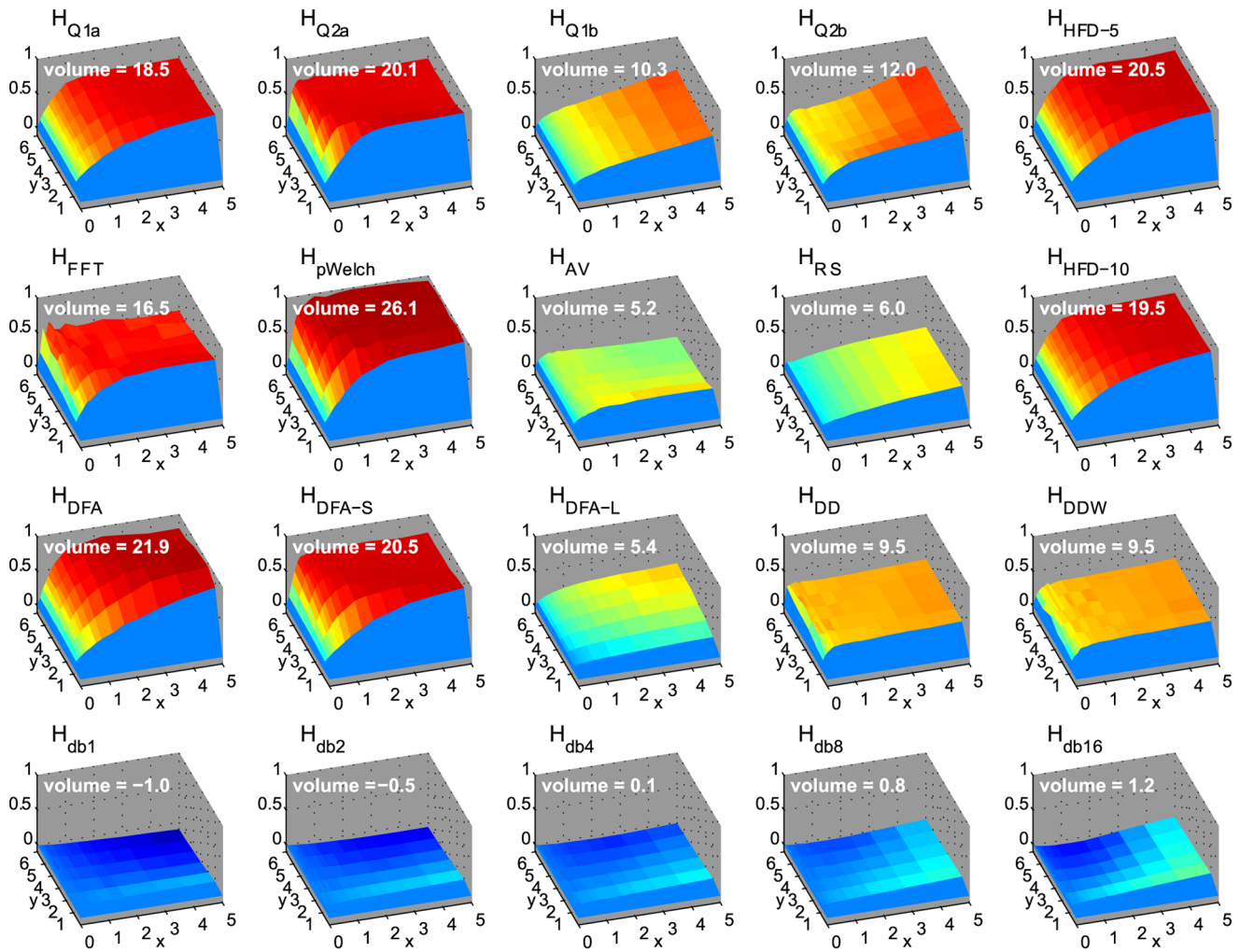
Figure 2. Number of spikes (as a percentage of points in a time series, x-axis) and their magnitude (in units of standard deviation of original time series, y-axis) affect Hurst estimates differently. The z-axis represents normalized to [0, 1] range one-sample t-test differences from the H-estimates of time series without spikes (each point on the surface is the t-value difference in H resulting from the introduction of spikes to the time series; the set of all t-scores for all measures was linearly transformed to [0, 1] range to show relative change due to presence of spikes). Volume is calculated as the difference between the surface and the plane defined by H-values of unaltered time series (spike magnitude of zero).
doi:10.1371/journal.pone.0063448.g002

3.2. Correlation between Measures (Simulated Data)

The correlations between Hurst estimates of the same 1000 time series with H values distributed uniformly between 0 and 1 were very high, indicating reliability (**Figure 3**, values above diagonal). Among the least correlated measures were H_{db*}, H_{DFA}, H_{DFA-L}, H_{AV}, and H_{RS}. Although most measures produced accurate estimates of the Hurst exponent, the correlation between the measures computed over the same data with a specific exponent were significantly smaller (**Figure 3**, values below diagonal show the correlations for $H = 0.5$). The highest correlations occurred between H_{Q*}, H_{FFT}, H_{pWelch}, H_{HFD*}, and H_{DFA-S}, and between H_{AV}, H_{RS}, and H_{DFA-L}. H_{DD} and H_{DDW} were mostly correlated to each other and H_{db*} were only weakly correlated to each other and to other measures. These groupings reflect similarities in computation between various measures despite the markedly different algorithms used to compute the quantity of interest.

3.3. Decrease in Estimation Error with Time-series Length (Simulated Data)

The error drop-off rate α, which is the rate of decrease in variance of the H-estimates as the length of the time series increases, is reported in **Figure 4**. For most measures, α is ~0.5, which indicates a standard one over square root relationship, except for H_{Q*a}, H_{DFA*}, H_{AV}, and H_{RS}, which are less efficient (**Figure 5**).

The computation times for time series of varying length are shown in **Figure 4** as well. Although computation times depend on length and parallelizability of the implemented algorithms, the measures can roughly be separated into two categories – the ones whose compute times were within an order of magnitude of the fastest one and the ones that were not. The most efficient ones were H_{Q*a}, H_{DD*}, H_{HFD*}, H_{FFT}, H_{AV}, and H_{RS}. The time to compute Hurst estimate for an entire brain for a single person varied from minutes to hours depending on the choice of measure, with the overall computation time of 1.5–2 hours per person for all measures combined. Most of that time was taken up by the slower

	H_{Q1a}	H_{Q2a}	H_{Q1b}	H_{Q2b}	H_{FFT}	H_{pWelch}	H_{DD}	H_{DDW}	H_{db1}	H_{db2}	H_{db4}	H_{db8}	H_{db16}	H_{HFD-5}	H_{HFD-10}	H_{DFA}	H_{DFA-S}	H_{DFA-L}	H_{AV}	H_{RS}	
H_{Q1a}	1	1.000	0.997	0.997	0.993	0.994	0.979	0.976	0.983	0.978	0.973	0.973	0.970	0.997	0.999	0.982	0.993	0.912	0.981	0.981	H_{Q1a}
H_{Q2a}	0.964	1	0.997	0.998	0.993	0.995	0.980	0.977	0.982	0.978	0.973	0.974	0.971	0.997	0.999	0.982	0.994	0.912	0.981	0.981	H_{Q2a}
H_{Q1b}	0.897	0.871	1	0.999	0.994	0.995	0.979	0.977	0.979	0.979	0.973	0.974	0.972	0.996	0.998	0.977	0.991	0.905	0.982	0.981	H_{Q1b}
H_{Q2b}	0.875	0.910	0.954	1	0.994	0.995	0.980	0.977	0.979	0.978	0.973	0.975	0.973	0.996	0.998	0.977	0.992	0.905	0.982	0.981	H_{Q2b}
H_{FFT}	0.526	0.551	0.568	0.600	1	0.998	0.985	0.983	0.977	0.978	0.973	0.974	0.974	0.994	0.993	0.977	0.990	0.908	0.978	0.976	H_{FFT}
H_{pWelch}	0.623	0.655	0.687	0.716	0.724	1	0.987	0.984	0.978	0.979	0.974	0.975	0.974	0.996	0.995	0.978	0.991	0.906	0.977	0.973	H_{pWelch}
H_{DD}	-0.051	-0.053	-0.003	-0.011	0.360	0.461	1	0.997	0.960	0.962	0.954	0.957	0.959	0.984	0.981	0.956	0.967	0.889	0.963	0.961	H_{DD}
H_{DDW}	-0.076	-0.080	-0.034	-0.045	0.316	0.427	0.904	1	0.956	0.958	0.952	0.953	0.956	0.982	0.978	0.954	0.964	0.887	0.960	0.959	H_{DDW}
H_{db1}	0.500	0.438	0.380	0.343	0.245	0.246	-0.004	-0.005	1	0.964	0.964	0.964	0.954	0.978	0.981	0.976	0.984	0.908	0.970	0.968	H_{db1}
H_{db2}	0.447	0.438	0.414	0.398	0.261	0.310	0.013	0.005	0.276	1	0.964	0.966	0.958	0.976	0.977	0.959	0.977	0.877	0.956	0.955	H_{db2}
H_{db4}	0.444	0.442	0.398	0.404	0.261	0.293	-0.009	-0.021	0.296	0.238	1	0.979	0.957	0.971	0.972	0.953	0.971	0.872	0.954	0.949	H_{db4}
H_{db8}	0.434	0.446	0.419	0.439	0.252	0.307	-0.019	-0.021	0.219	0.280	0.624	1	0.960	0.972	0.973	0.954	0.973	0.874	0.955	0.951	H_{db8}
H_{db16}	0.247	0.251	0.288	0.306	0.219	0.215	0.053	0.056	0.095	0.113	0.201	0.247	1	0.971	0.972	0.947	0.966	0.876	0.952	0.947	H_{db16}
H_{HFD-5}	0.617	0.619	0.731	0.714	0.651	0.792	0.212	0.211	0.226	0.280	0.294	0.305	0.300	1	0.999	0.976	0.988	0.908	0.978	0.977	H_{HFD-5}
H_{HFD-10}	0.899	0.873	0.922	0.891	0.585	0.713	-0.004	-0.050	0.369	0.379	0.387	0.395	0.313	0.802	1	0.978	0.991	0.909	0.979	0.978	H_{HFD-10}
H_{DFA}	0.489	0.497	0.320	0.324	0.265	0.225	0.011	-0.001	0.401	0.143	0.098	0.045	-0.059	0.152	0.296	1	0.987	0.963	0.980	0.980	H_{DFA}
H_{DFA-S}	0.932	0.963	0.783	0.815	0.424	0.505	-0.124	-0.156	0.444	0.414	0.404	0.398	0.157	0.421	0.764	0.559	1	0.913	0.978	0.976	H_{DFA-S}
H_{DFA-L}	0.093	0.090	0.008	0.005	0.109	0.014	0.060	0.053	0.195	-0.102	-0.115	-0.173	-0.072	0.016	0.017	0.841	0.116	1	0.929	0.931	H_{DFA-L}
H_{AV}	0.457	0.474	0.353	0.363	0.308	0.255	0.017	-0.005	0.375	0.125	0.140	0.108	0.015	0.207	0.326	0.695	0.499	0.524	1	0.984	H_{AV}
H_{RS}	0.432	0.435	0.325	0.331	0.292	0.165	-0.016	-0.021	0.374	0.142	0.115	0.106	-0.011	0.145	0.273	0.759	0.472	0.610	0.711	1	H_{RS}
	H_{Q1a}	H_{Q2a}	H_{Q1b}	H_{Q2b}	H_{FFT}	H_{pWelch}	H_{DD}	H_{DDW}	H_{db1}	H_{db2}	H_{db4}	H_{db8}	H_{db16}	H_{HFD-5}	H_{HFD-10}	H_{DFA}	H_{DFA-S}	H_{DFA-L}	H_{AV}	H_{RS}	

Figure 3. Correlations between Hurst estimates of simulated time series. Correlations between Hurst parameter estimates show that they are consistent but not precise. *Below Diagonal*: correlations between the measures computed over the same 1000 time series with $H = 0.5$ show that for most measures, errors on estimating the same exponent are not strongly correlated (light green to white to blue) with the exception of a few positively correlated (dark green) measures. *Above Diagonal*: correlations between Hurst estimates of the same 1000 time series with H values distributed uniformly between 0 and 1 show that measure estimates are consistent across a range of exponents, with some measures showing stronger correlation (green) than others (yellow).
doi:10.1371/journal.pone.0063448.g003

measures (H_{Q*b}, H_{db*}, H_{pWelch}, and H_{DFA*}), which renders these measures less practical for exploratory fMRI analysis.

3.4. Effect of Signal Processing: Detrending; Regressing Out Motion Parameters and Global Mean; Filtering (fMRI Data)

We evaluated the effects of signal processing on Hurst estimates in terms of improvement of sensitivity to tissue type (assessed by grey-to-white matter contrast) and activation (assessed by AUC of ROC curves). Since the types of signal processing applied to the time series are not independent, we looked at the effect of adding a particular type of processing on top of all others. For example, to evaluate the effect of detrending, we compared all types of processing that involved linear detrending (12 types) to the ones that did not (24 types). Similarly, to evaluate the effect of high-pass filtering, we compared all types that involved high-pass filtering (12 types) to ones that were not high- or band-pass filtered (12 types). The results of comparisons are summarized in **Figure 6**.

In terms of increasing sensitivity to activation for the FACES tasks, for both t-contrast based and ISC based AUC values were most improved by high-pass filtering, which was significantly better than band-pass filtering for event-related design and only marginally worse than band-pass filtering for blocked design, suggesting that activation is best captured by dynamics above a certain frequency. For GRST-LOST, we used a five-minute segment three minutes into the show (GRST-LOST5) in order to make comparisons to other, shorter, tasks meaningful. For GRST-LOST5, both detrending, regressing out the global mean, and filtering increased AUC, but high-pass filtering still had the greatest effect. The agreement between t-value and ISC based results was not surprising, as there was over 60% spatial overlap for the top 3% of the values at the group level (66.5% for BL-

FACES, 63.2% for ER-FACES). High-pass filtering remained the dominant effect for these three tasks as well as for RST task in terms of grey-white matter contrast, although both detrending and regressing out motion and global signal helped too. Second-order motion correction did not improve any of the metrics except GRST-LOST5, where it had the weakest effect.

In order to establish the best combination of processing steps, we computed contrasts for every combination vs. unprocessed estimates. For GM>WM contrast the clear winner for BL-FACES, ER-FACES, GSRT-LOST5, and RST tasks, was the *d-m2-bcw-fbp* combination (*Wald* $\chi^2 = 251.5$, 208.4, 121.9, 152.3, respectively, $df = 1$, $p \ll 0.0001$), followed by the *d-m-bcw-fbp* combination (*Wald* $\chi^2 = 160.0$, 122.6, 74.8, 92.9, respectively, $df = 1$, $p \ll 0.0001$) and the *d-m-bcw-fhi* combination (*Wald* $\chi^2 = 119.4$, 107.1, 91.5, 87.8, respectively, $df = 1$, $p \ll 0.0001$). The contrasts were consistent for all tasks, with minimum correlation across processing combinations of 0.91. The combinations that gave the biggest improvements in the AUC-T and AUC-ISC values over the unprocessed time series involved high-pass and not band-pass filtering for ER-FACES and GRST-LOST5; the winner combination was *d-m2-bcw-fhi* (AUC-ISC *Wald* $\chi^2 = 29.5$, 63.6, respectively, $df = 1$, $p \ll 0.001$) followed closely by *d-m-bcw-fhi* (AUC-ISC *Wald* $\chi^2 = 26.9$, 60.8, respectively, $df = 1$, $p \ll 0.001$). For BL-FACES, unlike other tasks, including motion regressors seemed to hinder agreement between complexity and activation, and the best combination involved only detrending, regressing out global mean, and high-pass filtering (*d-bcw-fhi* AUC-T, AUC-ISC *Wald* $\chi^2 = 30.2$, 34.1, respectively, $df = 1$, $p \ll 0.001$).

Since the second order motion correction did not significantly improve AUC or GM>WM contrast, we chose to focus on the *d-m-bcw-fhi* combination for rating the measures. We ranked the performance of each measure in terms of sensitivity to activation

	Volume	Alpha	t256	t1024	t4096	t16384
H_{Q1a}	18.49	0.36	0.07	0.46	2.03	11.51
H_{Q2a}	20.14	0.34				
H_{Q1b}	10.32	0.51	4.15	4.52	6.17	13.45
H_{Q2b}	11.98	0.51				
H_{FFT}	16.53	0.53	0.02	0.09	0.36	1.50
H_{pWelch}	26.14	0.51	0.69	0.67	1.01	3.49
H_{DD}	9.52	0.51	0.18	0.24	0.71	3.08
H_{DDW}	9.49	0.52				
H_{db1}	-1.01	0.48	0.49	0.56	0.88	3.36
H_{db2}	-0.48	0.47	0.50	0.57	0.86	3.39
H_{db4}	0.09	0.52	0.49	0.58	0.87	3.16
H_{db8}	0.81	0.58	0.48	0.60	0.87	2.96
H_{db16}	1.15	0.58	1.14	1.25	1.61	3.49
H_{HFD-5}	20.51	0.51	0.21	0.26	0.68	2.66
H_{HFD-10}	19.47	0.51	0.19	0.25	0.61	2.88
H_{DFA}	21.89	0.29	0.86	3.20	25.53	156.88
H_{DFA-S}	20.50	0.44				
H_{DFA-L}	5.36	0.19				
H_{AV}	5.20	0.29	0.28	0.34	0.68	2.77
H_{RS}	5.98	0.32	0.05	0.29	1.50	8.77

Figure 4. Comparison of Hurst estimates of simulated data. Volume under the surface of **Figure 2**, error drop-off rate alpha of **Figure 5**, and computation time (in seconds) for 1000 time series of length 256, 1024, 4096, and 16384 points. The computations, optimized using MATLAB Parallel Computing Toolbox, were performed on a quad-core Intel i7-960 @ 3.6 GHz.
doi:10.1371/journal.pone.0063448.g004

measured by AUC and to tissue type measured by GM>WM contrast (**Figure 7**). AUC-T and AUC-ISC ranks were averaged for the FACES tasks to avoid double-counting. For sensitivity to activation, in order from best to worst, the Hurst estimates ranked: H_{pWelch}, H_{FFT}, H_{HFD10}, H_{Q2a}, H_{HFD5}, H_{Q1a}, H_{Q2b}, H_{DFA-S}, H_{Q1b}, H_{db8}, H_{DFA}, H_{db4}, H_{DD}, H_{DDW}, H_{db16}, H_{RS}, H_{db2}, H_{AV}, H_{db1}, H_{DFA-L}. Similarly, for sensitivity to tissue type, the estimates ranked: H_{pWelch}, H_{FFT}, H_{HFD10}, H_{DFA-S}, H_{Q2b}, H_{DFA}, H_{Q1b}, H_{HFD5}, H_{RS}, H_{Q2a}, H_{db4}, H_{Q1a}, H_{DFA-L}, H_{AV}, H_{db2}, H_{db8}, H_{db1}, H_{db16}, H_{DDW}, H_{DD}.

The performance of each of the measures is illustrated in **Figure 8**, which shows a close relation between activation and complexity in the BL-FACES task. The ROC curves constructed using the top 1% of GLM t- or ISC r-values show strong agreement for H_{Q*}, H_{FFT}, H_{pWelch}, H_{HFD*}, H_{DFA}, H_{DFA-S}, H_{db8} and H_{db16}. The AUC values are the same as those presented in **Figure 7**. The middle columns of the left and right panel of the figure qualitatively illustrate the distribution of complexity values in the brain, with H_{pWelch} and H_{FFT} showing the most discernible edges of the brain. The right columns show a cross-section through the middle slice of the top 10% of the values of the corresponding image. The top row shows that the bulk of the activation to the

task>rest contrast resides in the visual cortex and that the inter-subject correlation also picked up the co-activation in the visual cortex across subjects, though in this case high r-values extended further into the midbrain. Most of the Hurst estimates picked out the visual and the prefrontal areas; in fact, the complexity measures accentuated these areas much more strongly than the activation maps themselves. For the complexity measures, these cross-sections also explain the AUC values, especially the low ones: some measures such as H_{AV} and H_{DFA-L} fail to adequately cover the visual areas and instead show higher H-values in the prefrontal regions, some (H_{db*}) are grainy, and others (H_{DDW}, H_{DD}, H_{RS}) fall somewhere in between.

3.5. Relation to Emotional Content (Arousal and Valence)

Not only did the complexity measures show high sensitivity to activation in tasks in their entirety, but the complexity also fluctuated with the emotional content of the LOST episode. We used the inter-subject correlation across the entire GRST-LOST task to identify regions engaged in passive viewing of the TV-episode, which, unsurprisingly, were well-associated with audiovisual processing (**Figure 9**). We then extracted the time series from
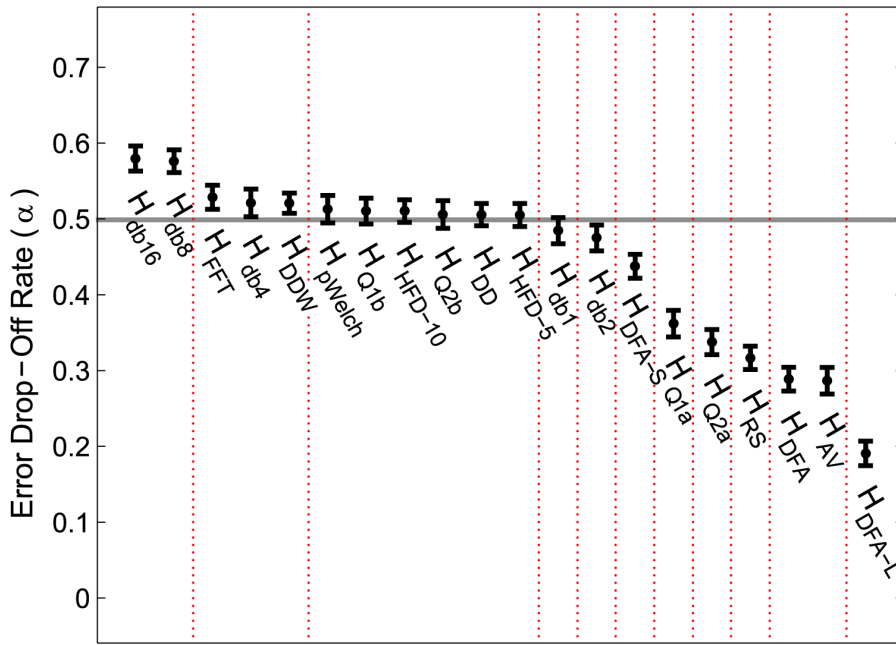
Figure 5. The complexity measures sorted by the average error drop-off rate α mostly follow the standard one over square root relationship (α = 0.5). The error drop-off rate was computed for "brown noise" time series of ten lengths ranging from 100 to 10000 points over 100 instances. The dotted vertical lines indicate which pairs of neighboring measures are significantly different at $p<0.001$ (uncorrected).
doi:10.1371/journal.pone.0063448.g005

AUC-T	BL-FACES			ER-FACES		
	F	p	Δ mean	F	p	Δ mean
d vs. all	3.92	0.048	-0.004	1.85	0.174	0.002
bcw vs. all	5.56	0.018	-0.005	4.33	0.037	0.004
m vs. all-m2	1.81	0.179	-0.003	0.87	0.350	0.002
m vs. m2	1.87	0.171	0.003	1.30	0.254	-0.002
fhi vs. all-fbp	225.15	0.000	0.037	255.69	0.000	0.035
fbp vs. all-fhi	318.31	0.000	0.041	30.32	0.000	0.011
fbp vs. fhi	6.23	0.013	0.005	125.15	0.000	-0.024

Legend:	
d	detrending
bcw	regressing out global mean
m	regressing out motion parameters
m2	regressing out motion parameters and their squares
fhi	high-pass filtering
fbp	band-pass filtering

AUC-ISC	BL-FACES			ER-FACES			GRST-LOST (5 min)		
	F	p	Δ mean	F	p	Δ mean	F	p	Δ mean
d vs. all	2.00	0.158	-0.003	0.50	0.478	0.001	5.49	0.019	0.005
bcw vs. all	2.30	0.130	-0.003	1.55	0.213	0.002	11.87	0.001	0.007
m vs. all-m2	0.85	0.355	-0.002	0.05	0.822	0.001	2.06	0.152	0.004
m vs. m2	1.19	0.274	0.003	0.37	0.542	-0.001	2.35	0.126	-0.004
fhi vs. all-fbp	219.89	0.000	0.037	201.42	0.000	0.033	258.36	0.000	0.041
fbp vs. all-fhi	298.02	0.000	0.042	5.25	0.022	0.005	127.47	0.000	0.026
fbp vs. fhi	4.25	0.039	0.004	152.82	0.000	-0.028	45.56	0.000	-0.016

GM>WM	BL-FACES			ER-FACES			GRST-LOST (5 min)			RST		
	F	p	Δ mean	F	p	Δ mean	F	p	Δ mean	F	p	Δ mean
d vs. all	10.13	0.001	1.580	10.19	0.001	1.394	14.98	0.000	2.154	14.04	0.000	2.044
bcw vs. all	16.22	0.000	1.999	15.55	0.000	1.722	24.54	0.000	2.754	22.47	0.000	2.584
m vs. all-m2	5.83	0.016	1.445	4.80	0.028	1.151	7.22	0.007	1.783	6.66	0.010	1.675
m vs. m2	1.96	0.161	-0.860	2.16	0.142	-0.798	3.71	0.054	-1.338	3.47	0.063	-1.270
fhi vs. all-fbp	409.93	0.000	12.543	434.21	0.000	11.406	376.43	0.000	13.502	345.16	0.000	12.653
fbp vs. all-fhi	499.02	0.000	12.843	465.04	0.000	10.771	272.11	0.000	9.613	328.81	0.000	10.818
fbp vs. fhi	0.30	0.581	0.299	1.75	0.186	-0.635	38.60	0.000	-3.890	9.46	0.002	-1.835

Figure 6. Comparison of combinations of signal processing. Results of an ANOVA of processing by measure (20) by subject (22 for FACES, 11 for GRST-LOST) show significant improvements in sensitivity and grey-white matter contrast of Hurst estimates (df = 1 for each comparison). The AUC-T values were derived from ROC curves constructed using top 1% of t-values of task vs. rest contrast. The AUC-ISC values were derived from ROC curves constructed using top 1% of z-transformed inter-subject correlations (ISC). The GM>WM values were derived from differences in Hurst estimates for grey and white matter voxels. Positive results significant at p<0.05 are highlighted.
doi:10.1371/journal.pone.0063448.g006

	AUC-T		AUC-ISC		GRST-LOST	GM>WM		GRST-LOST	RST
	BL-FACES	ER-FACES	BL-FACES	ER-FACES		BL-FACES	ER-FACES		
H_{Q1a}	0.97	0.67	0.95	0.63	0.59	98.95	72.99	28.97	27.02
H_{Q2a}	0.97	0.67	0.95	0.63	0.60	99.62	72.88	32.92	30.82
H_{Q1b}	0.96	0.56	0.94	0.50	0.62	109.75	90.09	67.10	65.25
H_{Q2b}	0.96	0.57	0.94	0.51	0.63	112.61	93.57	72.50	70.69
H_{FFT}	0.93	0.81	0.91	0.79	0.90	133.69	122.38	127.22	119.11
H_{pWelch}	0.96	0.83	0.94	0.80	0.92	140.10	128.75	136.63	127.10
H_{DD}	0.60	0.61	0.59	0.66	0.53	14.78	7.47	-4.16	-6.68
H_{DDW}	0.60	0.65	0.59	0.68	0.53	23.14	14.08	0.97	-1.16
H_{db1}	0.57	0.52	0.57	0.49	0.55	39.91	50.40	20.55	18.91
H_{db2}	0.64	0.52	0.62	0.48	0.58	49.38	41.73	32.94	33.96
H_{db4}	0.82	0.54	0.79	0.52	0.59	69.18	57.74	42.59	42.45
H_{db8}	0.95	0.64	0.93	0.62	0.59	81.98	44.33	23.60	22.71
H_{db16}	0.91	0.56	0.89	0.57	0.50	62.11	1.85	-2.37	-2.82
H_{HFD-5}	0.97	0.66	0.95	0.62	0.61	97.92	75.10	47.59	45.55
H_{HFD-10}	0.97	0.56	0.96	0.51	0.64	109.41	94.03	77.69	75.87
H_{DFA}	0.89	0.49	0.87	0.44	0.64	101.97	91.31	82.40	80.30
H_{DFA-S}	0.97	0.55	0.95	0.49	0.63	113.55	95.19	71.65	69.64
H_{DFA-L}	0.25	0.45	0.25	0.43	0.58	23.18	47.14	48.51	47.38
H_{AV}	0.38	0.53	0.39	0.50	0.58	40.32	49.86	42.67	38.80
H_{RS}	0.50	0.50	0.49	0.47	0.60	65.87	47.70	50.03	48.60

Figure 7. Sensitivity of Hurst estimates to activation and tissue type. The presented measures of sensitivity of Hurst estimates to activation and tissue type show that power-spectrum based measures are the most sensitive. The AUC-T values were derived from ROC curves constructed using top 1% of t-values of *task>rest* contrast. The AUC-ISC values were derived from ROC curves constructed using top 1% of z-transformed inter-subject correlations (ISC). The GM>WM values were derived from differences in Hurst estimates for grey and white matter voxels. The signal processing pipeline applied to the time series was *d-m-bcw-fhi* (see text). The values presented are averages across subjects (FACES: N = 22, RST & GRST-LOST: N = 11).
doi:10.1371/journal.pone.0063448.g007

the top 5% of the values in the brain (r>0.15), subdivided them into fifteen overlapping 10-min segments, and computed both complexity and average arousal and valence ratings of each segment. Complexity was strongly associated with the change in the emotional content for most of the Hurst estimates, with a slightly stronger association with the valence (**Figure 10**). Although H_{AV}, H_{DFA-L}, and H_{db*} preserved the overall shape of the response, their relation to self-report measures was either marginally or non significant. The second-derivative based estimates H_{DD*} showed large variance in estimates across subjects, deeming these measures less robust.

In addition to reflecting emotional content, Hurst estimates also varied with the average inter-subject correlation computed over the same 10-min segments (**Figure 10**) within the areas identified as active (**Figure 9**). The majority of measures showed strong correlation between H and ISC, with H_{AV} and H_{db*} being least responsive in this context as well.

3.6. Optimal Task Length

One of the most often raised concerns regarding the computation of complexity measures is the number of points in a time series necessary for a robust estimate of the Hurst exponent. The unusually long for fMRI GRST-LOST task was designed, in part, to address this issue. **Figure 11** shows the increase in grey-to-white matter contrast and AUC for time series from five to forty minutes in length, starting from the beginning and the end of the task. For the majority of measures, both GM>WM contrast and AUC increase monotonically with the number of points. Most of the measures plateau past the 20-min mark, suggesting that increasing time series length past a certain point ceases to improve estimation of complexity. The largest yield in performance per length of task occurs around the 10-minute mark, which means that in order to properly estimate H, the fMRI task has to be at least that long.
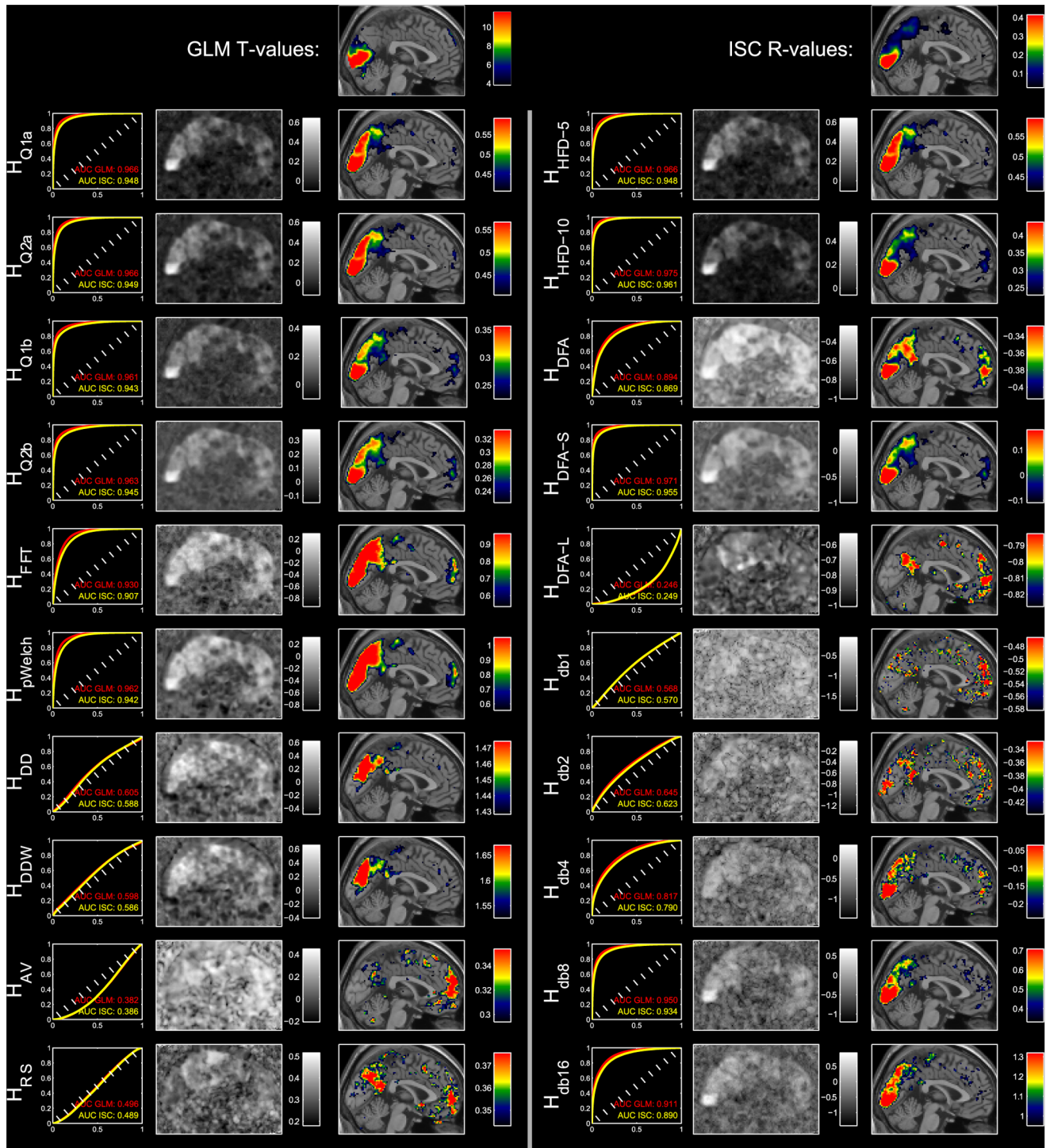
Figure 8. Hurst parameter estimates for the block faces task closely follow activation shown in the top row of the figure. The top left pane reflects activation (*t*-scores) for the contrast of *task>rest*, which in this case is presentation of emotionally valent faces vs. a fixation cross. The top right pane illustrates inter-subject correlations; scores presented as *r*-values for illustrative purposes. Rows below the first one show, for each measure: the ROC curves for agreement between complexity and activation with corresponding AUC values, the sagittal middle slice showing the distribution of *H*-values within and outside the brain, and the same slice overlaid with the heat map of highest *H*-values. All heat maps present the top 10% of the values in the entire image.
doi:10.1371/journal.pone.0063448.g008

3.7. Scanner Differences (fMRI data)

Given the increasing interest in shared, multi-institutional, data sets, as well as the development of neurodiagnostics, the degree to which complexity measures are sensitive to scanner differences is important. The BL-ANT task was run on two 3T scanners at the same institution – N = 22 on a Siemens and N = 12 on a Philips –
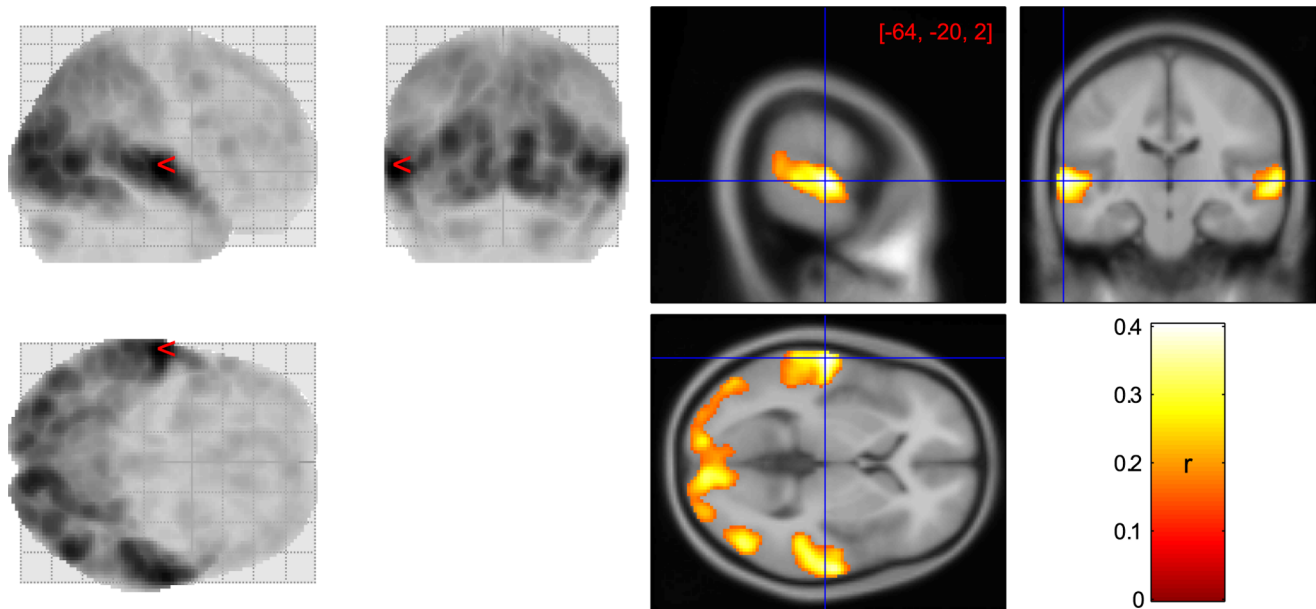
Figure 9. Inter-subject correlation across eleven participants for GRST-LOST task shows that most activation is restricted to areas involved in audiovisual processing. The glass brain (left) shows distribution of ISCs throughout the brain, while the heat map (right) outlines the areas deemed active at $r > 0.15$. Though the activation appears symmetric, the maximally activated voxel is located on the left (MNI $= -64, -20, 2$).
doi:10.1371/journal.pone.0063448.g009

using the same acquisition parameters. Activation areas and levels were equivalent across scanners [26]. Although we used independent populations for the different scanners, the groups were equivalent in terms of age ($t = 1.98$, $p = 0.06$, $df = 32$, mean age diff $= 4.5$ yrs), sex ($t = 1.17$, $p = 0.25$, $df = 32$), handedness ($t = 0.64$, $p = 0.53$, $df = 32$), or trait anxiety ($t = 0.62$, $p = 0.54$, $df = 32$). Despite this, principal component analysis (PCA) of Hurst exponents clearly separated the two groups across the first principal component (**Figure 12**, x-axis). The differences were highly significant for all estimates except H_{AV}. The principal components contained information about Hurst estimates derived from average time series from every region of the WFU pickAtlas [47,48] for each participant.

Discussion

4.1. Relation to Activation

One of the claims made in the past was that the fractal nature of the fMRI signal is connected to activation [20], but the question remains as to whether results obtained from complexity versus statistical measures are different only in sensitivity or also in kind. Task designs are usually optimized to elicit maximal activation as determined by applying a general linear model to the data. This means that regions of the brain that are active irrespective of the task are ignored (e.g. maintenance of homeostasis). Because scale-invariance measures require hundreds of points, deriving condition-specific assessments of scale-invariance in fMRI is currently not viable. This discrepancy in the approach makes conceptualizing the relationship between the two types of analysis nontrivial. However, there are several avenues we can take in order to make inferences about this relation.

We chose to validate against a simple paradigm that used well-established stimuli (i.e. Ekman and Karolinska faces), so that the activated brain areas would be previously established. For these tasks (BL-FACES and ER-FACES), the bulk of activity is associated with processing of a cohort of emotional faces picked

out by the *task>rest* contrast that compares it to a null condition that simply shows a fixation cross. To identify active voxels common to every subject, we correlated time series across pairs of subjects (**Figure 8**; results for ER-FACES very similar, not shown). These heat maps give different weighting to brain activity, stressing the voxels that fluctuated together more than those showing greatest difference in response to the stimuli.

The heat maps of the Hurst estimates accentuate activity in the visual cortex much more strongly than the activation maps themselves, indicating that they are, in a sense, more sensitive. This is consistent with the previous publications [19,20] that imaged four slices along the calcarine sulcus and showed that active voxels exhibit distinctly different behavior across nonlinear measures (such as Hurst exponent and its multifractal counterpart) even when applied to the residuals of GLM regression. However, the Hurst estimate maps extend beyond activation maps into the medial surfaces of precuneus, posterior parietal cortex, and posterior cingulate gyrus and also show a distinct "hot spot" in the medial prefrontal cortex. This indicates that although Hurst measures may be more sensitive to some aspects of brain activity, they certainly do not show one-to-one correspondence to "activation" defined by the GLM, as referred to by most of the fMRI literature.

4.2. Relation to Emotional Content (Arousal and Valence)

Given the nature of the show, the participants' ratings of arousal and valence of the content were highly correlated. Across the fifteen windows, the correlation between arousal and inverted valence was very high ($r = 0.92$, $p \ll 0.001$), which meant that most exciting clips were unpleasant and the less exciting ones were found to be pleasant. The positive correlation to complexity suggests that active regions had higher complexity when processing the more calm/pleasant content rather than the exciting/unpleasant one. Given the positive correlation to ISC, periods of higher complexity also had higher synchronized (across subjects) activity.
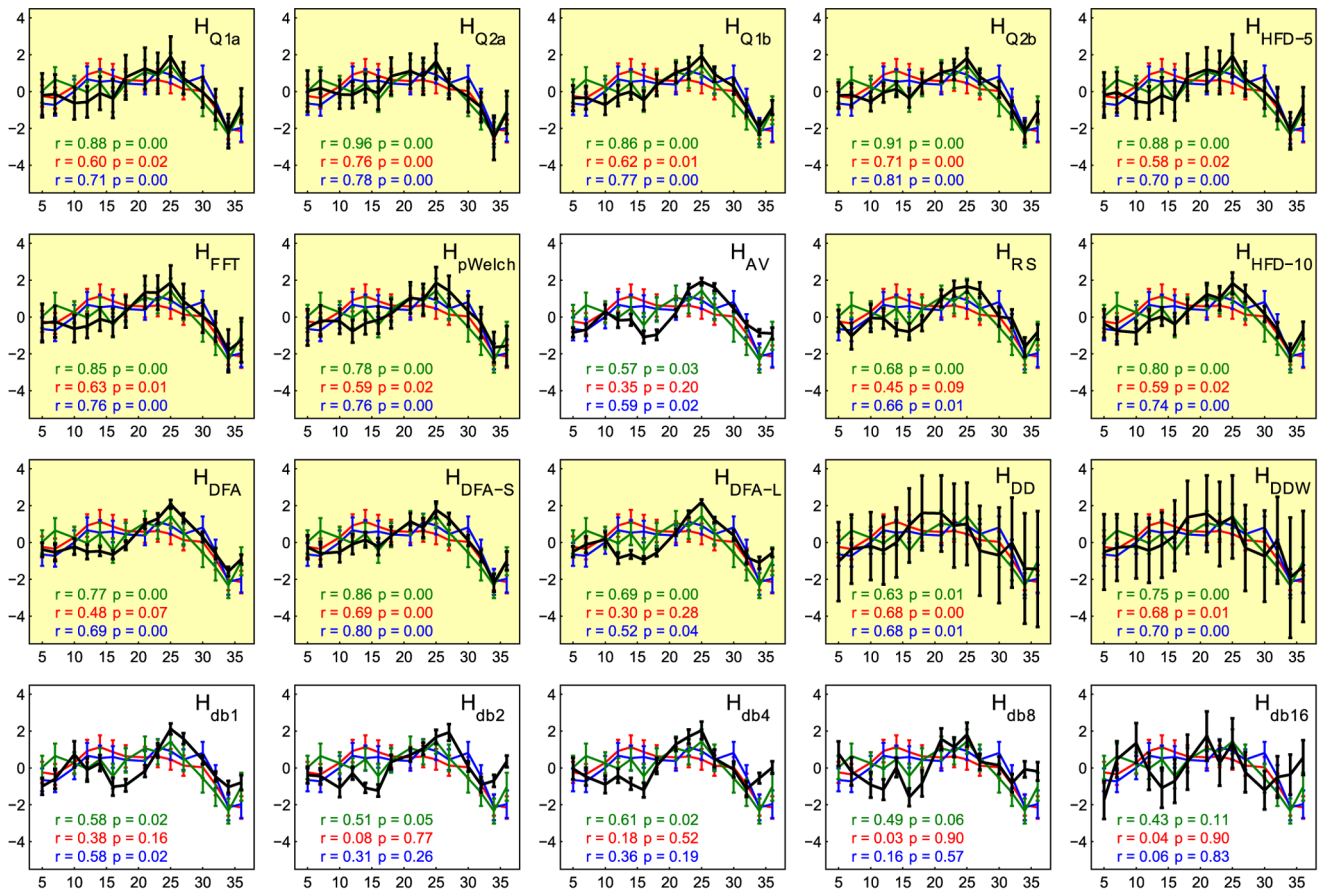
Figure 10. Variation in Hurst exponent estimates (black) correlates with arousal (red), valence (blue), and ISC-derived activation (green) throughout the GRST-LOST task. All estimates were computed over fifteen 10-minute windows evenly distributed throughout the task with 77% overlap, with the window centers plotted on the x-axis and z-scores of normalized variables on the y-axis. Error bars represent standard errors across the 11 participants. For arousal and valence, the positive values represent segments of the show rated as more calm and pleasant, while negative ones represent excited and unpleasant. The self-report measures were highly correlated ($r = 0.92$, $p \ll 0.001$). Highlighted plots represent complexity measures that correlated with either valence, arousal, or activation at $p < 0.01$ level.
doi:10.1371/journal.pone.0063448.g010

Although both self-report measures correlated with ISC-based activation ($r = -0.73$, $p = 0.002$ for arousal; $r = -0.70$, $p = 0.004$ for inverted valence), they showed less variation with complexity than ISC potentially due to the fact that they are discrete, limited in range, and prone to inaccurate representation of the true nature of emotional content of the show as the scenes are separated by a considerable amount of time and other emotional content.

4.3. Scanner Differences

Because the first principal component contained information about the whole brain divided into regions according to the WFU PickAtlas [47,48] for each person, it is unlikely that group differences arose due to local differences in brain activity. Furthermore, the differences in Hurst exponents between groups are very strong, arguably much stronger than between any two random samples of healthy human participants. This serves as an indication that the main variable driving the effect is the choice of scanner. Although the scan parameters were the same (TR, TE, flip angle, number of slices, gap, FOV, reconstruction matrix), the implicit differences between scanners contributing to background noise were strong enough to be picked up by the Hurst estimates and clearly separate a uniform group of people. The exact causes that give rise to differences in properties of time series across

scanners require a more rigorous investigation, but overall, this comparison shows that while combining datasets across scanners may in some cases be acceptable for standard linear model analyses, it is not for complexity investigations, unless the scanner differences are explicitly modeled or corrected for.

Conclusions

5.1

The comparisons based on simulated and experimental data show that complexity measures are highly correlated to each other and indeed measure the same quantity, but with differing sensitivity and susceptibility to artifacts (**Figure 13**). The poorest performing measures overall were H_{db*}, H_{DFA*} (except H_{DFA-S}), H_{AV}, and H_{RS}. Daubechies wavelet based computations (H_{db*}) have long computation times, are not sensitive to spikes, and show poor sensitivity to activation, tissue type, and emotional content; for these Daubechies wavelet based estimates the overall performance increases with the wavelet order up to a point (H_{db8}), and then deteriorates. H_{RS} and H_{AV}, performed poorly across the board. In terms of image contrast, overlap with activation, and group differences, H_{DFA*} performed poorly as well, with H_{DFA-S} outperforming H_{DFA} and H_{DFA-L}, suggesting that the bulk of useful information is found at shorter lags. The computation times
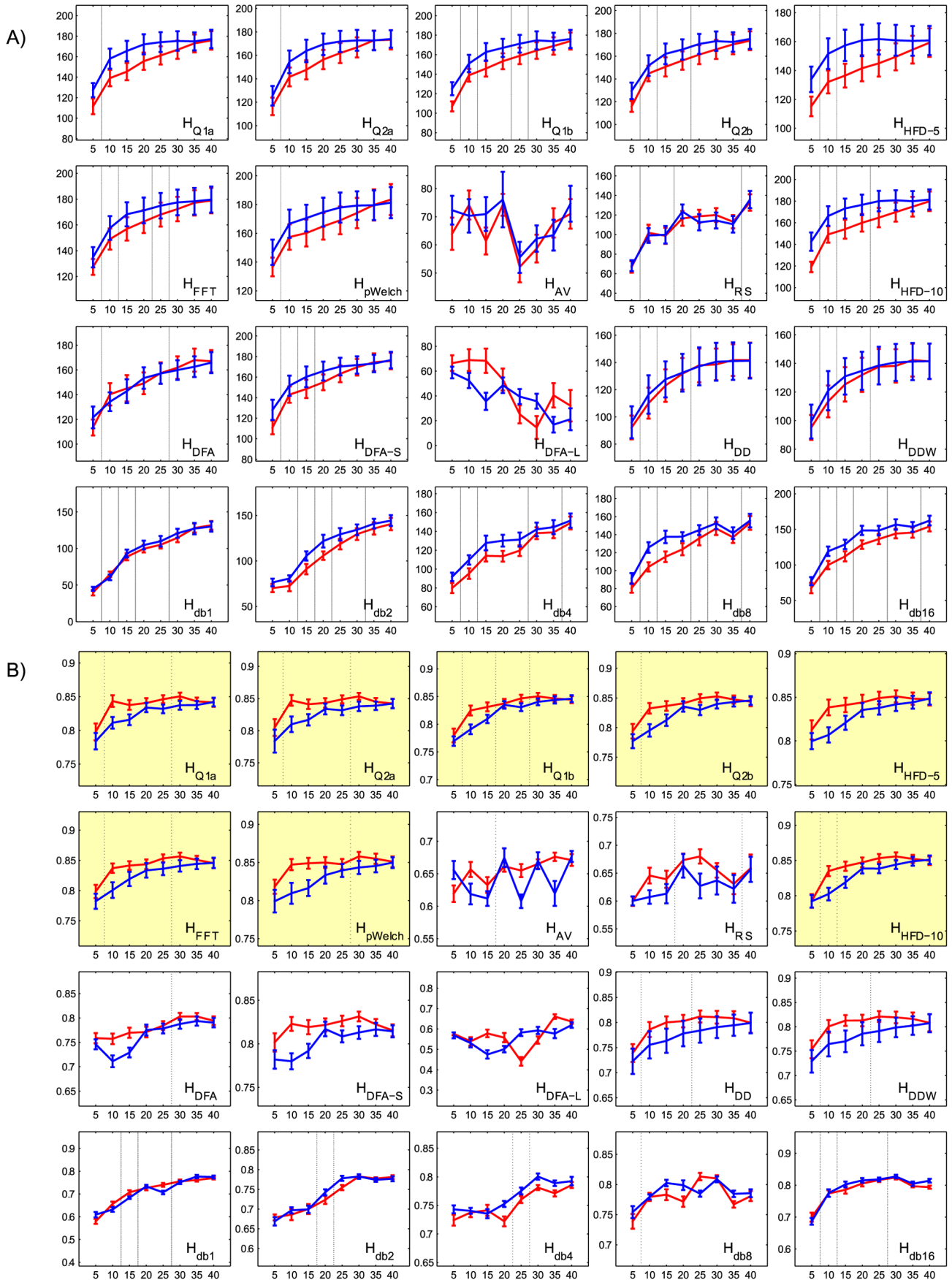
Figure 11. Mean (and standard error across 11 subjects) (A) GM>WM contrast (T-values on y-axis) and (B) overlap with activation for top 10% of the values (AUC on y-axis) increase for most of the Hurst estimates as the length of the time series (x-axis, minutes) increases. The two lines represent windows that start at the beginning (red) and the end (blue) of the task. Vertical lines between pairs of window sizes indicate that the subsequently larger window size significantly improves GM>WM contrast (A) or AUC (B) as compared to the smaller one for both forward and backward counting windows (p<0.025 each). The highlighted plots in (B) indicate measures that attained an AUC of 0.85 or more.
doi:10.1371/journal.pone.0063448.g011

were also very high for this set of measures. The H_{Q*b} measures are naturally very similar to H_{Q*a}, but are the longest to compute, with the error drop-off rate as the only distinguishable characteristic.

The most consistently successful measures were the power-spectrum based measures H_{FFT} and H_{pWelch}, with the latter slightly outperforming the former while taking much longer to compute. H_{HFD*} did not stand out in either a positive or a negative way, but performed consistently well across all the comparisons, ranking second overall behind the power-spectrum based measures, with H_{HFD-10} being slightly better than H_{HFD-5}. H_{Q*a} took the third place, having failed only one comparison.

5.2. The Bottom Line

The results outlined above recommend certain steps for estimating complexity of fMRI time series. First, it is crucial to either collect data using a single scanner or control for this effect. Second, it appears that detrending, regressing out the global mean, and excluding low frequencies improves agreement between complexity and activation. First order motion correction also has a small, but significant improvement on tissue contrast. Third, given the typical two-second acquisition time for fMRI, the tasks should last about 20 and definitely more than 10 minutes. Finally, the best measures to use are either the power-spectrum based ones (H_{FFT} or H_{pWelch}) on a restricted frequency range (above ~ 0.01 Hz), or H_{HFD*} or H_{Q*a} on filtered data.

Although the comparisons of numerical implementations of various algorithms over the simulated data give a "clean" comparison, the real usefulness of Hurst estimates is, however, assessed by their relation to an interpretable quantity, such as activation or tissue type. The better estimates, as outlined in this paper, are more sensitive to some aspects of brain activity than linear regression and correlation approaches. Despite the potential utility of complexity in neuroimaging, the underlying neuronal and BOLD features that give rise to the differences in Hurst estimates are not mechanistically well understood, and future investigation will be necessary to address this important question.
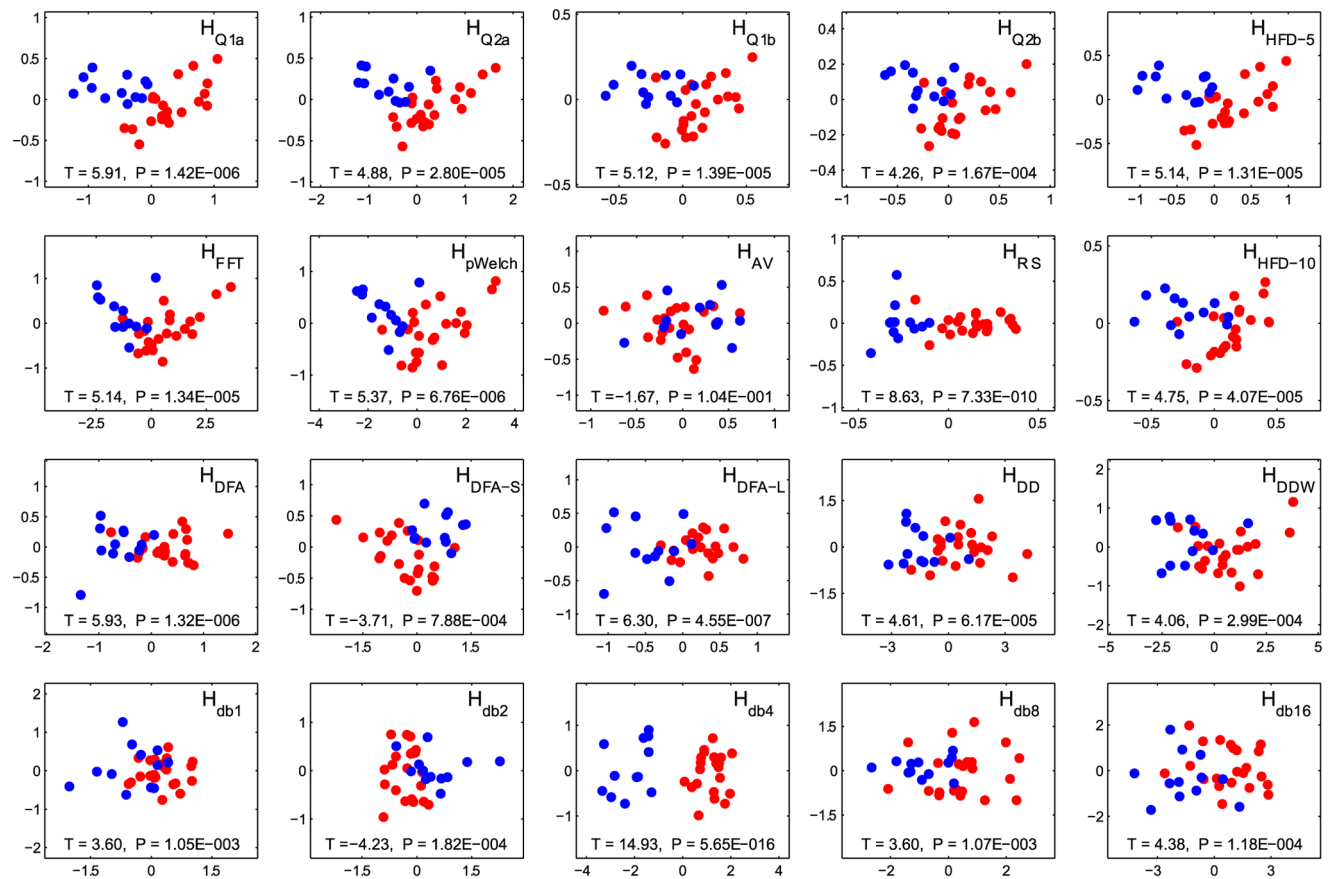
Figure 12. Strong scanner differences illustrated through principal components analysis. The red and blue points correspond to sets of subjects scanned on different scanners with the same task (BL-ANT). Each of the plots show the distribution of subjects along the first (x-axis, arbitrary units) and second (y-axis, arbitrary units) principal components of a matrix of Hurst region averages by subject. Each of the Hurst estimates was computed for the d-m-bcw-fhi processed average time series of every region of WFU pickAtlas. The t- and p-values represent differences between scanners along the first principal component (x-axis).
doi:10.1371/journal.pone.0063448.g012

Measure	Consistency with other measures (Figure 3)	Error drop-off rate (Figure 4)	Sensitivity to spikes (Figure 4)	Computation Time (Figure 4)	Sensitivity to activation (Figure 7)	Sensitivity to tissue type (Figure 7)	Relation to emotional content (Figure 10)	Monotone increase with time series (Figure 11)	Scanner differences (Figure 12)
H_{Q1a}	Good	Poor	Good	Good	Acceptable	Acceptable	Good	Good	Good
H_{Q2a}	Good	Poor	Good		Acceptable	Acceptable	Good	Good	Good
H_{Q1b}	Good	Good	Acceptable	Poor	Acceptable	Acceptable	Good	Good	Good
H_{Q2b}	Good	Good	Acceptable		Acceptable	Acceptable	Good	Good	Good
H_{HFD-5}	Good	Good	Good	Good	Acceptable	Acceptable	Good	Good	Good
H_{HFD-10}	Good	Good	Good	Good	Acceptable	Acceptable	Good	Good	Good
H_{DD}	Acceptable	Good	Acceptable	Good	Poor	Poor	Good	Good	Good
H_{DDW}	Acceptable	Good	Acceptable	Good	Poor	Poor	Good	Good	Good
H_{db1}	Acceptable	Good	Poor	Acceptable	Poor	Poor	Poor	Acceptable	Good
H_{db2}	Acceptable	Good	Poor	Acceptable	Poor	Poor	Poor	Acceptable	Good
H_{db4}	Acceptable	Good	Poor	Acceptable	Poor	Poor	Poor	Acceptable	Good
H_{db8}	Acceptable	Good	Poor	Acceptable	Acceptable	Poor	Poor	Acceptable	Good
H_{db16}	Acceptable	Good	Poor	Poor	Poor	Poor	Poor	Acceptable	Good
H_{FFT}	Good	Good	Good	Good	Good	Good	Good	Good	Good
H_{pWelch}	Good	Good	Good	Acceptable	Good	Good	Good	Good	Good
H_{DFA}	Acceptable	Poor	Good	Poor	Poor	Acceptable	Acceptable	Acceptable	Good
H_{DFA-S}	Good	Acceptable	Good		Acceptable	Acceptable	Good	Acceptable	Good
H_{DFA-L}	Poor	Poor	Poor		Poor	Poor	Acceptable	Poor	Good
H_{RS}	Acceptable	Poor	Poor	Acceptable	Poor	Poor	Acceptable	Poor	Good
H_{AV}	Acceptable	Poor	Poor	Good	Poor	Poor	Poor	Poor	Poor

Figure 13. Summary of performance of Hurst estimates. Summary of measure comparisons highlights H_{FFT}, H_{pWelch}, and H_{HFD*} as the most reliable methods overall.
doi:10.1371/journal.pone.0063448.g013

Author Contributions

Conceived and designed the experiments: LRMP DR TF. Performed the experiments: DR TF. Analyzed the data: DR. Contributed reagents/materials/analysis tools: DR TF. Wrote the paper: DR.

References

1. Gisiger T (2001) Scale invariance in biology: coincidence or footprint of a universal mechanism? Biol Rev Camb Philos Soc 76: 161–209.
2. Eke A, Herman P, Kocsis L, Kozak LR (2002) Fractal characterization of complexity in temporal physiological signals. Physiological Measurement 23: R1–R38–R31–R38.
3. Bullmore E, Barnes A, Bassett DS, Fornito A, Kitzbichler M, et al. (2009) Generic aspects of complexity in brain imaging data and other biological systems. Neuroimage 47: 1125–1134.
4. Levina A, Herrmann JM, Geisel T (2007) Dynamical synapses causing self-organized criticality in neural networks. Nature Physics 3: 857–860.
5. Rubinov M, Sporns O, Thivierge J-P, Breakspear M (2011) Neurobiologically Realistic Determinants of Self-Organized Criticality in Networks of Spiking Neurons. PLoS Comput Biol 7: e1002038-e1002038.
6. Bédard C, Destexhe A (2009) Macroscopic Models of Local Field Potentials and the Apparent 1/f Noise in Brain Activity. Biophysical Journal 96: 2589–2603.
7. Freyer F, Aquino K, Robinson PA, Ritter P, Breakspear M (2009) Bistability and Non-Gaussian Fluctuations in Spontaneous Cortical Activity. The Journal of Neuroscience 29: 8512–8524.
8. He BJ, Zempel JM, Snyder AZ, Raichle ME (2010) The Temporal Structures and Functional Significance of Scale-free Brain Activity. Neuron 66: 353–369.
9. He BJ (2011) Scale-Free Properties of the Functional Magnetic Resonance Imaging Signal during Rest and Task. The Journal of Neuroscience 31: 13786–13795.
10. Goldberger AL, Amaral LA, Hausdorff JM, Ivanov P, Peng CK, et al. (2002) Fractal dynamics in physiology: alterations with disease and aging. Proc Natl Acad Sci U S A 99 Suppl 1: 2466–2472.
11. Peng CK, Havlin S, Stanley HE, Goldberger AL (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos: An Interdisciplinary Journal of Nonlinear Science 5: 82–82.
12. Perkiömäki JS, Mäkikallio TH, Huikuri HV (2005) Fractal and complexity measures of heart rate variability. Clinical And Experimental Hypertension (New York, NY: 1993) 27: 149–158.
13. Daneshyari M, Kamkar LL, Daneshyari M (2010) Epileptic EEG: A Comprehensive Study of Nonlinear Behavior. In: Arabnia HR, editor. Advances in Computational Biology. New York, NY: Springer New York. 677–683.
14. Stam CJ (2005) Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clinical Neurophysiology 116: 2266–2301.

15. Hu X-S, Hong K-S, Ge SS (2011) Recognition of stimulus-evoked neuronal optical response by identifying chaos levels of near-infrared spectroscopy time series. Neurosci Lett 504: 115–120.
16. Lai M-C, Lombardo MV, Chakrabarti B, Sadek SA, Pasco G, et al. (2010) A Shift to Randomness of Brain Oscillations in People with Autism. Biological Psychiatry 68: 1092–1099.
17. Maxim V, Sendur L, Fadili J, Suckling J, Gould R, et al. (2005) Fractional Gaussian noise, functional MRI and Alzheimer's disease. Neuroimage 25: 141–158.
18. Radulescu AR, Rubin D, Strey HH, Mujica-Parodi LR (2011) Power spectrum scale invariance identifies prefrontal dysregulation in paranoid schizophrenia. Hum Brain Mapp.
19. Thurner S, Windischberger C, Moser E, Barth M (2002) Fractal noise maps reveal human brain activity: a key for unbiased fMRI analysis. cond-mat/0208415.
20. Shimizu Y, Barth M, Windischberger C, Moser E, Thurner S (2004) Wavelet-based multifractal analysis of fMRI time series. Neuroimage 22: 1195–1202.
21. Barnes A, Bullmore ET, Suckling J (2009) Endogenous human brain dynamics recover slowly following cognitive effort. PLoS One 4: e6626–e6626.
22. Bandettini PA, Bullmore E (2008) Endogenous oscillations and networks in functional magnetic resonance imaging. Hum Brain Mapp 29: 737–739.
23. Tolkunov D, Rubin D, Mujica-Parodi LR (2010) Power spectrum scale invariance quantifies limbic dysregulation in trait anxious adults using fMRI: adapting methods optimized for characterizing autonomic dysregulation to neural dynamic timeseries. Neuroimage 50: 72–72.
24. Wink AM, Bernard F, Salvador R, Bullmore E, Suckling J (2006) Age and cholinergic effects on hemodynamics and functional coherence of human hippocampus. Neurobiol Aging 27: 1395–1404.
25. Suckling J, Wink AM, Bernard FA, Barnes A, Bullmore E (2008) Endogenous multifractal brain dynamics are modulated by age, cholinergic blockade and cognitive performance. J Neurosci Methods 174: 292–300.
26. Carlson JM, Greenberg T, Rubin D, Mujica-Parodi LR (2011) Feeling anxious: anticipatory amygdalo-insular response predicts the feeling of anxious anticipation. Social Cognitive and Affective Neuroscience 6: 74–81.
27. Ekman P, Friesen W (1976) Pictures of Facial Affect. Palo Alto, CA: Consulting Psychologists Press.
28. Lundqvist D, Flykt A, Öhman A (1998) The Karolinska Directed Emotional Faces - KDEF. CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.
29. Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8: 109–114.
30. Preis T, Virnau P, Paul W, Schneider JJ (2009) Accelerated fluctuation analysis by graphic cards and complex pattern formation in financial markets. New Journal of Physics 11: 093024–093024.
31. Generalized Hurst exponent of a stochastic variable Available: http://www.mathworks.com/matlabcentral/fileexchange/30076-generalized-hurst-exponent.Accessed 2012 November.
32. Di Matteo T (2007) Multi-scaling in finance. Quantitative Finance 7: 21–36.
33. Di Matteo T, Aste T, Dacorogna MM (2003) Scaling behaviors in differently developed markets. Physica A 324: 183–188.
34. Matteo TD, Aste T, Dacorogna MM (2005) Long term memories of developed and emerging markets: using the scaling analysis to characterize their stage of development.
35. Taqqu MST, Willinger W (1995) Estimators for Long-Range Dependence: An Empirical Study. FRACTALS 3: 785–798.
36. Blok HJ (2000) On the nature of the stock market: Simulations and experiments. arXiv:cond-mat/0010211.
37. Hurst exponent - File Exchange - MATLAB Central Available: http://www.mathworks.com/matlabcentral/fileexchange/9842-hurst-exponent.Accessed 2012 November.
38. Higuchi T (1988) Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena 31: 277–283.
39. Flandrin P (1992) Wavelet analysis and synthesis of fractional Brownian motion. IEEE Transactions on Information Theory 38: 910–917.
40. Polychronaki GE, Ktonas PY, Gatzonis S, Siatouni A, Asvestas PA, et al. (2010) Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. Journal of Neural Engineering 7: 046007–046007.
41. Complete Higuchi fractal dimension algorithm Available: http://www.mathworks.com/matlabcentral/fileexchange/30119-complete-higuchi-fractal-dimension-algorithm/content/hfd.m.Accessed 2012 November.
42. Parameter estimation of fractional Brownian motion (wfbmesti) Available: http://www.mathworks.com/help/toolbox/wavelet/ref/wfbmesti.html.Accessed 2012 November.
43. Istas J, Lang G (1997) Quadratic variations and estimation of the local Hölder index of a Gaussian process. Annales de l'Institut Henri Poincare (B) Probability and Statistics 33: 407–436.
44. Flandrin P (1989) On the Spectrum of Fractional Brownian Motions. IEEE Transactions on Information Theory 35: 197–199.
45. Wilson SM, Molnar-Szakacs I, Iacoboni M (2008) Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension. Cerebral Cortex 18: 230–242.
46. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject Synchronization of Cortical Activity During Natural Vision. Science 303: 1634–1640.
47. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15: 273–289.
48. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19: 1233–1239.