

Medical Image Analysis

Instructor: Moo K. Chung

mchung@stat.wisc.edu

Lecture 15.

Image similarity measures

April 12, 2007

Similarity measures

- Correlation based
- Statistical (goodness-of-fit)
- Information theoretic (mutual information)
- Geometric (kappa index)

Correlations

- Pearson correlation
- Rank correlation
- Partial correlation
- Multiple correlation
- Partial multiple correlation
- Canonical correlation

Pearson product moment correlation coefficient

Given paired measurements $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, the Pearson product-moment correlation coefficient (Fisher, 1915) is a measure of association given by

$$r_P = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

where $\bar{X}$ and $\bar{Y}$ are the sample mean of $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ respectively.

Inference on correlation

$$H_0 : \rho = 0 \text{ vs. } H_1 : \rho \neq 0$$

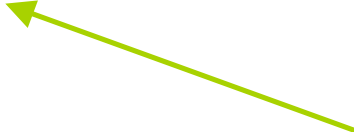
Under normality,

$$T = \frac{r_P \sqrt{n-2}}{\sqrt{1-r_P^2}}$$

T-statistic with n-2 degrees of freedom

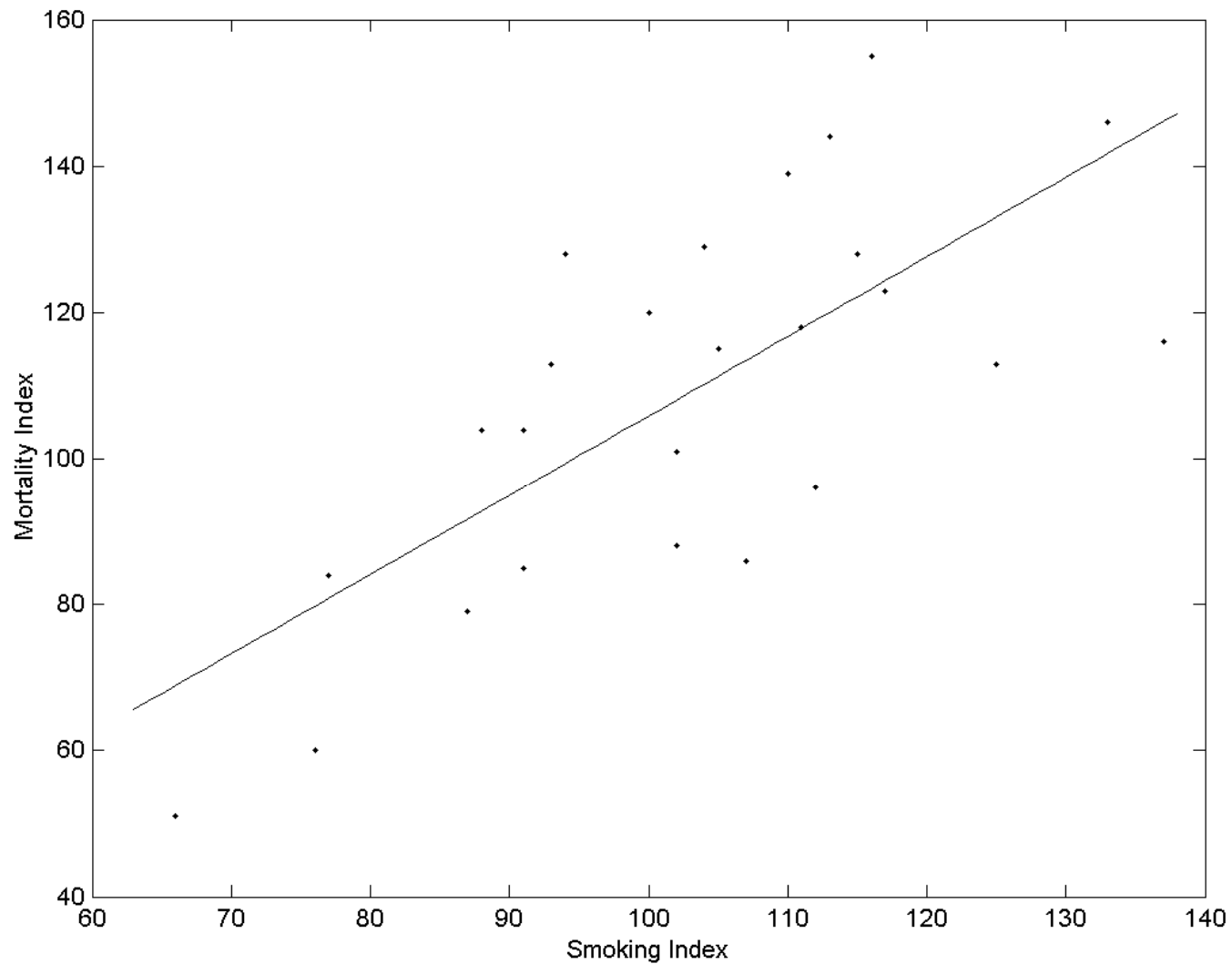
If we can not assume
normality,

Fisher transform: $F(r_p) = \frac{1}{2} \ln \left(\frac{1 + r_p}{1 - r_p} \right)$

$$Z = \sqrt{n-3} \left(F(r_p) - F(\rho) \right) : \text{standard normal}$$


Note: under the null hypothesis, this term vanishes.

Correlation vs. GLM



Are there any relationship between correlation analysis and GLM ?

Slope of regression vs. correlation

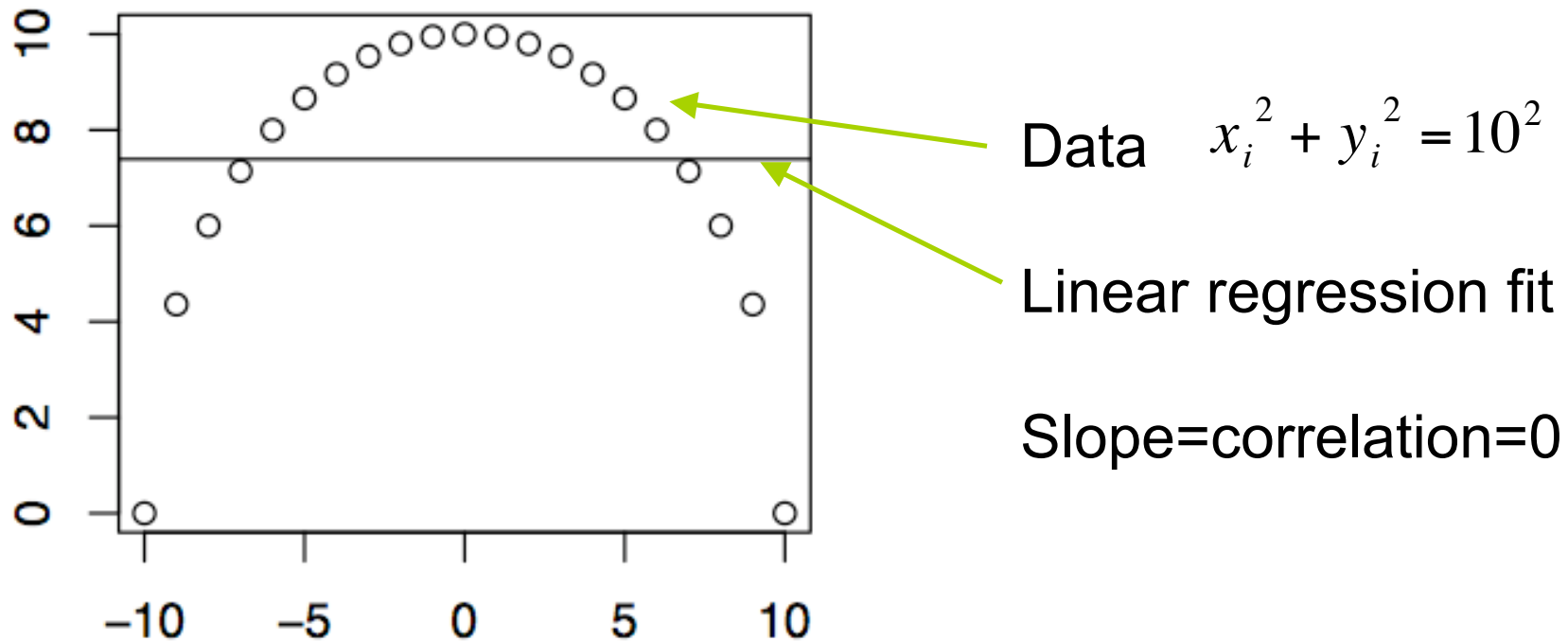
$$Y = \beta_0 + \beta_1 \cdot X$$

If (X,Y) are bivariate normal,

$$\beta_1 \cong \frac{\sigma_Y}{\sigma_X} r_P$$

Under normality, correlation analysis = linear regression

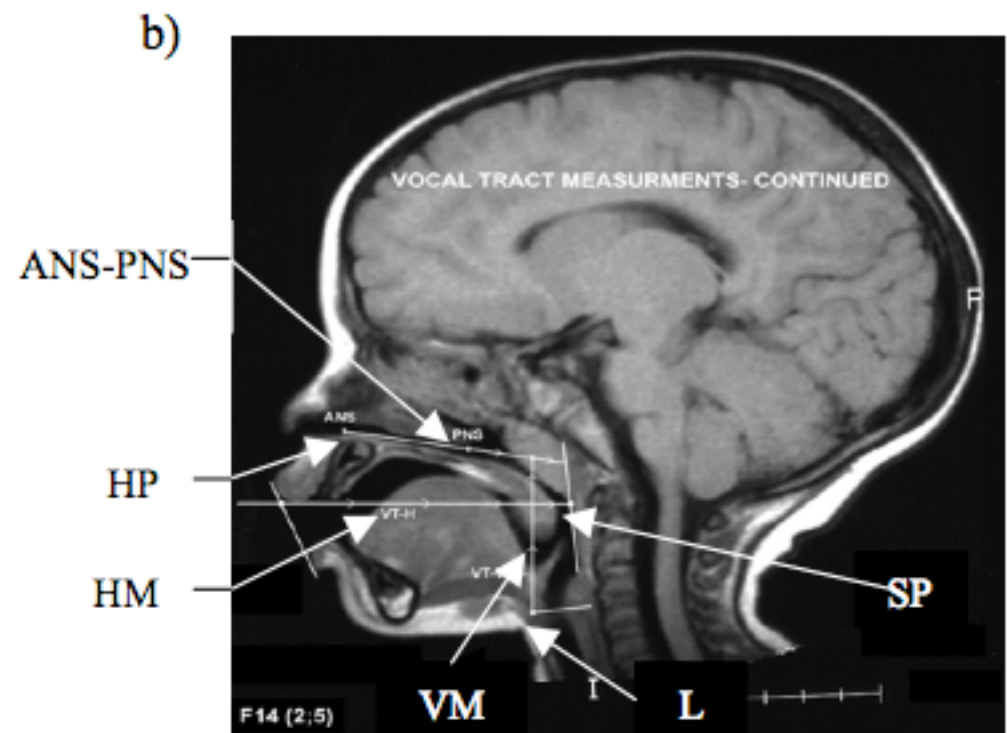
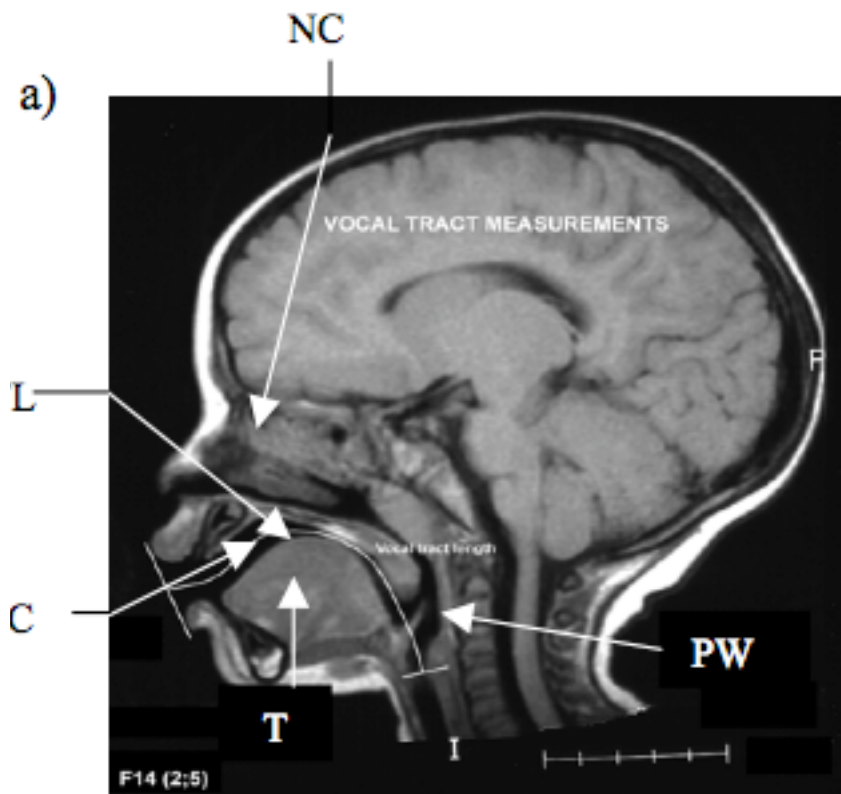
What correlation really measures?



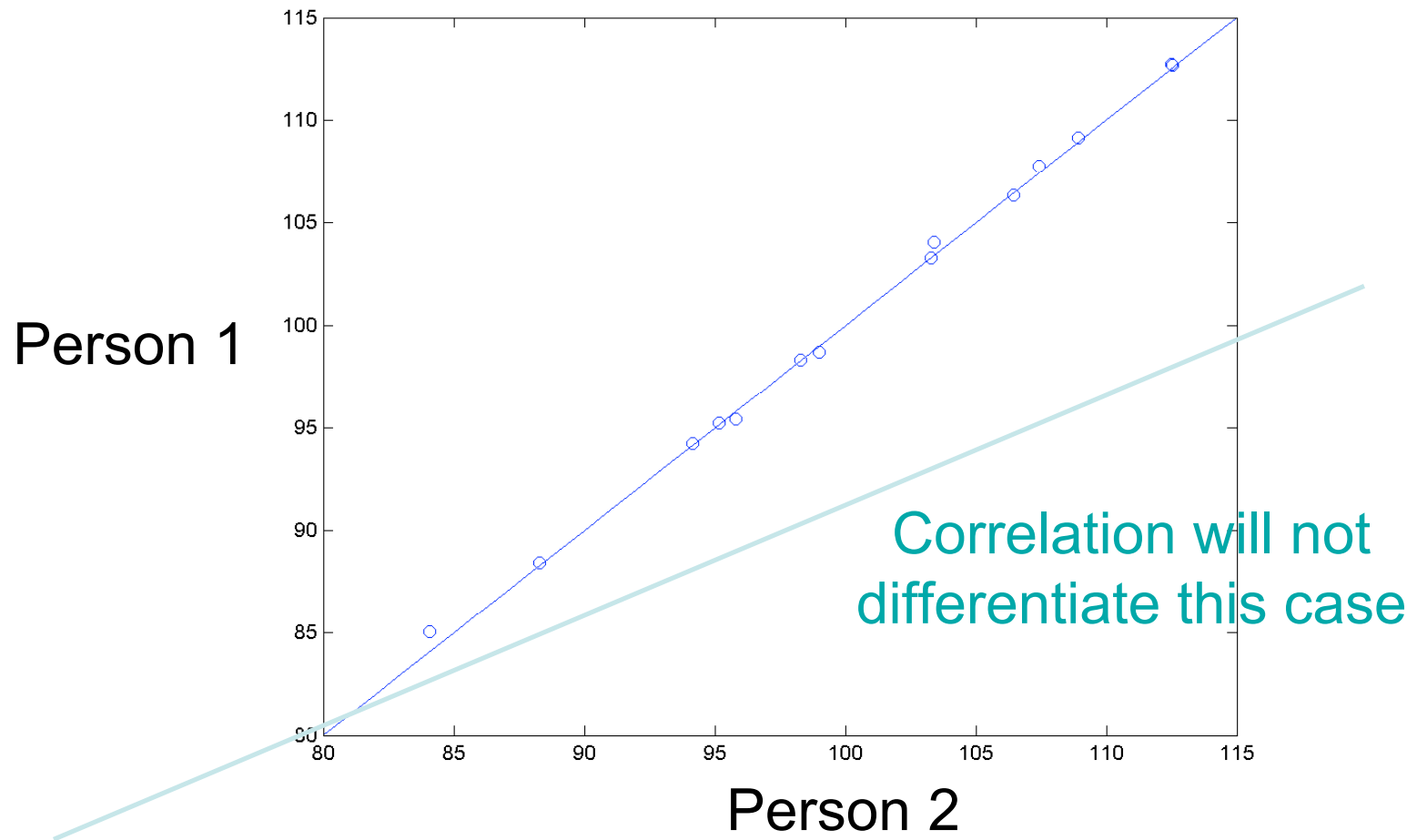
Correlation only measures linear dependency.
It does not measure nonlinear dependency.

Quiz: using correlation, is it possible to measure nonlinear dependency?

Application: length measurements
are obtained semi-automatically by
two people



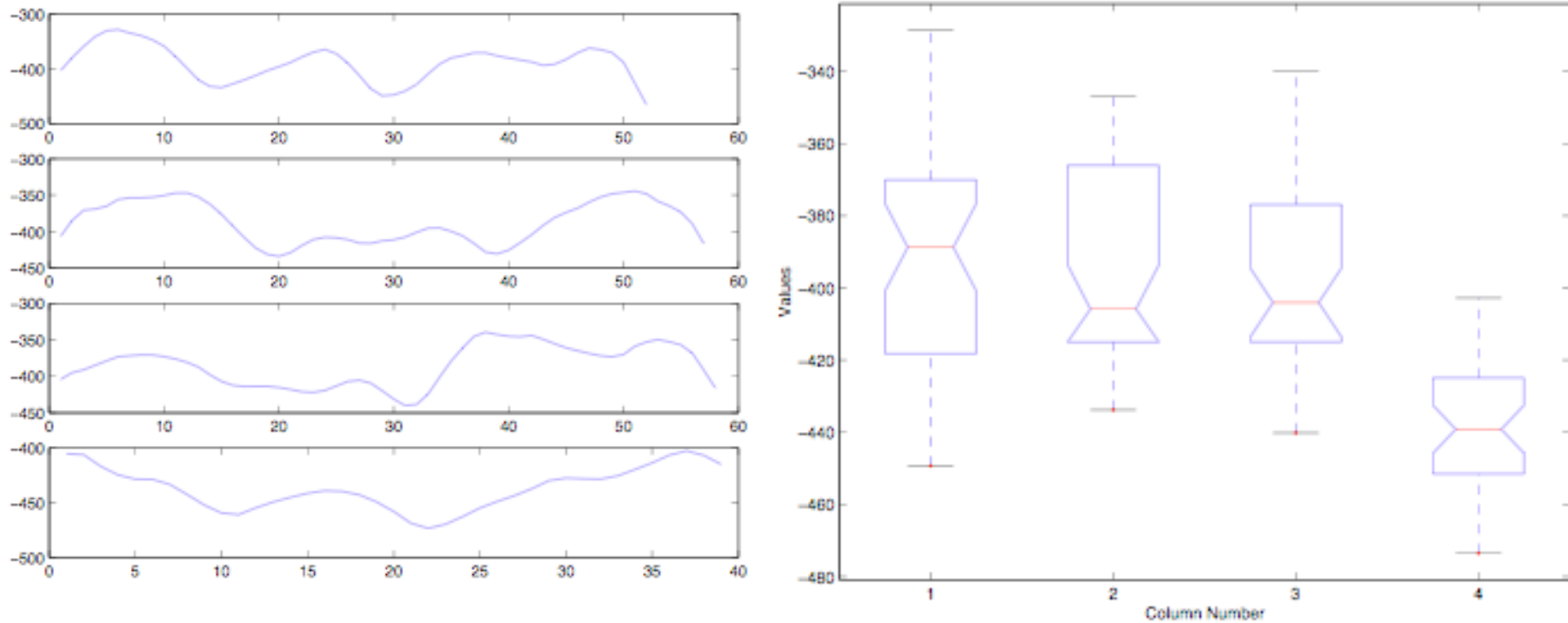
Checking consistency of manual measurements



Under H_0 , measurements will pass through $y=x$ line

Consistency for 4 measurements

Example: mouse breast cancer



Four line probe measurements that measures the amount of cancer

Question: Are they consistent?

One way ANOVA

(it can be formulated as GLM)

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

$$\bar{Y} = \frac{n_1 \bar{Y}_1 + n_2 \bar{Y}_2 + n_3 \bar{Y}_3 + n_4 \bar{Y}_4}{n_1 + n_2 + n_3 + n_4}$$

$$BGV = \frac{1}{4-1} \sum_{i=1}^4 n_i (\bar{Y}_i - \bar{Y})^2$$

$$WGV = \frac{1}{n_1 + n_2 + n_3 + n_4 - 4} \sum_{i=1}^4 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

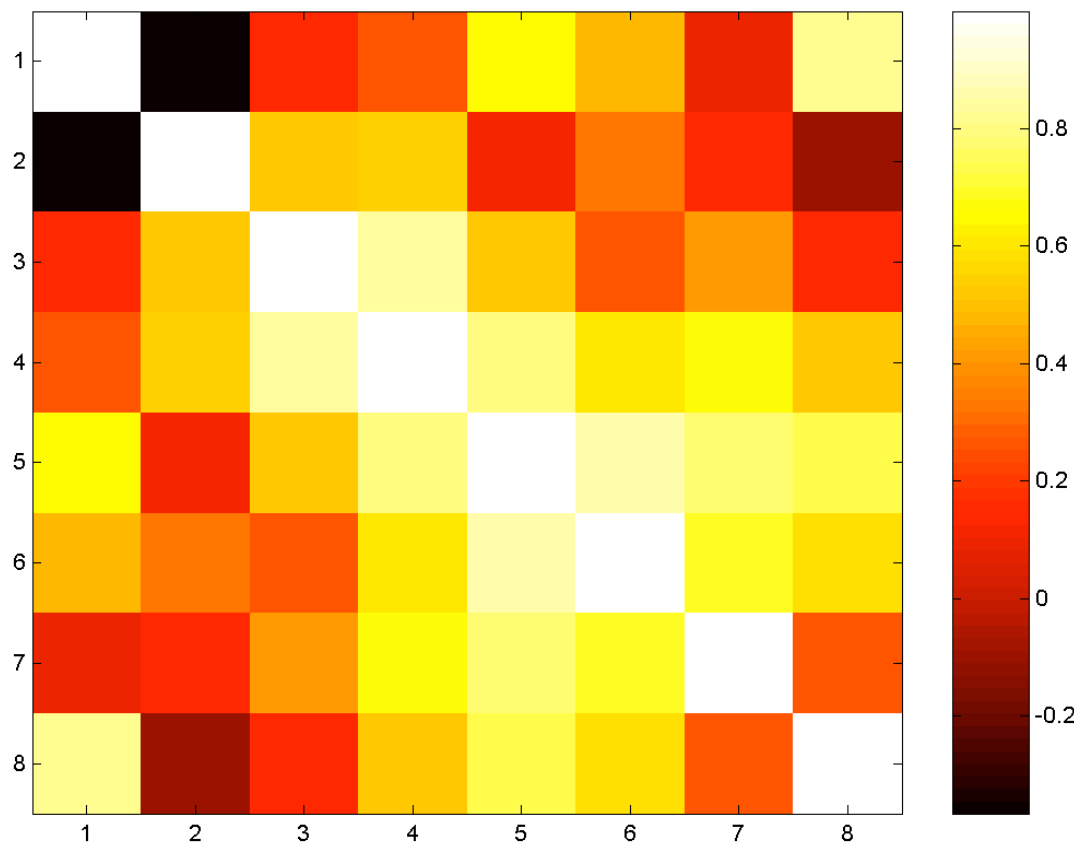


BGV/WGV

$$F_{4-1, n_1+n_2+n_3+n_4-4}$$

Correlation mapping technique

CT intensity values for 8 particular regions
in mouse breast cancer study

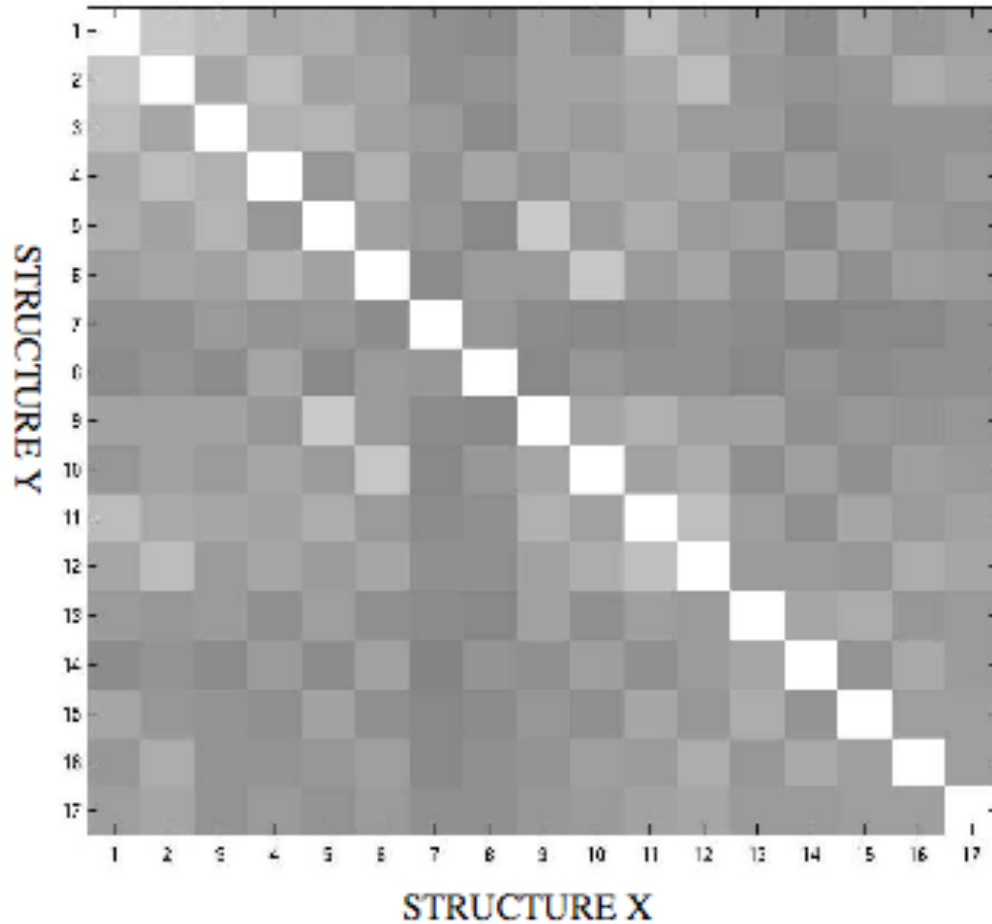


Regions denoted by numbers

Sorted correlation map
can be used to display the
pairwise correlation

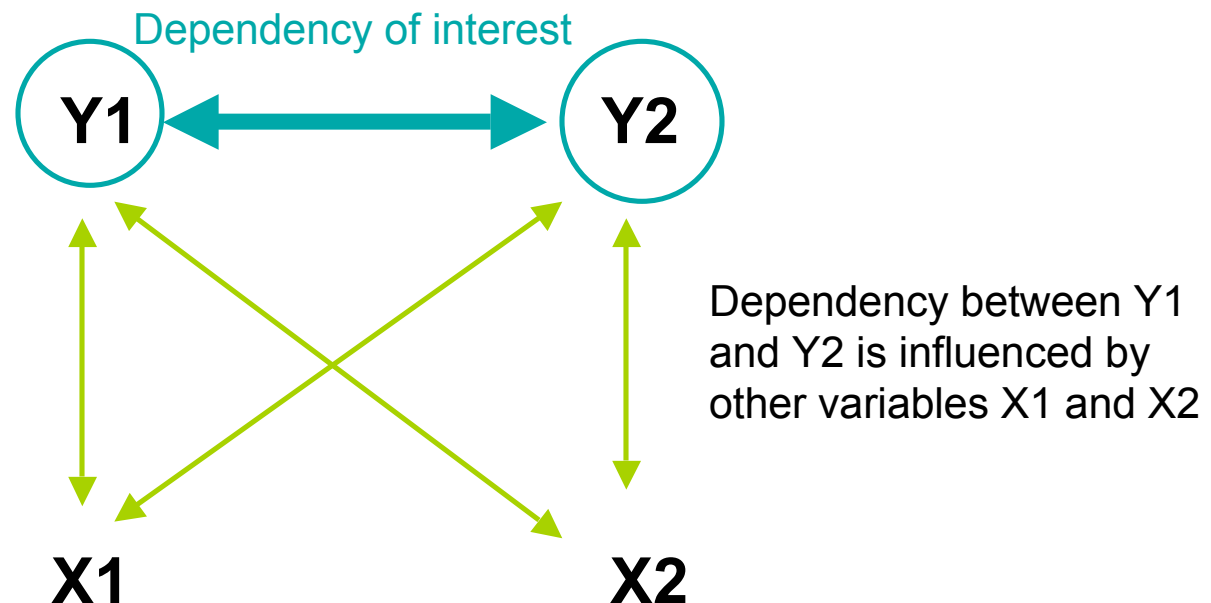
MATLAB:
Save correlations into
matrix format and display
using “imagesc” command

Correlation mapping technique (anatomical connectivity analysis)



Partial correlation

Measure of dependency while removing the effect of other variables.



Case study:

We are interested in relation between

Y1 = cortical thickness

Y2 = behavioral measure

Fact:

- Cortical thickness related to age (X1)
- Cortical thickness related to brain size (X2)

What we want:

$\text{Corr}(Y1, Y2)$ while accounting for X1 and X2

$$Y = (Y_1, Y_2), X = (X_1, X_2)$$

Covariance matrix: $\mathbb{V}(Y, X)' = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$

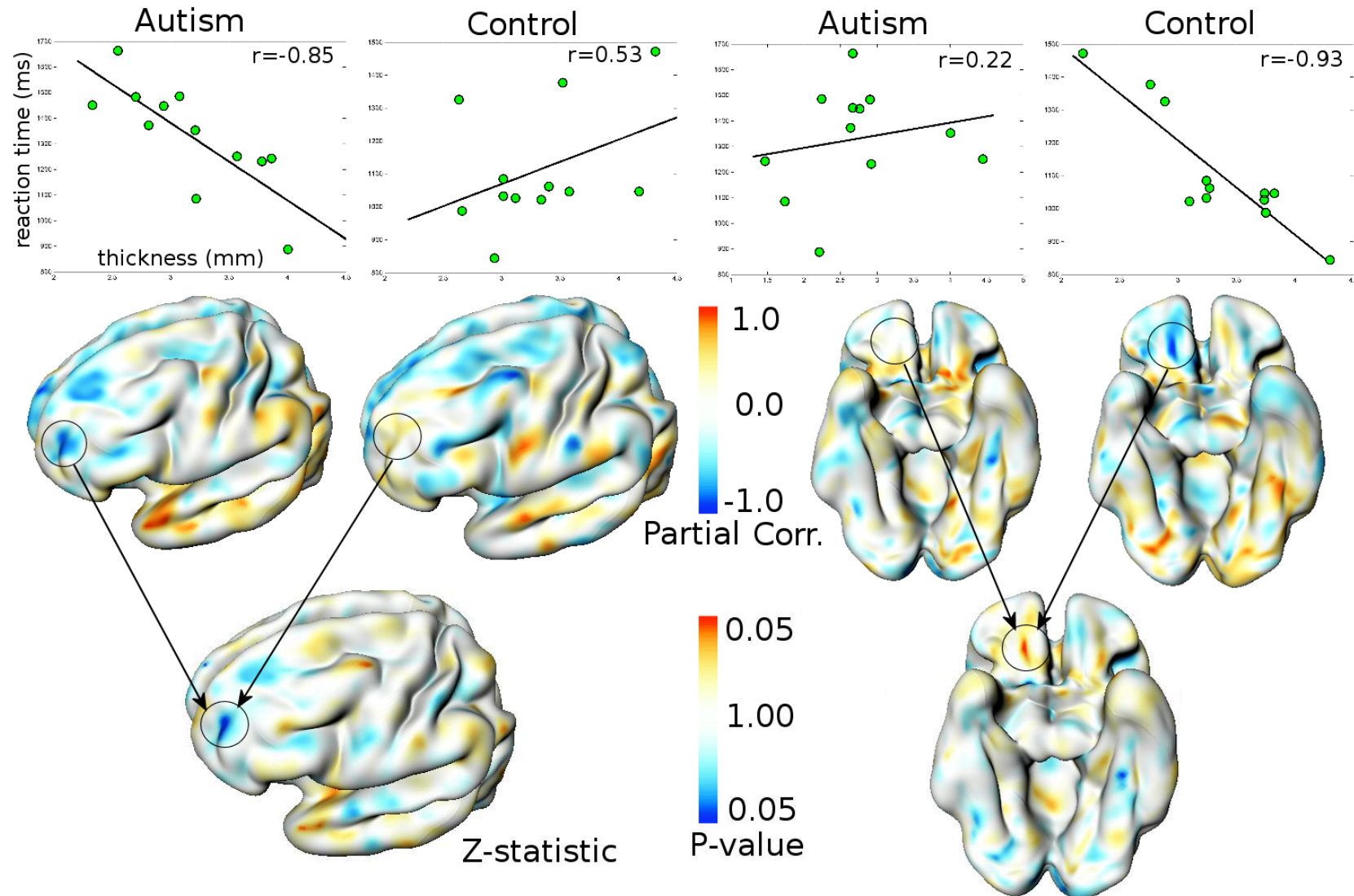
Partial covariance of Y given X:

$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = (\sigma_{ij})$$

Under normality,
this is equivalent to the conditional covariance:

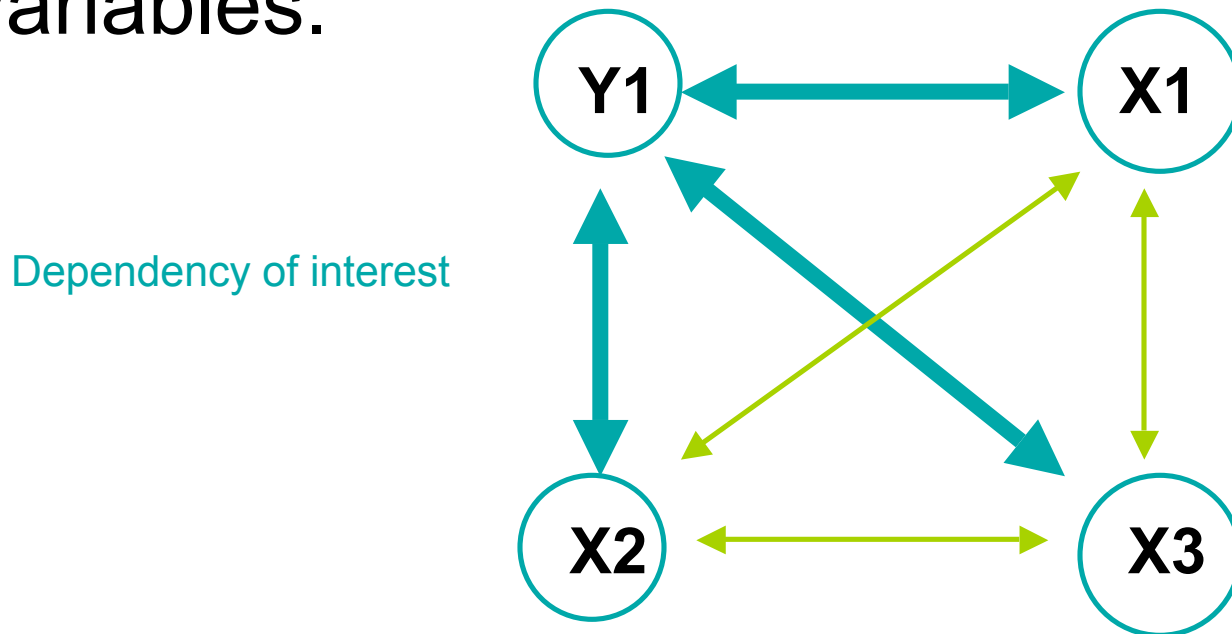
$$\text{Cov}(Y|X)$$

Correlation of emotion recognition response time and cortical thickness while accounting for age and brain size



Multiple correlation

Measure of dependency between one variable and remaining multiple variables.



Possible scenario: correlating a reference fMRI time series (amygdala) to fMRI time series in other regions

Definition of multiple correlation

- Given X_1, X_2, X_3 , consider weighted average $w_1X_1 + w_2X_2 + w_3X_3$

- The multiple correlation is defined as

$$\rho = \max_{w_1, w_2, w_3} \text{corr}(Y, w_1X_1 + w_2X_2 + w_3X_3)$$

Meaning: linear combination of maximal correlation

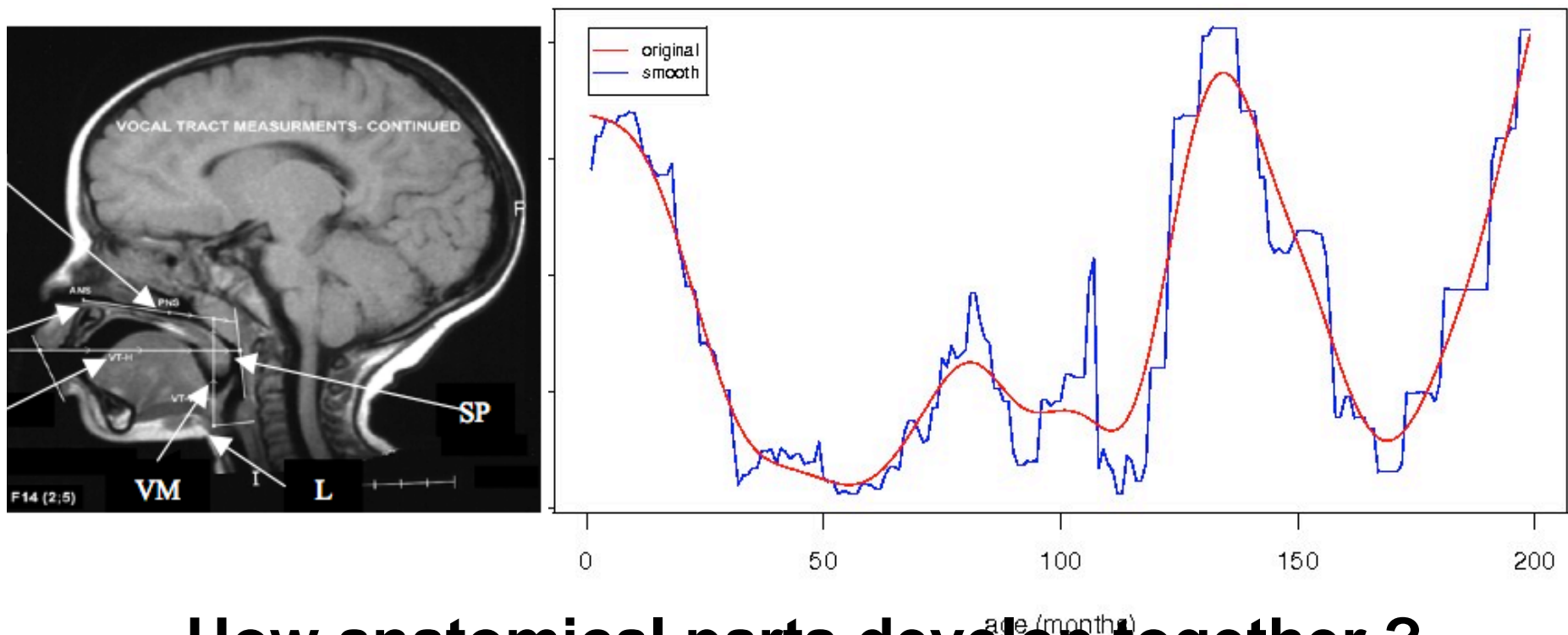
Computing multiple correlation

- No need for numerical optimization
- Multiple correlation can be obtained from multiple linear regression analysis

$$Y = w_1X_1 + w_2X_2 + w_3X_3$$

Coordinated growth of vocal tract and other anatomy

Multiple correlation curve of HardP + AntTong + MAndL

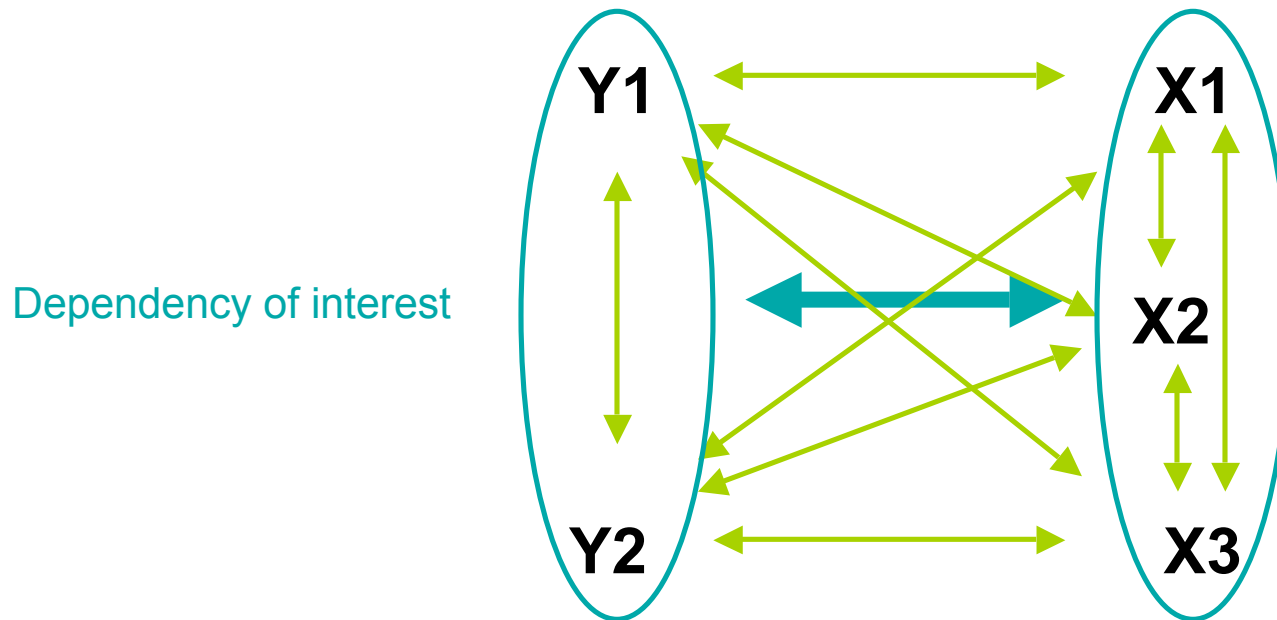


How anatomical parts develop together ?

Source: S. Wang, H. Vorperian

Canonical correlation

Measure of dependency between two vectors of unequal size



Possible scenario: network modeling, anatomical and functional connectivity in brain

Computing canonical correlation

$$X = w_1 X_1 + w_2 X_2 + w_3 X_3$$

$$Y = v_1 Y_1 + v_2 Y_2$$

$$\rho = \max_{w,v} \text{corr}(X,Y)$$

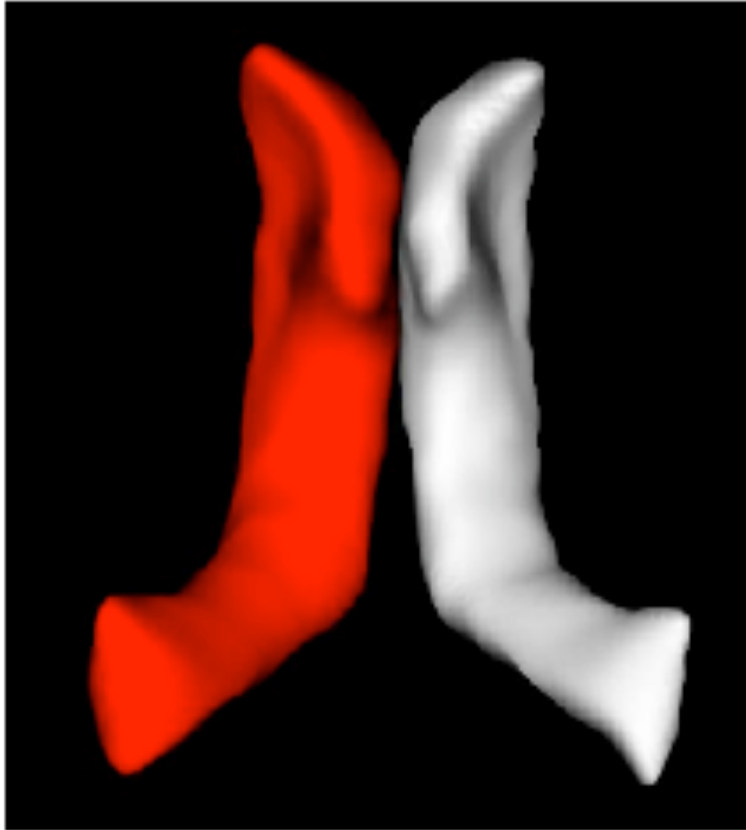
$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} w = \rho^2 w$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} v = \rho^2 v$$



Eigenvalues are identical. The principal eigenvalue is the canonical correlation

Application to anatomical shape analysis (Cootes, 2006 MMBIA)

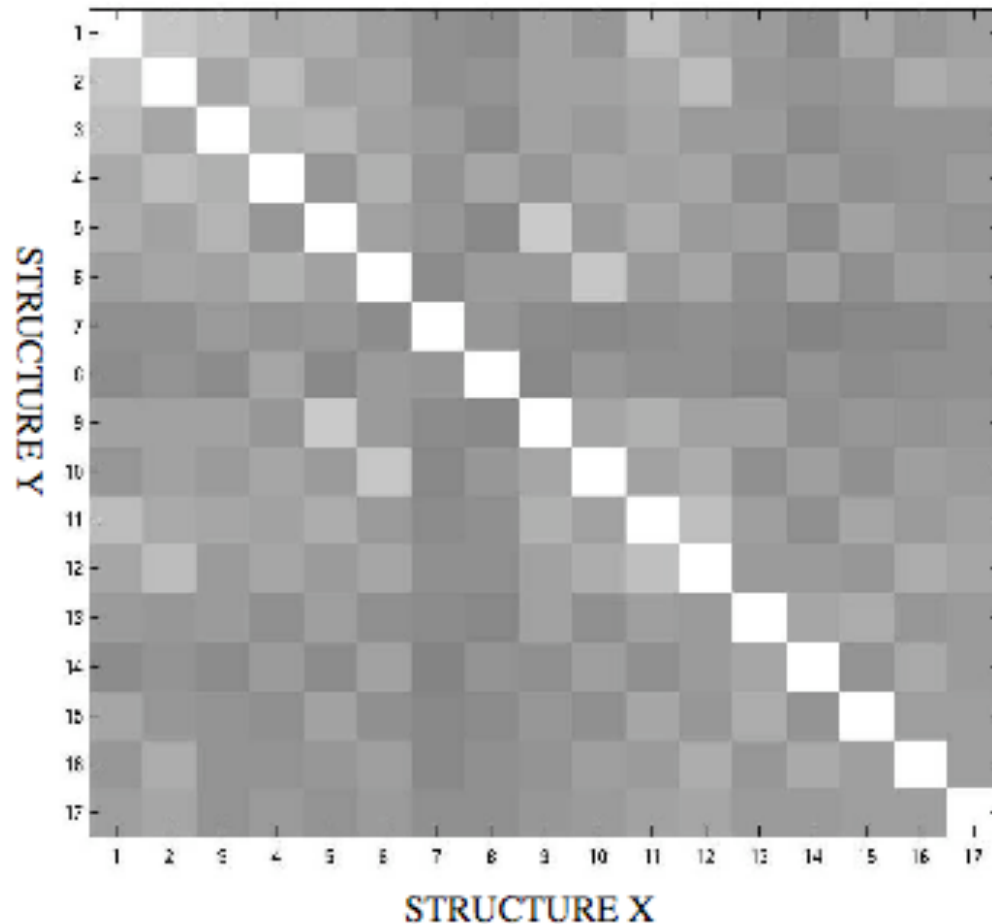


Ventricle surface in brain

- Surface coordinates are taken as a vector data $x=(x_1, x_2, \dots, x_n)$ of size 93.
- The vector data is canonically correlated with the surface coordinates of other anatomical objects.

Canonical correlation map showing pairwise correlation

- 1 = Left Lateral Ventricle
- 2 = Right Lateral Ventricle
- 3 = Left Caudate
- 4 = Right Caudate
- 5 = Left Putamen
- 6 = Right Putamen
- 7 = Left Accumbens
- 8 = Right Accumbens
- 9 = Left Pallidum
- 10 = Right Pallidum
- 11 = Left Thalamus
- 12 = Right Thalamus
- 13 = Left Amygdala
- 14 = Right Amygdala
- 15 = Left Hippocampus
- 16 = Right Hippocampus
- 17 = Brain Stem

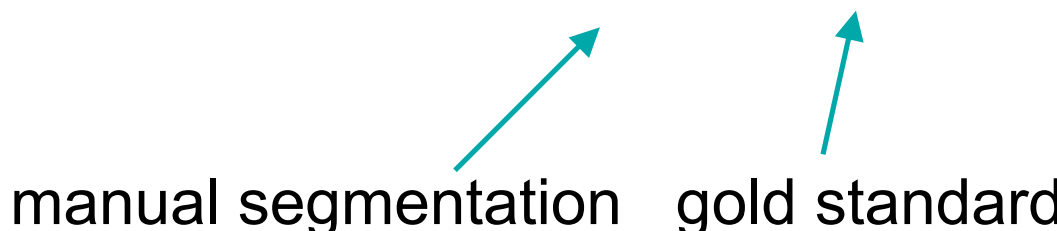


Set theoretic similarity measures

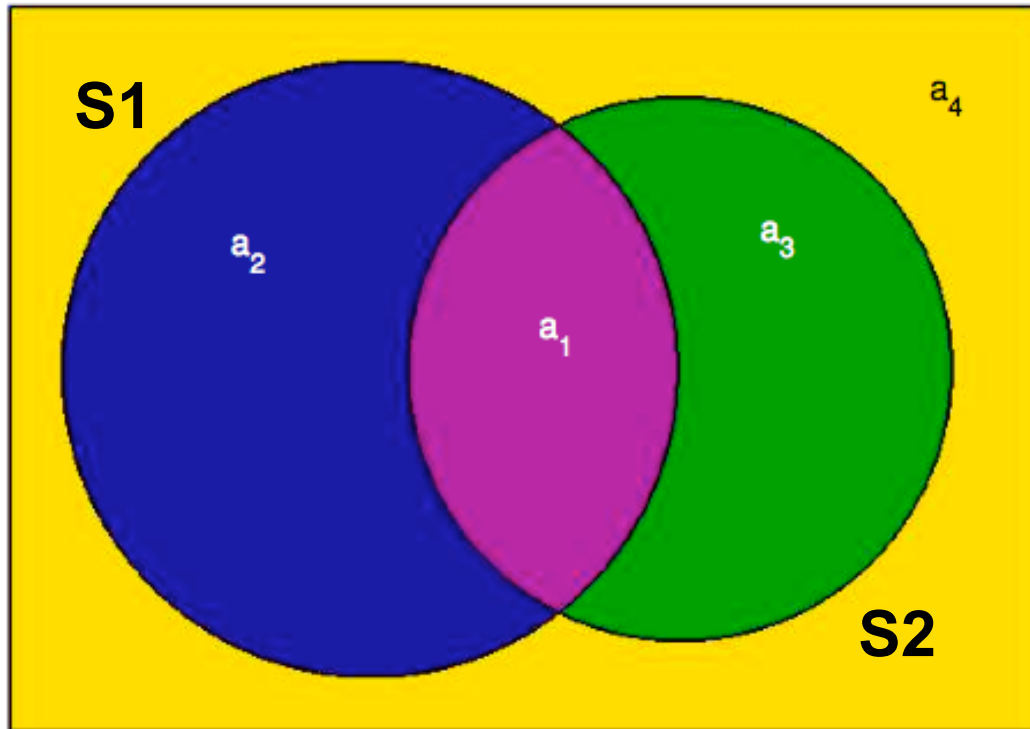
- Possible motivation: validation study on image segmentation
- *Kappa index* measures the proportion of areal overlap:

$$\kappa(S_1, S_2) = 2 \frac{|S_1 \cap S_2|}{|S_1| + |S_2|}$$

manual segmentation gold standard

The diagram consists of two teal arrows. The first arrow originates from the text 'manual segmentation' and points diagonally upwards and to the right, terminating at the term $|S_1|$ in the denominator of the Kappa index formula. The second arrow originates from the text 'gold standard' and points diagonally upwards and to the left, terminating at the term $|S_2|$ in the denominator of the same formula.

Range of Kappa index



- Minimum is obtained when there is no intersection:

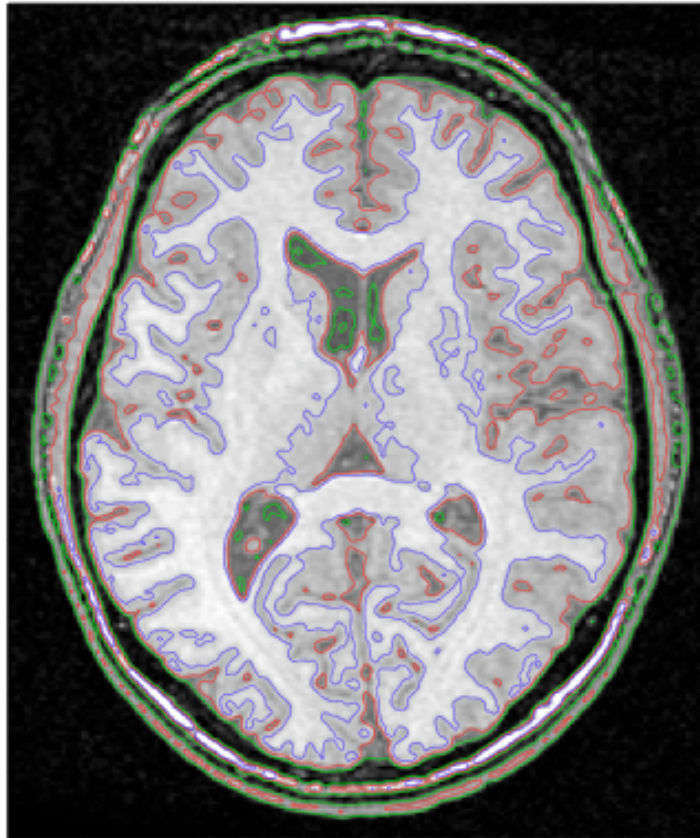
$$a_1=0, \kappa(S_1, S_2) = 0$$

- Maximum is obtained when every points intersect:

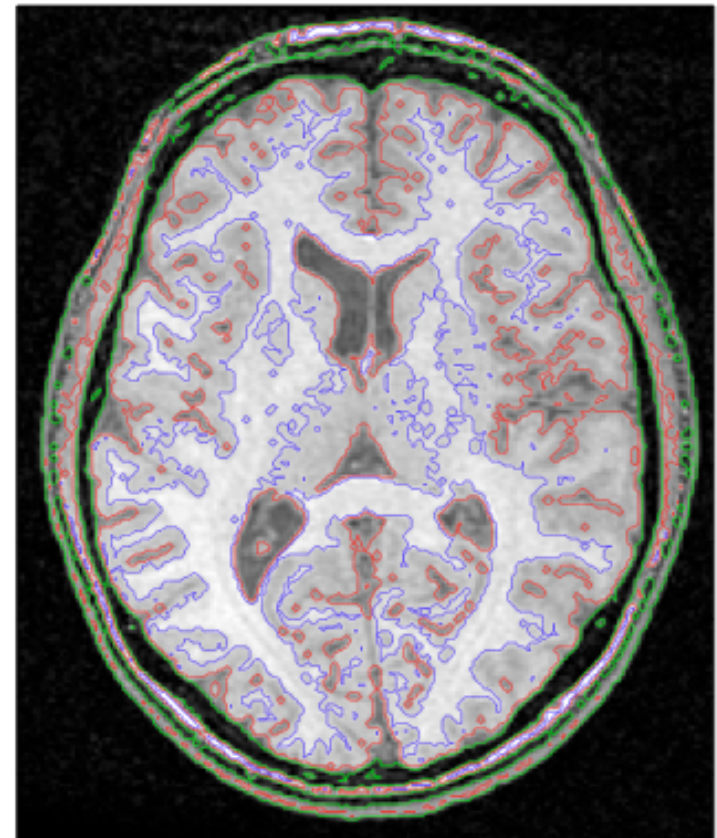
$$a_1=a_2=a_3, \kappa(S_1, S_2)=1$$

$$\kappa(S_1, S_2) = \frac{2a_1}{2a_1 + a_2 + a_3}$$

Comparing different segmentation techniques



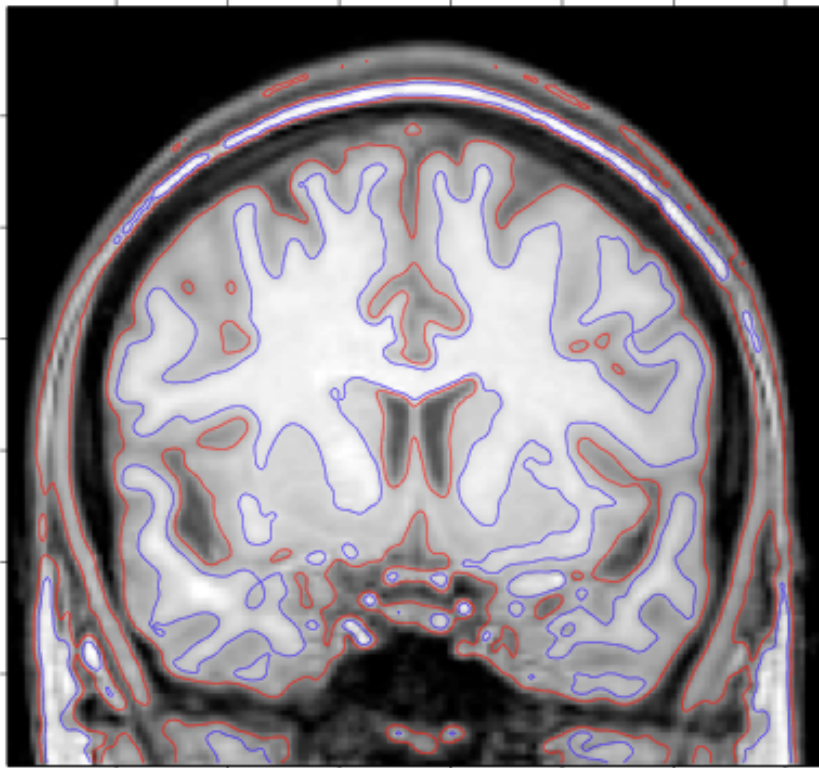
Thin plate spline (TPS)
segmentation



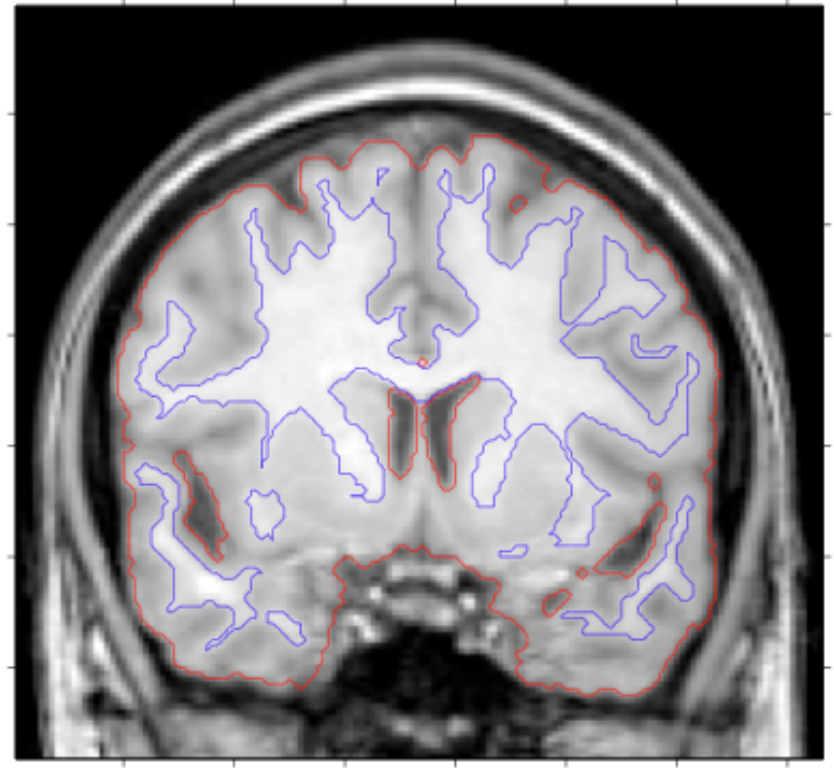
Neural network
segmentation

Source: PhD work of X. Xie

Automatic vs. manual (NIH gold standard)



TPS segmentation



manual segmentation

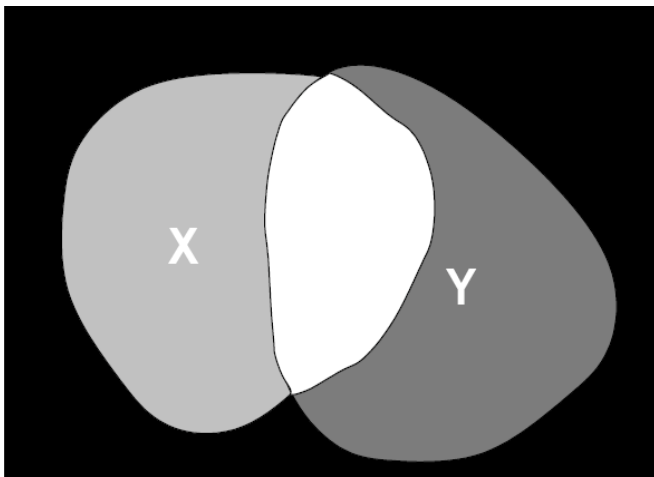
Correlation vs. Kappa index

	subject no.	Corr. Coef.		Kappa Index	
		GM	WM	GM	WM
TPS vs Manual	1	0.660	0.827	0.836	0.872
	2	0.702	0.757	0.841	0.827
	3	0.654	0.787	0.811	0.850
	4	0.410	0.678	0.723	0.770
	5	0.612	0.791	0.776	0.838
mean (sd)		0.608(0.115)	0.768(0.056)	0.798(0.049)	0.831(0.038)
SPM vs Manual	1	0.675	0.846	0.883	0.866
	2	0.686	0.839	0.887	0.880
	3	0.637	0.810	0.863	0.842
	4	0.091	0.672	0.679	0.753
	5	0.450	0.803	0.825	0.824
mean (sd)		0.518(0.250)	0.794(0.071)	0.827(0.087)	0.833(0.050)
TPS vs SPM	1	0.806	0.883	0.848	0.900
	2	0.626	0.759	0.794	0.824
	3	0.734	0.822	0.808	0.861
	4	0.426	0.767	0.730	0.793
	5	0.645	0.800	0.785	0.836
mean (sd)		0.647(0.143)	0.806(0.050)	0.793(0.043)	0.843(0.040)

Pearson correlation is not a good measure for binary output

Other set theoretic similarity measures

- Willams' index (Bouix et al., 2005)
- Housdorff distance
- Other measures:



Jaccard

$$\frac{|X \cap Y|}{|X \cup Y|}$$

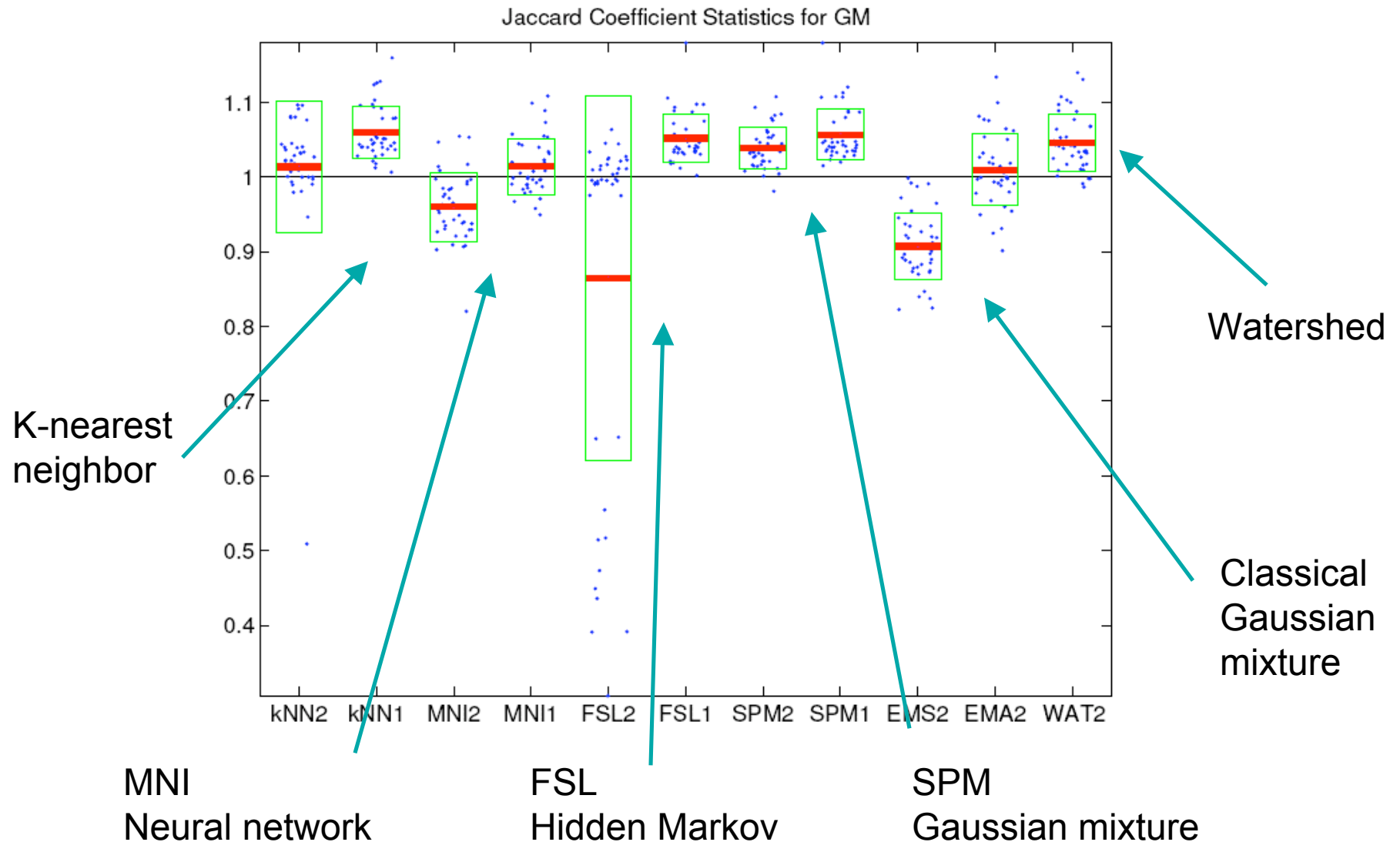
Simple matching:

$$\frac{|X \cap Y| + |\bar{X} \cap \bar{Y}|}{n}$$

Volume similarity:

$$1 - \frac{||X| - |Y||}{|X| + |Y|}$$

Performance analysis of brain image segmentation algorithms

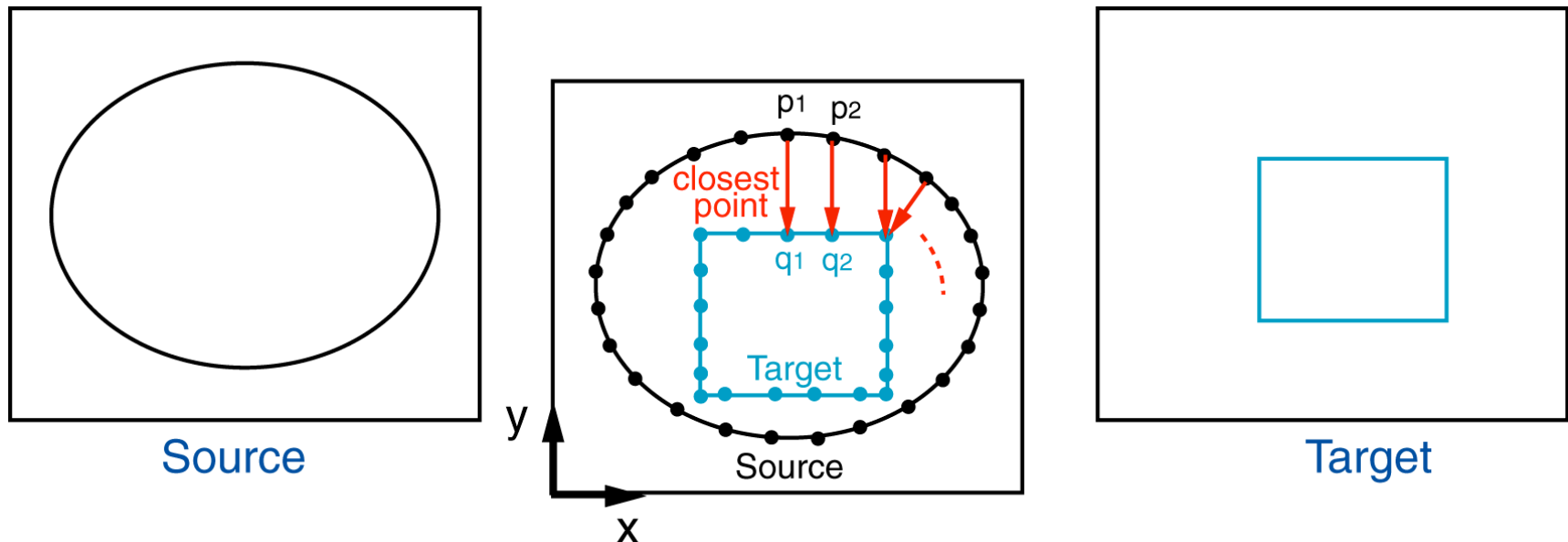


Geometric similarity measures

- Measure of surface/curve similarity

$$e(X_1, X_2) = \int \min_{p' \in A'} \|X_1(p') - X_2(p)\|^2 dA / \int \|X_2(p)\|^2 dA$$

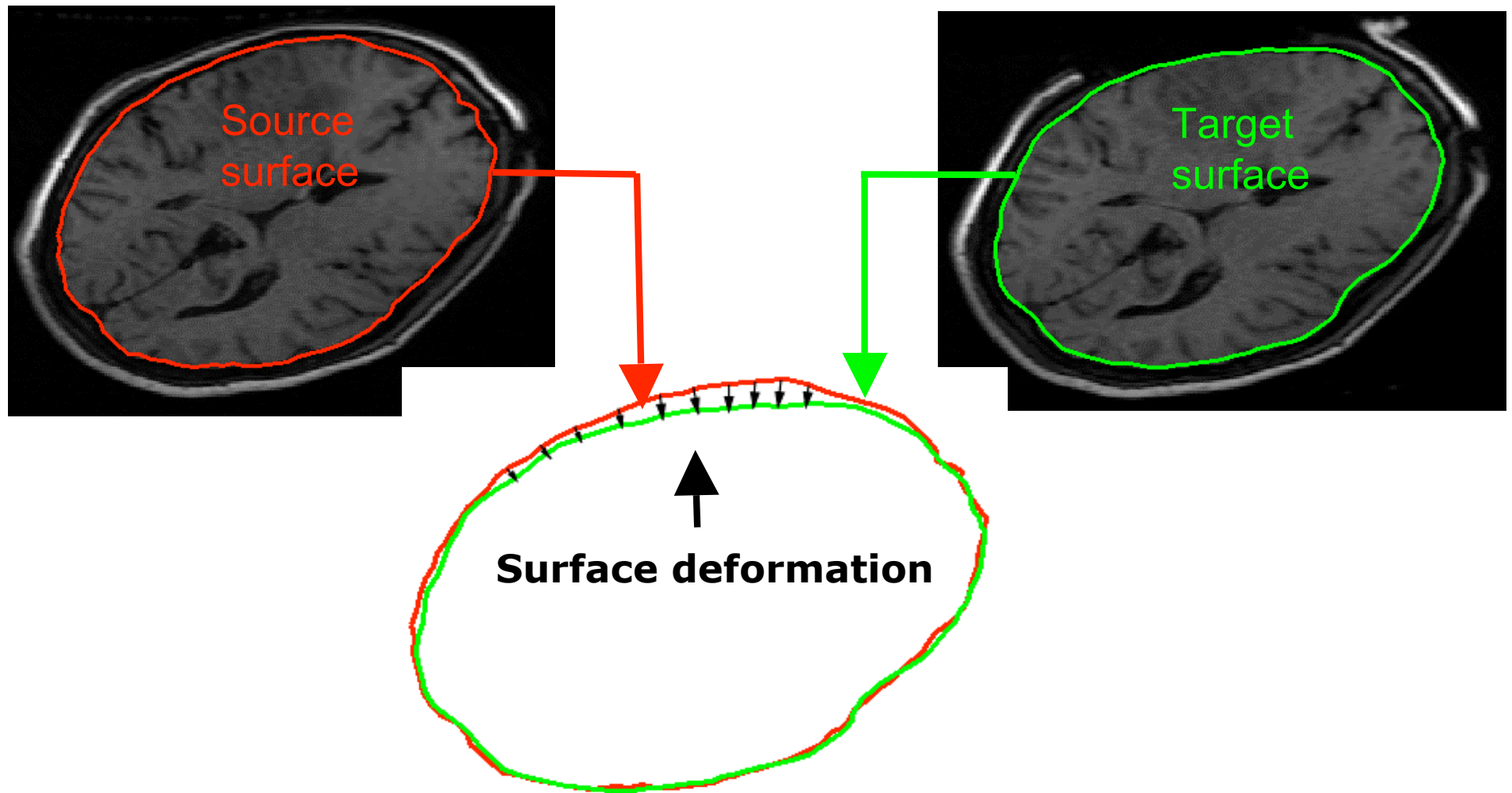
Staib (1996)



Requires surface correspondence

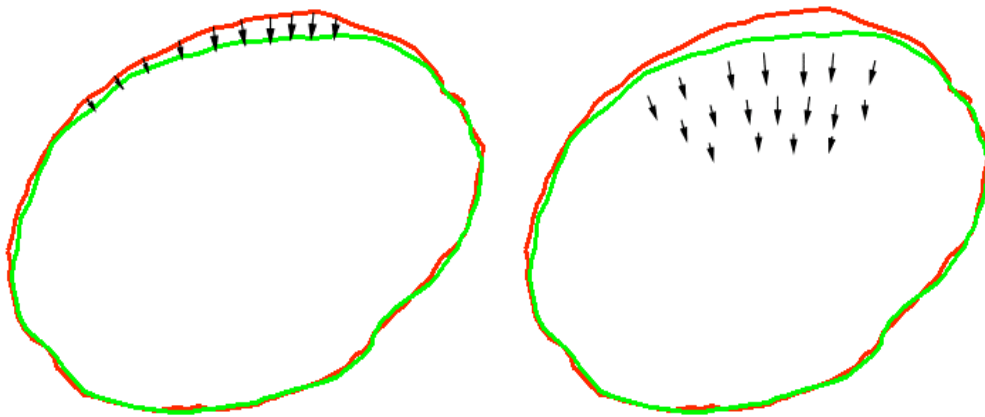
Source: L. Vigneron

Application of geometric similarity measures

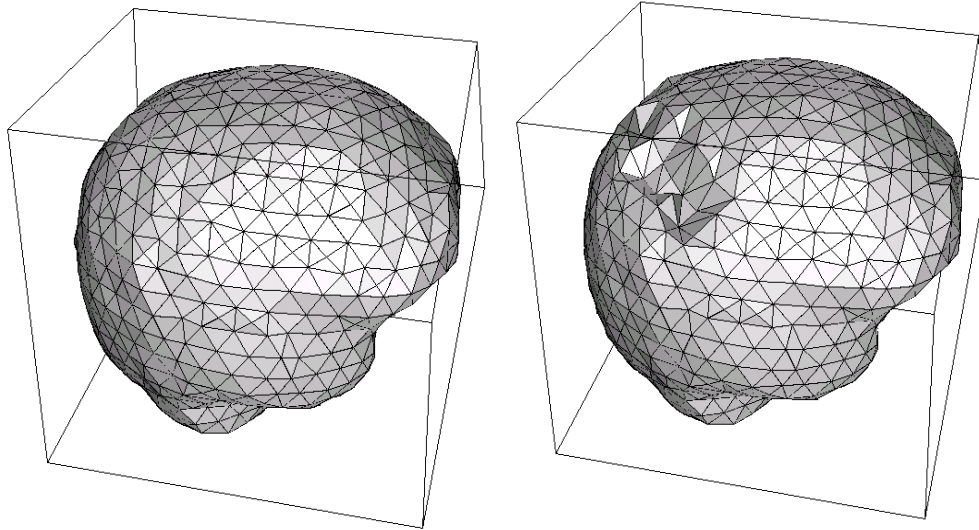


Deformable surface modeling: surface-to-surface registration

Deformable surface modeling

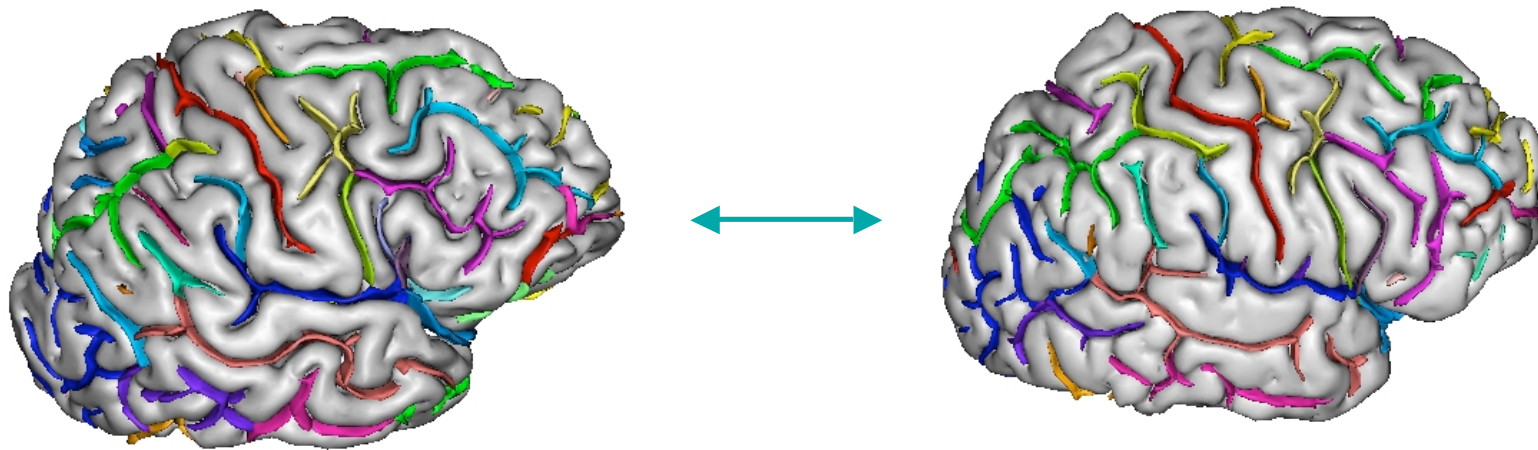


Surface-based image similarity measures have been used in surface segmentation, registration



Source: L. Vigneron

Application of surface similarity measure



Sulcal pattern matching and brain shape analysis

BrainVisa (J.-F. Mangin) output

Information theoretic similarity measure

- Entropy: thermodynamic concept of measuring the amount of chaos in a system
- Shannon (Bell lab, 1948) introduced it as the measure of complexity in information
- Kullback used it in defining the distance between two random variables

Kullback-Leibler divergence

- For probability density f and g , the Kullback-Leibler divergence (or KL distance) is defined as

$$H(f, g) = \mathbb{E}_f \ln \frac{f(X)}{g(X)} = \int \log \frac{f(x)}{g(x)} f(x) .dx$$

$$H(f, g) \geq 0$$

Equality is obtained when $H(f, f) = 0$

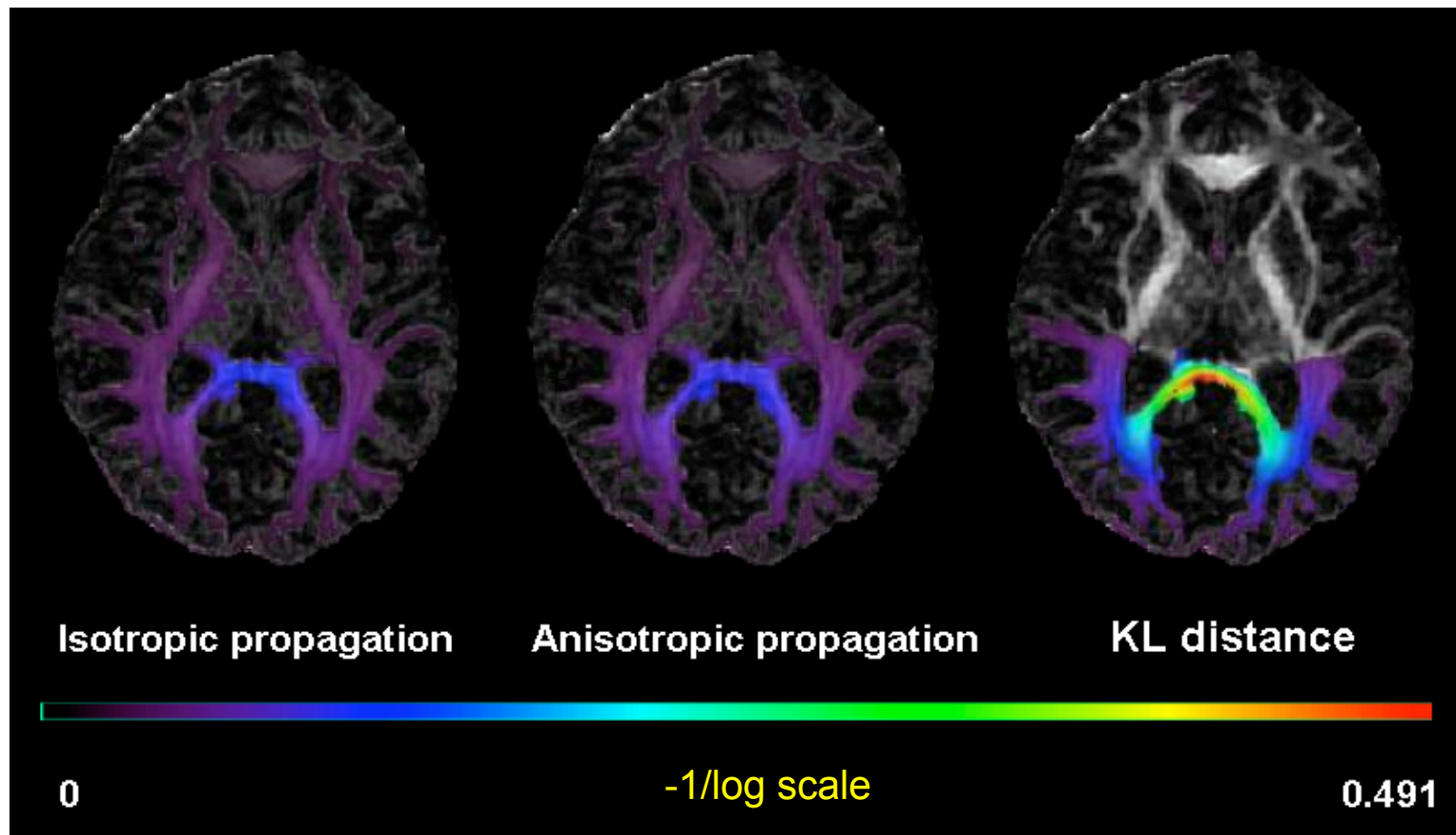
H is called ***relative entropy***

Computing KL-distance

- In practice, we estimate the distribution of a random variable via the empirical distribution (histogram).
- Discretization:

$$\int \ln \frac{f(x)}{g(x)} f(x) dx = \sum_i \ln \frac{f(x_i)}{g(x_i)} f(x_i)$$

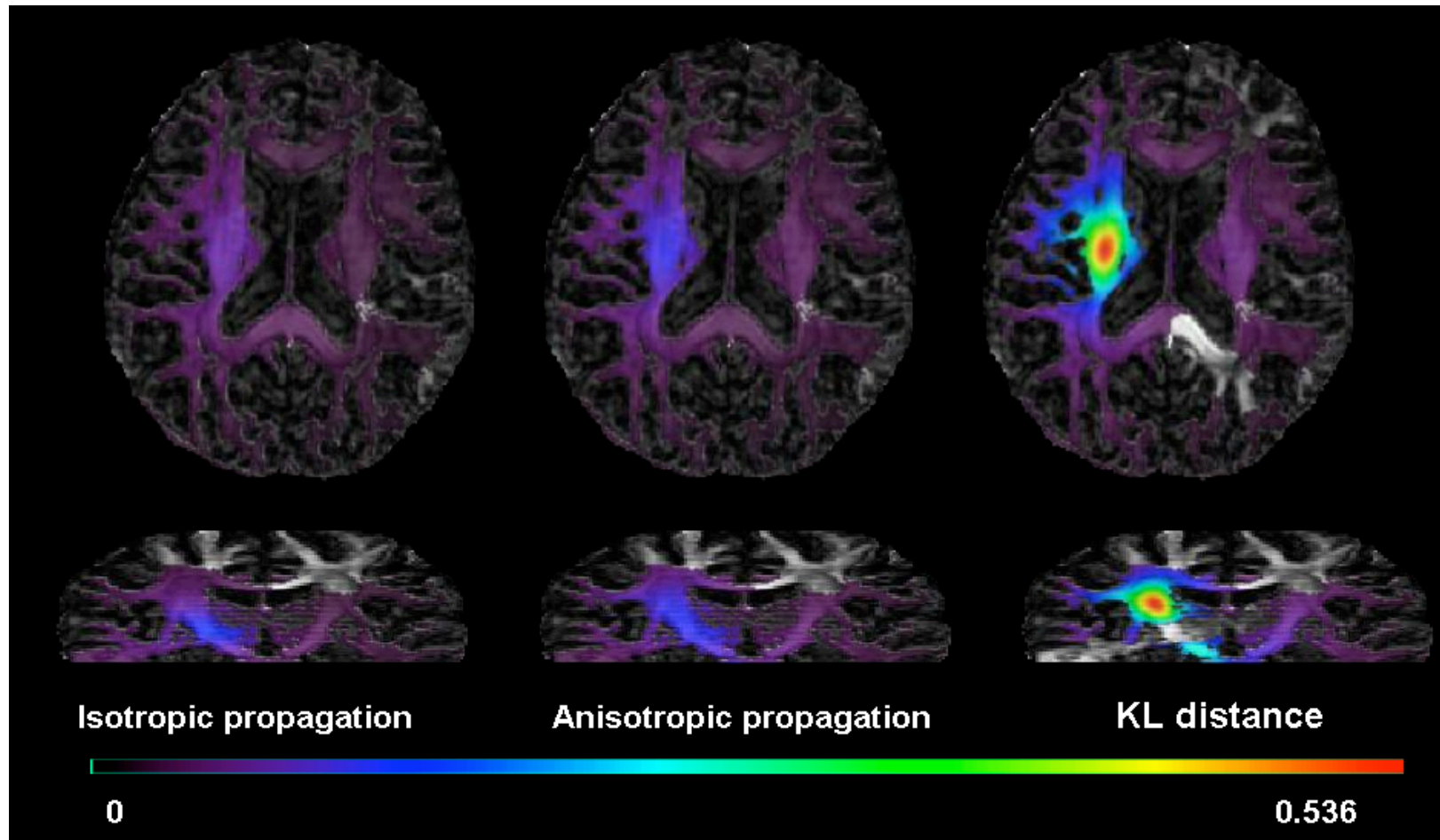
DTI connectivity via KL-distance



KL-distance is computed within a small window around each voxel

Source: PhD work of J.E. Lee

KL-distance in DTI



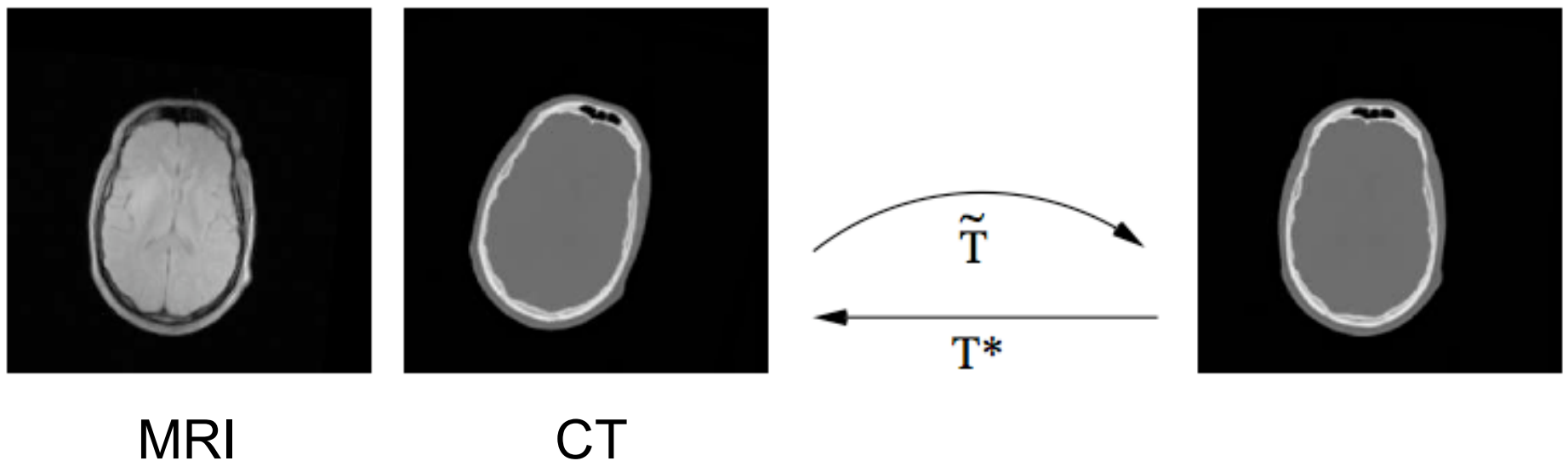
Null model

Alternate model

Significance of connectivity can be also tested.

Other use of KL-distance

- Registration cost function for locally varying image intensity histogram



Widely used similarity measure in multimodal image registration (SPM cosine-basis approach)

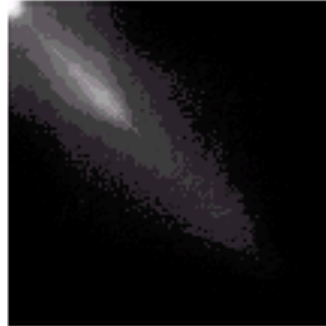
Source: Zollei (assigned paper)

KL-distance as image similarity

- Image intensities f_1 , f_2 are sorted
- Sorted image intensities are plotted in 2D
- KL-distance measures the dispersion of 2D histogram



3.82



6.79



6.98



7.15

Mutual information

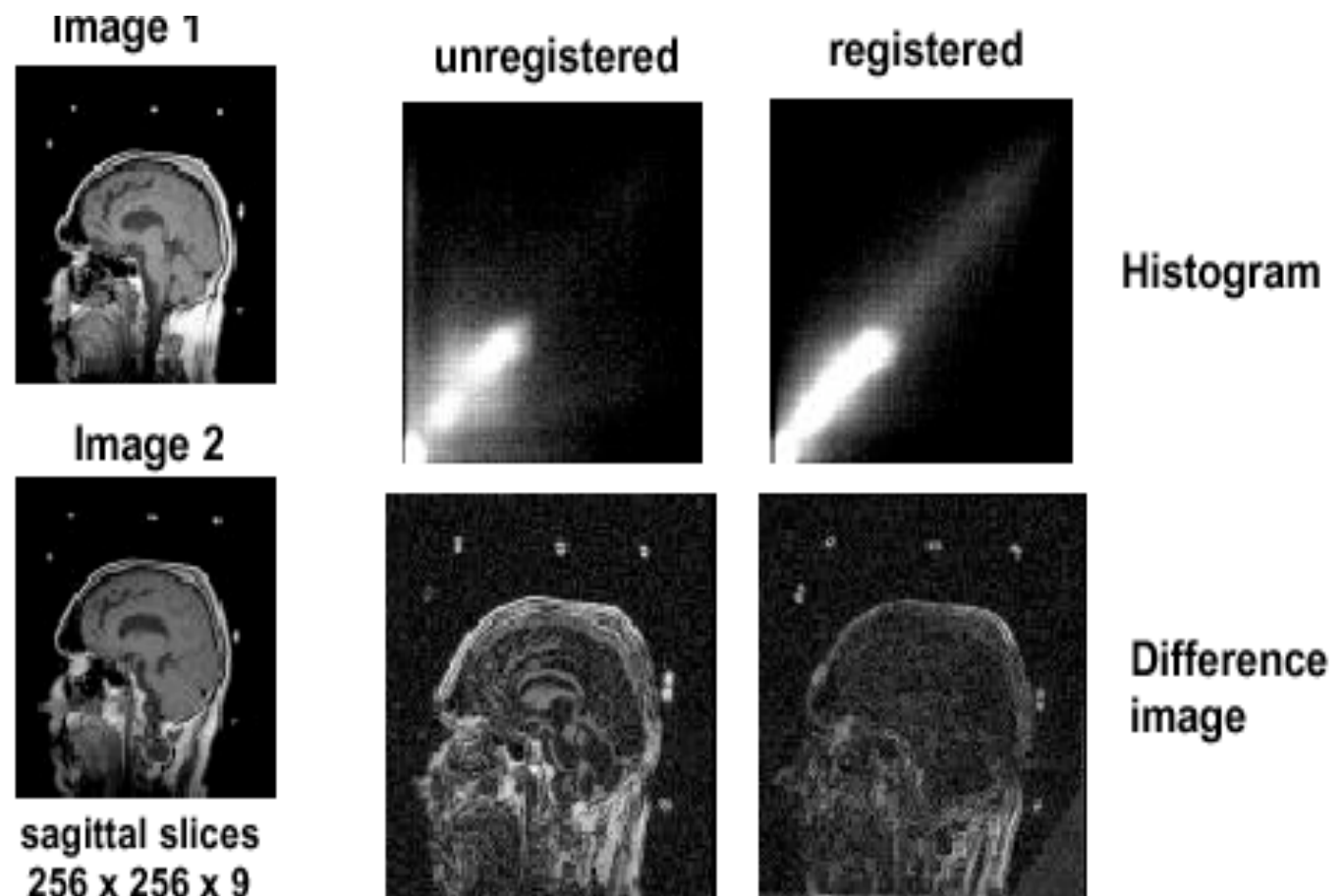
- Let f be the joint density of f_1 and f_2 . The mutual information of f_1 and f_2 is defined as

$$I(f_1, f_2) = H(f, f_1 f_2)$$

- If $f=f_1 f_2$, two random variables are independent. If random variables are independent, $I(f_1, f_2)=0$.

Note: If two random variables are independent, $\text{corr}(f_1, f_2)=0$. So mutual information behaves like correlation.

- Image registration is done by maximizing mutual information (minimizing joint entropy)



Source: Mangin