UCLA
UCLA Previously Published Works

Title
A wavelet-based statistical analysis of fMRI data I. Motivation and data distribution modeling

Permalink
https://escholarship.org/uc/item/3t6956d9

Journal
Neuroinformatics, 3(4)

ISSN
1539-2791

Authors
Dinov, Ivo D
Boscardin, John W
Mega, Michael S
et al.

Publication Date
2005

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Dinov ID, Boscardin, JW, Mega MS, Sowell EL, Toga AW. A Wavelet-Based Statistical Analysis of fMRI data: I. Motivation and Data Distribution Modeling, NeuroInformatics, Humana Press, 3(4), 319-343, 2005, PMID: 16284415.

A Wavelet-Based Statistical Analysis of fMRI data: I. Motivation and Data Distribution Modeling

Ivo D. Dinov[1,2], John W. Boscardin[3], Michael S. Mega[4], Elizabeth L. Sowell[1] and Arthur W. Toga[1]

[1]Laboratory of Neuro Imaging, Department of Neurology,
[2]Department of Statistics, [3]Department of Biostatistics,
UCLA, Los Angeles, CA 90095.
[4]Pacific Health Research Institute, Honolulu, HI 96813.

Correspondence:
Ivo D. Dinov, PhD
Department of Statistics
8130 Mathematical Sciences Bldg.
UCLA, Box 951554
Los Angeles, CA 90095-1554
E-mail: dinov@stat.ucla.edu
Tel. 310-825-8430
Fax: 310-206-5658

Abstract

We propose a new method for statistical analysis of functional magnetic resonance imaging (fMRI) data. The discrete wavelet transformation is employed as a tool for efficient and robust signal representation. We use structural MRI and functional fMRI to empirically estimate the distribution of the wavelet coefficients of the data both across individuals and across spatial locations. An anatomical sub-volume probabilistic atlas is used to tessellate the structural and functional signals into smaller regions each of which is processed separately. A frequency-adaptive wavelet shrinkage scheme is employed to obtain essentially optimal estimations of the signals in the wavelet space. The empirical distributions of the signals are computed on all regions in compressed wavelet space. These are modeled by *heavy-tail* distributions because their histograms exhibit slower tail decay than the Gaussian. We discovered that Cauchy, Bessel K-Forms and Pareto distributions provide the most accurate asymptotic models for the distribution of the wavelet coefficients of the data. Finally, we propose a new model for statistical analysis of functional MRI data using this atlas-based wavelet-space representation. In the second part of our investigation we will apply this technique to analyze a large fMRI data set involving repeated presentation of sensory-motor response stimuli in young, elderly and demented subjects.

Key Words: fMRI, wavelets, statistical analysis, brain mapping, brain atlas

Introduction

This is part one of a two-part report. Here we justify the utilization of wavelets for representation of fMRI functional brain data. In addition, we study empirically the distributions of the wavelet coefficients, the corresponding leptokurtic distribution models and the atlas-based technique for statistical analysis of the fMRI timeseries. In the second part of this manuscript we will present an extended analysis of a complex fMRI dataset (Buckner R. 2000) representing the BOLD response to paired and isolated sensory-motor trials in Alzheimer's disease subjects, age-matched normal controls and a similar number of young adults.

Our basic idea is to perform the statistical analysis of fMRI data in the wavelet domain. The reason for that is a two-fold: First, there are only a few wavelet coefficients that survive the wavelet shrinkage preprocessing, which reduces the rate of *type I* errors, and; Second, we effectively assess the variations in the *patterns* of activation in different groups, as opposed to studying the differences in the fMRI BOLD signal intensities, across subjects, anchored at specific voxel locations.

fMRI Imaging: Functional Magnetic Resonance Imaging (fMRI) uses blood oxygen level dependent (BOLD) signal to measure hemodynamic alterations of the volume and flow of the blood intravascular susceptibility (Ogawa 1990). There are still issues about the relation of the observed BOLD signal, glucose metabolism and neuronal activation. Various neuro-physiological studies have been undertaken to validate the direct association between the detected fMRI signal and cellular patterns of activation. Some of these use optical intrinsic signal imaging (Hess 2000), electroencephalography (Krakow 1999; Krakow 2000) or microelectrode recordings (Logothetis 2001).

It is commonly accepted that there are 6-8 seconds of delay between stimulus onset and the peak of the observed BOLD fMRI signal (Bandettini 1993; Bandettini 1997). In addition, the percent modulation change in the data between rest and activation states is only in the order of 1-4%. Presence of noise, caused by normal physiological processes, the imaging equipment, between- and within-subject motion and habituation (Hajnal 1994), requires the use of robust statistical techniques for analyzing fMRI data. One successful approach in the process of enhancing the signal at the time of data acquisition is to choose certain on/off block design with specific time-length and stimulus conditions (Bandettini 1993; Bullmore 1996). However, this method makes a critical assumption that the BOLD effect remains

constant across epochs. Poor signal-to-noise ratio can also be addressed by designing better data acquisition protocols, including event related designs, pulse-sequences, novel radio-frequency coil models, different time, space and resolution strategies, etc. (Logothetis 2001). For example, power lost in the signal in event-related fMRI can be counter balanced by increasing the numbers of trials for event-related averaging.

Analyses of fMRI data: Once the data is acquired there is only so much one can do to find significant signal temporal changes and their spatial locations. Various techniques have been proposed for post acquisition data analysis based on the general linear model (Friston 1995), mutual information (Tsai 1999), fuzzy logic (Golay 1998), neural networks (Chang and Chung 2000) and Bayesian statistics (Kershaw 1999).

A number of other statistical approaches for analyzing fMRI data have been proposed in the literature, as well. These involve Student's *T* test (Worsley 1994), χ^2 and ***F*** tests (Cao 1999), Kolmogorov-Smirnov test (Aguirre G. 1998), correlation analysis (Kuppusamy K. 1997) and non-parametric tests (Bullmore 1996). Most techniques, however, are applied in the spatial domain and either treat each voxel intensity as independent (an assumption oftentimes too artificial) or introduce a spatial smoothing by convolving the native data with a large (Gaussian) kernel. The latter scheme increases the statistical power, since it takes advantage of the fact that near-by voxels are highly correlated and decreases the noise by smoothing; however, it also lowers the already low fMRI signal gradient (Desco 2001).

What are Wavelets? Wavelets are scaled and shifted versions of a *mother wavelet* function, whose properties are discussed below. Wavelet analysis is similar to Fourier analysis (Xu and Chan 2002) in the sense that it decomposes a signal down into its basic constituent components. Whereas the Fourier transform represents the signal as a series of sine waves of different frequencies, the wavelet transform represents signals as linear combinations of wavelets. There are differences between the trigonometric and wavelet base signal representations. In comparison to the smooth and infinitely periodic sine wave decomposition, the wavelet representation is irregular (often fractal) in shape and uses compactly supported base functions. These properties of irregularity in shape and compact supportedness make wavelets a useful tool for analyzing non-stationary signals. Their irregular shape lends them to

analyzing signals with discontinuities or sharp changes, while their compactly supported nature enables temporal localization of signal features.

Continuous and Discrete Wavelet Transforms: A wavelet basis is constructed using an oscillatory, compactly supported, (mother) wave function, usually C^k differentiable on the real line, which has rapidly decaying tails and satisfies an admissibility condition

$$\int \psi(s)ds=0. \qquad (1)$$

Equation (1) ensures the invertiblility of the wavelet transform. Regularity of the wave ψ is determined by the number of the continuously vanishing moments. Hence, $Reg(\psi)$=M, if and only if

$$\int s^r \psi(s)ds=0,\ 0 \le r \le M\text{-}1 \ \text{ and } \ \int s^M \psi(s)ds \neq 0. \qquad (2)$$

Figure 1

Daubechies (Daubechies 1988) defined a class of mother wavelet functions, Figure 1, starting with a (father) wavelet, ϕ, such that integral translations of ϕ form an orthonormal basis for the space of the piecewise step-functions (characteristic functions on intervals of length one). Then, we obtain the following wavelet function

$$\psi(s) = \sqrt{2}\sum_{k=0}^{2M-1} g_k \phi(2s-k). \qquad (3)$$

The basic problem is to find a non-trivial sequence of parameters $\{g_k\}_{k=1}^{K}$, such that ψ has a compact support, proper regularity and vanishing moments of order M ($K \le 2M-2$). The Daubechies wavelet basis (Daubechies 1993) has the fewest number of vanishing moments of any class of orthogonal wavelets.

Generally, wavelet bases with higher-order non-trivial moments have larger supports. Daubechies wavelets, ψ, for example, are compactly supported on $[-(M-1); M]$. Therefore the dyadically-scaled and integrally-translated wavelets, $\psi_{j,k}=2^{j/2}\ \psi\ (2^{-j}\ -\ k)$, are supported on $[(k+1-M)\ 2^{-j};\ (k+M)\ 2^{-j}]$. Mallat (Mallat 1989) proved that the set of the scaled and translated wavelets, induced by the mother wavelet, form an orthonormal basis for $L^2(R)$. This means that linear combinations of such wavelets may be used to represent any square integrable function. An important characteristic of this type of signal decomposition is that the signal representation is frequency and space localized. So, a domain-restricted perturbation of the original signal only affects the wavelets whose supports intersect the

region of change of the function. Then, for *f(x)* in $L^2(R)$, the wavelet transform (WT) is defined by the collection of inner-products:

$$\left\{ w_{j,k} = \left\langle f, \psi_{j,k} \right\rangle = \int f(s)\psi_{j,k}(s)ds : j,k \in Z \right\}. \qquad (4)$$

The discrete WT is analogous to its continuous counterpart described above, except that the signal is split into dyadic blocks (the shifting and scaling operators are based on powers of 2). The continuous wavelet transform still uses discretely sampled data, however the shifting operator smoothly varies across the length of the sampled data, and the scaling operator allows much finer encoding resolution as it is defined on a range, where the minimum is the original signal scale and the maximum is user specifiable. This improved resolution of the continuous WT comes at the expense of increase in time, memory and computational complexity required to calculate the continuous wavelet coefficients. In practice, the discrete WT is computed using Mallat's pyramidal algorithm (Daubechies 1988; Mallat 1989; Daubechies 1991; Daubechies 1993; Donoho, Johnstone et al. 1996; Xu and Chan 2002). For example, in the case of the Daubechies 4-coefficient basis the discrete WT matrix can be thought of as a mixture of a smooth (**S**) and a detail (**D**) filter (Mallat 1989). Figure 2 illustrates the matrix of the discrete wavelet transform using the Daubechies 4-coefficient filter bank, the forward and inverse discrete WTs for a phantom signal consisting of 8 time observations. Odd rows in the WT matrix correspond to convolutions of the coefficients and the data (moving average or smoothing). The even rows correspond to *difference-convolutions*, hence details. The algorithm proceeds iteratively by applying this transform to the discretized data. At each stage half the results of the S and D filters are discarded (decimation) and the signal (of size N) is represented by N/2 smooth components and N/2 detail components. Hence, we have the necessity that the signal size is an exact power of two.

Figure 2

There exist straight forward extensions of the wavelet transformation in higher dimensions. Two ways in which this can be accomplished are either by tensor products (Chan 1999) or by carrying out one dimensional wavelet decompositions for each variable (index of dimensionality) separately (Bayoumi 2003). Examples of wavelet base functions in 2D and 3D are shown on Figure 3. This discrete version of the multidimensional wavelet transform (Dinov and Sumners 2001) is implemented in C and freely available for download from http://www.loni.ucla.edu/~dinov/WAIR.html.

Figure 3

Why Use Wavelets? The advantages of employing wavelet analysis over other conventional spatial-domain approaches include efficient signal representation (Dinov, Mega et al. 2002), noise reduction and decorrelation (Donoho and Johnstone 1998) and time-frequency localization (Daubechies 1991). From an empirical point of view, thresholding the wavelet coefficients provides a way to extract the essence of the data content of a signal. In Figure 4, we have the *HeavySine* function, thin red curve, its wavelet transform (WT), thin blue curve, and a wavelet estimate of the original signal (thick green smooth curve). A couple of interesting phenomena quickly emerge from this 1D plot. First, it is clear that very few of the wavelet coefficients have magnitudes larger than 0.3, across the range of frequencies and locations. A natural question then is: *If we set to zero all wavelet coefficients smaller than 0.3 and then synthesize the image back (by inverting the wavelet transform), will we get a reasonable representation of the original function?*

Figure 4

Figure 4 illustrates heuristically the answer to that question – the original and the reconstructed curves are similar. The protocol for selecting a subset of all wavelet coefficients that yield a robust and efficient signal representation will be addressed in the methods section, where we present frequency-adaptive wavelet shrinkage. In this example we used only the largest 2% of the wavelet coefficients to recover the *HeavySine* function. Even though the new estimate is not a perfect approximation of the original data it does capture the main trend of the signal, at 50:1 compression ratio. Another empirical observation worth pointing out now is the fact that the original signal is often times spatially correlated, whereas its WT is uncorrelated in the wavelet-space, see the blue curve [WT(*HeavySine*)] in Figure 4.

Prior Studies of fMRI Using Wavelets: Bullmore and coworkers presented a wavelet-based methodology for characterizing the noise structure in short to medium length fMRI timeseries (Fadili 2002). The same group also introduced wavelet-generalized least squares (WLS), which are used to obtain the best linear unbiased estimator of regression model parameters in the context of long-term memory errors. These theoretical developments were compared with results obtained by ordinary least squares and estimators based on auto regressive moving averages. The authors show superiority of the generalized WLS in terms of *type I* error control.

Wavelet-based analysis of fMRI data has been proposed as early as 1996, by Hilton and coworkers (Hilton 1996) who were the first to suggest signal denoising of fMRI data using a non-linear soft-thresholding. The group of Ruttimann (Ruttimann, Unser et al. 1998) also employed a multiresolution approach for increasing the sensitivity without sacrificing the probability for committing a *type I* error. They utilized a uniform thresholding in the wavelet domain followed by a Z-test, Bonferroni corrected for multiple comparisons. Yet another study, by Melke and coworkers (Mekle 2000) used multi-scale singularity detection in the temporal domain. Their data analysis is based on the constructed maxima lines, using wavelet expansions at integral scales to trace the wavelet modulus maxima. From the maxima lines the signal singularities are then localized by the beginning and ending of the stimulus.

Another method for detection of activation in fMRI data using wavelet analysis was introduced by Feilner's group (Feilner 2000). The extension proposed in this work was to incorporate non-stationary Gaussian noise and the subsequent T-testing, required by the assumption of unstable variances at different locales. No wavelet-space shrinkage is applied in their analysis, however, they use of a post-inverse wavelet transform signal thresholding, based on the residual noise-level. The same study investigated the effects of the choice of a WT relying on various information measures (e.g., Kullback-Leibler information, which measures the similarity between a statistical model and the true distribution).

Desco and coworkers investigated the performance of the wavelet analysis of fMRI data on a phantom functional volume by making use of different wavelet decomposition schemes, location, size and level of activation and the presence of smoothness (Desco, Hernandez et al. 2001). When measuring *sensitivity* (percentage of true activation areas detected) and *specificity* (percentage of locations correctly detected as non-activated), the authors clearly identified dependences in the results of the fMRI wavelet analysis, based on the statistical tests, the shape of the model-activation and the wavelet-family used. More recently, Meyer (Meyer 2003) introduced a new wavelet-based method for analysis of fMRI timeseries based on correlation of the observed BOLD signal and the stimulus temporal patterns. Bullmore's group proposed a Bayesian framework for multiresolution hypothesis testing on spatially extended statistic maps (Bullmore 2003; Bullmore 2004).

Methods

In this section we describe a frequency adaptive wavelet shrinkage strategy, which may be employed prior to statistical analysis. This scheme induces almost theoretically optimal signal estimators. Then we present the basics of stable and heavy-tail distributions. The latter ones we choose to use as models for the distributions of the wavelet coefficients of the functional MRI data in 3D and 4D. Finally, we introduce the framework for statistical analysis of fMRI timeseries utilizing a functional and anatomical sub-volume probabilistic atlas (F&A SVPA atlas) of the brain.

Frequency-Adaptive Wavelet Shrinkage: The process of wavelet shrinkage begins with a collection of observations $\{y_k\}$, $1 \le k \le N$, obtained by adding noise to some unknown signal $f(t_k)$, for each time point t_k. We assume that $y_k = f(t_k)+\varepsilon_k$, where $\varepsilon_k \sim \mathbf{N}(0,\sigma^2)$ are independent identically distributed Gaussian random variables. To recover the unknown function $f(t)$, having the data y_k alone, we construct function estimators $\hat{f}$. The *Risk* function

$$R(f,\hat{f}) = \frac{1}{N} E\left(\left\| f - \hat{f} \right\|^2 \right) = \frac{1}{N} E\left(\sum_{k=0}^{N-1} \left\| f(t_K) - \hat{f}(t_K) \right\|^2 \right) \qquad (5)$$

measures the performance of the estimator $\hat{f}$ by the average quadratic loss at the sample points where $\boldsymbol{E}$ is the expectation of the misfit between the function and its estimator. Notice the similarity between this *Risk* function (5) and the variance of unbiased estimators. Preference will be given to estimators having small *Risk* measures (i.e., small "variances").

If W represents the matrix of the discrete wavelet transform (Donoho and Johnstone 1995; Dinov and Sumners 2001), like in Figure 2, the time-domain model $y=f+\varepsilon$ becomes $w=\theta+z$ in wavelet space, where $w = W y$, $\theta = W f$ and $z = W\varepsilon$. Suppose U is a subset of the set of all wavelet coefficients $\{w\}$ of the data, one defines *selective-wavelet (SW) reconstruction* estimators by

$$\hat{f} = \sum_{j,k \in U} w_{j,k} W_{j,k} \qquad (6)$$

where $w=\{w_n\}= \{w_{j,k} \mid \ 0 \le j \le J-1;\ 0 \le k \le 2^j-1\}$, U is an index subspace, j codes for the frequency and k indexes the location of the corresponding wavelet coefficients, $W_{j,k}$ represents the $(j, k)^{th}$ row of the orthogonal wavelet transform matrix W and $N=2^J-1$ is the sample size of the data (Dinov, Mega et al. 2002). The *ideal Risk* for a function estimate using the SW reconstruction scheme is defined by

$$\mathsf{R}_{N,\sigma}(f, SW) = \inf_{\hat{f}} R(f, \hat{f}) \qquad (7)$$

Suppose now *f(t)* is the true 4D BOLD signal and $y(t) = f(t)+\varepsilon$ be the observed fMRI timeseries. Wavelet shrinkage is applied by filtering the wavelet transform $w = W(y)$, through the soft thresholding filter

$$\eta_{\lambda j}(w_{j,k}) = sign(w_{j,k}) \times \mathbf{max}\{ (|w_{j,k}| - \lambda j),\ 0 \} \quad (8).$$

The major challenge is to determine a meaningful threshold level λj, which will induce function estimators that are optimal and efficient across the space of functions of interest (these spaces include L^2, Besov spaces, Sobolev spaces, Hilbert spaces, etc.) (Donoho, Johnstone et al. 1996). One very promising example is

$$\lambda_j = \sigma\sqrt{2\ln(2n_j+4)}, \qquad (9)$$

where n_j is the number of wavelet coefficients at the *j*-frequency band (Dinov, Mega et al. 2002). *This particular thresholding scheme is appealing because noise is most likely to be present in the high-frequency bands of the signal; hence the steadily increasing (with the frequency index j) shrinkage imposed in this case*. From the theoretical point of view this scheme also induces an almost-optimal estimator $\left(\hat{f} = W^{-1}\,\eta_{\lambda j}\,W(y)\right)$ in L^2 of the unknown signal *f(t)*. First, we have the following upper bound on the risk measure

$$R(f, \hat{f}) \le (1+2\ln(N+4)) \times \left(\sigma^2 \ln N / N + \mathsf{R}_{N,\sigma}(f, SW)\right) \qquad (10)$$

indicating that this estimator is within $ln^2(N)/N$ of the *ideal risk* (7), where *N* is the size of the signal, σ is the signal noise level and *SW* is selective wavelet reconstruction scheme indicating a few of the wavelet parameters are used to synthesize the estimator. Secondly, we have a lower bound of any function estimator, again in terms of the ideal risk:

$$\inf_{\hat{\theta}} \sup_{\theta} \left\{ \left(1/N\, E(\|\theta - \hat{\theta}\|^2)\right) \times \left(\sigma^2 \ln N / 2N + 1/N\, \mathsf{R}_{N,\sigma}(f, SW)\right)^{-1} \right\} \sim 2\ln N \qquad (11)$$

In other words, no estimator can be closer to the ideal risk than $ln^2(N)/N$ across all possible fMRI signals. Therefore, the threshold strategy (9) that naturally increases with the increase of the frequency index of the wavelets yields a robust and optimal wavelet shrinkage representation.

Heavy-Tail Distributions: There is a volume of work on heavy-tail distributions in the field of signal processing. We consider some of the most popular and practically useful *leptokurtic* distributions

(distributions with tails heavier as compared to a Normal (Gaussian) curve with similar variance) (Gourieroux 1992).

As described earlier, we carry our statistical analyses in the *compressed wavelet space* (on wavelet coefficients that survive the frequency-adaptive shrinkage). Because these wavelet coefficients are large in magnitude it is extremely important to have a distribution model that is asymptotically accurate. In the results section we provide evidence that the empirical distributions of the wavelet coefficients of the MRI and fMRI data appear to be leptokurtic[1]. This motivates our interests in utilizing various heavy-tail distributions as proper models for the distribution of these wavelet parameters. The double-exponential distribution is defined by the following probability density function

$$f(x) = \frac{1}{2\beta} \times \exp\left(-\left| \frac{x-\mu}{\beta} \right| \right). \qquad (12)$$

It is one of the most commonly used leptokurtic distributions with tails decaying at the rate of exp(-|x|). The density function for the double-Pareto distribution with a tail-size parameter α and a scale parameter c is defined by

$$f(x) = \frac{\alpha c^{\alpha}}{|x-\mu|^{1-\alpha}}, \qquad (13)$$

for |x-μ|>c. The only problem with the double-Pareto distribution is that it has a singularity at μ. This is not a serious issue when one does statistics of extremes, as these are far away from μ. Another theoretically important heavy tail distribution is the Cauchy distribution with density function

$$f(x) = \frac{\gamma}{\pi} \times \frac{1}{\gamma^2 + (x-\mu)^2}. \qquad (14)$$

Cauchy distribution is frequently used to provide counterexamples to various phenomena, which appear quite natural for light-tail distributions. One transition between light- and heavy-tail distributions is provided by the ***T*** distribution which serves as a one-parameter homotopy family connecting the heavy-tailed Cauchy ($T_{df=1}$) and the standard normal distribution ($T_{df=\infty}$). The density function for a ***T*** distribution with df=n is given by

$$f(x) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi} \times \Gamma\left(\frac{n}{2}\right)} \times \left(1 + x^2 / n\right)^{-\frac{1}{2}(n+1)}, \text{ where } \Gamma(\alpha) = \int_0^{\infty} \exp(-y) \times y^{\alpha-1} dy. \qquad (15)$$

[1] In fact, the distribution of the wavelet coefficients of the 3D structural, 3D and 4D functional MRI data appears to be leptokurtic both across subjects at fixed location-scale indices, and within subjects across location-scale indices.

The last large class of leptokurtic distributions we consider as possible models for the wavelet coefficient distribution of the fMRI functional data is referred to as Bessel K forms (Grenander 2001). There exists an asymptotic analytic approximation of the density of Bessel K forms, which can be written by

$$\hat{f}(x;p,c)=\frac{1}{\Gamma(p)}\left(\frac{2}{c}\right)^{\frac{p}{2}}\left|x-\mu\right|^{p-1}\exp\left(-\sqrt{\frac{2}{c}}\left|x-\mu\right|\right). \qquad (16)$$

The parameters *p* and *c* can be estimated from the data by

$$\hat{p}=\frac{3}{skewness(WT(f))-3} \quad \text{and} \quad \hat{c}=\frac{\hat{\sigma}^2_{WT(f)}}{\hat{p}}, \qquad (17)$$

where the skewness and the sample variance of the wavelet transformation of the data (*WT(f)*) can be empirically obtained. It is known that this approximation is sub-optimal near the mean, μ. However, the asymptotic behavior of $\hat{f}(x;p,c)$ is very similar to that of the corresponding Bessel K form.

Atlas-based Statistical Analysis of fMRI data: The first step in the construction of an anatomical sub-volume probabilistic atlas (ASVPA) is to obtain a deterministic structural representation of the brain anatomy across the population (Thompson, Mega et al. 2000; Mega 2005). This was achieved by elastically warping the T-1 weighted fast SPGR MRI volumes of 26 subjects diagnosed with mild to moderate Alzheimer's disease, to an *average* template. This *average* template represents the collection of the averages of 36 gyral curves, manually drawn on all 26 subjects following rigid body alignment. A fully stereotactic fiducial-based warp was employed to register all 26 MRI volumes to the pseudo-average template, by requiring that the individual models for the 36 gyri were perfectly matched onto the corresponding *average* gyral curves. This way we minimized the global deformation needed to align all 26 subjects into a "common" space. The voxel-by-voxel mean of these elastically-warped volumes represents the deterministic brain atlas of the elderly. This rigid atlas framework is then used as a template for the construction of the induced anatomical sub-volume probabilistic atlas. The stereotactic magnetic resonance imaging data of 10 elderly AD and 10 age-matched normal control subjects were acquired on a 1.5 Tesla Siemens scanner. The T1-weighted MRI's were manually segmented into 2 hemispheres, with 30 regions of interest (ROI's) each composed of 3 different tissue types (white matter, WM, gray matter, GM, and cerebro-spinal fluid, CSF), for a total of 180 sub-volumes of interest (Mega, Cummings et al. 2001). All the data were aligned into the space of the deterministic elastic-warp template, described above using a 12 parameter polynomial warp. The

warped tissue segmented maps were then used to construct the anatomical sub-volume probabilistic atlases (ASVPA) in the following manner: Every probabilistic sub-volume of interest is represented as a cloud-like region with domain consisting of all voxels labeled part of this region across subjects, Figure 5. The intensity at a given voxel location represents the likelihood (probability) that this voxel is part of the region, based on the subject data. In other words, the intensities of the structural sub-volume probabilistic atlas reflect the chance (proportion) that each voxel-location belongs to the region of interest for that population.

Figure 5

Developing an atlas of *functional* brain data has been challenging due to image registration, intensity normalization, physiological brain variation and other unresolved signal localization issues. We have previously developed a functional brain atlas for Alzheimer's Disease (AD) in the frequency space using the wavelet signal representation. This atlas, AD functional and anatomical sub-volume probabilistic atlas (F&A SVPA) (Dinov, Mega et al. 2001), was based on rest-state PET data and employed in conjunction with our probabilistic anatomical atlas for the AD population (Crabtree EC 2000; Mega, Dinov et al. 2000; Mega, Cummings et al. 2001). Analogously, one can design a framework for construction of paradigm- and population-specific stochastic atlases of brain activation using fMRI data.

Figure 6 illustrates schematically the protocol for the construction of a new fMRI functional atlas, in general. The structural MRI and functional fMRI data are first anatomically tessellated to obtain the individuals' regions of interest. This involves delineating a number of regions of interest, which can be done manually, semi-manually or completely automatically (Mega, Thompson et al. 2000; Mega 2005). The next step involves applying the 4D WT to the fMRI data over each anatomical region separately. In the wavelet domain, we then threshold the wavelet coefficients to get the denoised and decorrelated representation of the timeseries in compressed wavelet space. Next, we calculate the joint cross-subject-and-time distribution of the 4D functional data. Finally, the induced functional and anatomical atlas (which is population and paradigm specific) is employed to statistically analyze new fMRI data. Note that the frequency-adaptive wavelet shrinkage is applied in the atlas-construction stage of the analysis.

Figure 6

The development of the F&A SVPA is directly coupled to the construction of an anatomical SVPA atlas; however, it is accomplished in the wavelet-space. We are interested in quantifying the paradigm-specific variability in functional data (fMRI) for diverse populations. More precisely, quantify functional variability, as measured by differences in the observed BOLD signal between subjects, at different spatial locations. Suppose we have N subjects scanned in an on/off block-design paradigm using fMRI. To simplify the presentation assume we are using our anatomical AD SVPA atlas, which contains 60 regions of interest. Every individual's fMRI timeseries over an ROI is analyzed separately by transforming it into frequency-space using the discrete WT. For each ROI, the *mean* and *variance* of the transformed wavelet coefficients, across subjects, will be computed at each location-frequency index, $w=(j,k)$.

Together with the probability maps encoded in the anatomical SVPA atlas, these distribution maps of functional variability in frequency space represent the desired functional and anatomical F&A SVPA atlas. If $P_n(v,t)$ is the intensity of subject n, at voxel location $\boldsymbol{v}$ and time t, $1 \le n \le N$, and A_k is the indicator function

$$A_k(v) = \begin{cases} 1, v \in ROI_k \\ 0, v \notin ROI_k \end{cases} \qquad (18)$$

of the k^{th} ROI of the ASVPA, then $P_{n,k}(v, t) = P_n(v, t) \times A_k(v)$ represents the value of the BOLD fMRI signal at location v and time t, for subject n over the ROI_k, $1 \le k \le 60$, $1 \le n \le N$. Denoting by $\hat{P}_{n,k} = \eta_{\lambda_j}(WT(P_{n,k}))$ the frequency-adaptively shrunk WT of the regional 4D fMRI signal we obtain an estimate of the sample mean and standard deviation of the intensities in frequency-space, over ROI_k:

$$\mu_k(w) = \frac{1}{N}\sum_{n=1}^{N} \hat{P}_{n,k}(w) \quad ; \quad \sigma_k(w) = \sqrt{\frac{1}{(N-1)}\sum_{n=1}^{N}\left(\hat{P}_{n,k}(w) - \mu_k(w)\right)^2} \ . \qquad (19)$$

Finally, the collection of volumes $\{\mu_k\}$ and $\{\sigma_k\}$, $1 \le k \le 60$, represents the *functional component* of the F&A SVPA atlas, for each scale-shift wavelet-space index w.

Suppose P is a new fMRI timeseries, with $P_k(v,t) = P(v,t) \times A_k(v)$ being the restriction of P over ROI_k. Then $\hat{P}_{n,k} = \eta_{\lambda_j}(WT(P_{n,k}))$ is the shrunk WT of P_k, using (8), and we employ the following secondary wavelet *shrinkage* strategy, $\hat{T}$:

$$\hat{T}\left(\hat{P}_k(w)\right)=\begin{cases}\hat{P}_k(w)\,, & \mathbf{P}\left(\left|\hat{P}_k(w)\right|>d_{\alpha,k}\right)<\alpha,\\ 0\,, & \mathbf{P}\left(\left|\hat{P}_k(w)\right|>d_{\alpha,k}\right)\geq\alpha\end{cases} \qquad (20)$$

where $\mathbf{P}$ is the probability and $d_{\alpha,k}$ is the critical value for a chosen heavy-tail model distribution $D_{\alpha,k}$. The sub-indices α and k represent the desired test significance level and the ROI, respectively. In particular, if one chooses a T_{df} distribution from the family of heavy-tail distribution models then the equation (20) becomes:

$$\hat{T}\left(\hat{P}_k(w)\right)=\begin{cases}\hat{P}_k(w)\,, & \left|\hat{P}_k(w)-\mu_k(w)\right|\geq t_{df,\alpha}\times\sigma_k(w),\\ 0\,, & \left|\hat{P}_k(w)-\mu_k(w)\right|<t_{df,\alpha}\times\sigma_k(w)\end{cases} \qquad (21)$$

where μ_k and σ_k are the mean and standard deviation of the distribution encoded in the F&A SVPA at location-frequency index w, and t_α is the T-score at significance level α. In our experiments we will investigate thresholds corresponding to α ={0.05, 0.001, 0.0001}. In general, one needs to use a heavy-tail model distribution of the data that has mean and variance approximately equal to the sample mean and variance, $\overline{\mu}_k$ and $\hat{\sigma}_k^2$.

The operator $\hat{T}$ in equation (20) either *shrinks* a wavelet coefficient or leaves it unchanged, depending on whether it is significantly different from the corresponding F&A SVPA atlas mean at w. Note that (20) acts only on the wavelet coefficients that survive the frequency-adaptive shrinkage (9). Thus, minimizing the number of statistical tests performed in the wavelet domain. This will be the first step of the statistical assessment.

The second step will involve inverting the WT and *constraining* the resulting wave patterns to the corresponding ROI's in space and time. This procedure is followed by applying another thresholding, which selects only the highest intensities (0.01-0.0001%, precise assignment will have to be closely studied) within each separate ROI. Finally, we reconstitute the statistical image in anatomical 3D space from the *collage* of regional statistical maps representing the spatial-domain thresholded differences, which were significant in the wavelet-space. This volume can be expressed symbolically as

$$\hat{\hat{P}}(v)=\bigcup_{k=1}^{60}\hat{\hat{P}}_k(v)\,,\quad \text{where}\ \hat{\hat{P}}_k=T\left(WT^{-1}\left(\hat{T}\left(\eta_{\lambda_j}\left(WT(P_k)\right)\right)\right)\right)\ \text{and}\ T\left(P_k(v,t)\right)=\begin{cases}P_k(v,t)\,, & \left|P_k(v,t)\right|\geq\beta_\alpha^{(k)}\\ 0\,, & \left|P_k(v,t)\right|<\beta_\alpha^{(k)}\end{cases} \qquad (22)$$

In this expression, α is a significance level, $\beta_\alpha^{(\kappa)}$ is the threshold separating the highest α x 100 % from the lowest $(1-\alpha)$x100% image intensities (uniform threshold over ROI_k) and WT^{-1} is the inverse wavelet transform operator. This methodology, is clearly directly extendible to, and applicable for, various anatomical tessellation schemes, disease populations and fMRI acquisition protocol. One needs to exercise caution when employing an F&A SVPA atlas constructed for one population to study another unrelated fMRI timeseries. This is because the empirical distributions of the wavelet coefficients may be significantly different across populations or study paradigms.

Results

We now present the empirical evidence that the distributions of the fMRI and MRI data in the wavelet domain have heavy tails. This fact is used in determining a robust method for statistical analysis of the wavelet coefficients trying to identify significantly different patterns of activation across the 4D fMRI volume, instead of mapping statistical differences at individual *votixels* (volume-time-elements).

Data: The large structural MRI database (Mazziotta, Toga et al. 2001), part of the International Consortium for Brain Mapping (ICBM), was utilized to obtain 578 anatomical volumes for testing the characteristics of the distribution of the wavelet coefficients within and between subjects. These data represents young healthy adult MRI brains drawn from diverse ethnic, gender and socio-economic populations. The functional fMRI data used in our experiments consisted of randomly selected subjects from the pool of 41 participants (14 young adults, 14 non-demented older adults and 13 demented older adults), Buckner *et al*., 2000. In that study the task paradigm consisted of presentation of a 1.5 second visual stimulus. Subjects were required to press a key with their right index fingers upon stimulus onset. The visual stimulus was an 8-Hz counterphase flickering checkerboard subtending approximately 12^o of the visual angle. Runs were structured so that for every eight-image acquisition, one of two kinds of trial condition was presented (15 trials per run for a total of 60 trials per subject). The trials either involved stimuli presented in isolation or in pairs with and inter-trial interval of 5.36 seconds. One trial and two-trial conditions were pseudo-randomly intermixed so that eight trials of one type and seven trials of the other appeared in each run. All subjects were English speaking, normal visual acuity right-

handers.

Wavelet coefficient distribution of the MRI data: Figure 7 indicates the frequency histogram of the average magnitude, across all 578 MRI volumes part of the ICBM database (Mazziotta, Toga et al. 1995), of the wavelet coefficients at 1,000 randomly selected locations in the wavelet domain (random indices of $w_{j,k}$). On the horizontal axis is the magnitude of the average wavelet coefficient between [-4; 4] and on the vertical axis (range [0; 578]) are the frequencies of these average magnitudes[2]. Clearly one observes an empirical distribution with heavy tails.

Figure 7

Would the average histogram on Figure 7 significantly change if we focus on a single subject and plot the histogram of the wavelet coefficients across 1,000 randomly selected locations in the wavelet domain (random indices of $w_{j,k}$)? Figure 8, shows these frequency distributions for three separate MRI scans (randomly selected from the pool of 578). There is little variation between these three individuals, and the overall shape of the distribution of the wavelet coefficients for one individual across 1,000 random locations is more regular, still heavy-tailed, than the averaged distribution in the previous Figure 7.

Figure 8

The 2D image in Figure 9 illustrates a small portion of all of the individual distributions. The horizontal axis again represents the magnitude of the wavelet coefficients in the range [-4; 4], vertical axis labels the random voxel indices and the color map on each row indicates the frequencies of occurrence of a wavelet coefficient of certain magnitude across all subjects MRIs. Bright colors indicate high, and dark colors represent low, frequencies. Note that Figure 8 effectively represents the across-subjects (row) average of these individual frequency histograms shown in Figure 9.

Figure 9

[2] There is nothing special with the range [-4 ; 4] which we have chosen to work with. The same phenomena were observed as we extended the range on which we calculate the distribution of the wavelet coefficients – the distribution remains *heavy-tailed*.

Wavelet coefficient distribution of the functional MRI data: Similar results are obtained if one studies functional fMRI instead of the structural MRI volumes. Figure 10 indicates the frequency histogram of the average wavelet coefficient magnitude, across all 128 fMRI 3D time-volumes, part of the fMRI study of Buckner and colleagues (Buckner R. 2000), at 1,000 randomly selected locations in the wavelet domain (random indices of $\boldsymbol{w}_{j,k}$).

Figure 10

And the 2D image in Figure 11 shows all of the individual distributions for all 128 time points (one run, six trials). Horizontal axis represents the magnitude of the wavelet coefficients in the range [-4; 4], the vertical axis labels the random voxel indices and the color map on each row indicates the frequencies of occurrence of a wavelet coefficient of certain magnitude across all time point of the fMRI epochs. Bright colors again indicate high, and dark colors represent low, frequencies. Figure 10 effectively represents the across row average of these individual frequency histograms shown in Figure 11.

Figure 11

In both cases, MRI and fMRI data, we have strong evidence that heavy-tail distributions should be used as models to account for the slow decay of the probability mass functions at the tails. We also investigated the empirical distribution of the wavelet coefficients of the 4D fMRI volume consisting of 128 time points. The 4D WT was applied to the complete timeseries volume and Figure 12 illustrates the frequency distribution of the wavelet coefficients of this large volume. Once again, one observes the slow asymptotic decay of the tails of this distribution, indicating heavy-tail distribution models are necessary to study these data (data-size: 64x64y16z128t, floating point). Here the x-axis represents the range [-10; 10] for the histogram of the wavelet coefficients. Sporadic behavior of the large-magnitude wavelet coefficients is clearly visible at the tails of this empirical distribution.

Figure 12

We fitted several heavy-tail distribution models to this frequency histogram of 4D wavelet coefficients. The results are shown on Figure 13. Figure 14 depicts the extreme left trail of the data distribution (note that the data is symmetric). The double-exponential distribution and the *T* distribution models appear to be asymptotically underestimating the tails of the data, but provide good fits around the mean. Cauchy, Bessel K forms and double-Pareto distribution models, in that order, provide increasingly heavier tails with Cauchy being the most likely candidate for the best fit to the observed wavelet coefficients across the entire range. The double-Pareto and the Bessel K forms densities provide the heaviest tails, however, they are inadequate in the central range [-3; 3] and undefined near the mean.

Figure 13

The parameters for these models are estimated using maximum likelihood. Here are the estimates for the models shown on Figures 13 and 14, see equations (12) through (17): Double-Exponential($\mu = 0$, $\beta = 15$), Double-Pareto($\alpha = 0.001$, $c = 8$, $\mu = 0$), Cauchy($\gamma = 13$, $\mu = 0$), T($df = 2$), Bessel K Form($p = 0.0001$, $c = 3 \times 10^6$, $\mu = 0$).

Figure 14

Table 1 illustrates different quantitative measures of goodness-of-fit to the histogram of the wavelet coefficients of the 4D fMRI data of a Gaussian and the five models shown on Figures 13 and 14. A smaller numeric value in this table indicates a better fit. It is important to point out that these numeric measures can not be used as a single decision making tool because the models are fit globally, where as their utilization is usually limited to the extreme tails. Hence, the ranking of the quality of fit reported in this table may be misleading. For example, according to the (KL2) Kullback-Leibler measure (Kullback 1951; Kullback 1959), the Gaussian model is a better (global) fit to the data than the heavy-tail models of Double-Pareto and Bessel K Forms. The same results are obtained by using the Bhattacharyya resistor-average distance (Bhattacharyya 1943; Kailath 1967), which is a symmetric alternative to the Kullback-Leibler divergence. To avoid the singularities for the Double-Pareto and the Bessel K Forms families at the origin these measures are computed on $|x|>0.1$. There is no best threshold value one can use to resolve these singularities, as this will depend on the *type I* error rate, in general. We have used 0.1 as a means to simply resolve the singularities, not to obtain a best tail

distribution model fit. Based on these results we have used the Cauchy and T distributions as heavy-tail models in practice.

Table 1

Discussion and Conclusions

Atlas-based studies: The main advantage in using a frequency-spatial F&A SVPA atlas is that small across-subjects spatial intensity variations, within each ROI, will have less of an effect on the statistical maps, as compared to other time-domain approaches (Friston 1995; Dinov, Mega et al. 2000). Note that for any two subjects the bumps and valleys of the signal may have similar frequency-distributional characteristics but be slightly offset in the spatial-domain. This approach also avoids the low-pass filtering of the data required by other such techniques (Friston 1995). The process of our frequency-spatial analysis begins by employing the N-dimensional discrete wavelet transform provided in the wavelet analysis of image registration package (Dinov and Sumners 2001). Then an appropriate distribution model is selected in a data-driven manner. And finally, the regional distribution of the wavelet parameters of the data is compared against the distribution of the across subject wavelet coefficients saved in the F&A SVPA atlas. One can perform both individual vs. atlas and group/population comparisons in this atlas-based framework.

Heavy-tailness: In general, heavy tail signals are more likely to exhibit large observations and oftentimes may have an impulsive nature. Our empirical results indicate that 3D MRI/fMRI and 4D fMRI data have leptokurtic sporadic nature. The term *heavy tail* refers to the fact that the probability density functions (histograms) of the signals have relative large mass in the extreme *tail*s. Because, the tails of leptokurtic distributions decay much slower than the Gaussian density many of the commonly used statistical tests are not appropriate for analyzing such data. Notably, in Figure 13 we see that the *P*-values corresponding to a statistical score of say –10.0 for the *T* and *Cauchy* distributions will be different by several orders of magnitude. This means that one is bound to commit *type I* error by not using the correct heavy-tail distribution model.

The quantitative assessment of the model fits presented in Table 1 may be augmented by looking at the quantile-quantile plots for the data vs. the model distributions. Quantile-quantile (QQ) plots (Evans 2000) are used to graphically depict the closeness of distributions. In our case we show in Figure 15 the similarity between the 4D sample wavelet coefficient distribution (vertical axes) and four of the heavy-tail distribution models used in the results section (horizontal axes). Cauchy, double-exponential, Gaussian and T distribution models are only shown as the double-Pareto and Bessel K Forms have singularities at their centers. These plots illustrate that, when studying the statistical properties of wavelet coefficients derived from MRI and fMRI data, heavy-tail distributions are better alternatives to Gaussian models, but still not perfect. One possible alternative approach will be to attempt fitting a mixture model of two or more components, some Gaussian and some heavy-tail.

fMRI Data Analysis Plan: In the second part of this investigation we will apply this methodology to analyze the fMRI data acquired and processed by (Buckner R. 2000). This dataset consisted of 41 right-handed, English-speaking gender-matched volunteers separated in 3 groups: 14 young participants (5 males mean age 21.1 ± 3 years); 14 old subjects (five males) were non-demented with (mean age of 74.9 ± 9 years) and; 13 old demented patients (six males) were demented (mean age of 77.2 ± 8 years). Older adults were excluded if they had neurologic, psychiatric, or medical illnesses that could manifestate as dementia. All older adults had normal (corrected) visual acuity. The Buckner's group employed the *T* and *F* statistics to assess for group effects (young adults, normal elderly and normal demented). These investigators found a significant difference between the groups of old demented and young adult subjects using Mann–Whitney nonparametric test. The individual subject and group analyses were performed in an event-related fashion using selective-averaging methods. To determine the added contribution of the second event in the two-trial conditions, the one-trial conditions were simply subtracted from the two-trial conditions. Secondly, Buckner and coworkers extracted the evoked hemodynamic responses for each subject from *a priori* defined regions and compared these using a random-effects statistical model. Peak activations in visual and motor cortex were then derived from this composite activation map. The hemodynamic response and variance were obtained for each region and each subject and entered into subsequent analyses based on analysis of variance.

In the atlas-based method we are proposing, an F&A SVPA atlas of one population will be generated and the individual fMRI data from the remaining two groups will be tested against the distributions of the wavelet coefficients of the atlas. The statistical test we employ will be quite distinct from the

general linear model used by others, because first our statistics will be computed on the data in compressed wavelet space. And secondly, our statistical tests will involve one (or several) heavy-tail distribution models as discussed in the Methods and Results sections.

Information Sharing Statement

All data presented in this manuscript are freely available from the ICBM (http://www.loni.ucla.edu/ICBM/) and fMRIDC (http://www.fmridc.org/) database resources. The software tools developed to model, analyze and visualize the data are also publicly available at http://www.loni.ucla.edu/CCB/Software/ & http://socr.stat.ucla.edu/. Specific requests and inquiries regarding the algorithms, processing and analysis protocols should be addressed to the corresponding author.

Acknowledgements

This research is supported by grants from NIA P50 AG16570, K08 AG100784; NLM R01 2R01 LM05639-06; NIH/NCRR 2 P41 RR13642, NIH/NIMH 5 P01 MN52176, NSF DUE 0442992 and NCBC U54 021813. We are also indebted to De Witt L. Sumners, Dimitre Stefanov, Anuj Srivastava, Jussi Tohka and Dushyanth Krishnamurthy for helpful hints and discussions. The authors also wish to acknowledge the contributions from the anonymous reviewers that made this manuscript more focused and complete.

References

Aguirre G., Z. E., D'Esposito, M. (1998). "A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data." Magn Reson Med **39**: 500-505.

Bandettini, P., Jesmanowicz, A., Wong, E. and Hyde, J. (1993). "Processing strategies for time-course data sets in functional MRI of the human brain." Magnetic Resonance in Medicine **30**: 161-173.

Bandettini, P., Wong, E. (1997). "A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI." NMR in Biomedicine **10**(4-5): 197-203.

Bayoumi, M. W. a. M. (2003). "Discrete Wavelet Transform: Architectures, Design and Performance Issues." Journal of VLSI Signal Processing Systems **35**(2): 155 - 178.

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions." Bull. Calcutta Math. Soc. **35**: 99–109.

Buckner R., S. A., Sanders A., Raichle M. and Morris J. (2000). "Functional Brain Imaging of Young, Nondemented, and Demented Older Adults." Journal of Cognitive Neuroscience **12(suppl. 2)**: 24-34.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996). "Statistical methods of estimation and inference for functional MR image analysis." Magnetic Resonance in Medicine **35**: 261-277.

Bullmore, E. a. F., J and Breakspear, M and Salvador, R and Suckling, J and Brammer, M. (2003). "Wavelets and statistical analysis of functional magnetic resonance images of the human brain." Stat Methods Med Res **375-99**(5): 375-99.

Bullmore, E. a. F., J and Maxim, V and Sendur, L and Whitcher, B and Suckling, J and Brammer, M and Breakspear, M (2004). "Wavelets and functional magnetic resonance imaging of the human brain." Neuroimage **23**(Suppl 1): S234-49.

Cao, J. a. W., K. (1999). "The detection of local shape changes via the geometry of Hotelling's T2 fields." Annals of Statistics **27**: 925-942.

Chan, J. C. G. a. A. K. (1999). Fundamentals of Wavelets: Theory, Algorithms, and Applications, Wiley-Interscience.

Chang, C. Y. and P. C. Chung (2000). "Two-layer competitive based Hopfield neural network for medical image edge detection." Optical Engineering **39**(3): 695-703.

Crabtree EC, M. M., Linshield C, Dinov ID, Thompson PM, Felix J, Cummings JL, Toga AW. (2000). "Alzheimer grey matter loss across time: unbiased assessment using a probabilistic Alzheimer brain atlas." Soc. for Neurosci. Abs. **26**: 294.

Daubechies, I. (1988). "Orthonormal bases of compactly supported wavelets." Comm. Pure & Appl. Math. **41**: 909-996.

Daubechies, I. (1991). The wavelet transform: a method for time-frequency localization. Advances in spectrum analysis and array processing. S. Haykin, Prentice-Hall**:** 366-417.

Daubechies, I. (1993). "Orthonormal bases of compactly supported wavelets, II. Variations on a theme." SIAM J. Math. Anal. **24**: 499-519.

Desco, M., J. A. Hernandez, et al. (2001). "Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation." Human Brain Mapping **14**(1): 16-27.

Desco, M., Hernandez, J., Santos, A. and Brammer, M (2001). "Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation." Human Brain Mapping **14**(1): 16-27.

Dinov, I. D., M. S. Mega, et al. (2001). "Construction of the first rest-state functional subvolume probabilistic atlas of normal variability in the elderly and demented brain." Neurology **56**(8): A248-A248.

Dinov, I. D., M. S. Mega, et al. (2000). "Analyzing functional brain images in a probabilistic atlas: a validation of subvolume thresholding." J Comput Assist Tomogr **24**(1): 128-38.

Dinov, I. D., M. S. Mega, et al. (2002). "Quantitative comparison and analysis of brain image registration using frequency-adaptive wavelet shrinkage." IEEE Trans Inf Technol Biomed **6**(1): 73-85.

Dinov, I. D., M. S. Mega, et al. (2002). "Quantitative comparison and analysis of brain image registration using frequency-adaptive wavelet shrinkage." Ieee Transactions on Information Technology in Biomedicine **6**(1): 73-85.

Dinov, I. D. and D. W. Sumners (2001). "Applications of frequency dependent wavelet shrinkage to analyzing quality of image registration." Siam Journal on Applied Mathematics **62**(2): 367-384.

Donoho, D. L. and I. M. Johnstone (1995). "Adapting to unknown smoothness via wavelet shrinkage." Journal of the American Statistical Association **90**(432): 1200-1224.

Donoho, D. L. and I. M. Johnstone (1998). "Minimax estimation via wavelet shrinkage." Annals of Statistics **26**(3): 879-921.

Donoho, D. L., I. M. Johnstone, et al. (1996). "Density estimation by wavelet thresholding." Annals of Statistics **24**(2): 508-539.

Evans, M., Hastings, N. and Peacock, B. (2000). Statistical Distributions. New York, Wiley.

Fadili, M., Bullmore, E (2002). "Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors." NeuroImage **15**: 217-232.

Feilner, M., Blu, T, Unser, M (2000). "Optimizing wavelets for the analysis of fMRI data." Proceedings of the SPIE - The International Society for Optical Engineering, SPIE-Int. Soc. Opt. Eng., **4119**(1-2): 626-37.

Friston, K., Holmes, AP, Worsley, KJ, Poline, J-B, Frith, CD and Frackowiak, RSJ (1995). "Statistical parametric maps in functional imaging: A general linear approach." Human Brain Mapping **2**: 189-210.

Golay, X., Kollias, S., Stoll, G., Meier, D., Valvanis, A. and Boesiger, P. (1998). "A new correlation-based fuzzy logic clustering algorithm for fMRI." Magnetic Resonance in Medicine **40**: 249-60.

Gourieroux, C. a. M., A. (1992). "Qualitative threshold ARCH models. (Autoregressive Conditional Heteroscedasticity) (ARCH Models in Finance)." Journal of Econometrics **v52**(n1-2): 159-200.

Grenander, U. a. S., A. (2001). "Probability Models for Clutter in Natural Images." IEEE Trans. Pattern Analysis and Machine Intelligence **23**(4): 424-429.

Hajnal, J., Myers, R., Oatridge, A., Schwieso, J., Young, I. and Byder, G. (1994). "Artifacts due to stimulus correlated motion in functional imaging of the brain." Magnetic Resonance in Medicine **31**: 283-291.

Hess, A., Stiller, D., Kaulisch, T., Heil, P., Schliech, H. (2000). "New insights into the hemodynamic blood oxygenation level-dependent response through combination of functional magnetic resonance imaging and optical recordings in gerbil barrel cortex." J. Neurosci. **20**: 3328-3338.

Hilton, M., Odgen, T, Hattery, D, Eden, G, Jawerth, J (1996). Wavelet Denoising of Functional MRI Data. Wavelets in Medicine and Biology. A. Aldroubi, Unser, M. Boca Raton, CRC Press**:** 93-114.

Kailath, T. (1967). "The divergence and Bhattacharyya distance measures in signal selection." IEEE Trans. on Comm. Tech. **COM-15**(1): 52–60.

Kershaw, J., Ardekani, BA, Kanno, I (1999). "Application of Bayesian inference to fMRI data analysis." IEEE Trans Med Imaging **18**: 1138-53.

Krakow, K. e. a. (1999). "EEG-triggered functional MRI of interictal epileptiform activity in patients with partial seizures." Brain **122**: 1679-1688.

Krakow, K. e. a. (2000). "EEG recording during fMRI experiments: image quality." Hum. Brain Mapp. **10**: 10-15.

Kullback, S. (1959). Information Theory and Statistics. New York, Wiley.

Kullback, S., Leibler, R. A. (1951). "On information and sufficiency." Ann. Math. Stat. **22**: 79–86.

Kuppusamy K., L. W. a. H. E. (1997). "Statistical assessment of cross-correlation and variance methods and the importance of electrocardiogram gating in functional magnetic resonance imaging." Magn. Reson. Imaging **15**: 169 -181.

Logothetis, N., Pauls, J., Augath, M., Trinath, T. and Oeltermann, A. (2001). "Neurophysiological investigation of the basis of the fMRI signal." Nature **412**: 150-157.

Mallat, S. (1989). "A theory for multi-resolution signal decomposition: The wavelet representation." IEEE-TPAMI **11**: 674-693.

Mazziotta, J., A. Toga, et al. (2001). "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)." Philos Trans R Soc Lond B Biol Sci **356**(1412): 1293-322.

Mazziotta, J. C., A. W. Toga, et al. (1995). "A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM)." Neuroimage **2**(2): 89-101.

Mega, M., Dinov, ID, Thompson, P, Manese, M, Lindshield, C, Moussai, J, Tran, N, Olsen, K, Felix, J, Zoumalan, C, Woods, RP, Toga, AW, Mazziotta, JC. (2005). "Automated Brain Tissue Assessment in the Elderly and Demented Population: Construction and Validation of a Sub-Volume Probabilistic Brain Atlas." NeuroImage(in press).

Mega, M. S., J. L. Cummings, et al. (2001). "Cognitive and metabolic responses to metrifonate therapy in Alzheimer disease." Neuropsychiatry Neuropsychol Behav Neurol **14**(1): 63-8.

Mega, M. S., I. D. Dinov, et al. (2000). "Orbital and dorsolateral frontal perfusion defect associated with behavioral response to cholinesterase inhibitor therapy in Alzheimer's disease." J Neuropsychiatry Clin Neurosci **12**(2): 209-18.

Mega, M. S., P. M. Thompson, et al. (2000). "The UCLA Alzheimer's atlas: Structural and functional applications." Annals of Neurology **48**(3): 427-427.

Mekle, R., Laine, A, Perera, G, DeLaPaz, R (2000). "Activation detection in fMRI data via multi-scale singularity detection." Proceedings of the SPIE - The International Society for Optical Engineering, SPIE-Int. Soc. Opt. Eng., **4119**(1-2): 615-625.

Meyer, F. (2003). "Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series." IEEE Trans Med Imaging **22**(3): 315-22.

Ogawa, S. L., T M; Kay, A R; Tank, D W. (1990). "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." PNAS **87**(24): 9868-9872.

Ruttimann, U. E., M. Unser, et al. (1998). "Statistical analysis of functional MRI data in the wavelet domain." Ieee Transactions on Medical Imaging **17**(2): 142-154.

Thompson, P., M., M. Mega, S., et al. (2000). "Early cortical change in Alzheimer's disease detected with a disease-specific, population-based, probabilistic brain atlas." Neurology **54**(7 Supp. 3): A475-A476.

Tsai, A., Fisher, J., Wible, C., Wells, W., Kim, J. and Willsky, A. Analysis of Functional MRI Data Using Mutual Information. Presented at Medical Imaging and Computer-Assisted Intervention - MICCAI'99, C. Taylor, A. Colchester, Eds. Cambridge, England, 473-480 (1999) (1999). Tsai, A., Fisher, J., Wible, C., Wells, W., Kim, J. and Willsky, A. Analysis of Functional MRI Data Using Mutual Information. Presented at Medical Imaging and Computer-Assisted Intervention - MICCAI'99, Cambridge, England.

Worsley, K. (1994). "Local maxima and the expected Euler characteristic of excursion sets of X2, F and T fields." Advances in Applied Probability **26**: 13-42.

Xu, Z. H. and A. K. Chan (2002). "Encoding with frames in MRI and analysis of the signal-to-noise ratio." Ieee Transactions on Medical Imaging **21**(4): 332-342.

Figures and Tables

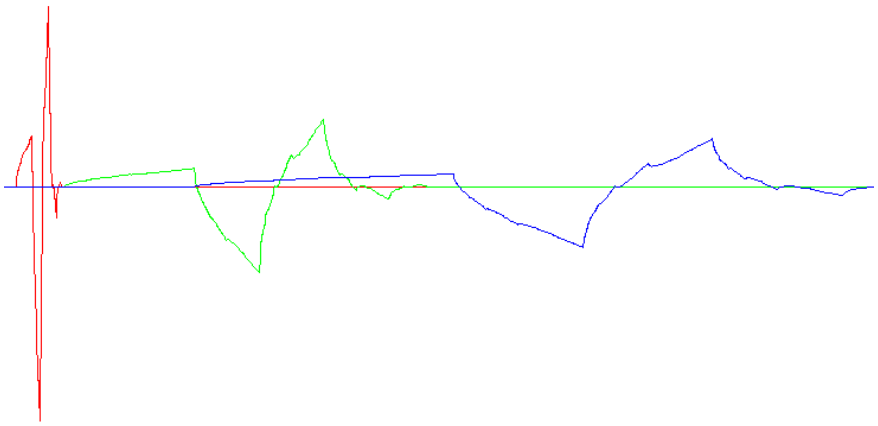

Figure 1. Shown here are three instances of Daubechies 4 wavelets. The differences in frequencies, spatial locations (domain) and range are clearly visible. The compact support, fast decaying and oscillatory properties of wavelets are common to all.

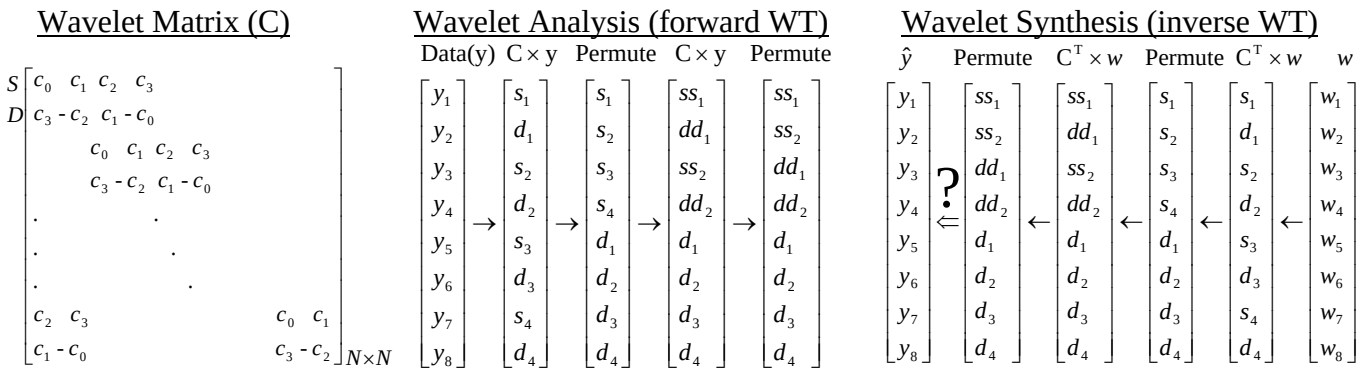

Figure 2: These three panels show the discrete WT matrix (left), the forward discrete WT (analysis) algorithm applied on a signal of size 2^3 (middle) and the inverse discrete WT (synthesis) used to obtain a reconstructed signal. The smoothing (S) and detail (D) factors are expressed as rows in the wavelet matrix (C). The iterative application of the matrix C, in the forward transform, produces the smoothing and detail coefficients s_i and d_i.

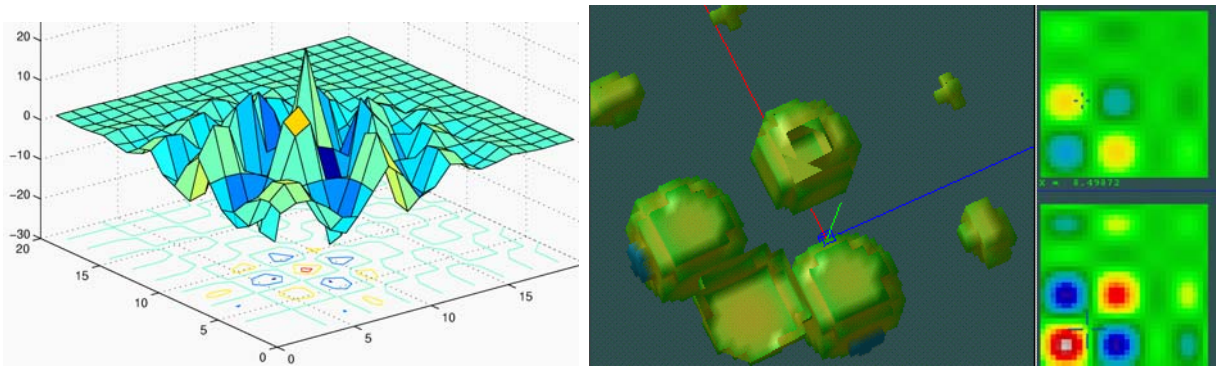

Figure 3. Examples of 2D (left panel) and 3D (right) renders of Daubechies wavelets illustrating the compact support, fast decaying and oscillatory properties of wavelets in general. The right panel shows a render of an iso-surface of the 3D wavelet with two cross-sections (axial and sagittal) shown on the side.

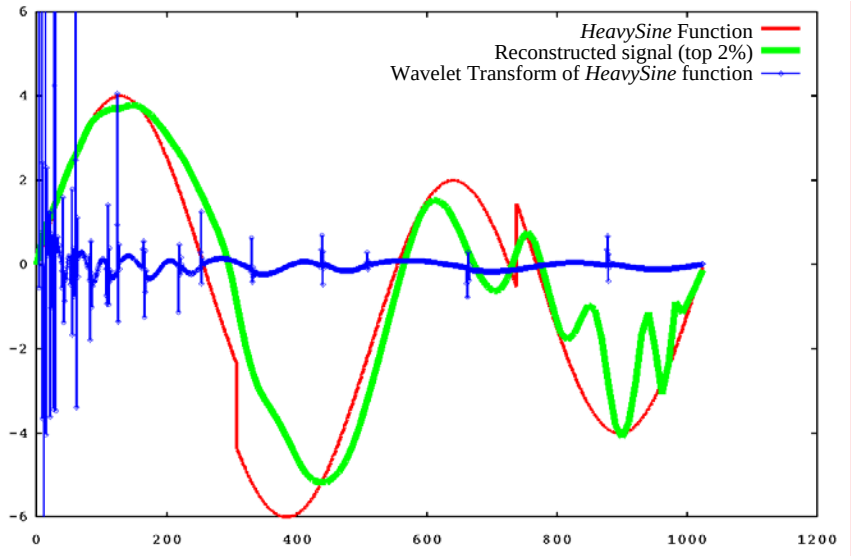

Figure 4. The three curves on this graph represent the original signal (*HeavySine* function, thin red curve), its wavelet transform (thin blue curve) and the reconstructed function estimator using only the largest 2% of the wavelet parameters (thick green curve). Note the space-frequency decorrelation of the original data in the wavelet-space (thin blue curve).

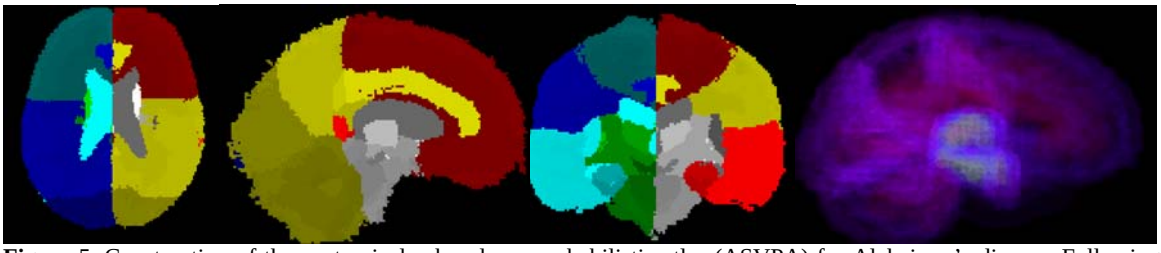

Figure 5. Construction of the anatomical sub-volume probabilistic atlas (ASVPA) for Alzheimer's disease. Following a co-registration of all data into a common space manual delineations are used to obtain subject-specific region labels. These individual labels are then combined to obtain the probability maps indicating likelihoods that voxels belong to any region of interest. Three left panels illustrate axial, sagittal and coronal slices through a one delineated brain and the right panel shows the probabilistic models in the anatomical SVPA for the main lobes, thalamus and hippocampus (hot colors indicate high probability and therefore larger chance that voxels belong to the corresponding region of interest.

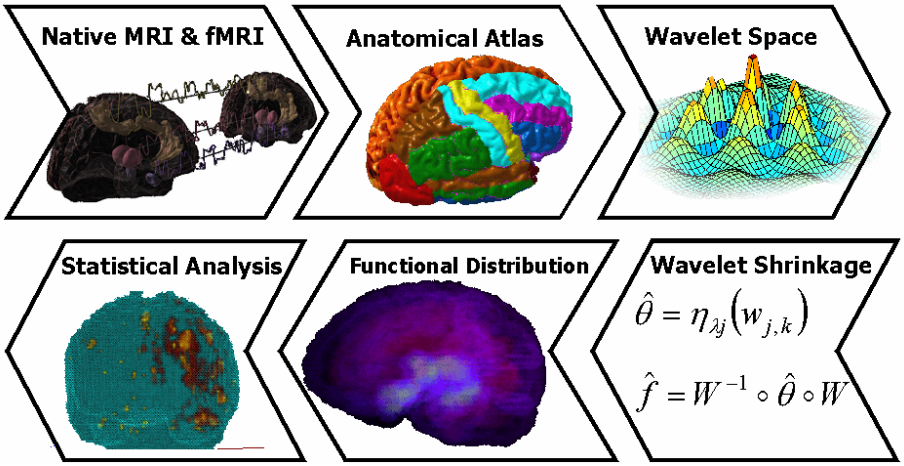

Figure 6. Design protocol for the wavelet-based construction and utilization of the fMRI functional atlas (F&A SVPA). The construction of the anatomical probabilistic atlas is augmented by including the regional distributions of the wavelet coefficients of the functional signals for one population. Statistical analysis of functional variability between a new fMRI volume and the F&A SVPA atlas is assessed following wavelet-space shrinkage.

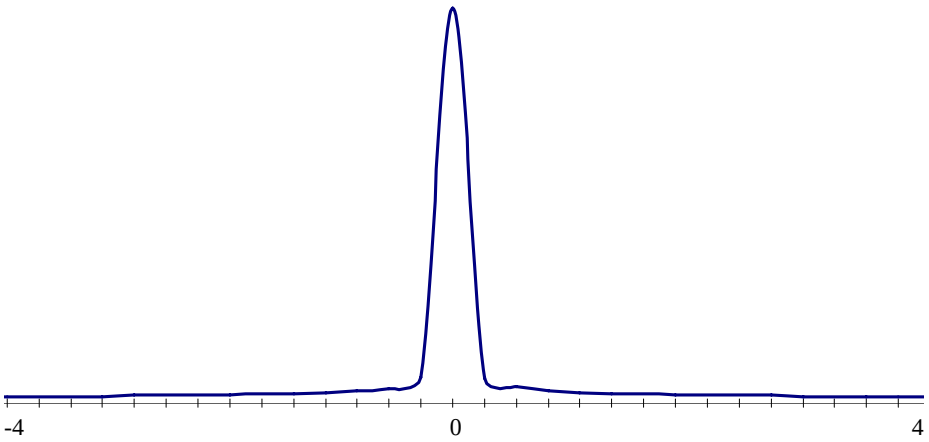

Figure 7. Displayed is the frequency histogram of the wavelet coefficients at 1,000 randomly selected locations (random indices of $w_{j,k}$), averaged across all 578 MRI volumes part of the ICBM database [Mazziotta et al., 1995]. Notably, most of the wavelet coefficients are near the origin, with some having sporadic, but large magnitudes. Heavy-tail distributions models will be appropriate for these data.

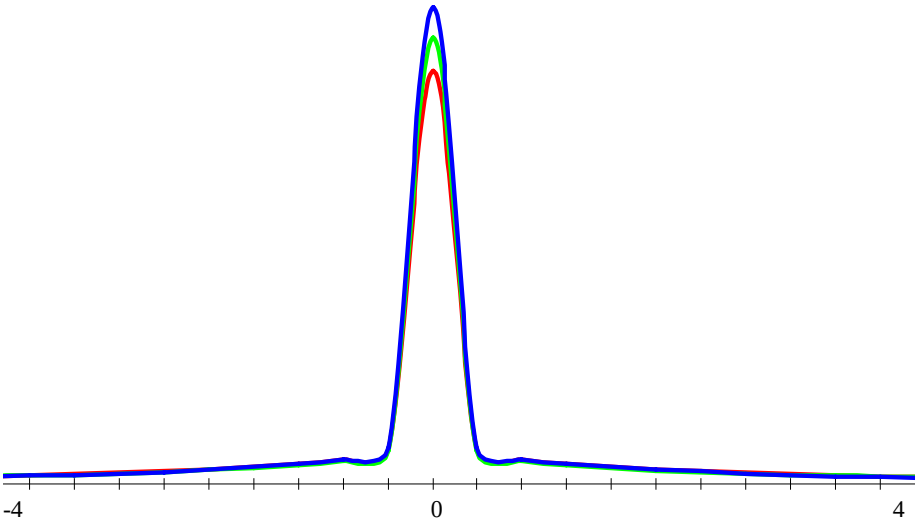

Figure 8. Shown here are the frequency distributions of the wavelet coefficients for three separate MRI volumes (randomly selected from the pool of 578). There is little variation between these three individuals, however the overall shape of the distribution of the magnitudes of the wavelet coefficients of each individual across the 1,000 random locations is more regular than the averaged distribution across subjects depicted in the previous **Figure** 6, still heavy-tailed.

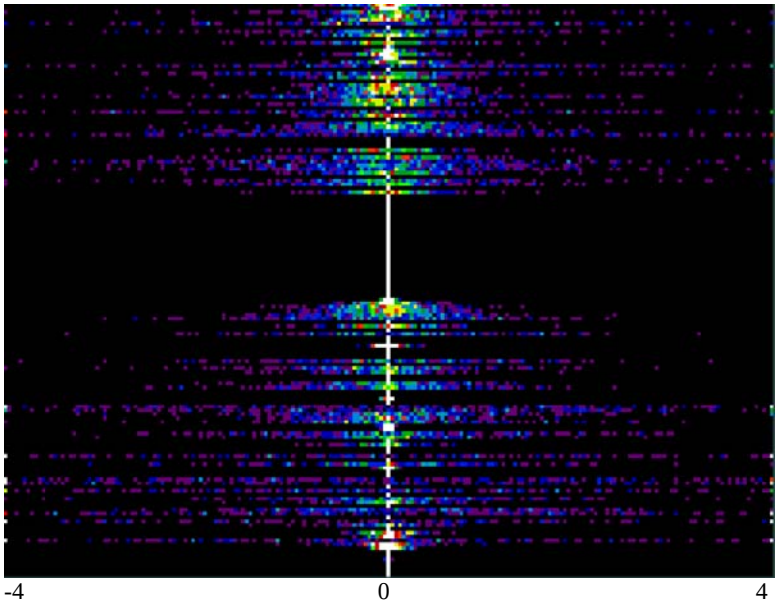

Figure 9. This image illustrates a portion of all of the individual distributions of the wavelet parameters for the 578 ICBM MRI volumes. The horizontal axis represents the magnitude of the wavelet coefficients in the range [-4 ; 4]. The vertical axis labels the voxel index and the row color map indicates the frequencies of occurrence of a wavelet coefficient of certain magnitude across all MRIs. Bright colors indicate high, and dark colors represent low, frequencies across subjects.

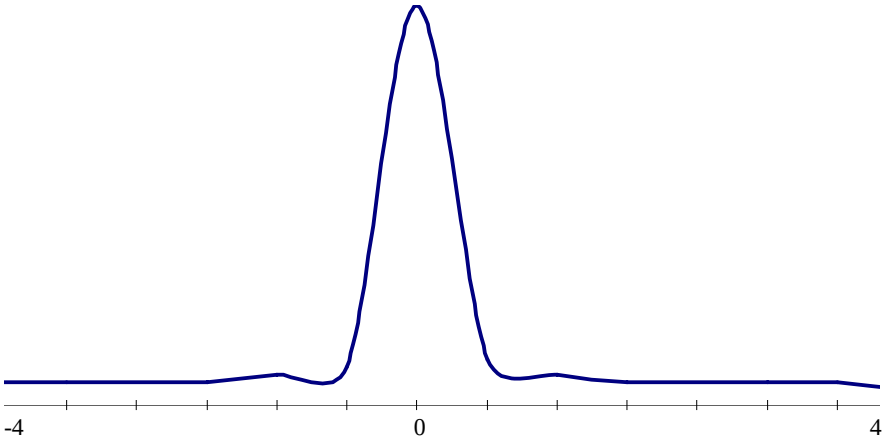

Figure 10. This diagram depicts the frequency histogram of the average magnitude of a wavelet coefficient, across all 128 fMRI 3D time-volumes of one subject fMRI run (part of the fMRI study of Buckner and colleagues [Buckner *et al*., 2000]) at 1,000 randomly selected locations (random indices of $w_{j,k}$). Again, we observe heavy-tailness of the data.

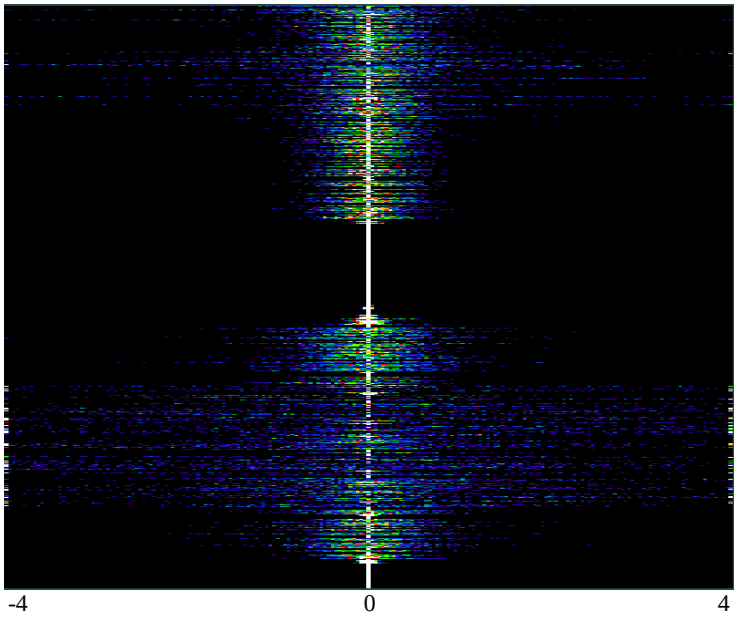

Figure 11. A 2D image displaying all of the individual distributions for each 128 3D time points (one run, six trials) of an fMRI timeseries. On the horizontal axis is the magnitude of the wavelet and the vertical axis labels the 1,000 voxel indices. Colors indicate the frequencies of occurrence of a wavelet coefficient of certain magnitude across all time point of the fMRI epochs. The across row average is effectively shown on Figure 9.

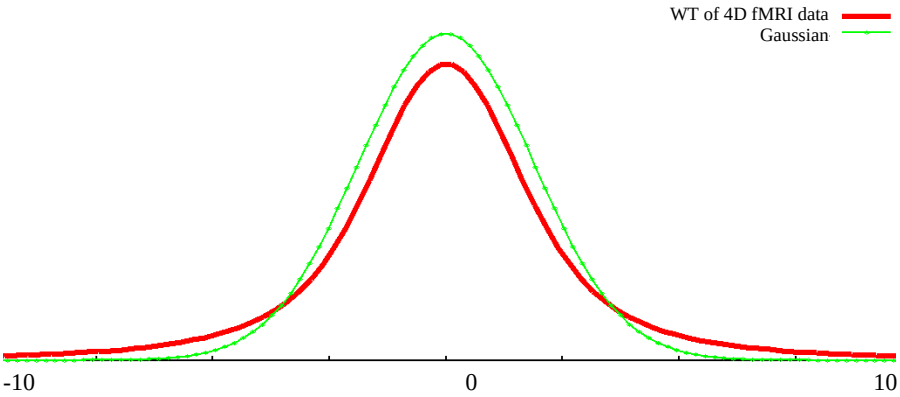

Figure 12. This graph shows the frequency distribution of the wavelet coefficients of the complete 4D WT of the fMRI timeseries. Note the slow asymptotic decay of the tails of this distribution (data-size: 64x64y16z128t, floating point). Here the x-axis represents the magnitude of the wavelet coefficients and the vertical axis indicates the frequency of occurrence of a wavelet coefficient of the given magnitude within the 4D dataset. Sporadic behavior of the large-magnitude wavelet coefficients is clearly visible at the tails of this empirical distribution.

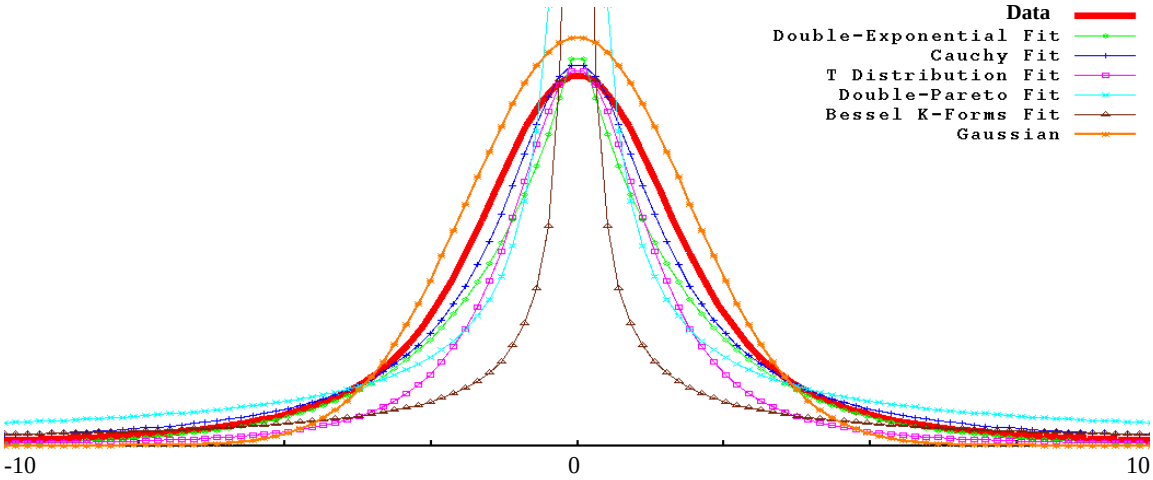

Figure 13. Several heavy-tail distribution models are fitted to the frequency histogram of 4D fMRI wavelet coefficients. These include *double-exponential*, *Cauchy*, *T*, *double-Pareto* and *Bessel K form* models. Because our statistical tests will be applied on the coefficients that survive wavelet shrinkage it is important to utilize a distribution model that provides accurate asymptotic approximation to the real data in the tail regions.

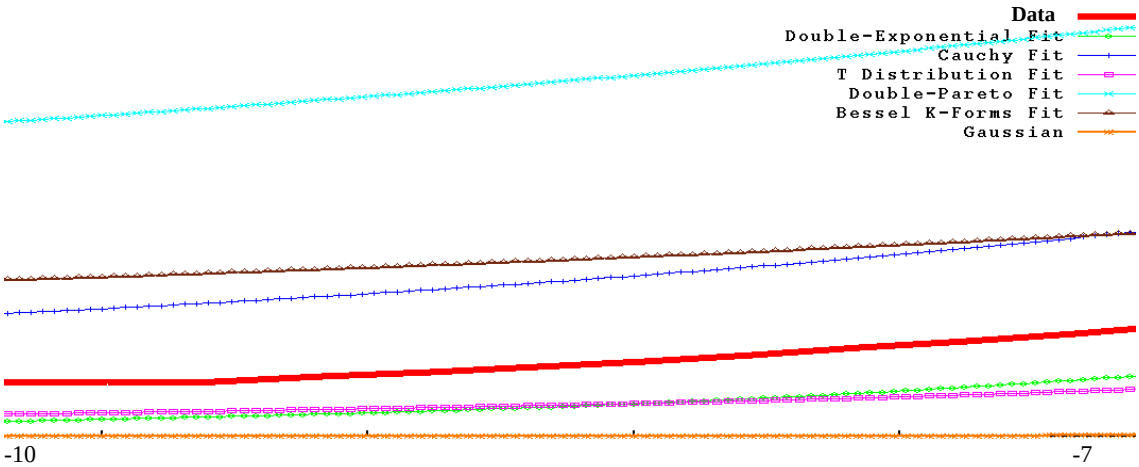

Figure 14. Shows the extreme left trail of the data distribution (data is symmetric). The *double-exponential* and the *T* distribution models underestimate the tails of the data asymptotically, but provide good fits around the mean. *Cauchy*, *Bessel K forms* and *double-Pareto* distribution models, in that order, provide increasingly heavier tails with *Cauchy* being the most likely candidate for the best fit to the observed wavelet coefficients across the entire range. The *double-Pareto* and the *Bessel K form* densities provide the heaviest tails, however, they are inadequate in the central range [-3 : 3] and undefined near the mean of zero.

Table 1: Quantitative assessment of the fit of five distributions models (and a Gaussian) to the histogram of the wavelet coefficients of the 4D fMRI data for one subject. Five different classifying functional are applied – Standard norms (L_1 and L_2), Kullback-Leibler (KL1 and KL2) divergence measures and Bhattacharyya resistor-average distance (BAT). A smaller numeric value indicates a better fit.

Models / Measure	Double Exponential	Cauchy	T	Double Pareto	Bessel K Form	Gaussian
L_1	1.3731	1.0619	2.2774	6.0515	5.0481	1.9669
L_2	0.1838	0.1133	0.2504	1.4800	0.7390	0.1941
KL1	0.0109	0.0187	0.0407	0.2579	0.2573	0.2957
KL2	0.0101	0.0217	0.0390	0.2980	0.2972	0.0729
BAT	0.0026	0.0050	0.0100	0.0703	0.0702	0.0283

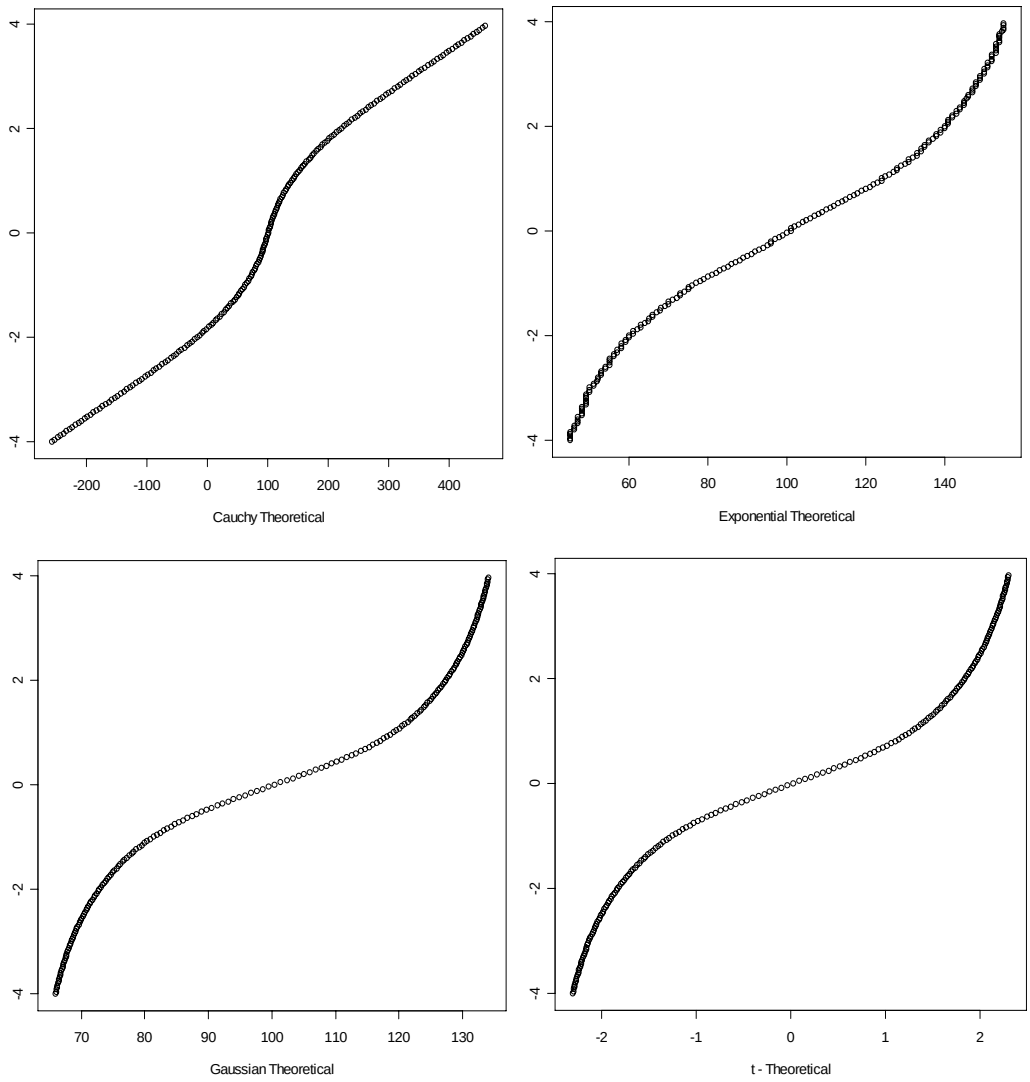

Figure 15: Quantile-quantile (QQ) plots illustrating graphically the quality of fit between the distribution models (horizontal axes) and the sample distribution of the 4D wavelet coefficient (vertical axes). Top row shows Cauchy and Double-Exponential and the bottom row shows the QQ plots for a Gaussian and T distribution models. None of these models yields a perfectly linear quantile relation indicating a mixture distribution may be another alternative for representing the wavelet coefficients of fMRI timeseries.