Neuroimage Processing

Instructor: Moo K. Chung

mkchung@wisc.edu

Lecture 2.

General Linear Models (GLM)

Multivariate General Linear Models (MGLM)

September 11, 2009

Research Projects

If you have your own imaging data, you can do your own project after consultation with me but you need to apply various techniques covered in class.

If you don’t have your own data, simply use the in-class data sets.

Least squares estimation (LSE)

A method for determining unknown parameters of an overdetermined linear system by minimizing the sum of squared residuals.

Properties of LSE

- Invented by Gauss in 1794.
- LSE can be done without any statistical assumption.
- LSE is the maximum likelihood estimation with Gaussian noise.
- Always try LSE first before you try anything more complicated.

Example.

Solve the following over determined system:

a= 1

a= 2

Sum of squared residual = (a-1)^2 + (a-2)^2.

The minimum is given by a=1.5, the sample mean.

Normal equations

a= 1
a= 2

$$\begin{pmatrix}1\\1\end{pmatrix}a=\begin{pmatrix}1\\2\end{pmatrix}$$

$$(1,1)\begin{pmatrix}1\\1\end{pmatrix}a=(1,1)\begin{pmatrix}1\\2\end{pmatrix}$$

$$2a=3$$

General Linear Model (GLM)

The general linear model (GLM) is a very flexible and general statistical framework encompassing a wide variety of fixed effect models such as the multiple regressions, the analysis of variance (ANOVA), the multivariate analysis of variance (MANOVA), the analysis of covariance (ANCOVA) and the multivariate analysis of covariance (MANCOVA)

The parameters of the model are mainly estimated by the least squares estimation and has been implemented in many statistical packages such as R or Splus and brain imaging packages such as SPM and fMRI-STAT.

Read section 2.9 of the textbook

Model

Let y_i be the response variable, which is mainly coming from images and $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})$ to be the variable of interest and $\mathbf{z}_i = (z_{i1}, \cdots, z_{ik})$ to be nuisance variables corresponding to the i-th subject. We assume there are n subjects. Then we have a GLM

$$y_i = \mathbf{z}_i\lambda + \mathbf{x}_i\beta + \epsilon_i$$

where $\lambda = (\lambda_1, \cdots, \lambda_k)'$ and $\beta = (\beta_1, \cdots, \beta_p)'$ are unknown parameter vectors to be estimated. We assume ϵ to be the usual zero mean Gaussian noise. Then we determine the significance of the variable of interests $\mathbf{x}_i$ by testing the null hypothesis

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

Inference

The fit of the reduced model corresponding to $\beta = 0$, i.e. $y_i = \mathbf{z}_i\lambda$, is measured by the sum of the squared errors (SSE):

$$\text{SSE}_0 = \sum_{i=1}^{n}(y_i - \mathbf{z}_i\widehat{\lambda}_0)^2,$$

where $\widehat{\lambda_0}$ is the least squares estimation obtained from the reduced model. Similarly the fit of the full model corresponding to $\beta \neq 0$, i.e. $y_i = \mathbf{z}_i\lambda + \mathbf{x}_i\beta$, is measured by

$$\text{SSE}_1 = \sum_{i=1}^{n}(y_i - \mathbf{z}_i\widehat{\lambda}_1 - \mathbf{x}_i\widehat{\beta}_1)^2,$$

where $\widehat{\lambda}_1$ and $\widehat{\beta}_1$ are estimated from the full model.

Then under H_0, the test statistic is the ratio

$$F = \frac{(\text{SSE}_0 - \text{SSE}_1)/p}{\text{SSE}_0/(n-p-k)} \sim F_{p,n-p-k}.$$

The larger the F value, it is more unlikely to accept H_0.

Estimation

The unknown parameters are estimated via the least squares method. The detailed exposition of the least squares estimation using the matrix inversion is necessary for numerical implementation in `MATLAB` and it is the basis of SPM and fMRI-STAT. The reduced model (2.1) can be written in a matrix form

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk} \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}_{\boldsymbol{\lambda}}. \tag{2.4}$$

By multiplying $\mathbf{Z}'$ on the both sides, we obtain

$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{Z}\boldsymbol{\lambda}.$$

Now the matrix $\mathbf{Z}'\mathbf{Z}$ is a full rank and can be invertible if $n \geq k$. Therefore, the matrix equation can be solved by performing a matrix inversion

$$\widehat{\boldsymbol{\lambda}}_0 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Similarly the full model can be written in a matrix form by concatenating the row vectors $\mathbf{z}_i$ and $\mathbf{x}_i$ into a larger row vector $(\mathbf{z}_i, \mathbf{x}_i)$, and the column vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ into a larger column vector $(\boldsymbol{\lambda}', \boldsymbol{\beta}')'$. Then the full model can be written in a matrix form and solved similarly.

Summary

$$Y = \mathbf{z}\lambda + \mathbf{x}\beta + e$$

$\mathbf{x} = (x_1, \cdots, x_p)$ Variables of interest

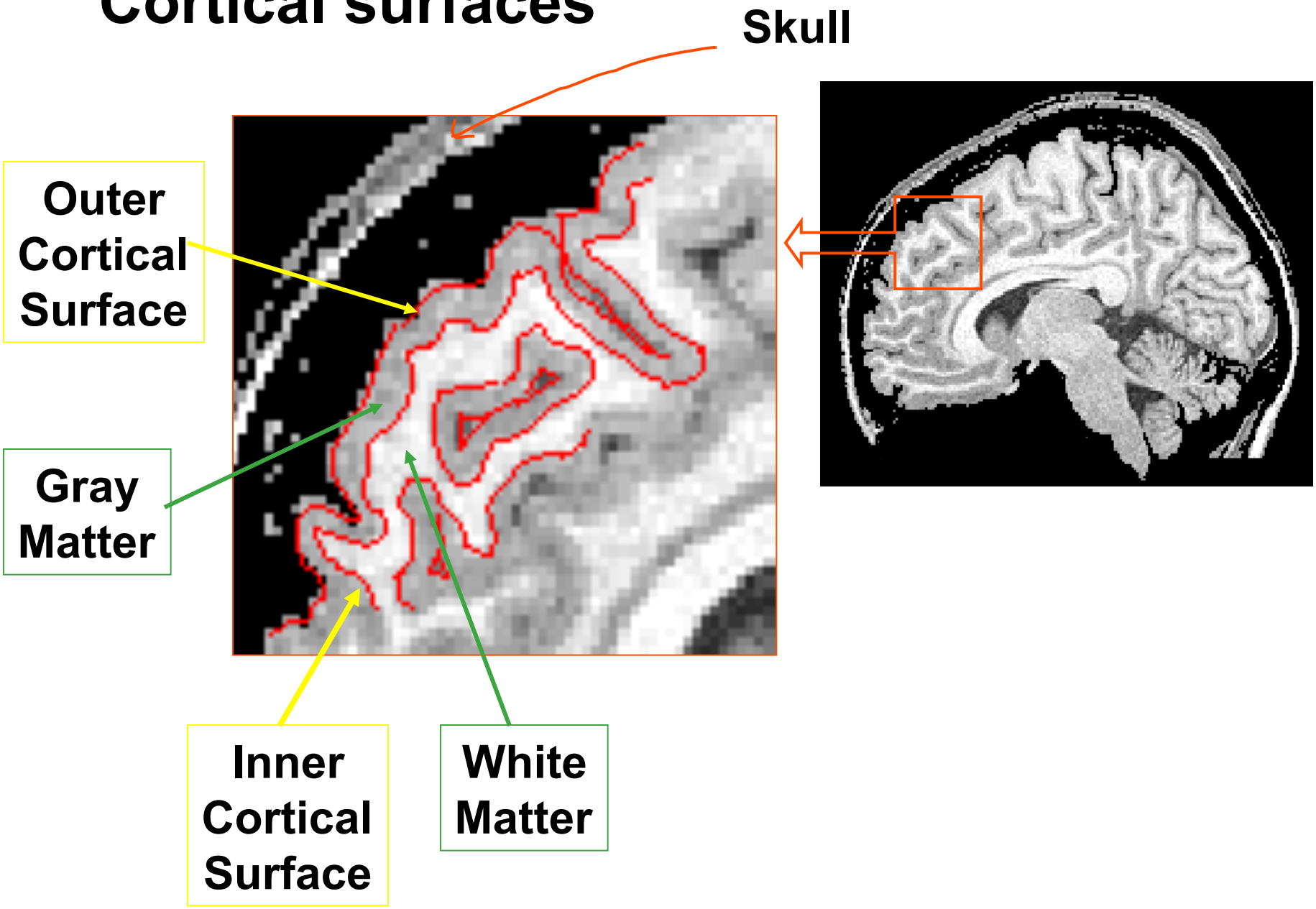
$\mathbf{z} = (z_1, \cdots, z_k)$ Nuisance covariates

$$F = \frac{(SSE_0 - SSE_1)/p}{SSE_0/(m+n-p-k)} \sim F_{p,m+n-p-k}$$

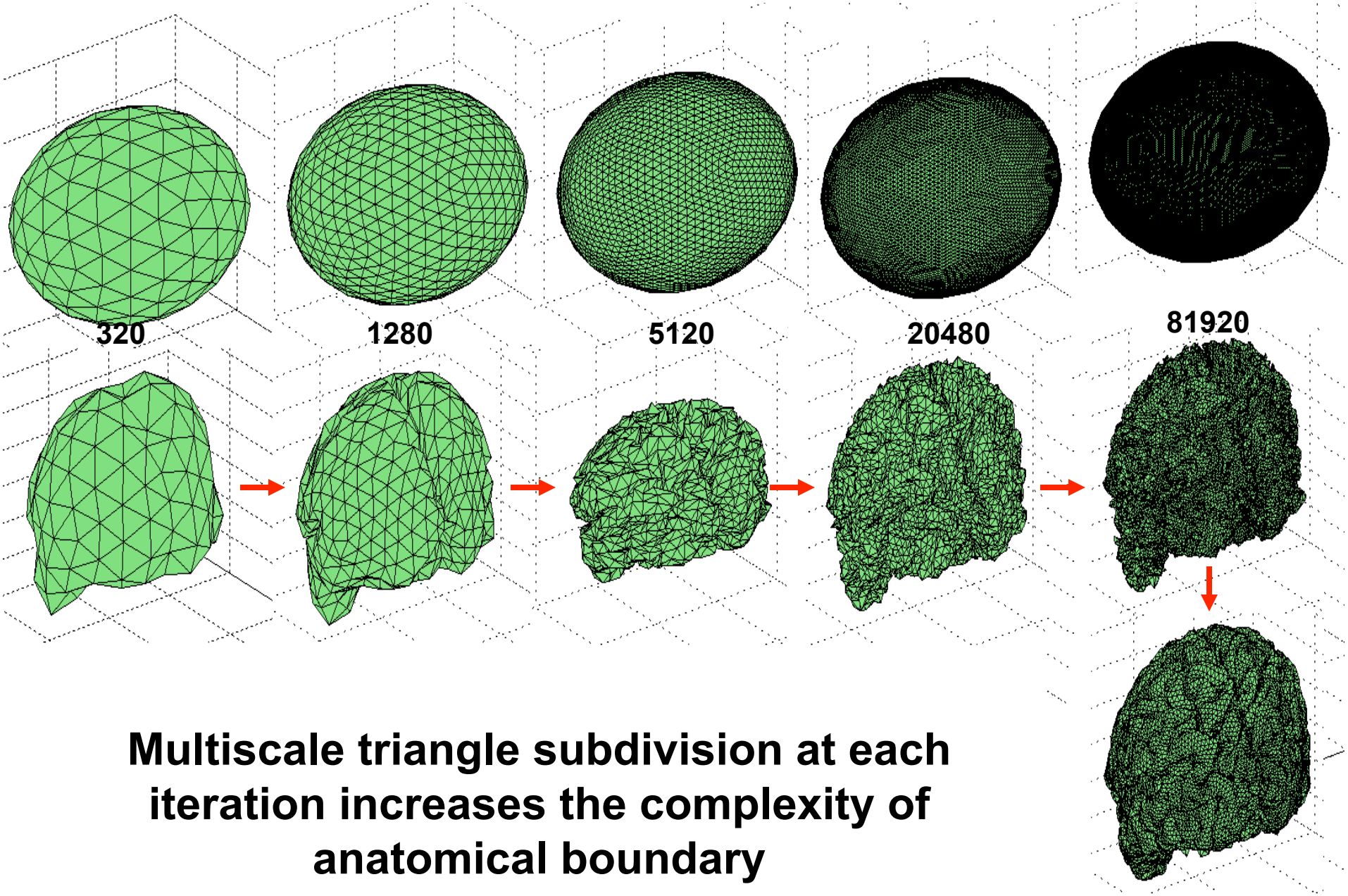
Computing the sum of squared errors (residuals) requires the least squares estimation (LSE) of unknown parameters.

Application:
GLM on cortical thickness

Cortical surfaces

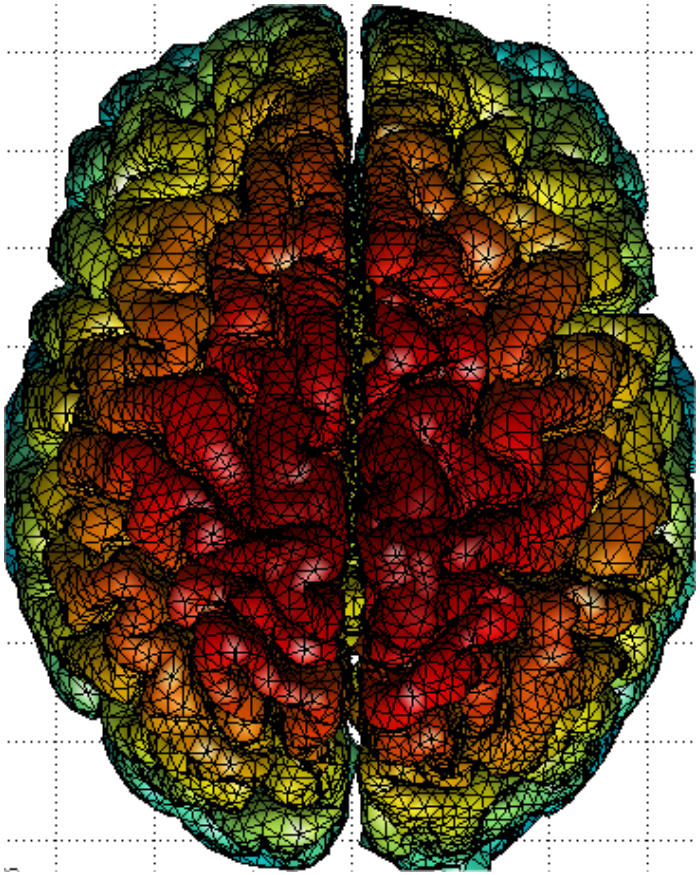


Deformable surface algorithm McDonalds *et al*. (2001) NeuroImage



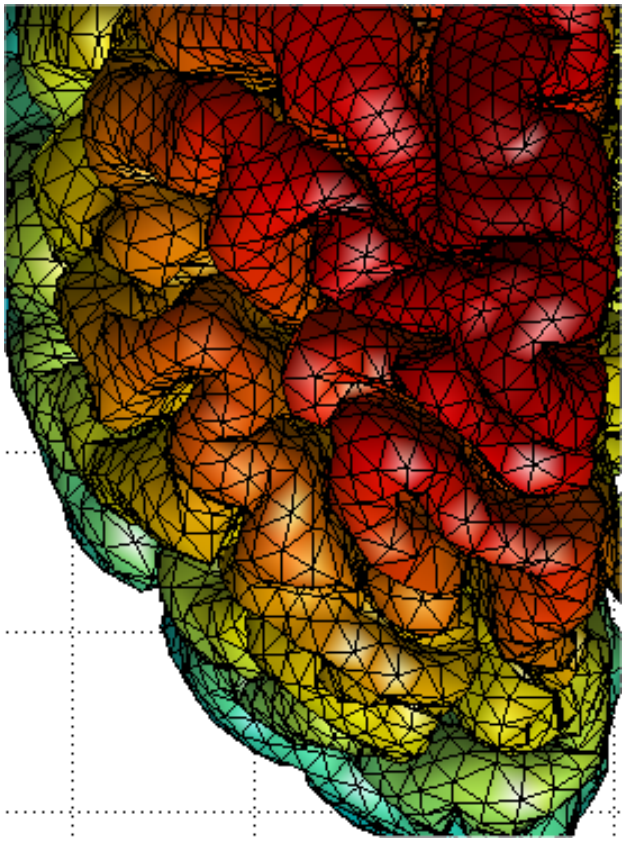
Multiscale triangle subdivision at each iteration increases the complexity of anatomical boundary

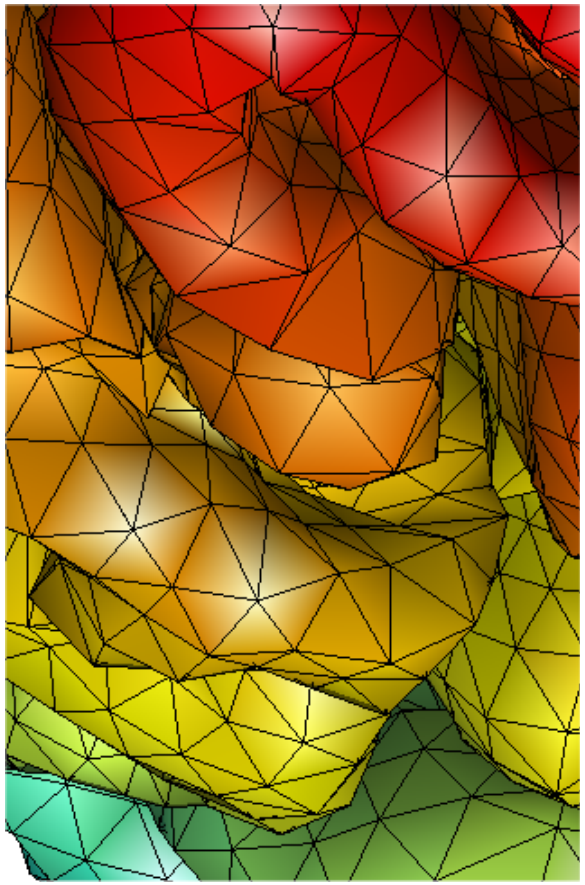
Cortical Surface
Polygonal mesh
Mesh resolution 3mm



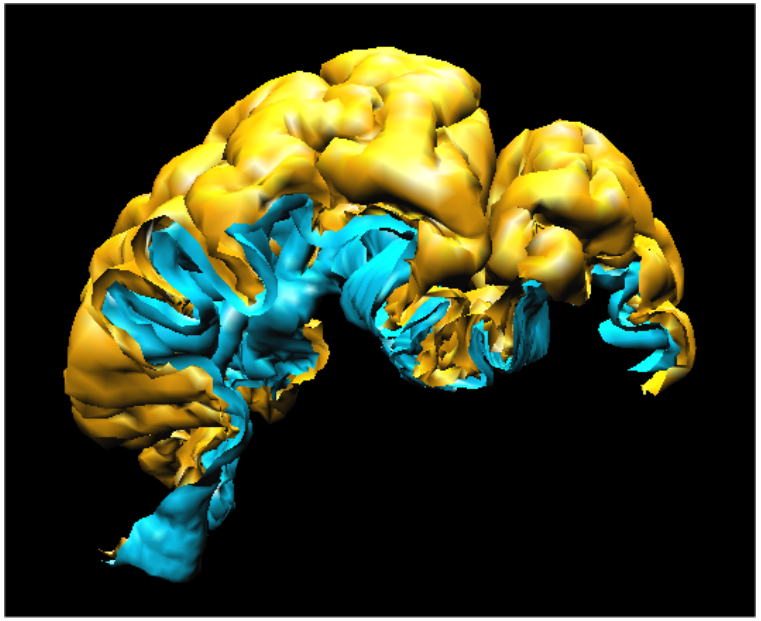
82,190 triangles

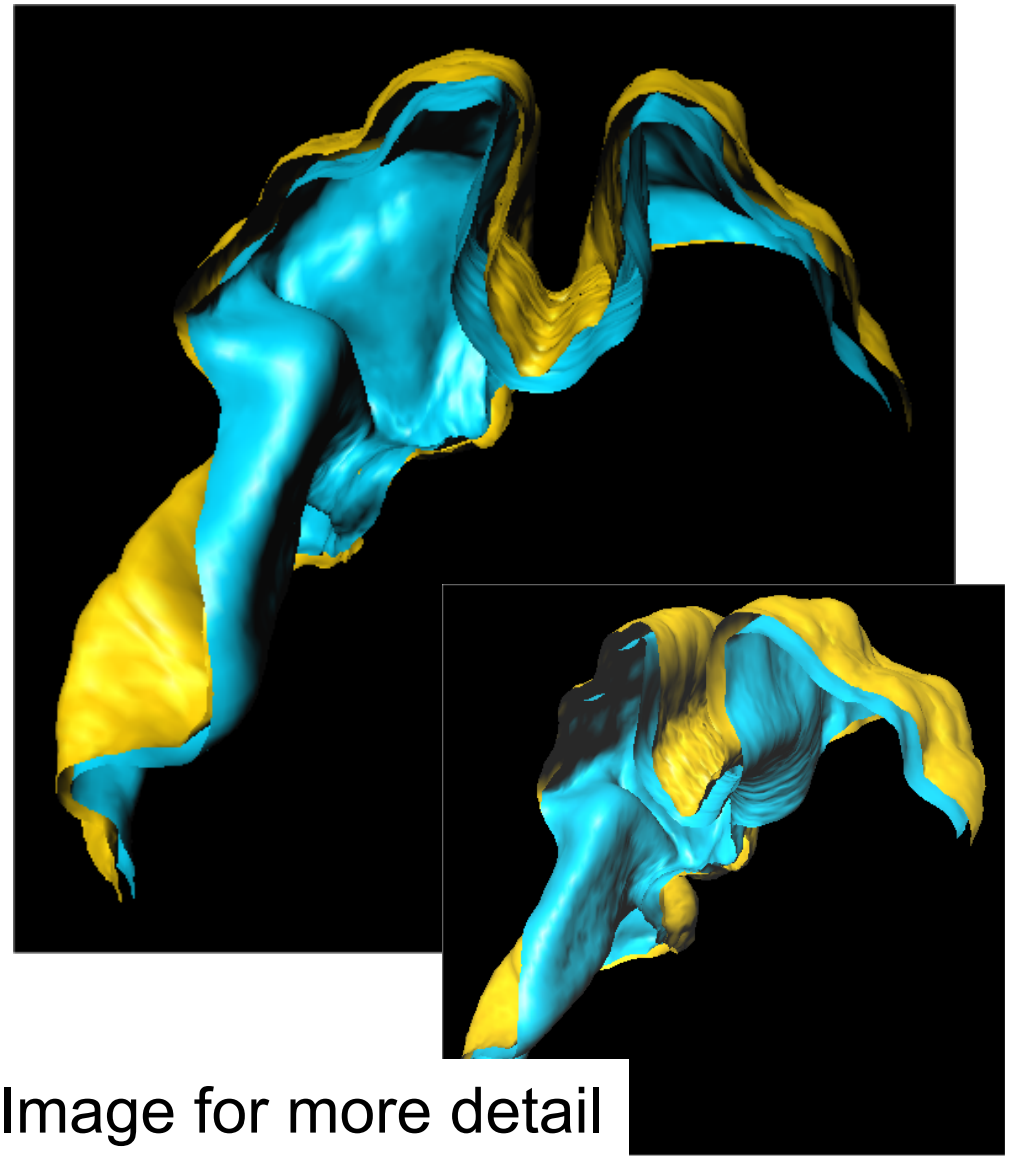
40,962 vertices





Outer and inner cortical surfaces





Read Chung et al., 2003 NeuroImage for more detail

GLM on thickness

Given two groups (autism vs. control), we are interested in testing the significance of group difference on cortical thickness. So we consider the following GLM:

$$\texttt{thickness}_i = \lambda_1 + \beta_1 \cdot \texttt{group}_i + \epsilon, \tag{2.5}$$

where the dummy variable `group` is 1 for autism and 0 for control. This the case for $k = 1$ and $z_{i1} = 1$, and $p = 1$. The reduced model in this case is

$$\texttt{thickness}_i = \lambda_1.$$

The least squares estimation of λ_1 is simply the sample mean given by

$$\widehat{\lambda}_1 = \frac{1}{n} \sum_{i=1}^{n} \texttt{thickness}_i.$$

The test statistic F is then distributed as $F_{1,n-1-k}$, which is the square of the student t-distribution with $n-1-k$ degrees of freedom, i.e. t^2_{n-1-k}. The advantage of using the t-statistic is that unlike the F-statistic, it has two sides so we can actually use it to test for one sided alternative hypothesis $H_1 : \beta_1 \geq 0$ or $H_1 : \beta_1 \leq 0$. Therefore, the t-statistic map can provides the direction of the difference (if autism is thicker or thinner) that the F-statistic map cannot provide.

The model (2.5) is not necessarily a proper model since the model did not incorporate the possible confounding effects of brain size and age variations for each subject. In order to control the possible confounding effect of age, we consider consider following GLM:

$$\texttt{thickness}_i = \lambda_1 + \lambda_2 \cdot \texttt{age}_\texttt{i} + \beta_1 \cdot \texttt{group}_\texttt{i} + \epsilon_\texttt{i}, \tag{2.6}$$

Why GLM?

The effect of nuisance covariates such as age and gender

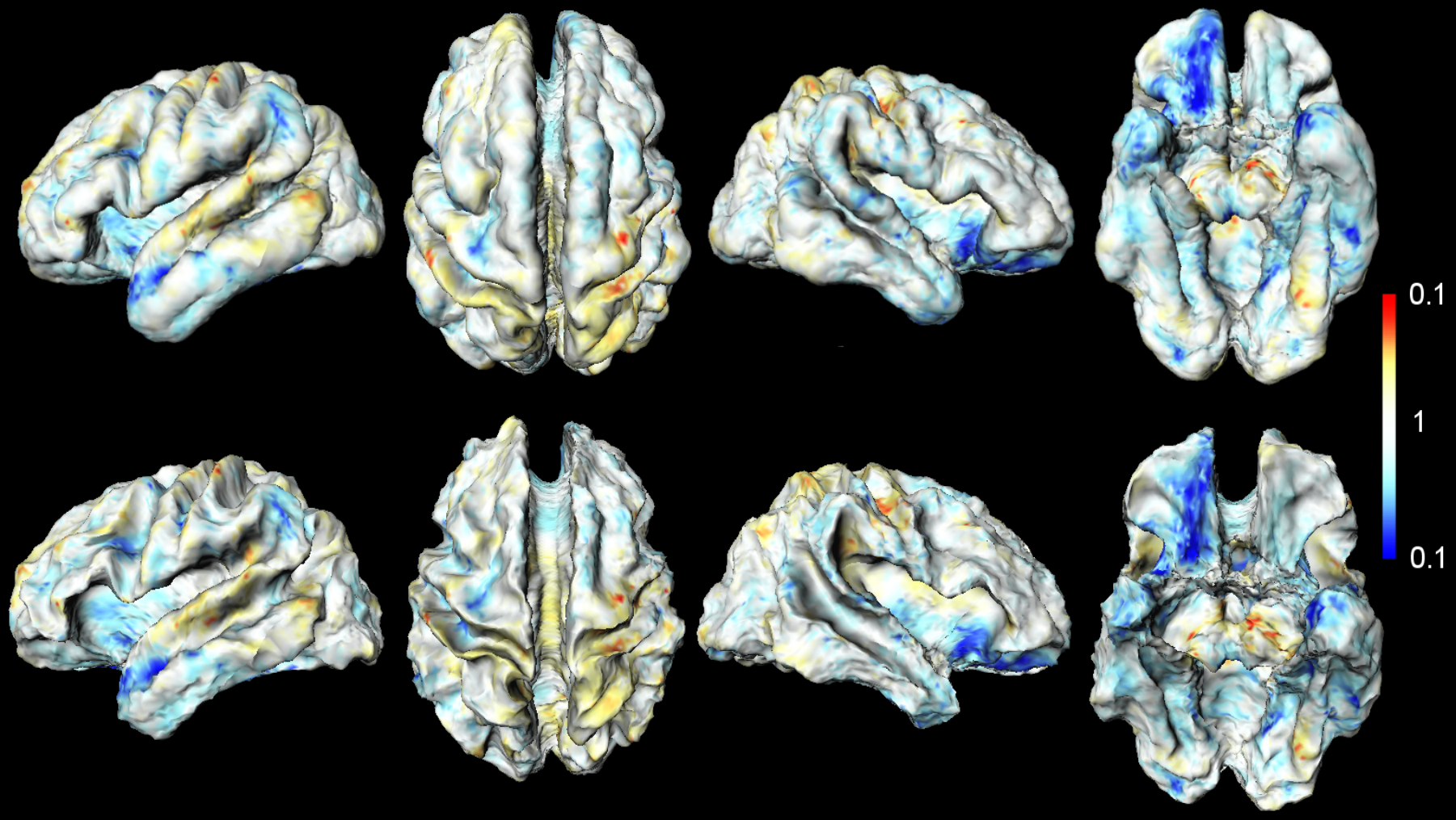
Age distribution for autistic subjects:

16.1 (s.d. 4.5)

Age distribution for control subjects:

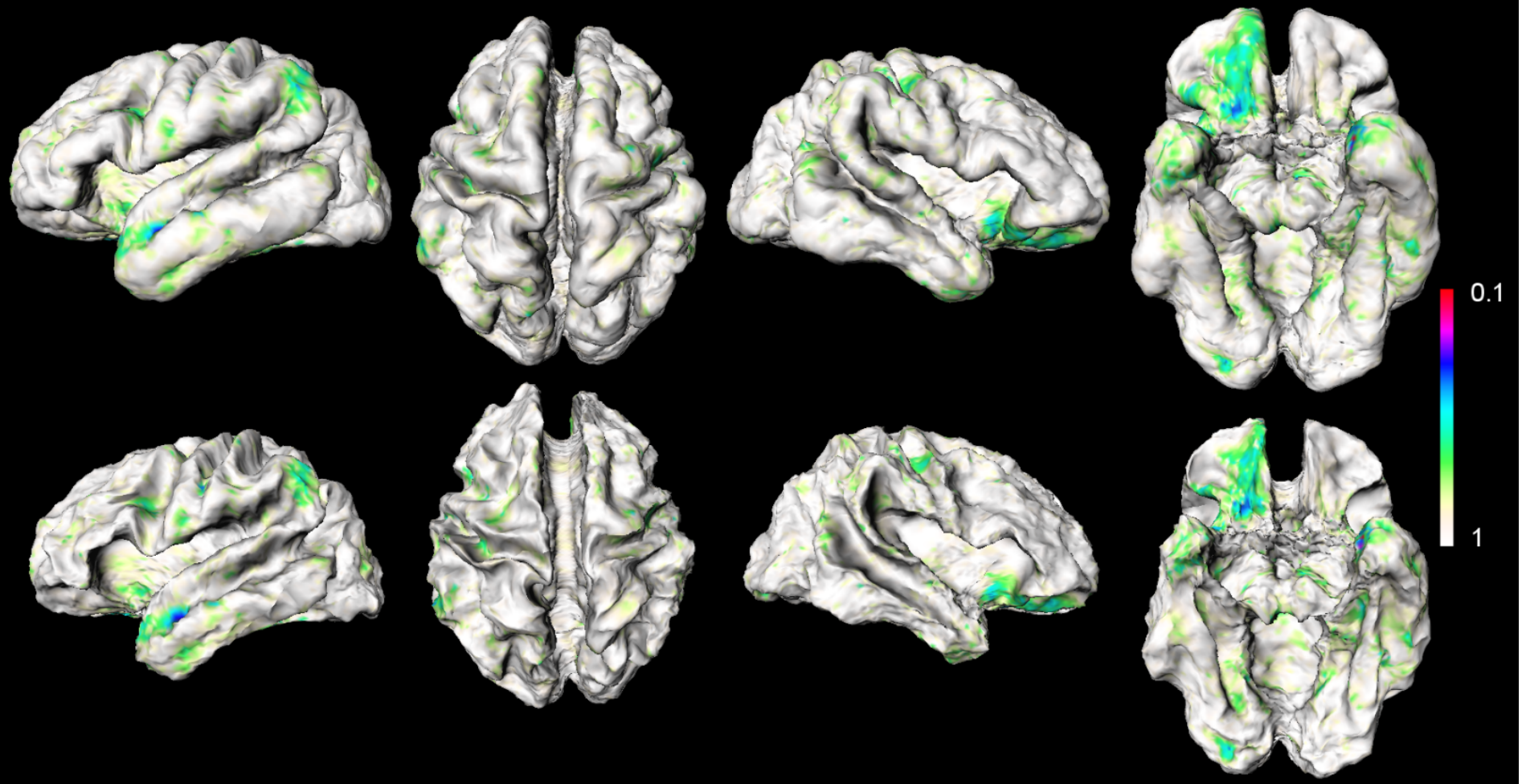
17.1 (s.d. 2.8)

p-value map for *t*-test on cortical thickness difference



Decrease: left superior temporal sulcus, left occipital-temporal gyrus, right orbital prefrontal
Increase: left superior temporal gyrus, left middle temporal gyrus, left and right postcentral sulci

p-value map for *F*-test removing age effect



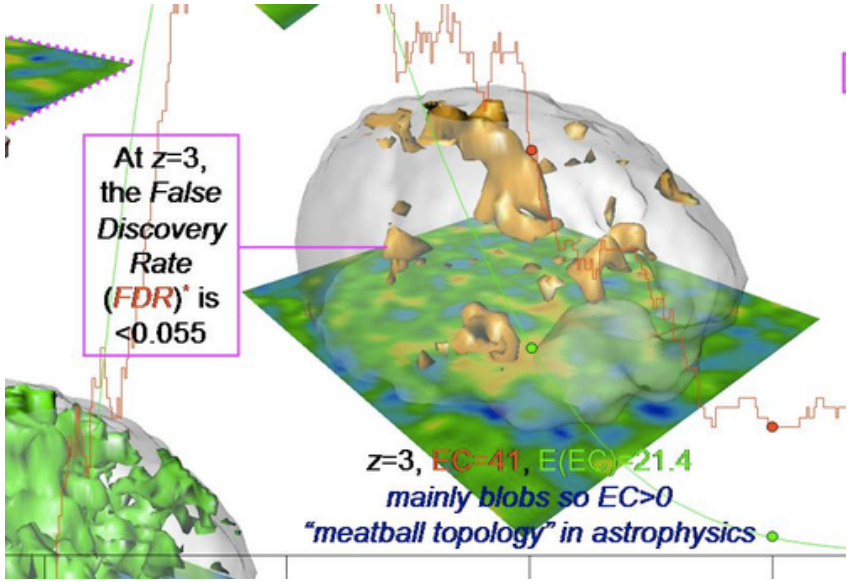
Decrease: left superior temporal sulcus
left occipital-temporal gyrus
right orbital prefrontal

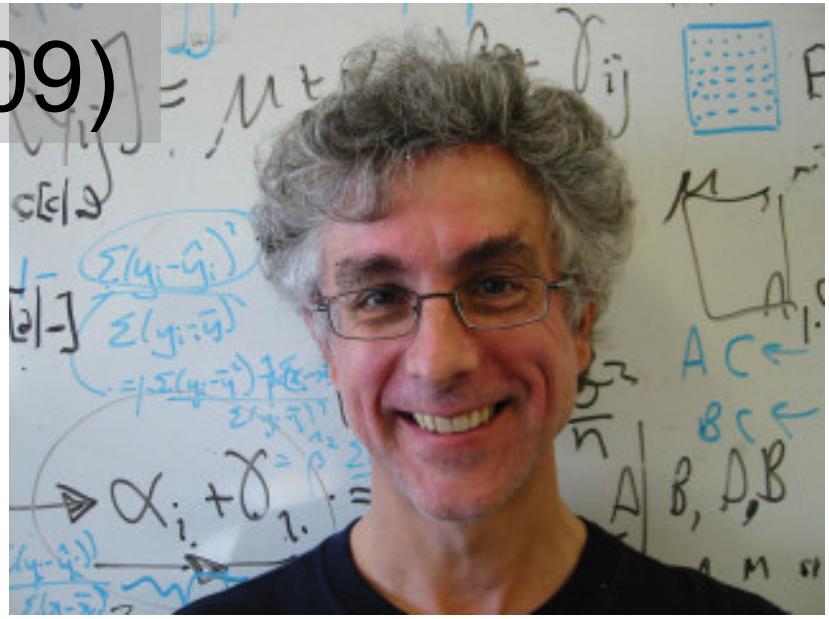
It is better to remove the age effect in anatomical data especially for developmental age range (10-20 years).

MATLAB demonstration of GLM

/matlab/CMN.lecture02.MGLM.09.10.2010.m

Keith Worsley (1951-2009)





Brain imaging research done for 21 years between 1987-2008. The first major publication in 1992.

Keith Worsley's SurfStat Package

http://www.math.mcgill.ca/keith/surfstat/

SurfStat is a Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory.

Brain substructure modeling:
http://www.stat.wisc.edu/~mchung/research/amygdala/

Reference:

Worsley, K.J., Taylor, J.E., Carbonell, F., Chung, M.K., Duerden, E., Bernhardt, B., Lyttelton, O., Boucher, M., Evans, A.C. (2009).
SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory.
NeuroImage, OHBM

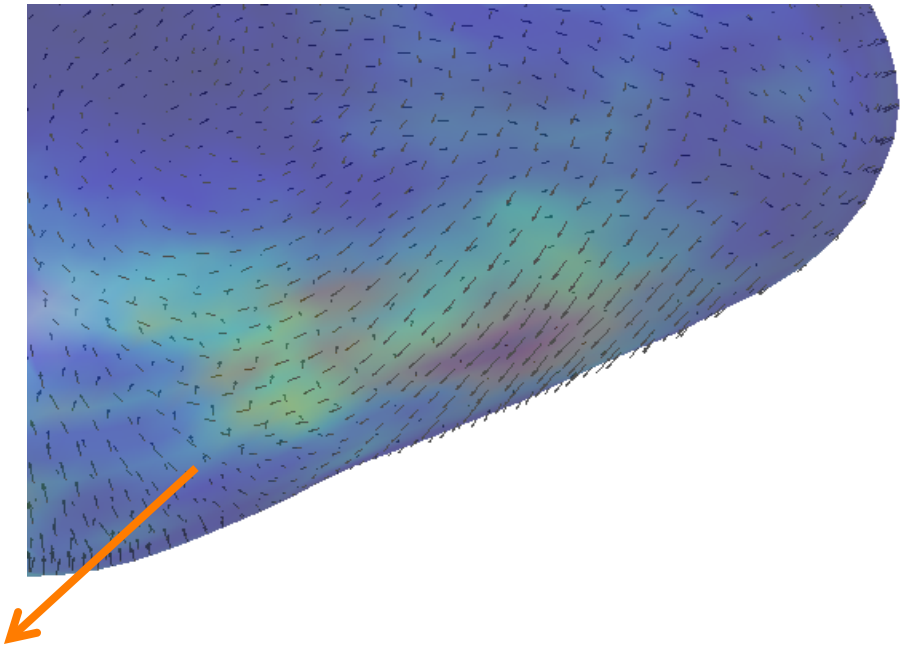
Chung, M.K., Worsley, K.J., Nacewicz, B.M., Dalton, K.M., Davidson, R.J. 2010. General multivariate linear modeling of surface shapes using SurfStat. 53:491-505.

Multivariate General Linear Model

Multivariate version of general linear model.
SPM, AFNI do not have it. Only SurfStat has it.

Read section 5.3 of the textbook

Given vector data P, we set up



covariance matrix

$$P_{n\times 3} = X_{n\times p}B_{p\times 3} + Z_{n\times r}G_{r\times 3} + U_{n\times 3}\Sigma_{3\times 3},$$

displacement vector

variable of interest

nuisance covariates

noise

$$P_{n\times 3} = X_{n\times p} B_{p\times 3} + Z_{n\times r} G_{r\times 3} + U_{n\times 3}\Sigma_{3\times 3}, \qquad (5.10)$$

where P is the matrix of coordinates, X is the matrix of contrasted explanatory variables, and B is the matrix of unknown coefficients. Nuisance covariates are in the matrix Z and the corresponding coefficients are in the matrix G. The subscripts denote the dimension of matrices. The components of Gaussian random matrix U are zero mean and unit variance. Σ accounts for the covariance structure of coordinates. Then we are interested in testing the null hypothesis

$$H_0 : B = 0.$$

For the reduced model corresponding to $B = 0$, the least squares estimator of G is given by solving $P = ZG$. i.e.

$$\widehat{G}_0 = (Z'Z)^{-1}Z'P.$$

We will assume that there is more sample size n than the number of parameters r to be estimated. The residual sum of squares of the reduced model is

$$E_0 = (P - Z\widehat{G}_0)'(P - Z\widehat{G}_0).$$

For the full model, the parameters are estimated by solving

$$P_{n\times 3} = XB + ZG = [X, Z]_{n\times(p+r)} \begin{bmatrix} B \\ G \end{bmatrix}_{(p+r)\times 3}.$$

The least squares estimation is given by

$$\begin{bmatrix} \widehat{B} \\ \widehat{G} \end{bmatrix} = ([X, Z]'[X, Z])^{-1}[X, Z]'P.$$

The corresponding residual sum of squared error is

$$E = (P - X\widehat{B} - Z\widehat{G})'(P - X\widehat{B} - Z\widehat{G}).$$

By comparing how large the residual E is against the residual E_0, we can determine the significance of coefficients B. However, since E and E_0 are matrices, we take a function of eigenvalues of E_0E^{-1} as a statistic. Since we expect the sample size n to be larger than 3, there are three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ satisfying

$$\det(E_0 - \lambda E) = 0.$$

This requires solving the generalized eigenvalue problem

$$E_0 v = \lambda E v$$

Lecture 3 topics

Least squares estimation
Fourier representation
Iterative Residual Fitting

Please read:
shen.2003
styner.2006
chung.2010.NI

Li Shen Seminar Next week

Thursday Mok-am Hall Bldg 500 2:00pm

Friday 11:00am This room