Statistical Analysis of fMRI Data using Wavelets in a Probabilistic Atlas Space

Ivo Dinov
Department of Statistics &
Laboratory of Neuro Imaging
UCLA School of Medicine
Los Angeles, CA 90095

dinov@loni.ucla.edu
http://www.loni.ucla.edu/~dinov

☺ IPAM BMI Summer School Quotes … ☺

I mean no disrespect with these quotes, on the contrary I find them impressive …
All quotes taken mostly out of context … Most quotes are paraphrased …

"Our Brain Mapping Approach is *Fuzzy …*"
"There are *no nice* flat-maps [preserving length/area]"
"All this is very *simple* and *clear* …"
"[One] can use Bayesian MAP estimation to solve *everything* [tissue, warp, bias] …"
"This is a *Brainless* talk …"
"[We] write *this* and *this* and apply *this* to get *this* …"
"I have *no idea* what this [slide] shows …"
"If *Matlab* can't do it it's not worth studying/trying …"
"We wanted to obtain an *MRI* and a *CRYO* volume of 1 *postdoc …*"
"Avoid any registration …"
"[Our] challenge is to keep a *baby quiet*, it'd *nothing to do with math/eng/neuroimg …*"

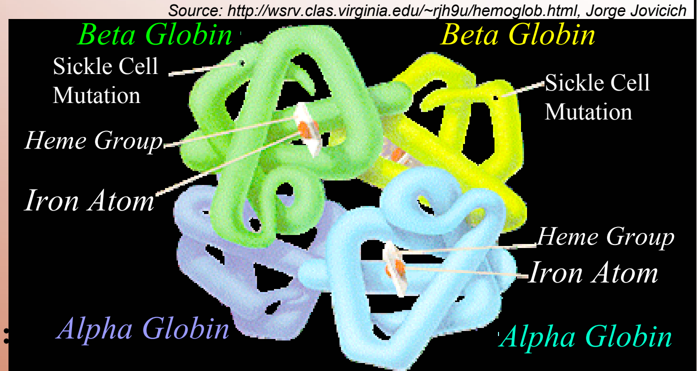
Work with:

- **W. John Boscardin**
  Department of Biostatistics
  UCLA School of Public Health
  http://rem.ph.ucla.edu/~johnb/
- **Elizabeth Sowell, Paul Thompson,**
  Roger Woods & Arthur Toga
  Department of Neurology, LONI, UCLA
  http://www.loni.ucla.edu/
- **Michael Mega,** Neural Net Research, Portland, OR
  http://www.loni.ucla.edu/~mega

UCLA, Ivo Dinov *Slide* **3**

Hemoglobin – a molecule to breathe with

Source: http://wsrv.clas.virginia.edu/~rjh9u/hemoglob.html, Jorge Jovicich



Hemoglobin (Hgb):
- four globin chains
- each globin chain contains a heme group
- at center of each heme group is an iron atom (Fe)
- to each heme group an oxygen atom (O_2) can attach
- oxy-Hgb (four O_2) is **diamagnetic** → **no** ΔB (net magnetization) effects
- deoxy-Hgb is **paramagnetic** → if [deoxy-Hgb] ↓ → local ΔB ↓.

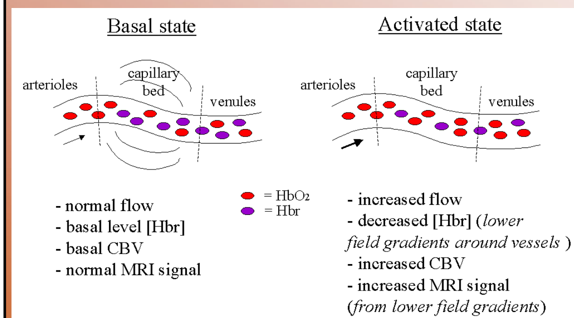
like aluminum/platinum, deoxy-Hgb, **has small positive magnetic susceptibility**

Slide **4** *UCLA, Ivo Dinov*

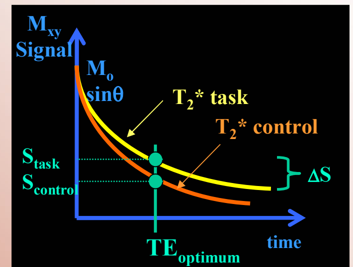
BOLD signal

Blood Oxygen Level Dependent signal

↑neural activity ➔ ↑ blood flow ➔ ↑ oxyhemoglobin ➔ ↑ T2* ➔ ↑ MR signal

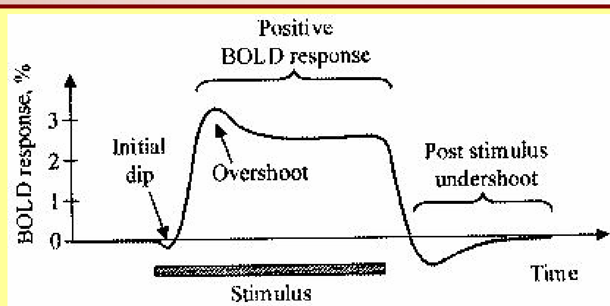


Source: fMRIB Brief Introduction to fMRI



Source: Jorge Jovicich

Hemodynamic Response Function



% signal change
= (point – baseline)/baseline
usually 0.5-3%

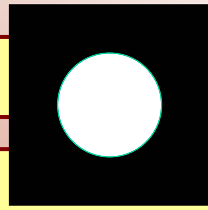
initial dip
-more focal and potentially a better measure
-somewhat elusive so far, not everyone can find it

time to rise
signal begins to rise soon after stimulus begins

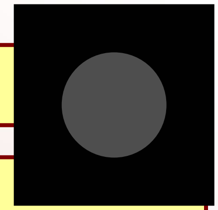
time to peak
signal peaks 4-6 sec after stimulus begins

post stimulus undershoot
signal suppressed after stimulation ends

BOLD Contrast



Image

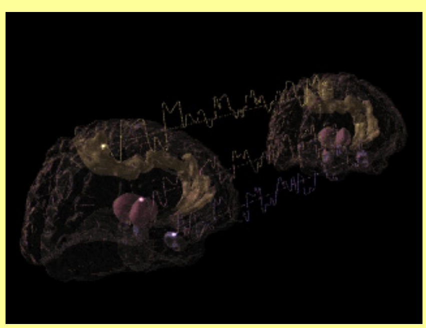


Image

- *Haemoglobin* acts as an endogenous intravascular contrast agent.
- As the level of oxygenation changes, so too does the contrast in the images.

Oxy-haemoglobin | Deoxy-haemoglobin | Oxy-haemoglobin | Deoxy-haemoglobin

fMRI Simulation movie



Problems with BOLD

- How localized is the BOLD response to the site of neuronal activity?
  - Is the signal from draining veins rather than the tissue itself?
- How is the signal change coupled to neuronal activity?
  - Do changes in timing of BOLD responses reliably tell us about changes in timing of neural activity?

Activation Imaging using BOLD

- **Resting** state versus **Active** state
  - e.g. Finger tapping, word recognition.
- Whole brain scanned in ~3-5 seconds using a high speed imaging technique (EPI).
- Perform analysis to detect regions which show a signal increase in response to the stimulus.

fMRI Data Analysis Tools (non exclusive!)

- Brain Voyager http://www.brainvoyager.de/
- CARET http://brainmap.wustl.edu/resources/caretnew.html
- FSL http://www.fmrib.ox.ac.uk/fsl/
- SCIrun http://software.sci.utah.edu/scirun.html
- LONI Pipeline http://www.loni.ucla.edu
- McGill BIC www.**bic**.mni.**mcgill**.ca/usr/local/matlab5/toolbox/fmri/
- AFNI http://afni.nimh.nih.gov/afni/
- SPM http://www.fil.ion.ucl.ac.uk/spm/

Slide **11** UCLA, Ivo Dinov

Hypotheses vs. Data driven Approaches

Hypothesis-driven

Examples: t-tests, correlations, general linear model (GLM)

- a priori model of activation is suggested
- data is checked to see how closely it matches components of the model
- most commonly used approach (e.g., SPM)

Data-driven – Independent Component Analysis (ICA)

- No prior hypotheses are necessary
- Multivariate techniques determine the patterns in the data that account for the most variance across all voxels
- Can be used to validate a model (see if the math comes up with the components you would've predicted)
- Can be inspected to see if there are things happening in your data that you didn't predict
- Can be used to identify confounds (e.g., head motion)
- Need a way to organize the many possible components
- New and upcoming

Slide **12** UCLA, Ivo Dinov

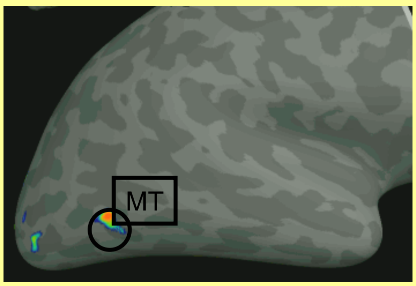
Two approaches: Global vs. Regional

A. Region-Of-Interest (ROI) approach (e.g., LONI's Sub-Volume Thresholding)

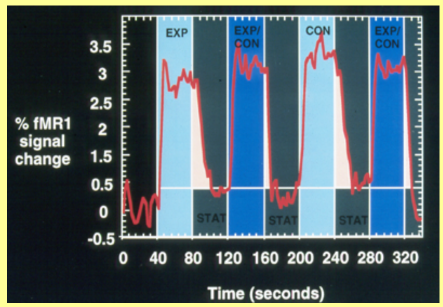
1. A localizer run(s) to find a region (e.g., show moving rings to find middle temporal MT area)
2. Extract time course information from that region in separate independent runs
3. See if the trends in that region are statistically significant
   The runs that are used to generate the area are independent from those used to test the hypothesis.

Example study: Tootell et al, 1995, Motion Aftereffect

Localize "motion area" MT (V5) in a run comparing **moving** vs. **stationary** rings



Extract time courses from MT in subsequent runs while subjects see illusory motion (motion aftereffect)



Two Approaches: Whole Brain Stats

B. Whole volume statistical approach

1. Make predictions about what differences you should see if your hypotheses are correct
2. Decide on statistical measures to test for predicted differences (e.g., t-tests, correlations, GLMs)
3. Determine appropriate statistical threshold
4. See if statistical estimates are significant

Statistics available

1. T-test
2. Correlation
3. Frequency-Based (Fourier/Wavelet/Fractal) modeling
4. General Linear Model

-overarching statistical model that lets you perform many types of statistical analyses (incl. correlation/regression, ANOVA)

Source: Tootell et al. 1995

Why do we need statistics?

- fMRI intensities essentially represent a random field
  - -variation caused by scanner (magnet, coil, time)
  - -variation in neurophysiology (area, task, subject)
- The rest-state scans used against activation tasks in *subtraction paradigm* setting to correct for some of these effects within the same run.
- Statistical analyses help us confirm whether the *eyeball tests of significance based on visual thresholding are real/correct*.
- Because we do so *many comparisons* (~10^6), we need a way to compensate (increase of Type I error).

fMRI of Young, Nondemented and Demented Adults

Paradigm: Event-related design to assess between-population differences of *amplitude* and *variance* of hemodynamic response to visual stimulus (14 young, 14 old nonAD and 13 AD subjects)

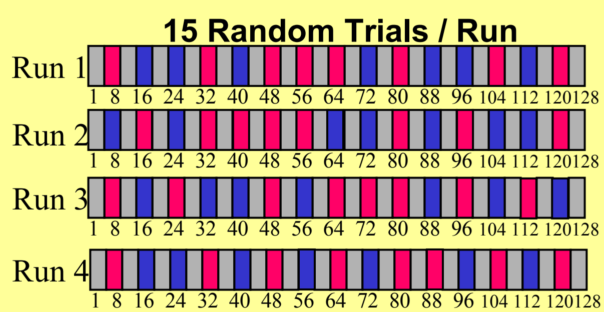
Buckner *et al.* *J. Cog. Neurosci.* 2000

- blank
- 1-trial condition–1.5 second 8Hz B/W checkerboard flicker
- 2-trial condition–2 visual stimuli 5.36 sec apart

Stimuli required a right hand index-finger click

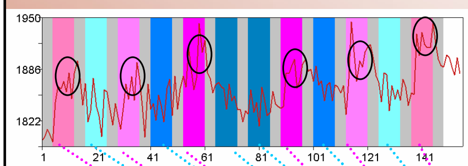
16 coronal slices
64x64 in-plane voxels
volume time = 2 sec
3 x 3 x 5 mm^3 voxels

4 runs per Subject
128 volumes per run
2 (randomly chosen) conditions (always 7+8)
1 Trial = 21.44 sec (8 vol's)
Total 60 Trials/subject



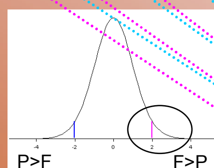
One Approach – Voxel-Based T-tests

Look for differences b/w One-Trial & Two-Trial Activation, for a given voxel:

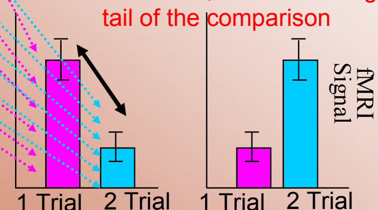


Measure average MR signal and SD for each volume in which 1-Trial stimuli were presented (7-8 epochs x 8 volumes/epoch = 56-64 volumes).

Measure average MR signal and SD for each volume in which 2-Trial stimuli were presented (56-64 volumes).



To look for OneTrial – TwoTrial Activation, look at the negative tail of the comparison

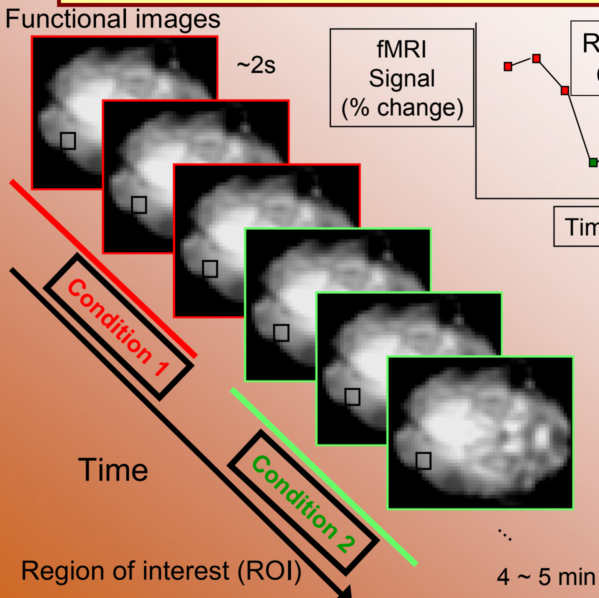


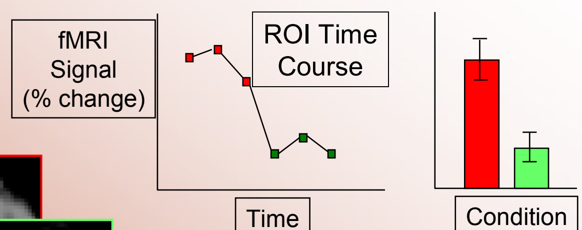
Stat-Significant Mean difference?
Calculate T_o value. Look up p value for that number of degrees of freedom (df >= 56 x 2 = 112).
e.g., For ~112 df $T_o > 1.98 \rightarrow p < 0.05$
$T_o > 3.39 \rightarrow p < 0.001$
Repeat this process 65,536 more times (64x64x16), once for each voxel.

If p=0.001 and 65,536 voxels, 65,536*0.001 ➔ 66 voxels could be significant purely by chance

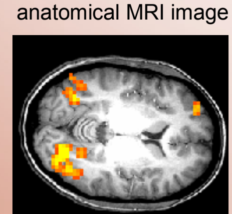
Slide **18** *UCLA, Ivo Dinov*

Activation Statistics

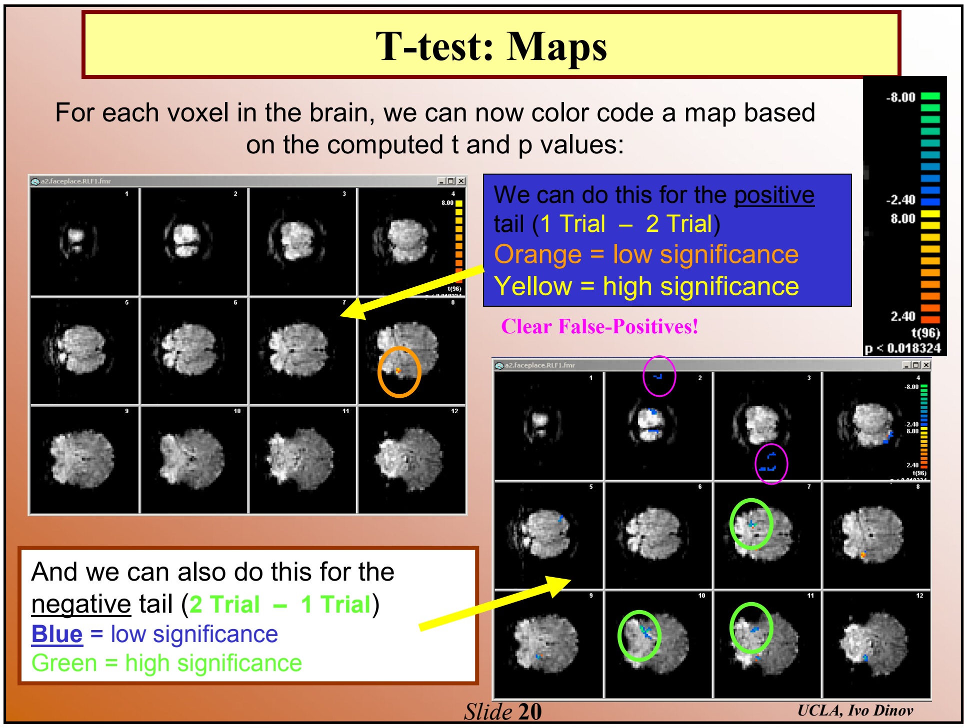


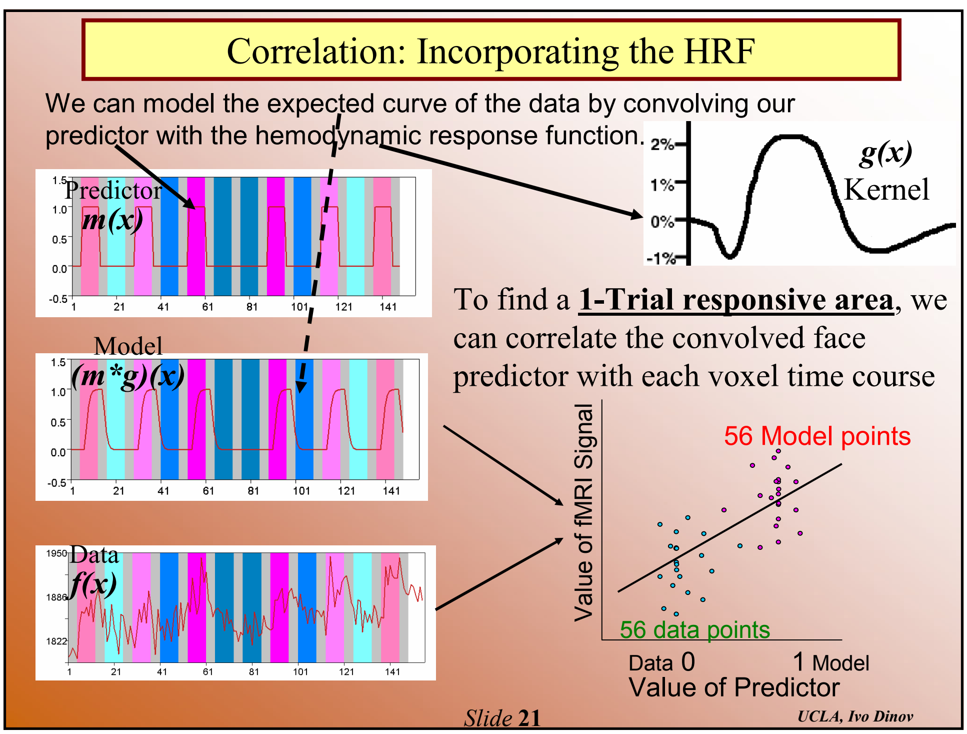


Statistical Map
superimposed on anatomical MRI image



Slide **19** *UCLA, Ivo Dinov*


T-test: Maps
For each voxel in the brain, we can now color code a map based on the computed t and p values:
We can do this for the positive tail (1 Trial – 2 Trial)
Orange = low significance
Yellow = high significance
Clear False-Positives!
-8.00
-2.40
8.00
2.40
t(96)
p < 0.018324
And we can also do this for the negative tail (2 Trial – 1 Trial)
Blue = low significance
Green = high significance
Slide 20
UCLA, Ivo Dinov


Correlation: Incorporating the HRF
We can model the expected curve of the data by convolving our predictor with the hemodynamic response function.
Predictor
m(x)
Model
(m*g)(x)
Data
f(x)
g(x)
Kernel
2%
1%
0%
-1%
To find a 1-Trial responsive area, we can correlate the convolved face predictor with each voxel time course
56 Model points
56 data points
Value of fMRI Signal
Data 0
1 Model
Value of Predictor
Slide 21
UCLA, Ivo Dinov

Two Main fMRI Designs

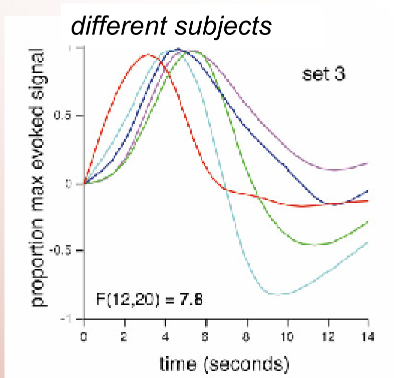
Block design
- Compare long periods (e.g., 16 sec) of one condition with long periods of another
- Traditional approach
- Most statistically powerful approach
- Less dependent on how well you model the HRF (hemo)
- Habituation effects?!?

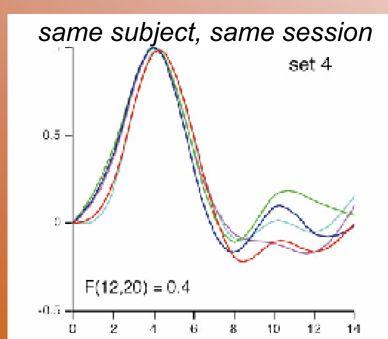
Event-related design
- Compare brief trials (1 sec) of 1 condition with brief of another
- Newer approach (1997+)
- Less statistically powerful but has some advantages
- Trials can be well-spaced to allow the MR signal to return to baseline b/w trials (e.g., 21+ sec apart) or closely spaced (e.g., every 2 sec)

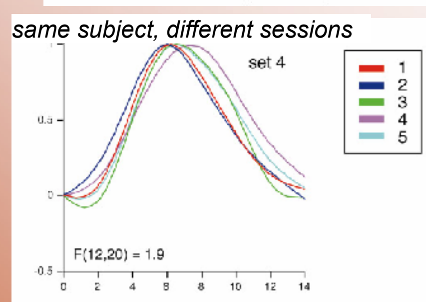
Variability of HRF: Evidence

Aguirre, Zarahn & D'Esposito, 1998
- HRF shows considerable variability between subjects

- Within subjects, responses are more consistent, although there is still some variability between sessions

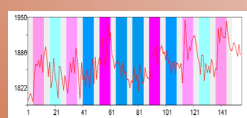






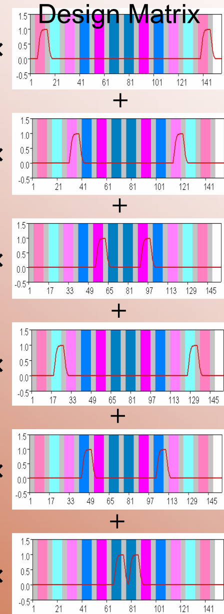
General Linear Model for fMRI

Parse out variance in the voxel's time course to the contributions of six predictors plus residual noise (what predictors can't account for).



fMRI signal

$=$



$\beta_1 \times$ + $\beta_2 \times$ + $\beta_3 \times$ + $\beta_4 \times$ + $\beta_5 \times$ + $\beta_6 \times$

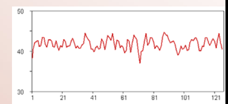
Examples:

Stimulus

Subject

Run

+



Trial

residuals

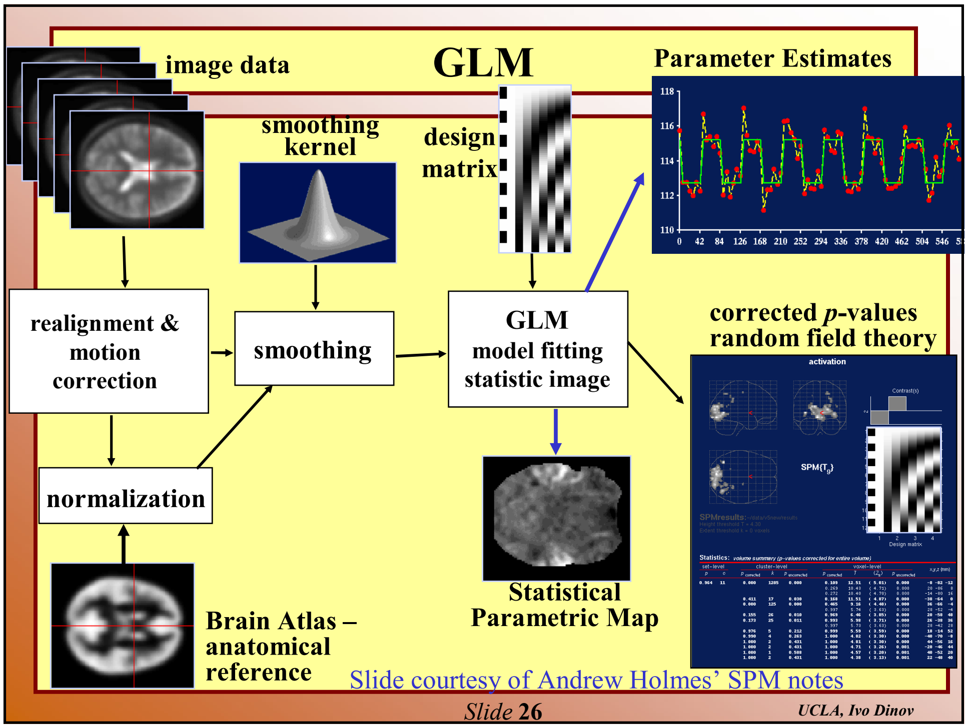
Group (AD/Young)

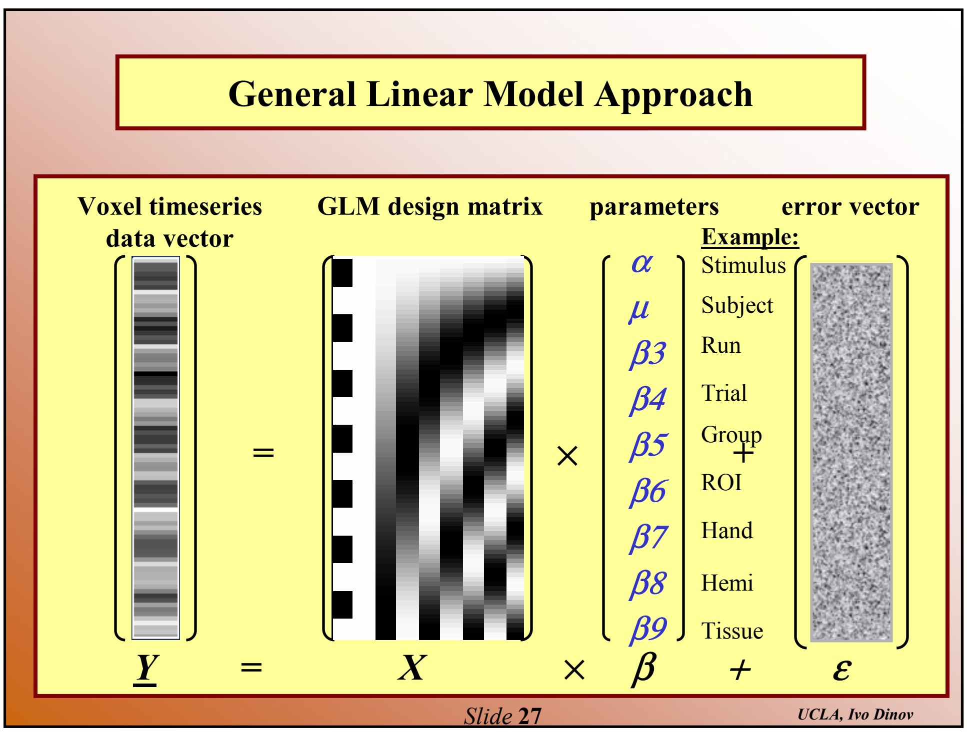
ROI

Adapted from Brain Voyager course slides

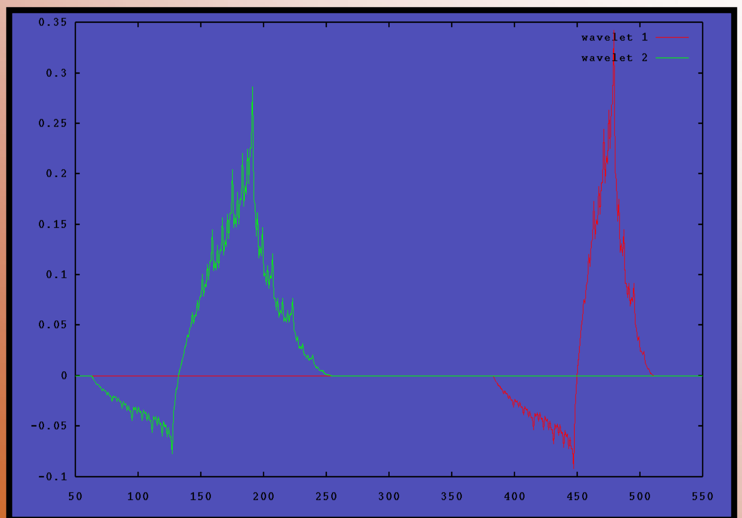
Advantages of General Linear Model (GLM)

- Can perform data analysis within and between subjects without the need to average the data itself
- Allows you to counterbalance random stimuli orders
- Allows you to exclude segments of runs with artifacts
- Can perform more sophisticated analyses (e.g., 2 factor ANOVA with interactions)
- Easier to work with (do one GLM vs. many T-tests and/or correlations)


GLM
image data
Parameter Estimates
smoothing kernel
design matrix
realignment & motion correction
smoothing
GLM model fitting statistic image
corrected p-values random field theory
normalization
Statistical Parametric Map
Brain Atlas – anatomical reference
Slide courtesy of Andrew Holmes' SPM notes
Slide 26
UCLA, Ivo Dinov

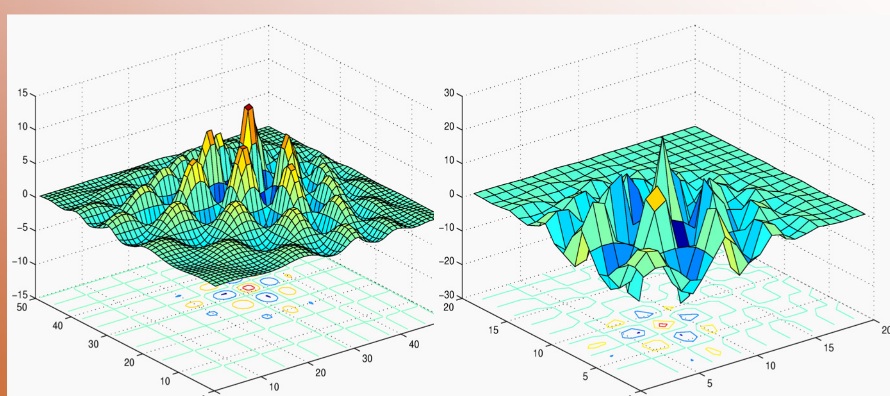

General Linear Model Approach
Voxel timeseries data vector
GLM design matrix
parameters
error vector
Example:
α
μ
β3
β4
β5
β6
β7
β8
β9
Stimulus
Subject
Run
Trial
Group
ROI
Hand
Hemi
Tissue
=
×
+
Y = X × β + ε
Slide 27
UCLA, Ivo Dinov

Completeness and Efficiency of Signal Representation – Wavelet functions 1D



Slide **28** *UCLA, Ivo Dinov*

Completeness and Efficiency of Signal Representation – Wavelet functions 2D



A 2D Daubechies wavelet illustrating the compact support, fast decaying and oscillatory properties of wavelets.

Slide **29** *UCLA, Ivo Dinov*

Visual Wavelets

If Time Permits:

C:\Ivo.dir\Research\movies\WaveletMovie.mpg

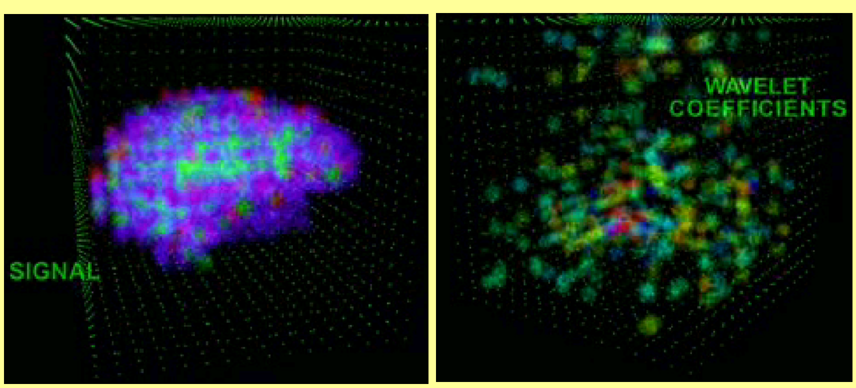
Online At:

http://www.loni.ucla.edu/Software/

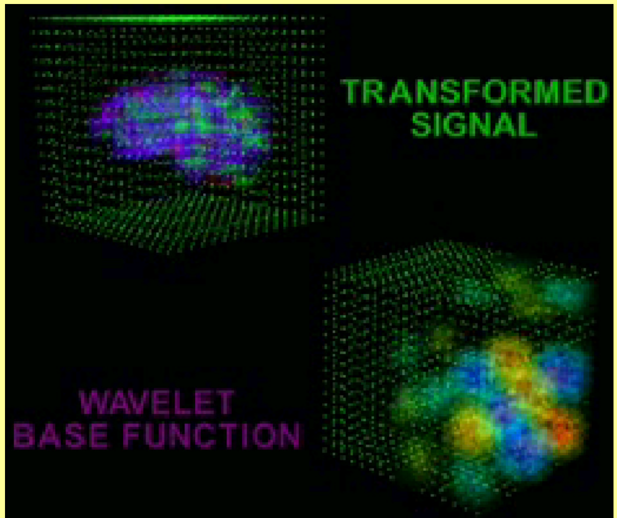
http://www.loni.ucla.edu/~dinov/WAIR.html

Completeness and Efficiency of Signal Representation – Wavelet functions 3D

Signal & it's 3D WT

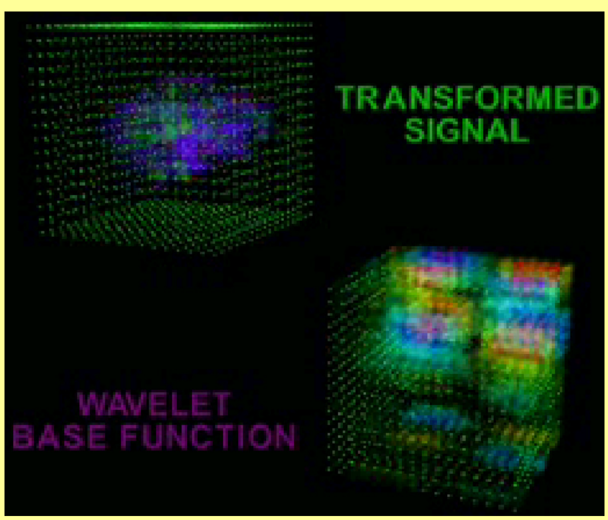


Completeness and Efficiency of Signal Representation – Wavelet functions 3D



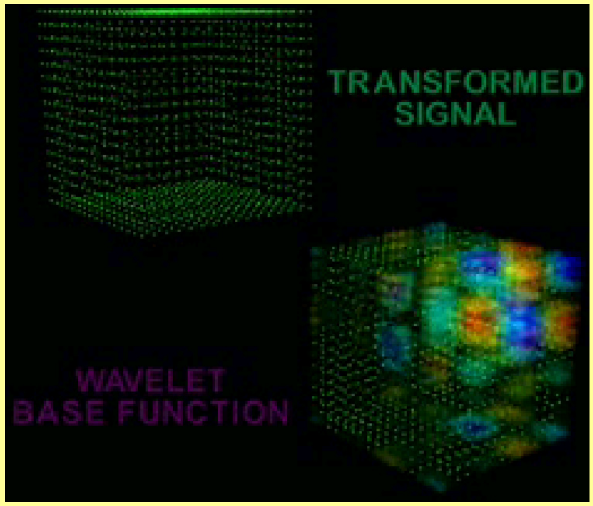
Slide **32** *UCLA, Ivo Dinov*

Completeness and Efficiency of Signal Representation – Wavelet functions 3D



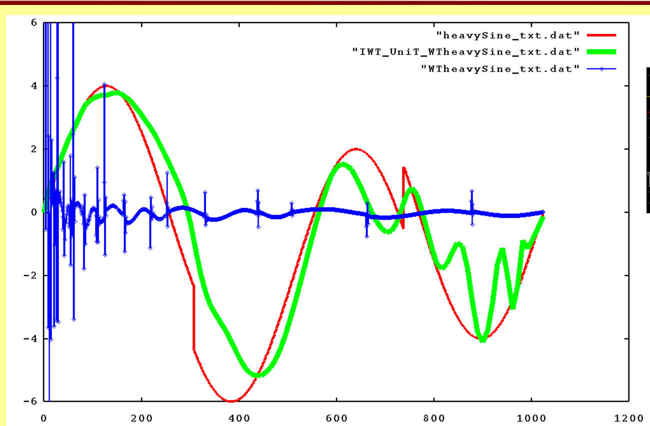
Slide **33** *UCLA, Ivo Dinov*

Completeness and Efficiency of Signal Representation – Wavelet functions 3D



Slide **34** *UCLA, Ivo Dinov*

Wavelets http://socr.stat.ucla.edu/



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. Note the space-frequency decorrelation of the original data in the wavelet-space (blue curve).

Slide **35** *UCLA, Ivo Dinov*

Wavelet-space Shrinkage

- **Wavelet Denoising:** $y_i = f(x_i) + e_i,\ 0 \le i \le N-1,\ \ e_i \sim N(0,\sigma^2) IID.$

If W is the WT operator and d is a subset of $\{w_{j,k}\}$, $w = Wy = Wf + We_i$ and $\hat{f} = T_{SW}(y,d) = \sum_{(j,k)\in d} w_{j,k} W_{j,k}(x)$, *Selected Wavelet* reconstruction.

Ideal Risk: For f an L-piece-wise poly of degree D:

$$\Re_{N,\sigma}(SW,f) = \inf_d \{R_{N,\sigma}(T_{SW}(y,d),f)\} =$$

$$\inf_d \left\{ \frac{1}{N} E\left(\sum_{i=0}^{N-1} (T_{SW}(y,d)(x_i) - f(x_i))^2 \right) \right\} =$$

$$= \begin{cases} \frac{L(D+1)\sigma^2}{N}, & known-breaks \\ O\left(\frac{\ln N \sigma^2}{N}\right), & unknown-breaks \end{cases}$$

Inf **attainable?**

Wavelet-space Shrinkage

- Spatial- vs. Frequency-Adaptive wavelet shrinkage

$$\eta_{\lambda_{j,k}}(w_{j,k}) = sgn(w_{j,k})\left[|w_{j,k}| - \lambda_{j,k}\right]_+, \text{ where } \lambda_{j,k} = \begin{cases} \lambda^{DJ} = & \sigma\sqrt{2\ln N} \\ \lambda^{DS}_{j,k} = & \sigma\sqrt{2\ln(2n_j+4)} \end{cases}$$

If $\hat{\theta}^{DJ}_{j,k} = \eta_{\lambda^{DJ}_{j,k}}(w_{j,k})$, $\hat{\theta}^{DS}_{j,k} = \eta_{\lambda^{DS}_{j,k}}(w_{j,k})$ and $\hat{f} = W^{-1}(\hat{\theta}(W(y)))$ then

$$R(\hat{f}_{DJ},f) = R(\hat{\theta}^{DJ},\theta) \le (1 + 2\ln N)\left(\frac{\sigma^2}{N} + \Re_{N,\sigma}(SW,f)\right)$$

$$R(\hat{f}_{DS},f) = R(\hat{\theta}^{DS},\theta) \le (1 + 2\ln(N+4))\left(\frac{\ln N\sigma^2}{2N} + \Re_{N,\sigma}(SW,f)\right)$$

in essence an *optimal* estimator.

- Optimal Function Estimators

If $\hat{\theta}^{DJ}_{j,k} = \eta_{\lambda^{DJ}_{j,k}}(w_{j,k})$, $\quad \inf_{\hat{\theta}} \sup_{\theta\in R^N} \frac{\frac{1}{N}E(\|\hat{\theta}-\theta\|^2)}{\frac{\sigma^2}{N} + \Re_{N,\sigma}(SW,f)} \sim 2\ln N.$

If $\hat{\theta}^{DS}_{j,k} = \eta_{\lambda^{DS}_{j,k}}(w_{j,k})$, $\quad \inf_{\hat{\theta}} \sup_{\theta\in R^N} \frac{\frac{1}{N}E(\|\hat{\theta}-\theta\|^2)}{\sigma^2\frac{\ln N}{2N} + \Re_{N,\sigma}(SW,f)} \sim \frac{4}{3}\ln N.$

Wavelet-space Shrinkage - What function spaces are these estimates optimal for?

<u>MiniMax</u> attaining estimators over scale of function classes, F, *Risk* $R(N,F) = inf_{f^\wedge}\, sup_{f \text{ in } F}\, R(f^\wedge,f)$.

<u>E.g., L^2 Sobolev spaces</u>

$$W_2^m(C) = \left\{ f \;:\; \|f\|_2^2 + \left\| \frac{d^m f}{dt^m} \right\|_2^2 \le C^2 \right\}$$

(parameter=m). More generally, for $L^p \rightarrow W^m_p(C)$.

Virtually all of our data is represented by signals belonging to L^2 Sobolev spaces with m<4

Wavelets & PDE's

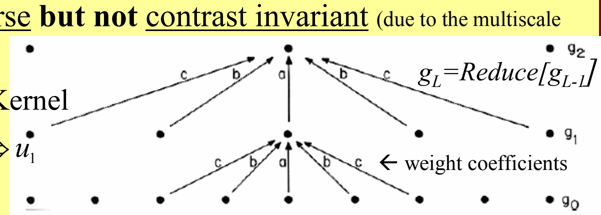
- Burt and Adelson introduced the pyramidal algorithm for computing the Laplacian [Burt and Adelson, *The Laplacian Pyramid as a Compact Image Code*. IEEE Trans. on Comm., 9(4): pp. 532-540, 1983].
- Procedure: Use **heat equation** to numerically (de)blur an image. Iteratively, convolve the image with a smoothing kernel, then subsample ($\downarrow 2$) and compute the difference, a Laplacian. The result is a faded version of the original signal. This pyramidal representation is <u>sparse</u> **but not** <u>contrast invariant</u> (due to the multiscale subsampling.)

$$\frac{\partial u}{\partial t} = \Delta u; \quad K = \text{Isotropic Kernel}$$

$$u_o \Rightarrow \hat{u}_1 = \underline{u_o * K} \Rightarrow \underline{\downarrow 2} \Rightarrow u_1$$

$$u_1 \underset{\text{Details}}{\overset{\text{KEEP}}{\longrightarrow}} u_o - u_1 \sim \Delta u_o$$

$$(wavelet \quad coeff's). \quad \text{Repeat!}$$



Ref. Gabor, 1960, Morel, 2004

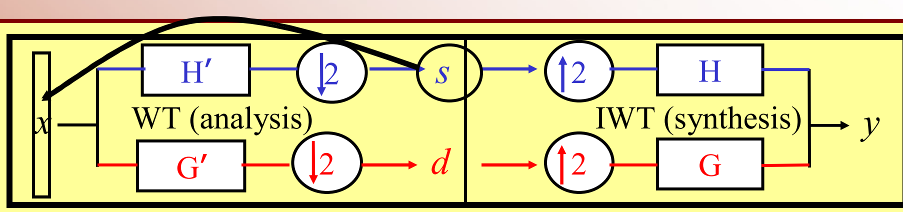
Wavelets & PDE's

- Marr's *edge detector* (1980's) is to use **second derivative** to *locate* the maxima of the first derivative (which the edge contours pass through).

- *Haar Basis* (1909) encodes the edges into image representation via the first derivative operator (i.e. moving average/difference):

$$\left(\dots x_{2n}, x_{2n+1}, \dots\right) \xleftrightarrow{\;Haar\;} \left(\dots a_n = \frac{x_{2n}+x_{2n+1}}{2}, d_n = \frac{x_{2n}-x_{2n+1}}{2}, \dots\right)$$

Wavelets & Engineering



H is a low-pass filter

G is a high-pass filter

↓2 is the down-sampling operator: (1 3 4 6 5 8 7) → (1 4 5 7)

↑2 is the up-sampling operator: (1 4 5 7) → (1 0 4 0 5 0 7)

Orthogonal filter bank is biorthogonal, ←→ both *analysis* filters H′ and G′ are the *time reversals* of the *synthesis* filters H & G: e.g., H=(1,2,3)→H′=(3,2,1)

Biorthogonal (perfect) filter bank: if $y = x$ for all inputs x.

Wavelets & Engineering ➔ Pyramidal Algorithm

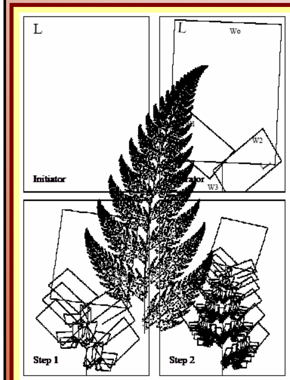
Alternating quardature mirror – Smooth (**S**) and Detail (**D**) filters for Daub4 filterbank.

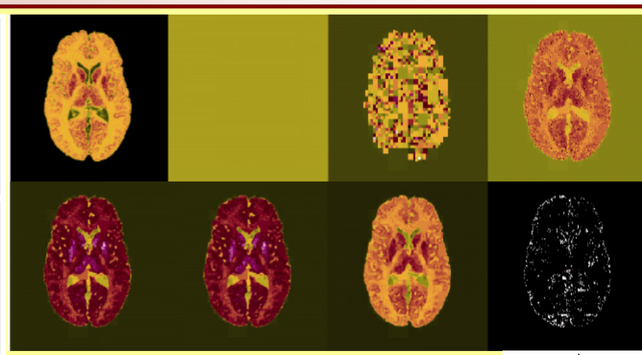
Forward DWT

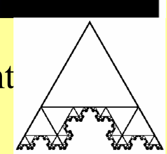
Data(y)	Apply Matrix: C_{xy}	Permute Elements	Apply Matrix: C_{xy}	Permute Elements

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} \rightarrow \begin{bmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \end{bmatrix} \rightarrow \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ \boxed{\begin{matrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{matrix}} \end{bmatrix} \rightarrow \begin{bmatrix} ss_1 \\ dd_1 \\ ss_2 \\ dd_2 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \rightarrow \begin{bmatrix} \left.\begin{matrix} ss_1 \\ ss_2 \end{matrix}\right\} \text{Smooth Components, Mother function coefficients} \\ \left.\begin{matrix} dd_1 \\ dd_2 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{matrix}\right\} \text{Details, wavelet coefficients} \end{bmatrix}$$

Wavelets & Fractals



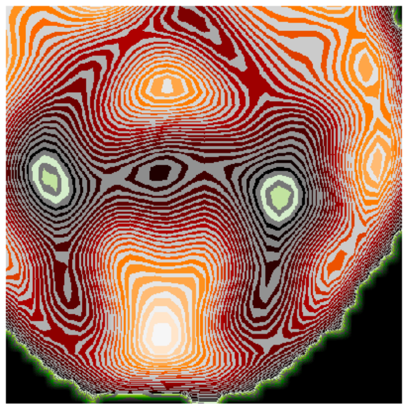


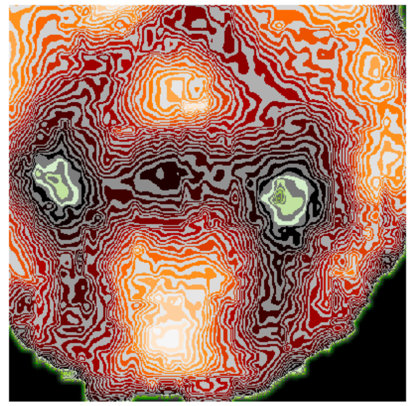


Fractal quadtree-based transformation is equivalent to Haar-filter wavelet transform.

Geoffrey M. Davis, *A Wavelet-Based Analysis of Fractal Image Compression*, IEEE Trans. Img. Proc., 1997
Arnaud Jacquin, *Image coding based on a fractal theory of iterated contractive image transformations*, IEEE TIP, 1992.
Michael Barnsley and Arnaud Jacquin, \Application of recurrent iterated function systems to images", SPIE, 1988.
Thierry Blu and Michael Unser, *Wavelets, Fractals, and Radial Basis Functions,* IEEE Trans Signal Proc., 2002.

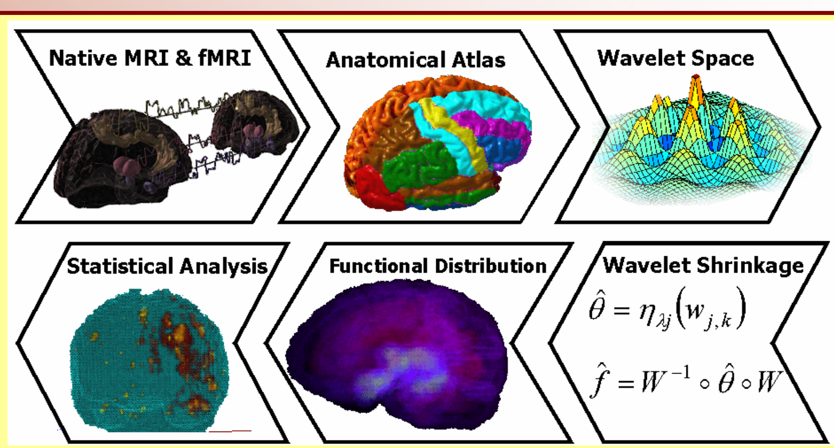
Wavelets & Fractals





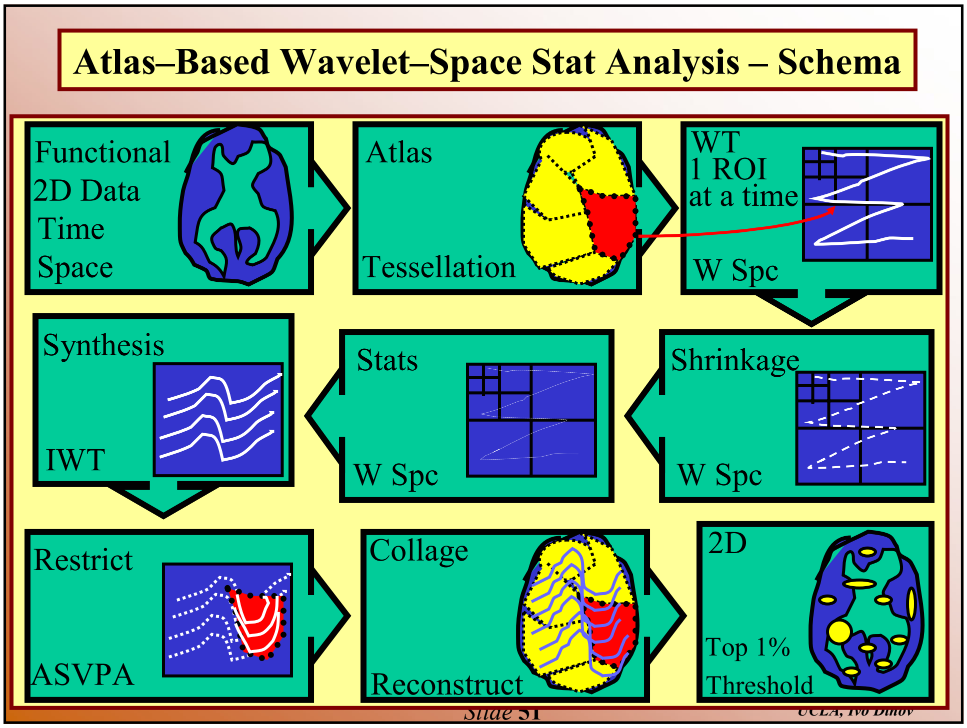
Slide 49 UCLA, Ivo Dinov

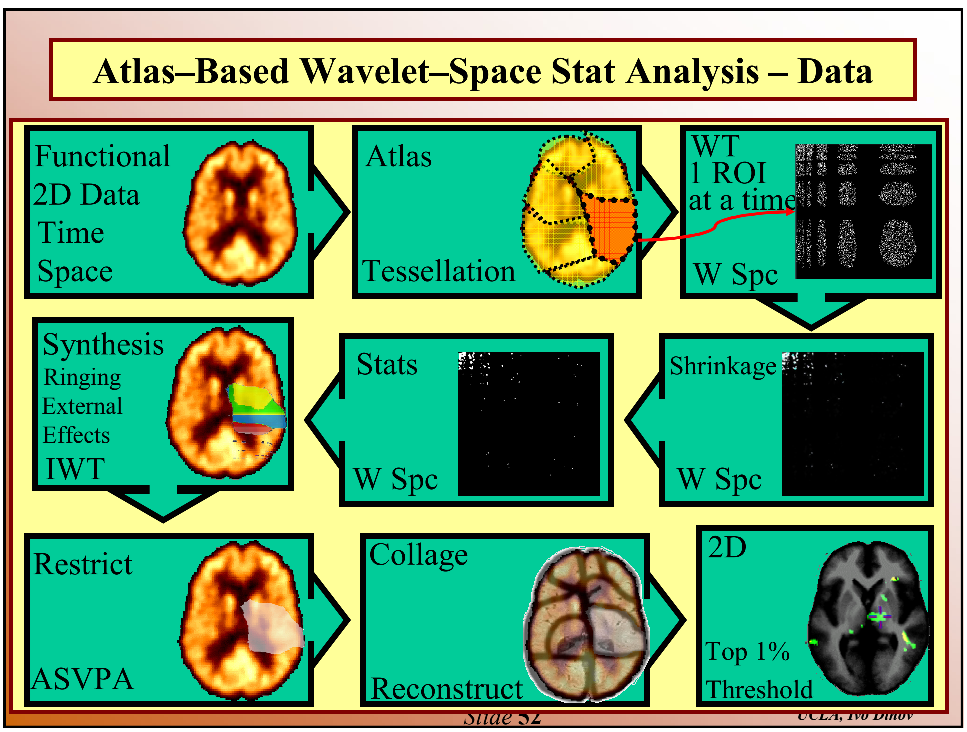
Atlas–Based Wavelet–Space Stat Analysis – Top Level



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.

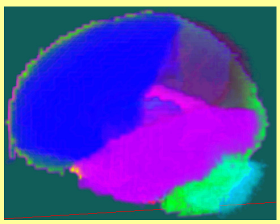
Slide 50 UCLA, Ivo Dinov

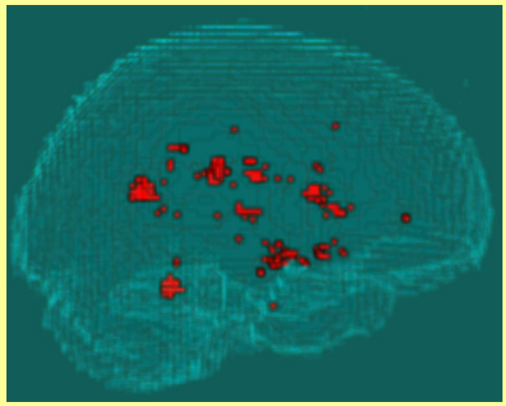

Atlas–Based Wavelet–Space Stat Analysis – Schema
Functional
2D Data
Time
Space
Atlas
Tessellation
WT
1 ROI
at a time
W Spc
Synthesis
IWT
Stats
W Spc
Shrinkage
W Spc
Restrict
ASVPA
Collage
Reconstruct
2D
Top 1%
Threshold
Slide 51
UCLA, Ivo Dinov


Atlas–Based Wavelet–Space Stat Analysis – Data
Functional
2D Data
Time
Space
Atlas
Tessellation
WT
1 ROI
at a time
W Spc
Synthesis
Ringing
External
Effects
IWT
Stats
W Spc
Shrinkage
W Spc
Restrict
ASVPA
Collage
Reconstruct
2D
Top 1%
Threshold
Slide 52
UCLA, Ivo Dinov

Atlas–Based Wavelet–Space Stat Analysis – Demo

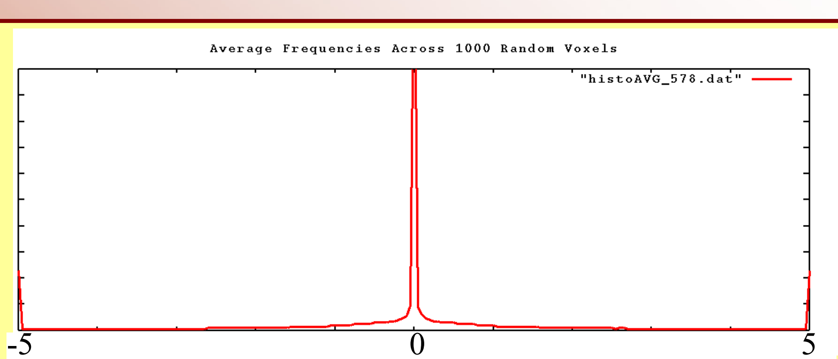
C:\Ivo.dir\LONI_Viz\LONI_Viz_VR_driver.bat (Sub-Sampling 2x2x2)
File → Load Data
(E:\Ivo.dir\Research\Data.dir\FA_SVPA_WaveletBasedAtlas\stat_S1_HR_VR_2_ADSVPA_mult10.img.gz)
File → Load Mask
(E:\Ivo.dir\Research\Data.dir\FA_SVPA_WaveletBasedAtlas\IWT_SVPA_20_lin_h_lobes30.img.gz)
Voxel (40,20,33)





Slide **55** *UCLA, Ivo Dinov*

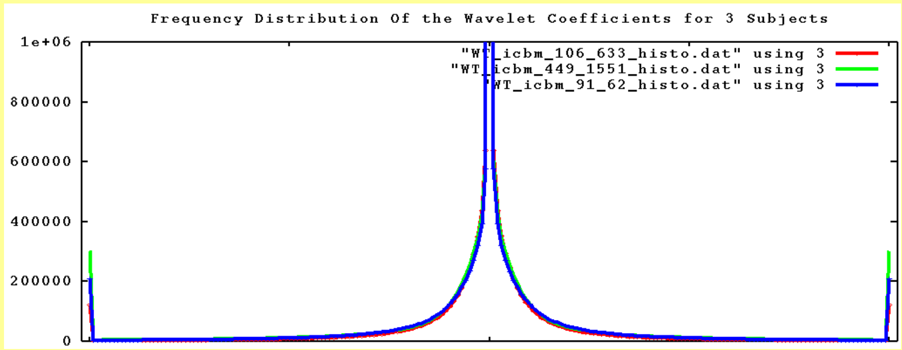
Distribution of the wavelet coefficients



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.

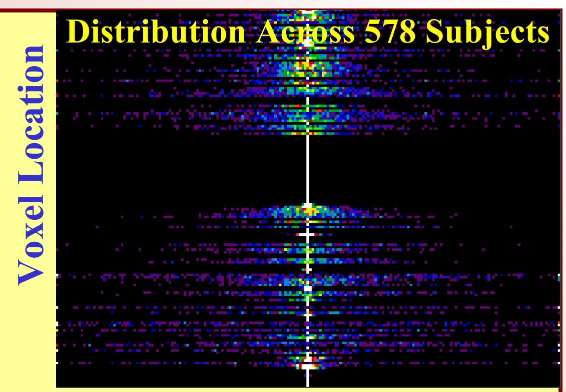
Slide **56** *UCLA, Ivo Dinov*

Distr’n of the wavelet coefficients – 3 volumes



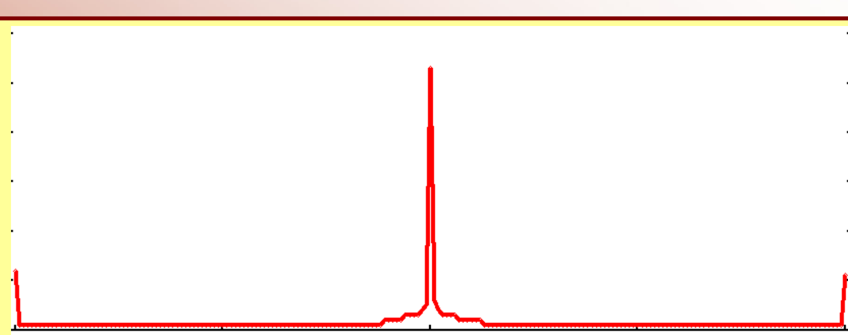
Shown here are the frequency distributions of the wavelet coefficients for three separate MRI volumes (randomly selected from the pool of 578). There is little variation between these three individuals, however the overall shape of the distribution of the magnitudes of the wavelet coefficients of each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

Wavelet Coefficient distributions MRI data



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], vertical axis labels the subject index and the row color map indicates the frequencies of occurrence of a wavelet coefficient of certain magnitude across all 1,000 voxels. Bright colors indicate high, and dark colors represent low, frequencies.

fMRI data – 1 subject wavelet coefficient distributions

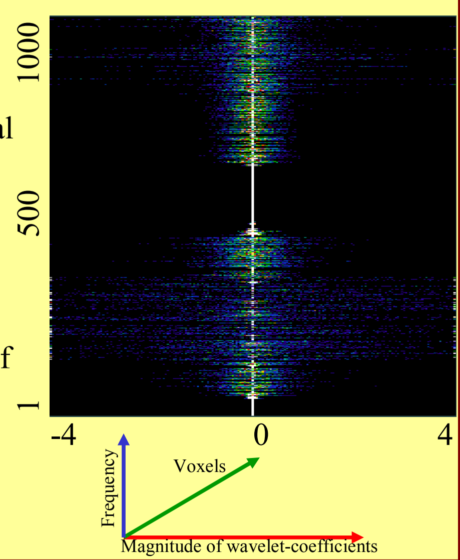


This diagram depicts the frequency histogram of the average magnitude of a wavelet coefficient, across all 128 fMRI 3D time-volumes (part of the fMRI study of Buckner and colleagues [Buckner *et al.*, 2000]) at 1,000 randomly selected locations (random indices of $w_{j,k}$). Again, we observe the heavy-tailness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

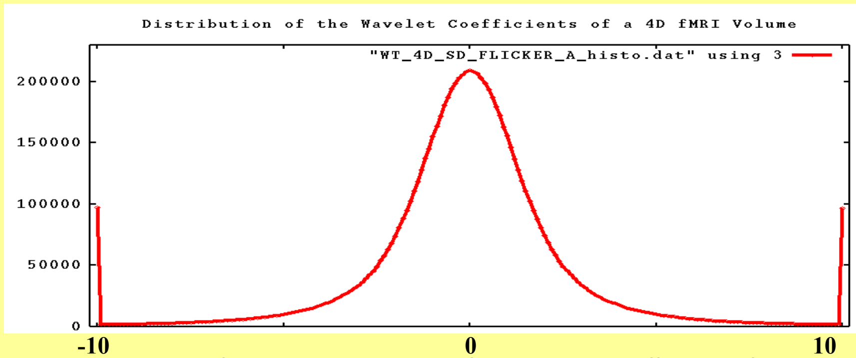
Wavelet coefficient Distribution – 1 run

A 2D image displaying all of the individual distributions for each 128 3D time points (1 run, 15 trials) of an fMRI timeseries. On the horizontal axis is the magnitude of the wavelet and the vertical axis are the voxel locations.

Colors indicate the frequencies of occurrence of a wavelet coefficient of certain magnitude across all 1,000 voxel locations. The across row average is effectively shown on previous Figure.



Distribution of 4D wavelet coefficients



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.

Heavy-Tail wavelet coefficient distribution models …

Leptokurtic Distribution Candidates:

1. **Double-exponential** distribution $f(x) = \frac{1}{2\beta} \times \exp\left(-\left|\frac{x-\mu}{\beta}\right|\right)$

2. **Double-Pareto** distr. $f(x) = \frac{\alpha c^{\alpha}}{|x-\mu|^{1-\alpha}}, \quad |x-\mu| > c$

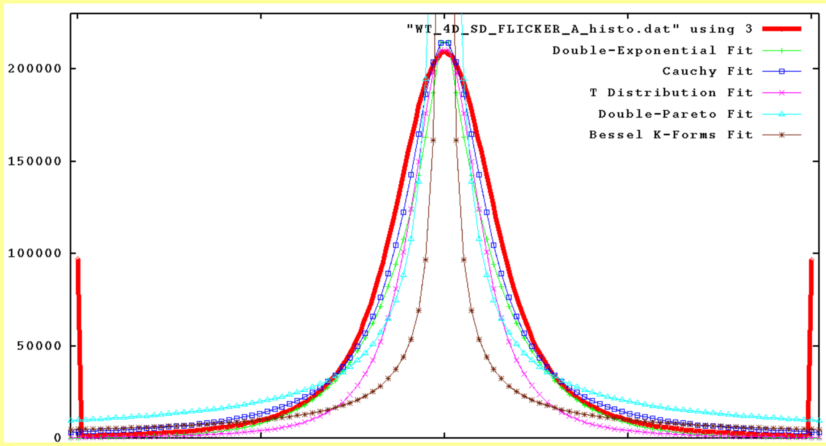
3. **Cauchy** distribution: $f(x) = \frac{\gamma}{\pi} \times \frac{1}{\gamma^2 + (x-\mu)^2}$

4. **T** distribution: $f(x) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi} \times \Gamma\left(\frac{n}{2}\right)} \times \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}$

5. $\hat{p} =$

http://socr.stat.ucla.edu/

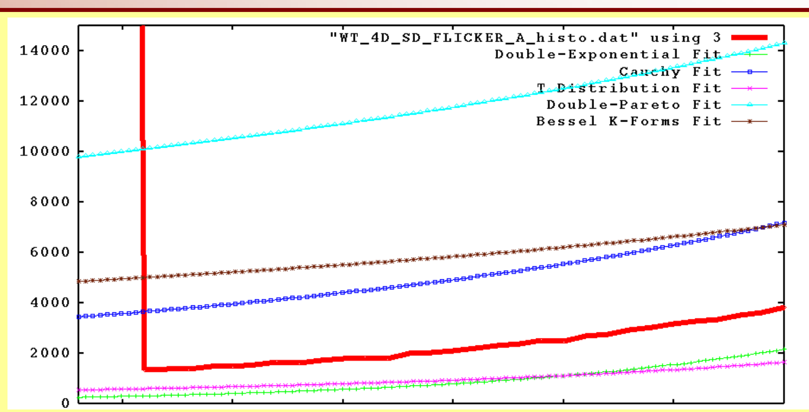
Distribution of (and Models for) the Wavelet Coefficients of the 4D fMRI volume



Several heavy-tail distribution models are fitted to the frequency histogram of 4D 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.

Slide **63** ***UCLA, Ivo Dinov***

Right Tail of Distribution of (and Models for) the Wavelet Coefficients of the 4D fMRI volume

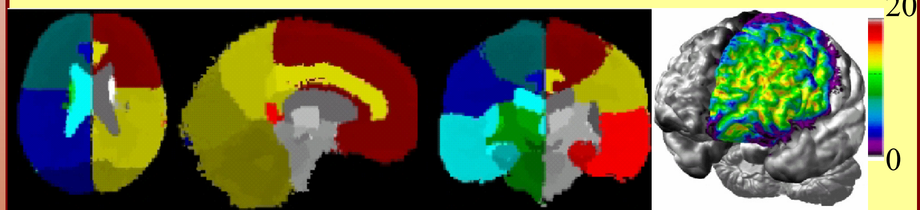


Shows the extreme left trail of the data distribution (data is symmetric). The *double-exponential* and the *T* distribution models underestimate the tails of the data asymptotically, but provide good fits around the mean. *Cauchy*, *Bessel K forms* and *double-Pareto* distribution models, in that order, 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.

Slide **64** ***UCLA, Ivo Dinov***

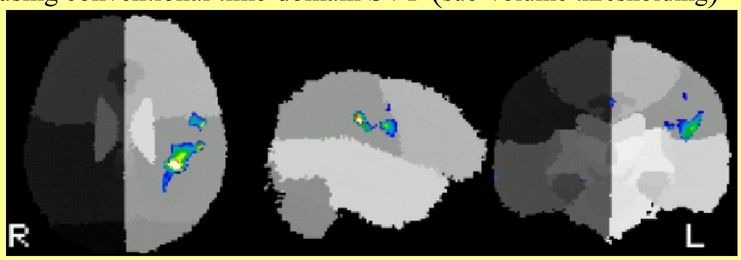
Wavelet-domain Atlas-based fMRI Analysis

- Anatomical Sub-Volume Probabilistic Atlas (SVPA)



- Analysis using conventional time-domain SVT (sub-volume thresholding)

Subtraction paradigm: AD-ElderlyNC



Wavelet-domain Atlas-based fMRI Analysis

- Raw results following analysis in the wavelet-domain (has ringing effects outside of the true ROI where the activation supposedly occurred).
- Post-processed wavelet domain statistics [final uniform thresholding (top 1%), reconstruction of all wavelet-space stats into a single volume – regional calculations]:

Bayesian Mixture Models for fMRI Analysis

- We use two component mixture prior distributions on the wavelet coefficients $\theta_{j,k}$ with

$$\theta_{j,k} \mid \pi_j \tau_j \sim \pi_j N\left(0, \tau_j^2\right) + (1-\pi_j)\delta(0)$$

where π_j is a proportion between 0 and 1, δ(0) is the Dirac point mass at zero and $\tau_j > 0$. In other words, there is a level-dependent positive probability π_j *a priori* that each wavelet coefficient will be exactly zero. If not, the coefficient will be normally distributed with mean zero and a level-specific standard deviation τ_j

Bayesian Mixture Models for fMRI Analysis

- Given the observed data over an ROI $y=f+\varepsilon$, the corresponding wavelet representation $w = \theta + z$, where W is the discrete WT, $w = Wy$, $\theta = Wf$ and $z = W\varepsilon$, and the above prior distribution for the true wavelet coefficients, the *posterior distributions of* $\theta_{j,k}$ are again independent two-component mixtures:

$$p\left(\theta_{j,k} \mid w_{j,k}\right) \sim \lambda_{j,k} N\left(\frac{w_{j,k}\tau_j^2}{\sigma^2+\tau_j^2}, \frac{\sigma^2\tau_j^2}{\sigma^2+\tau_j^2}\right) + (1-\lambda_{j,k})\delta(0)$$

Where the $\lambda_{j,k} = 1/(1+\rho_{j,k})$ are the posterior odds that $\theta_{j,k}$ is exactly zero are: $\rho_{j,k} = \frac{1-\pi_j}{\pi_j}\frac{\sqrt{\tau_j^2+\sigma^2}}{\sigma}\exp\left(\frac{-\tau_j^2 w_{j,k}}{2\sigma^2(\tau_j^2+\sigma^2)}\right)$

LONI Resource Collaborators

Slide **69** ***UCLA, Ivo Dinov***