Group Analysis

Alexa Morcom

Edinburgh SPM course, 2017

Thanks to Jesper Anderson, Tom Nichols, Jean Daunizeau, Stephan Kiebel & other SPM authors for slides

Overview

Making the group-level inferences we want

- Optimising the GLM
- The two-stage GLM
- Two methods of RFX inference

2-stage GLM

Single subject

Each subject's scans are modelled separately
Single subject parameter estimates

1st level

Single subject **contrasts of parameter estimates** represent different hypothesis tests

Group/s of subjects

A group model is made using the contrasts
Parameter estimates apply to group effect/s

2nd level

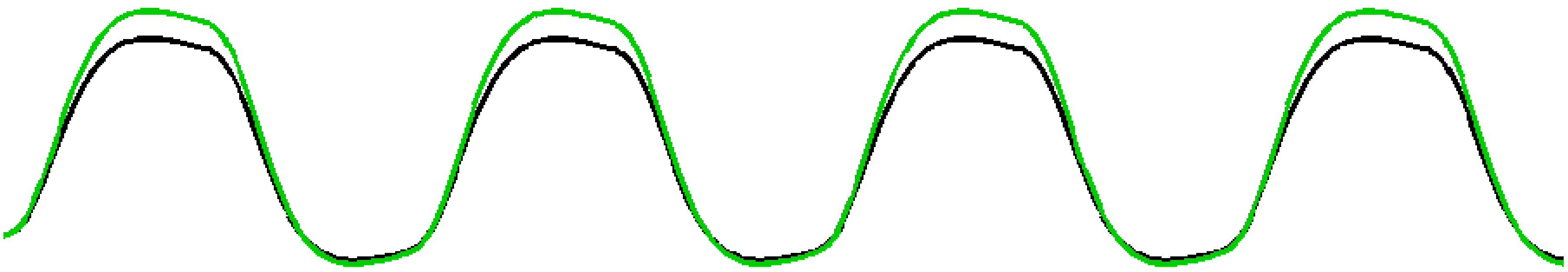
Group level **contrasts of 2nd level parameter estimates** are used to form statistics

- Hierarchical models
- Mixed-effects models
- Random effects (RFX) models
- Variance components

... All the same

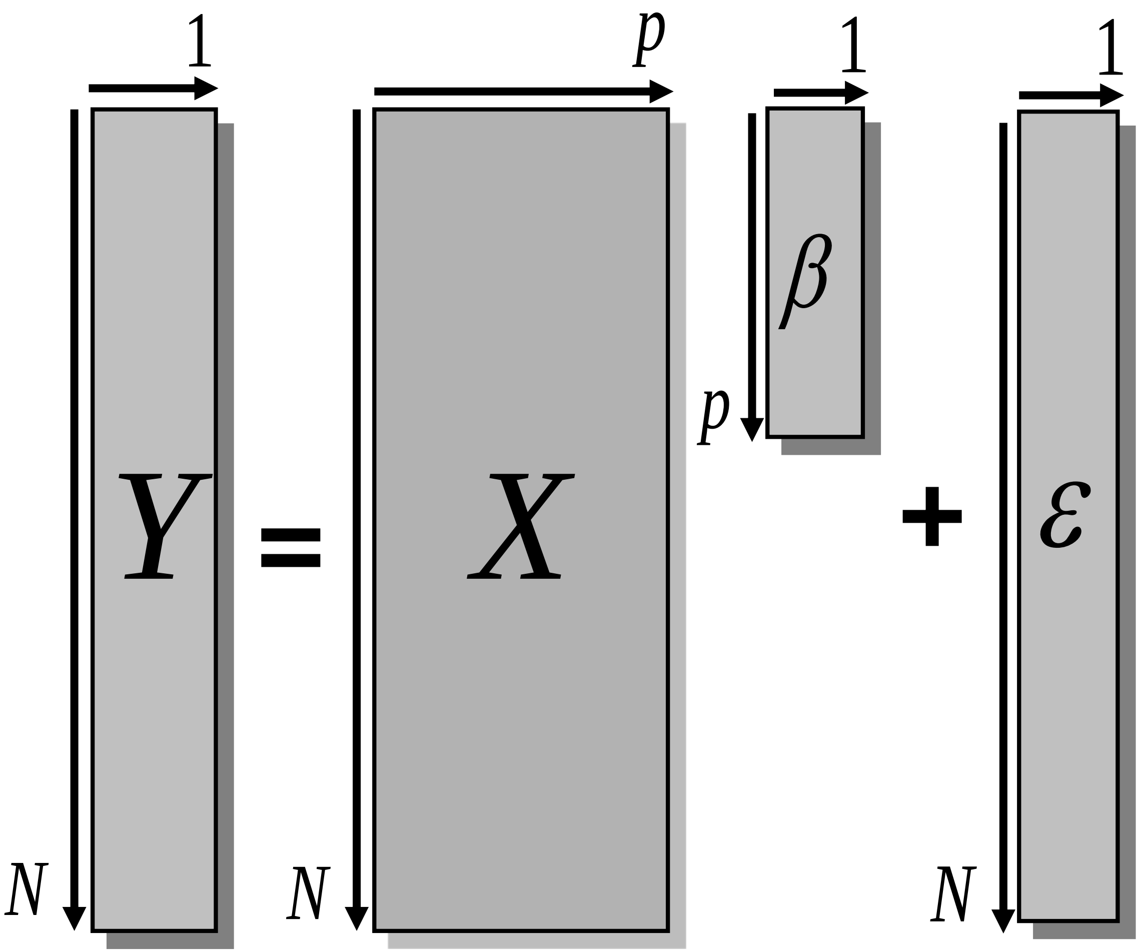
Refer to dealing with multiple sources of variance to make the inferences we want, i.e. generalising to a population

Why isn’t this easy?



- Multiple levels of model
- Multiple components of error in each
- To estimate a model’s parameters we need to know about the error

The GLM revisited



$$y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

Estimate with **Ordinary Least Squares** (OLS)

N: number of scans
p: number of regressors

Model is specified by

1. Design matrix X
2. **Assumptions about ε**

Ordinary Least Squares revisited

Find $\hat{\beta}$ that minimises $\boxed{\left\| y - X\beta \right\|^2 = \varepsilon^T \varepsilon}$

The Ordinary Least Squares parameter estimates are:

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

Estimation is direct – multiply data by the (pseudo) inverse of X

This is only valid (and is optimal) if errors are i.i.d. – if there is a single error covariance component, i.e., the variance s^2.

$$\boxed{\varepsilon \sim N(0, \sigma^2 I)}$$

This matters because covariance affects the statistics…

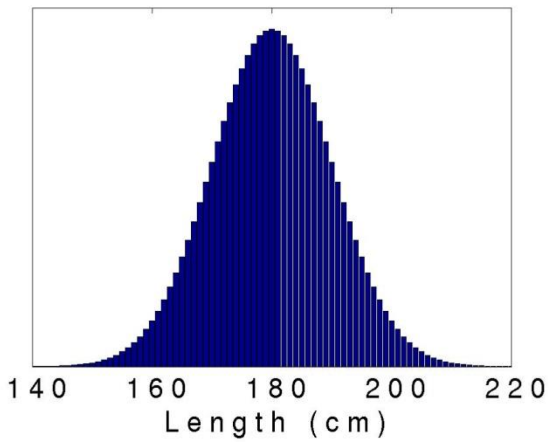
Error covariance and statistics

Classical inference is about what is surprising

- A statistic tests an effect's size relative to its expected behaviour under the null hypothesis
- The degrees of freedom must reflect how related (correlated) different observations are
- If observations covary, there are fewer independent observations than we think, so significance of statistics can be overrated

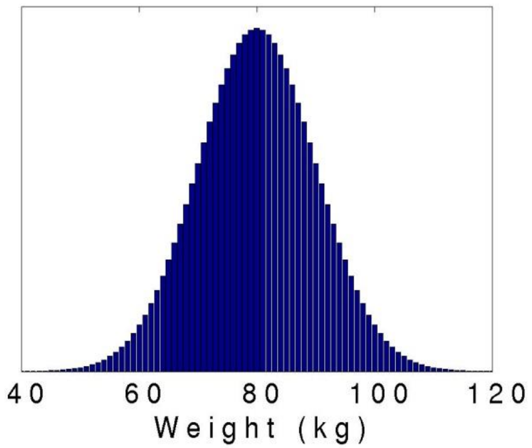
Variance

Length of men



μ=180cm, σ=14cm (σ^2=200)

Weight of men



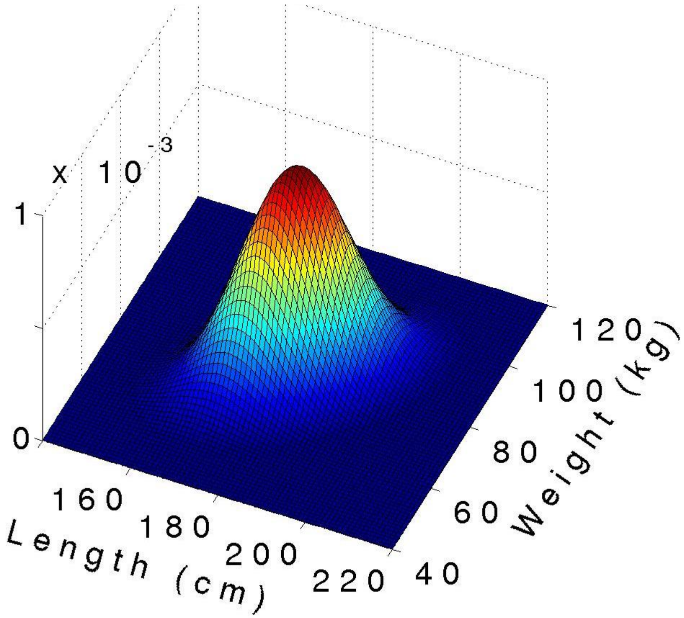
μ=80kg, σ=14kg (σ^2=200)

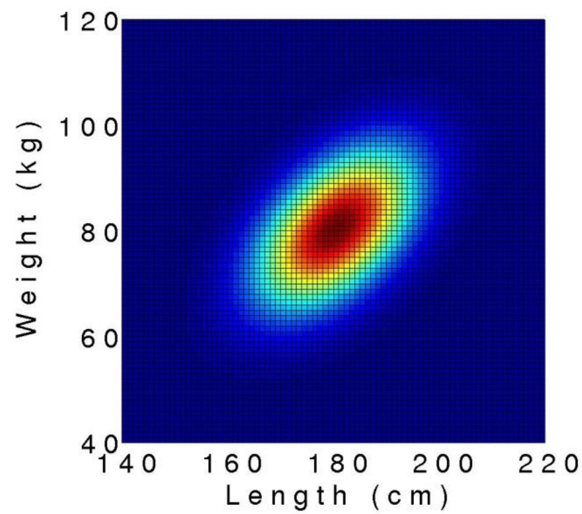
Each 1-dimensional variable is completely characterised by μ (mean) and σ^2 (variance)

i.e. can calculate $p(l|\mu,\sigma^2)$ for any l and $p(w|\mu,\sigma^2)$ for any w

Variance-covariance matrix

- Can also view length and weight as a 2-dimensional random variable (p(*l*,*w*)).





$p(l,w|\mu,\Sigma)$

$$\mu = \begin{pmatrix} 180 \\ 80 \end{pmatrix} \qquad \Sigma = \begin{bmatrix} 200 & 100 \\ 100 & 200 \end{bmatrix}$$

Length and weight are related – i.e., covary

Covariance and statistics again

$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}} \sim t_{N-p}$$

$$= \frac{\textbf{contrast of estimated parameters}}{\sqrt{\textbf{covariance estimate}}}$$

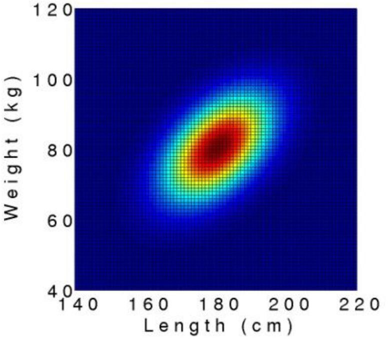
- How good (i.e., precise) is the contrast of betas $c^T\beta$ as an estimator of our effect?
- If it is precise (low covariance) this maximises *T*

- The df are also important…

Covariance and degrees of freedom

- Measure departure from sphericity (episilon)
- Evaluate significance of sum of squares ratios using F with fewer df (approx); Greenhouse-Geisser

Heights & weights



$$\Sigma = \begin{bmatrix} 200 & 100 \\ 100 & 200 \end{bmatrix} \quad \varepsilon = 0.8$$

= Satterthwaite correction (SPSS)

(in theory sl. liberal – but see Mumford & Nichols, 2009)

The rain in Bergen

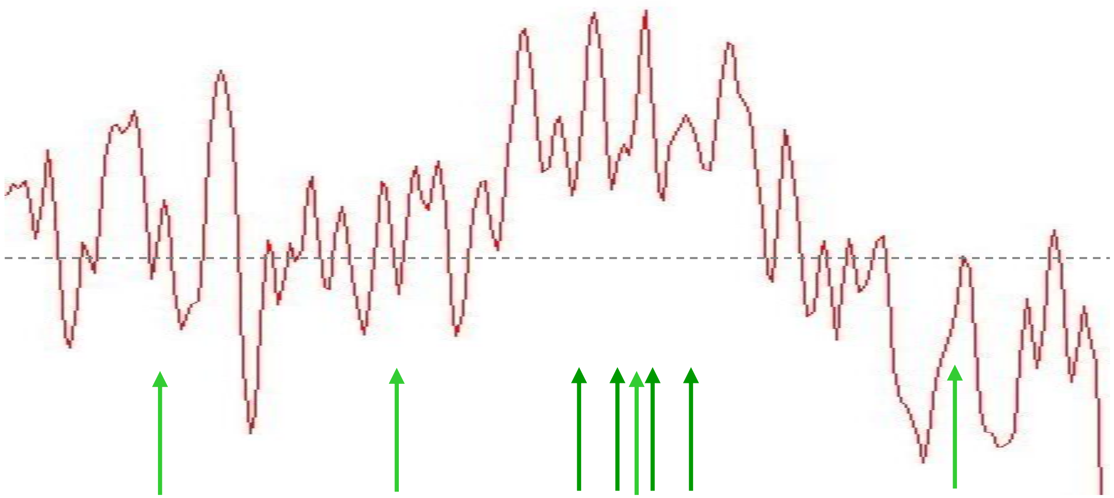
How much do the following observations tell us?

Rain on 4 consecutive days in June

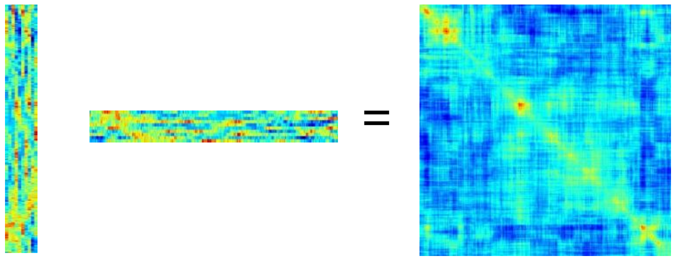
Rain on the same day in May, June, July and August

…which is more likely to indicate a wet summer?

Can we determine the patterns of temporal autocorrelation?

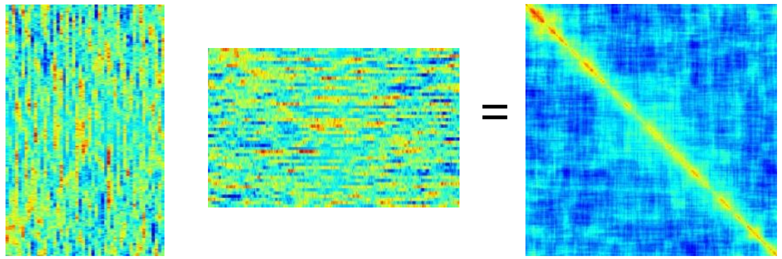


The rain in Bergen



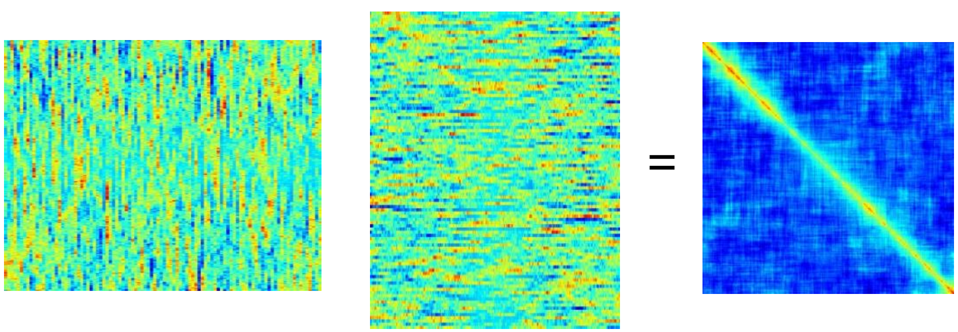
$\hat{\mathbf{E}}$ $\hat{\mathbf{E}}^T$ $\mathbf{S}$

Estimate based on 10 years



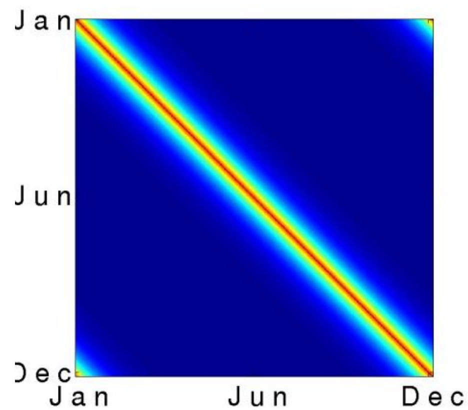
$\hat{\mathbf{E}}$ $\hat{\mathbf{E}}^T$ $\mathbf{S}$

Estimate based on 50 years



$\hat{\mathbf{E}}$ $\hat{\mathbf{E}}^T$ $\mathbf{S}$

Estimate based on 100 years

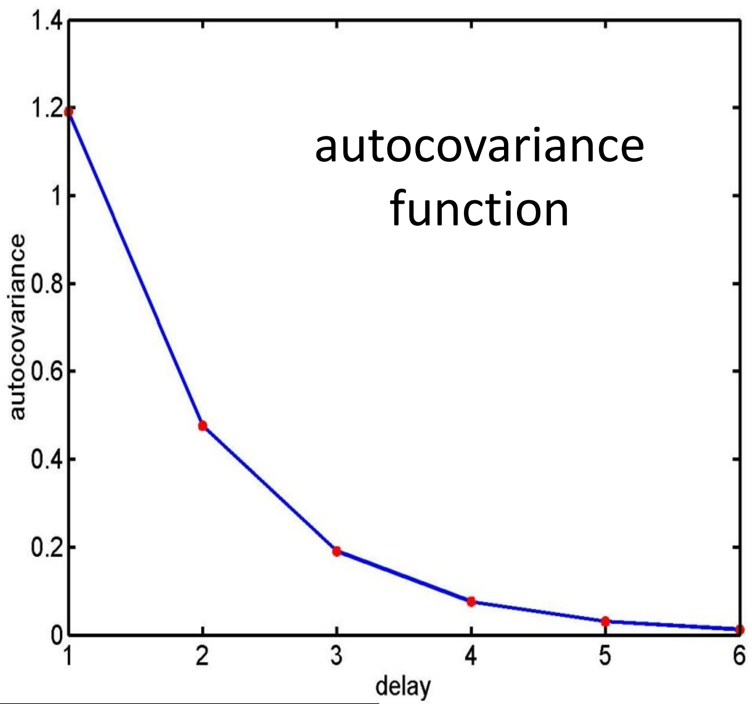


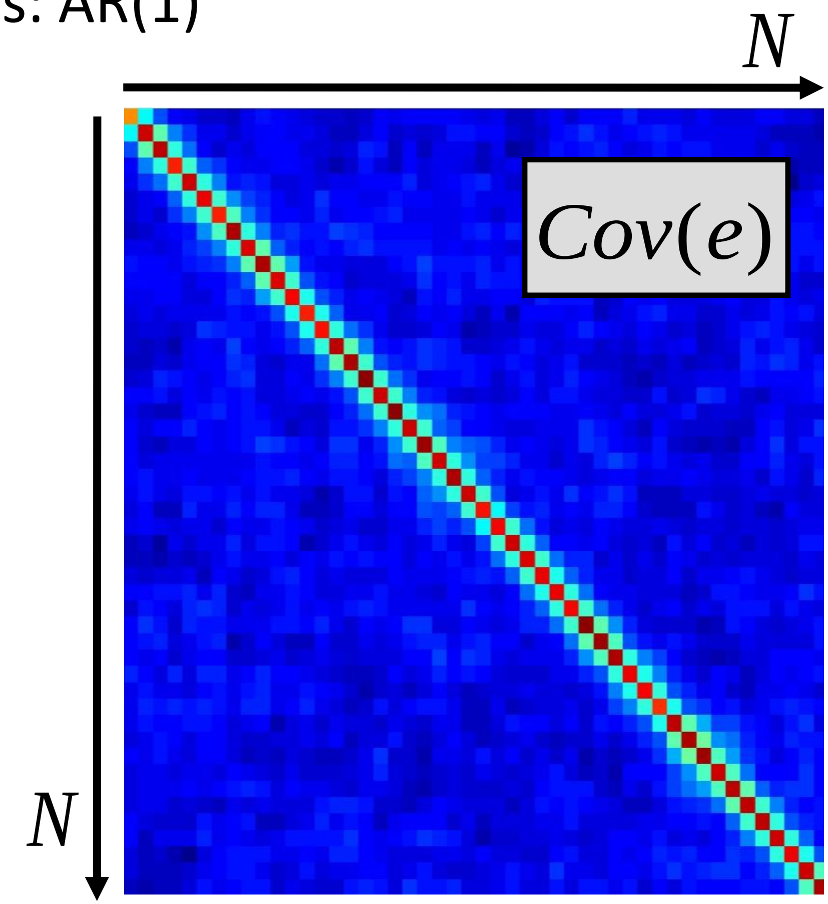
True $\mathbf{\Sigma}$ – as if there were not 100*365=36500 data points, but 2516!

Serial correlations in fMRI

$$e_t = ae_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0,\sigma^2)$$

1st order autoregressive process: AR(1)





BUT also high-pass filtering

Serial correlations in fMRI

Pre-whitening

- Use an enhanced noise model with multiple error covariance components
- Estimate components AR (1) + white noise
- Specify a filter matrix *W* to whiten the data – *'undoing'* the serial correlations

$$\boxed{Wy = WX\beta + We}$$

$$\boxed{We \sim N(0, \sigma^2 W^2 V)}$$

Serial correlations in fMRI

SPM12 prewhitening model: AR(1) + white noise*

- AR(1) cannot be estimated precisely at each voxel
- But precision is critical, or estimates are worse than OLS – biased AND imprecise
- Use spatial regularisation: pool estimation over ‘active voxels‘ (1st pass OLS estimate at $p < .001$)
- + White noise – voxel-specific variance s^2

*Bayesian estimation option: AR(3) with spatial priors

$$e_t = ae_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0, \sigma^2)$$

Serial correlations in fMRI

Once data are 'pre-whitened', estimation can proceed using Ordinary Least Squares

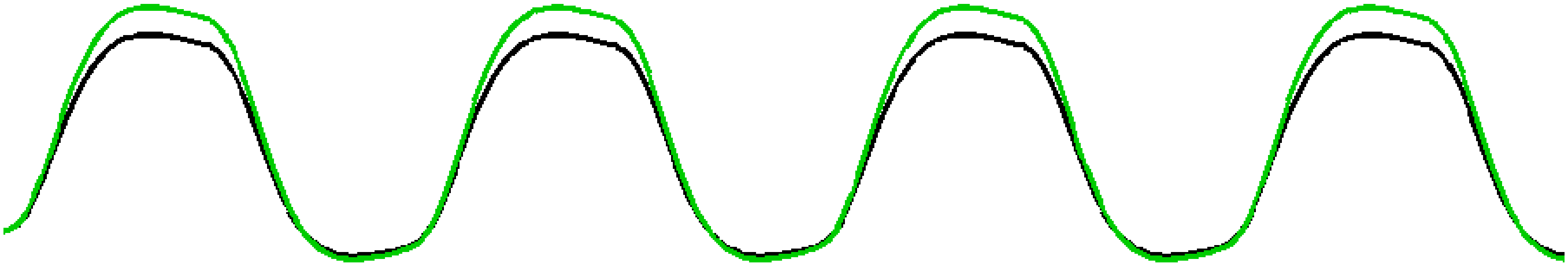
- The parameter estimates are again optimal – unbiased and minimum variance
- The df are also correct, if we want to do our statistical inference at the first level

Take-home message (1)

- If *'error structure'* is complex with multiple components of covariance – not just i.i.d. – our inference depends on modelling the error structure

- What does this have to do with 2-level models?

Why isn't this easy?



- Multiple levels of model
- Multiple components of error in each

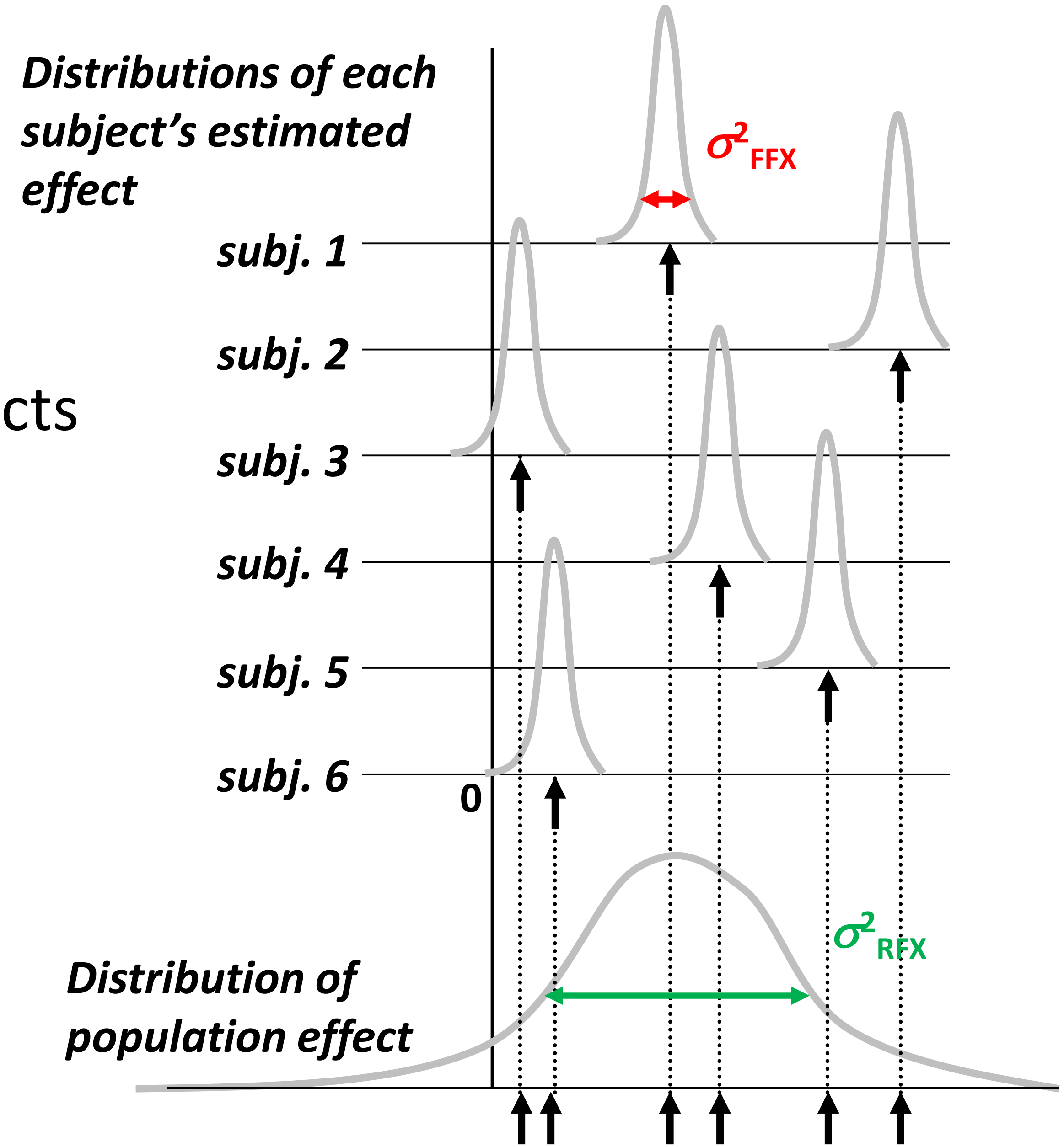
Fixed vs. random effects

Fixed effects:

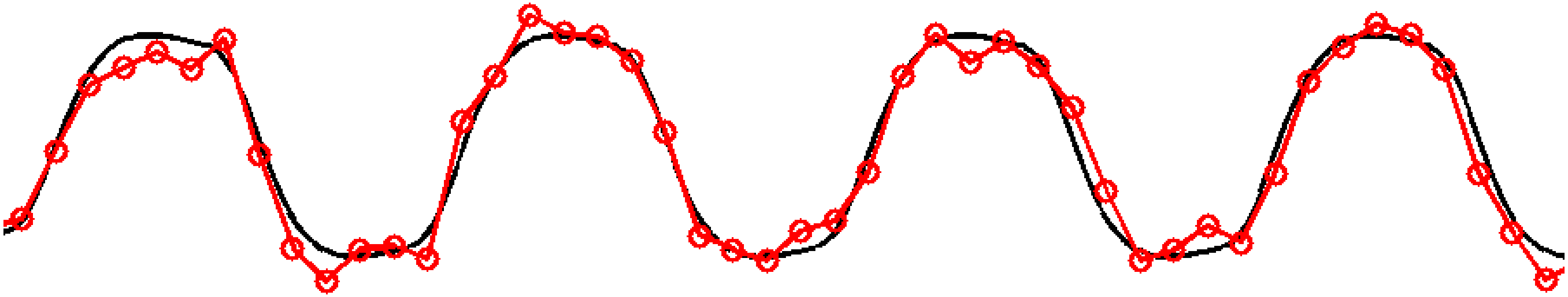
Intra-subjects variation suggests all these subjects different from zero

Random effects:

Inter-subjects variation suggests population not different from zero



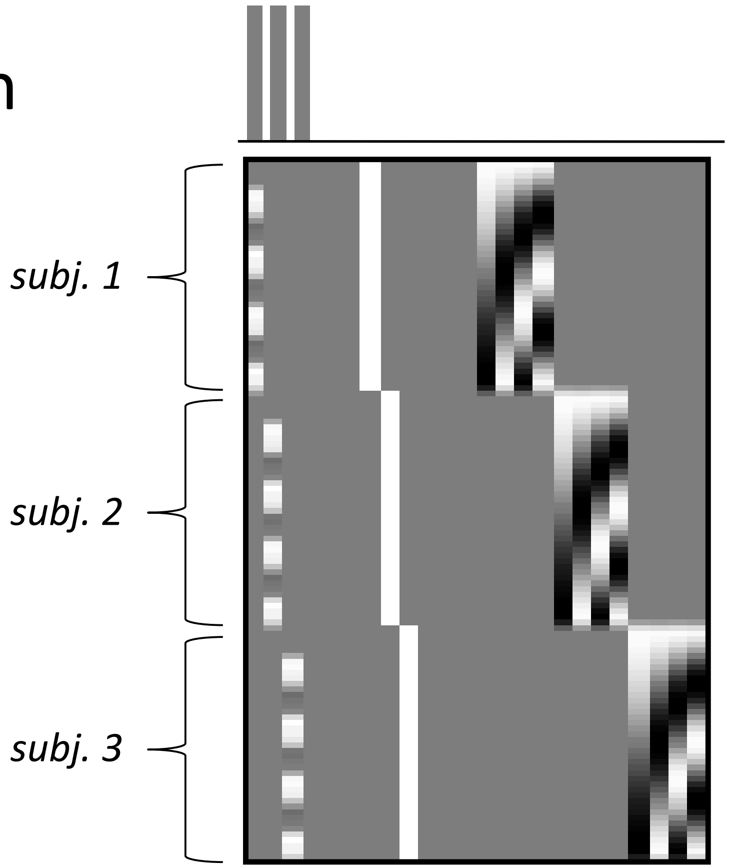
Fixed effect



- The only source of variation (over sessions) is measurement error
- The true response magnitude is *fixed*

Fixed effect modelling in SPM

- Grand GLM (single level) approach (model all subjects at once, pool estimates over subjects)
- Good:
  - *maximise df*
  - *simple model*
- Bad:
  - *assumes common variance*
  - *over subjects at each voxel*

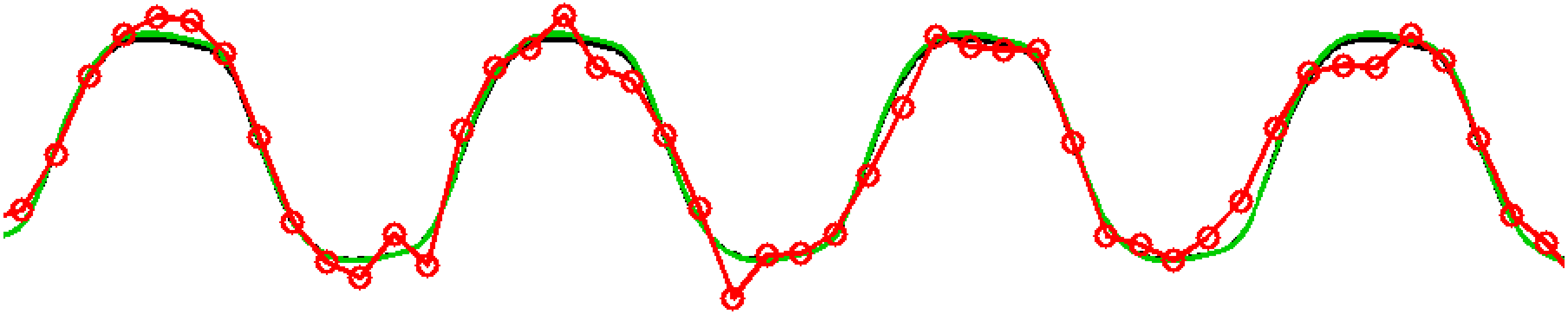


Fixed vs. random effects

Summary

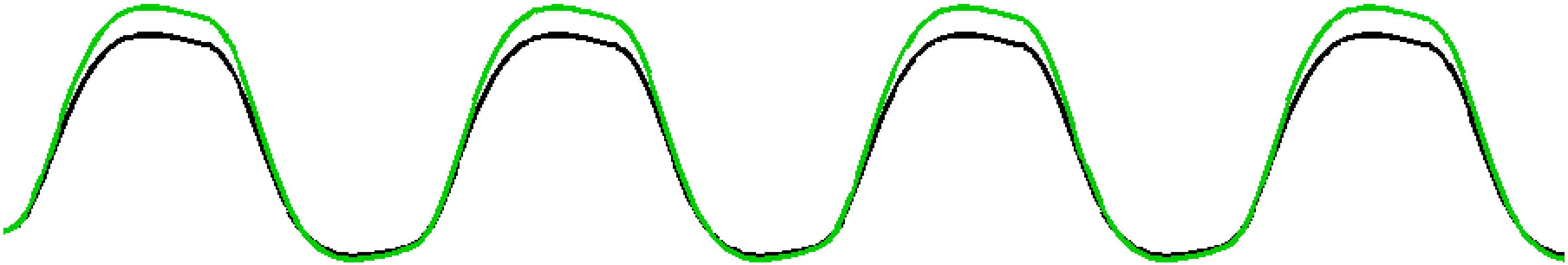
- Fixed effect inference: *“I can see this effect in this cohort”*
- Random effect inference: *“If I were to sample a new cohort from the same population I would get the same result”*
- Fixed isn’t ‘wrong’, but is not usually of interest

Random effects



- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)

Random effects

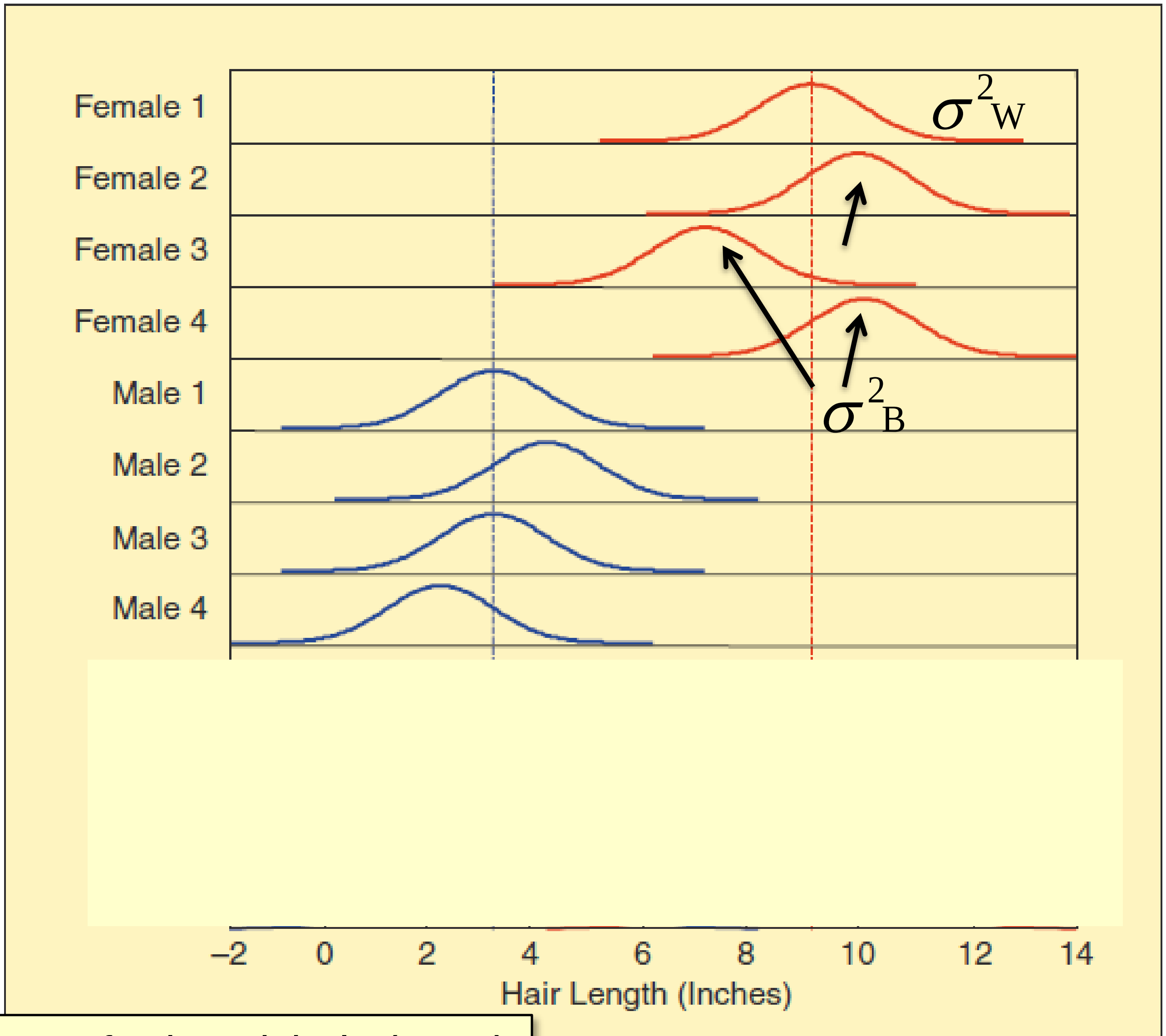


- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject/session has random magnitude
  - but note, population mean magnitude is *fixed*

Why bother with two stages?

- We want to make an inference to the population, not a single subject, so why do we care?
- Why can’t we just do group stats on data for each voxel, like a standard *t*-test?
- We could, if data Y were simple values per voxel – precisely known.
- Instead, we have estimates of individual subjects‘ effects – so more than 1 covariance component

Hierarchical models



Mumford & Nichols (2006)

Does hair length differ by gender?

2 sources of variability

Within-subject: $\boxed{\sigma^2_W}$

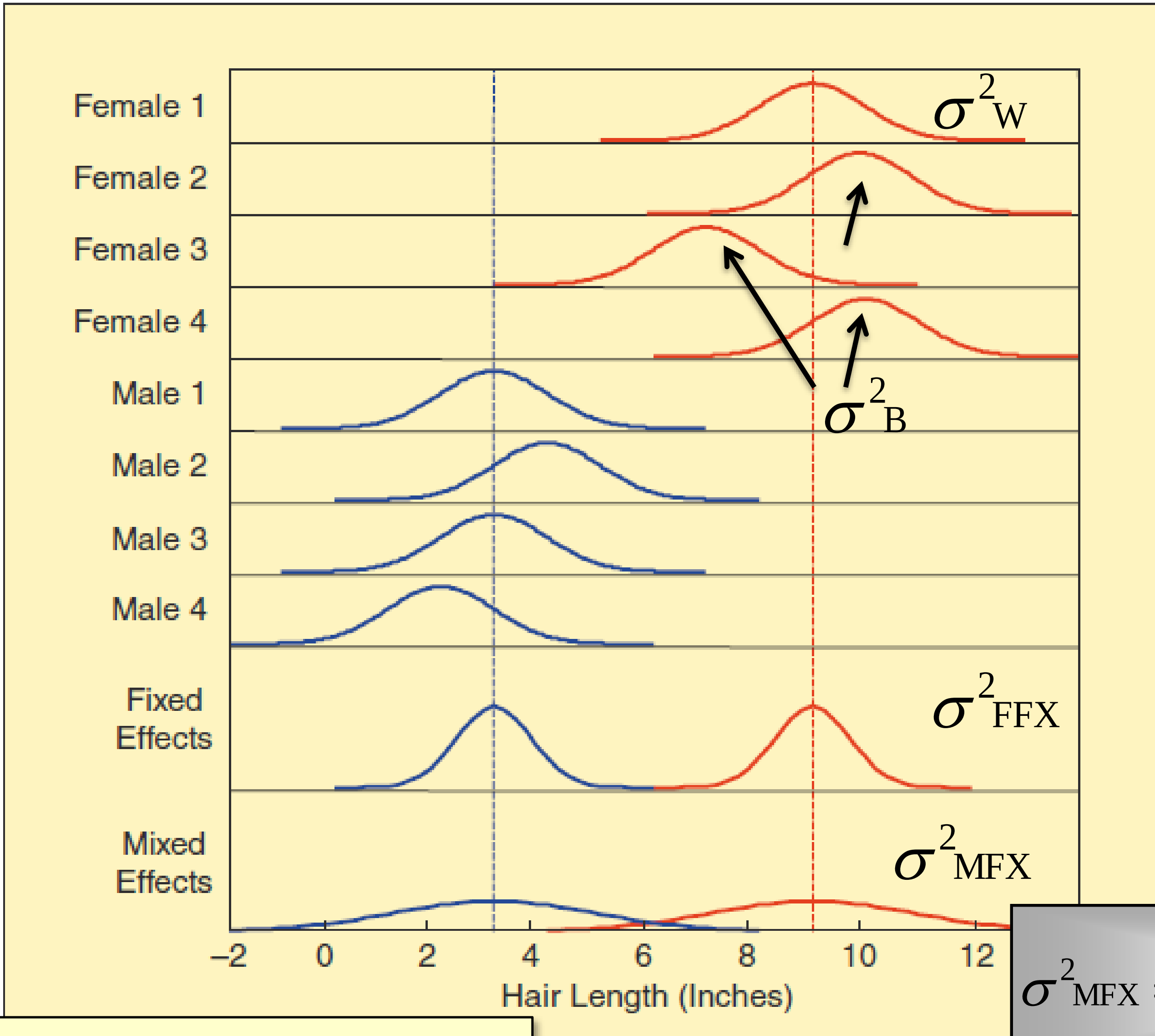
Between-subjects: $\boxed{\sigma^2_B}$

To generalise across this sample, combine data from hairs measured in all subjects, get σ^2_{FFX}

To generalise to population, use estimates of hair length for each subject, get σ^2_{MFX}

MIX of between/ within subject variability

Hierarchical models



Mumford & Nichols (2006)

Does hair length differ by gender?

2 sources of variability

Within-subject (1)

Between-subjects (49)

To generalise across this sample if p = 25 hairs per subject

$$\sigma^2_{FFX} = \frac{1}{4} * \frac{\sigma^2_W}{25} = 0.01$$

To generalise to population, given N = 4 subjects per group

$$\sigma^2_{MFX} = \frac{1}{4} * \frac{\sigma^2_W}{25} + \boxed{\frac{1}{4}} \sigma^2_B = 12.26$$

Why bother with two stages?

Why can't we just do group stats on the data from each voxel?

- ...that could be valid but would not be optimal
- Hierarchical models deal with the mixed sources of variance, not just between-subject variance
- Better to model both scan-to-scan and subject-to-subject variability
- There is therefore more than 1 variance component (nonsphericity) at the group level

Hierarchical models

Two* approaches in SPM

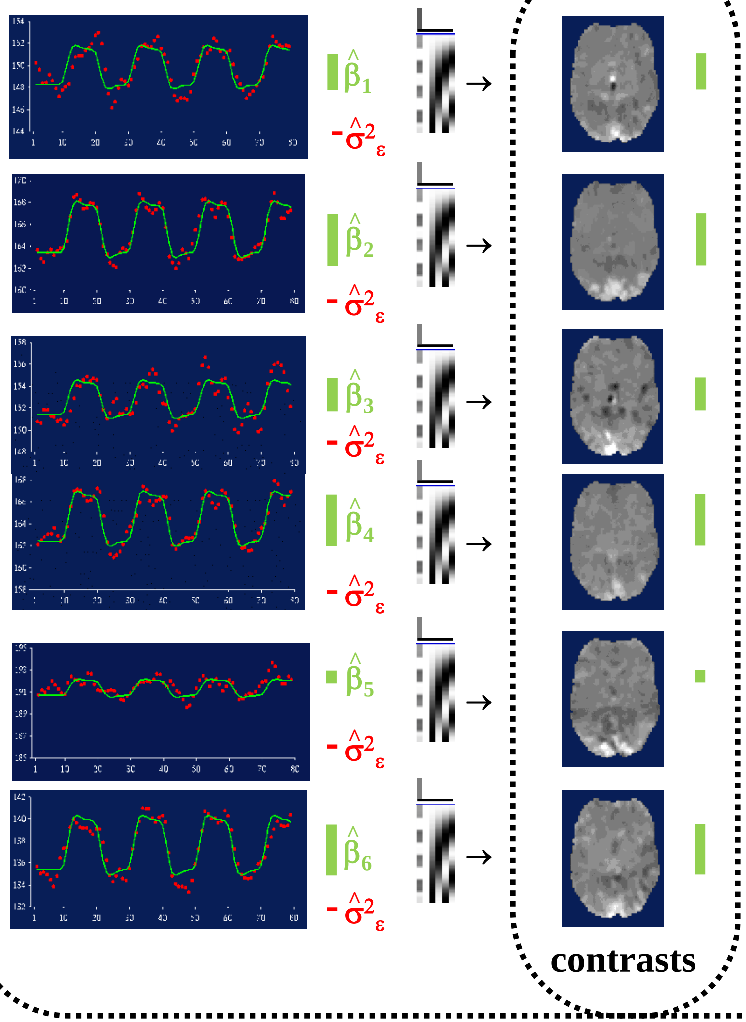
1. Simple summary statistic – Holmes & Friston
2. Non-sphericity modelling at group level

- Pros and cons – assumptions vs. flexibility
  - Subject variances equivalent
  - Subject design matrices equivalent
  - (2) enables a wide range of 2^{nd} level models

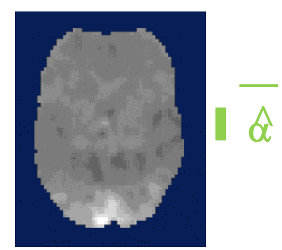
*Actually 3, but we talk about 2 here; see Friston et al., 2005 re. ‘full MFX’

Simple summary statistic approach ('HF')

1st level (within subjects) — 2nd level (between-subjects)



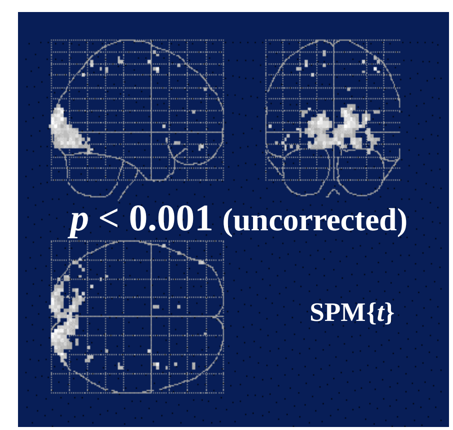
estimated mean activation image…



…to be compared with MFX variance:

$$\sigma^2 = \sigma^2_\alpha + \sigma^2_\varepsilon / w$$

no voxels significant at $p < 0.05$ (corrected)



Models within-subject variance implicitly

Simple summary statistic approach (‘HF’)

Assumptions

- Distribution normal over independent subjects
- Homogeneous variance
  - Subjects’ residual errors same
  - Subjects’ design matrices same

  - 2 covariance components can be collapsed into 1 if these elements of the group level covariance are homogenous over subjects

Simple summary statistic approach ('HF')

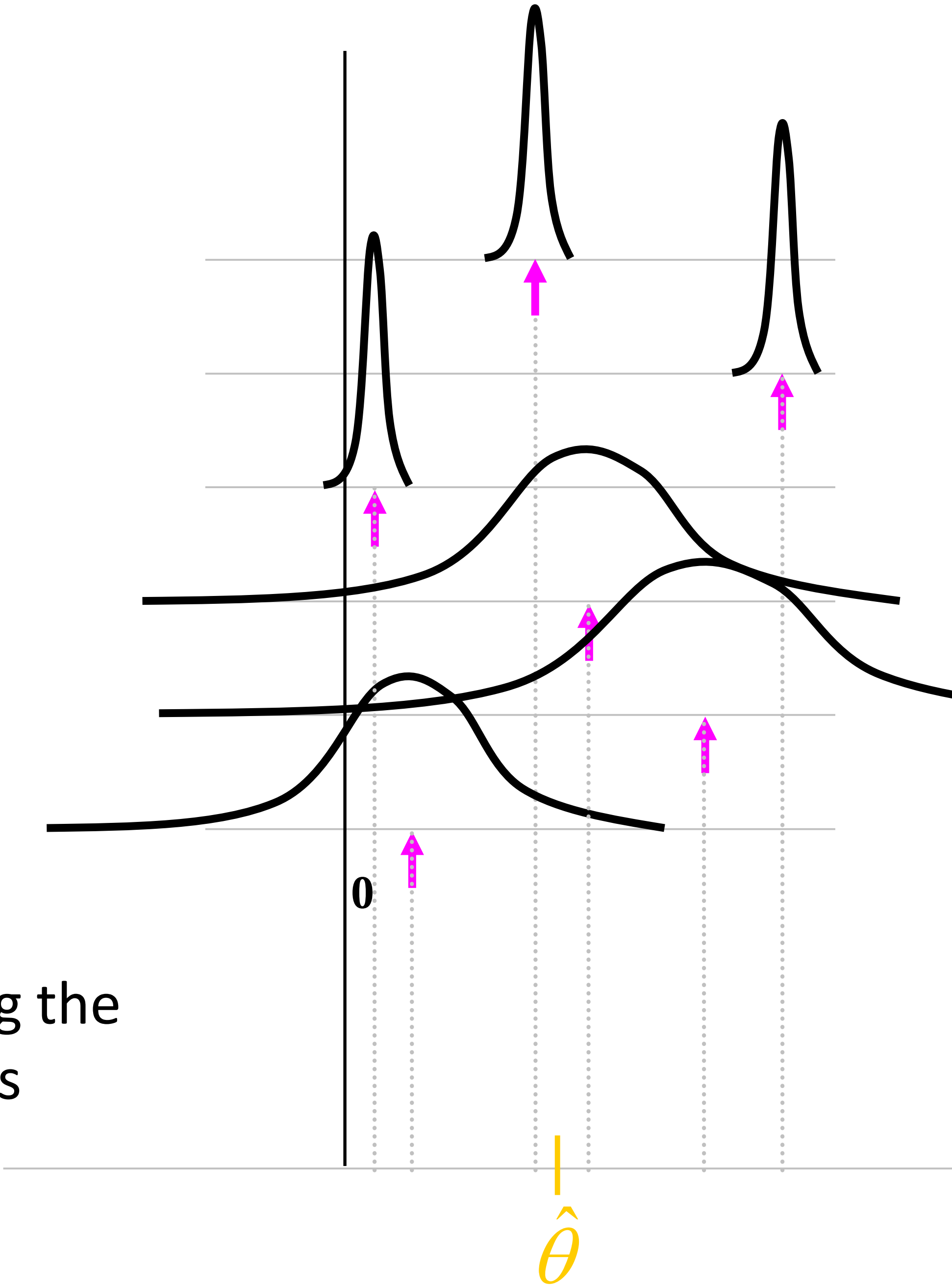
Use only a single image per subject

- Limited to 1- or 2-sample t-tests at the 2nd level
- Balanced designs

- Limitation = strength
  - No 2nd level sphericity assumption
  - 'Partitioned' error term @ 2nd level

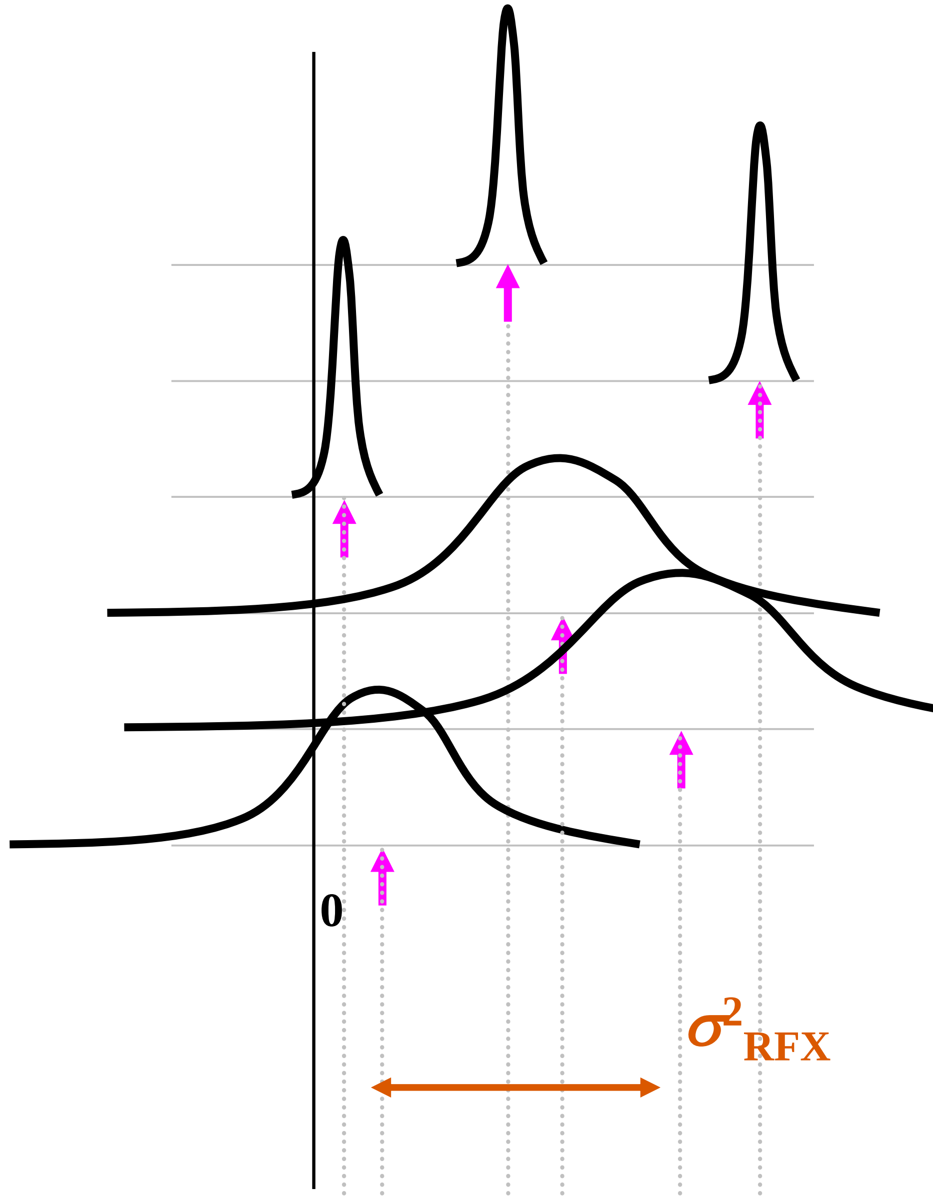
HF - efficiency

- If assumptions true
  - Optimal, fully efficient
- If σ^2_{FFX} differs between subjects
  - Reduced efficiency
  - Here, optimal group parameter estimate $\hat{\theta}$ requires down-weighting the 3 highly variable subjects



HF - validity

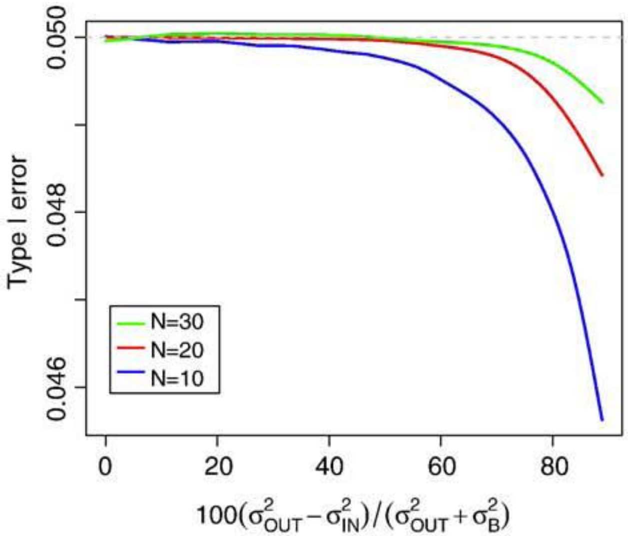
- If assumptions true
  - Exact *P*-values
- If σ^2_{FFX} differs btw subj.
  - Standard errors not OK
    - Estimate of σ^2_{RFX} may be biased
  - df not OK
    - Here, 3 subjects dominate
    - df < 5 = 6-1



HF – robustness

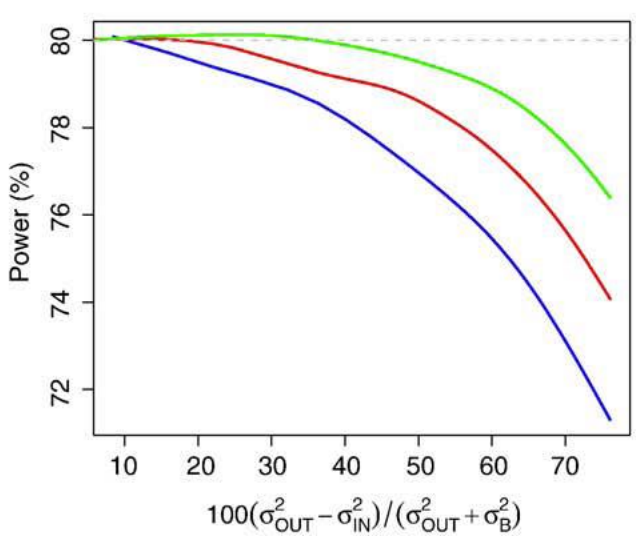
- In practice, validity & efficiency are excellent
  - For the one sample case, HF is very robust

False Positive Rate



(outlier severity)

Power Relative to Optimal



(outlier severity)

  - Potential concern with 2-sample or correlation if outliers/ large imbalance

Modelling 2nd level non-sphericity

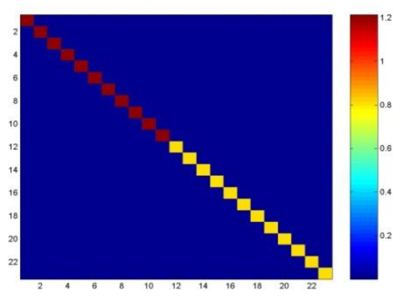
A more flexible summary statistic approach

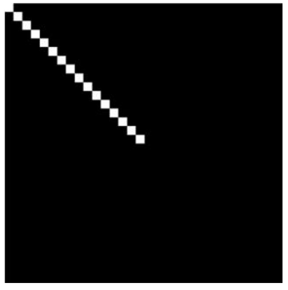
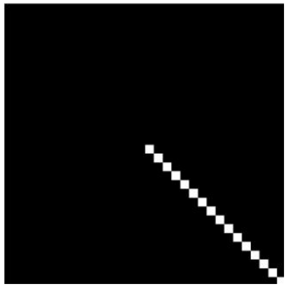
- 1st level model is just the same
- At 2nd level, linear combination of basis functions to represent different sources of covariance
- i.e., multiple covariance components
- Same estimation using prewhitening approach, and spatial regularisation (cross-voxel pooling)

Modelling 2nd level non-sphericity

Errors can be Independent but Non-Identical when…

1) Model includes one contrast but from different groups – 2-sample t-test e.g. patients and control groups



$Q_1 =$  $Q_2 =$ 

Modelling 2nd level non-sphericity

Error can be Non-Independent and Non-Identical when…

2) Several contrasts per subject are taken to 2nd level
   i.e., Repeated Measures/ Mixed ANOVA

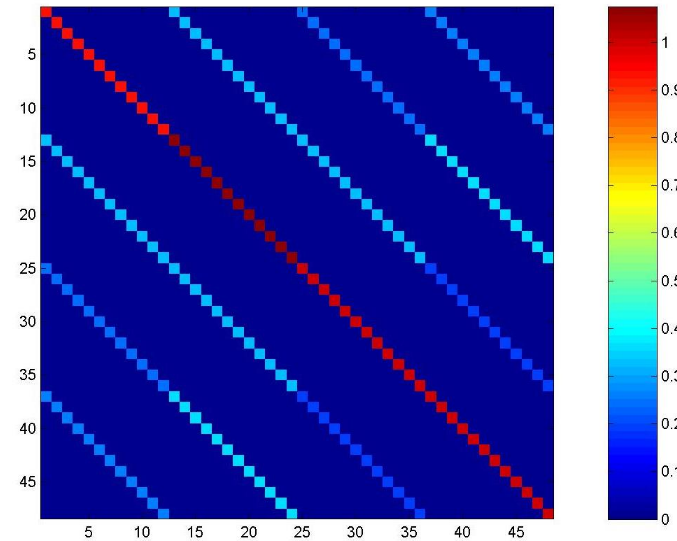
3) Omnibus test is needed across several basis
   functions characterising the hemodynamic response

e.g. F-test combining HRF, temporal derivative and dispersion regressors

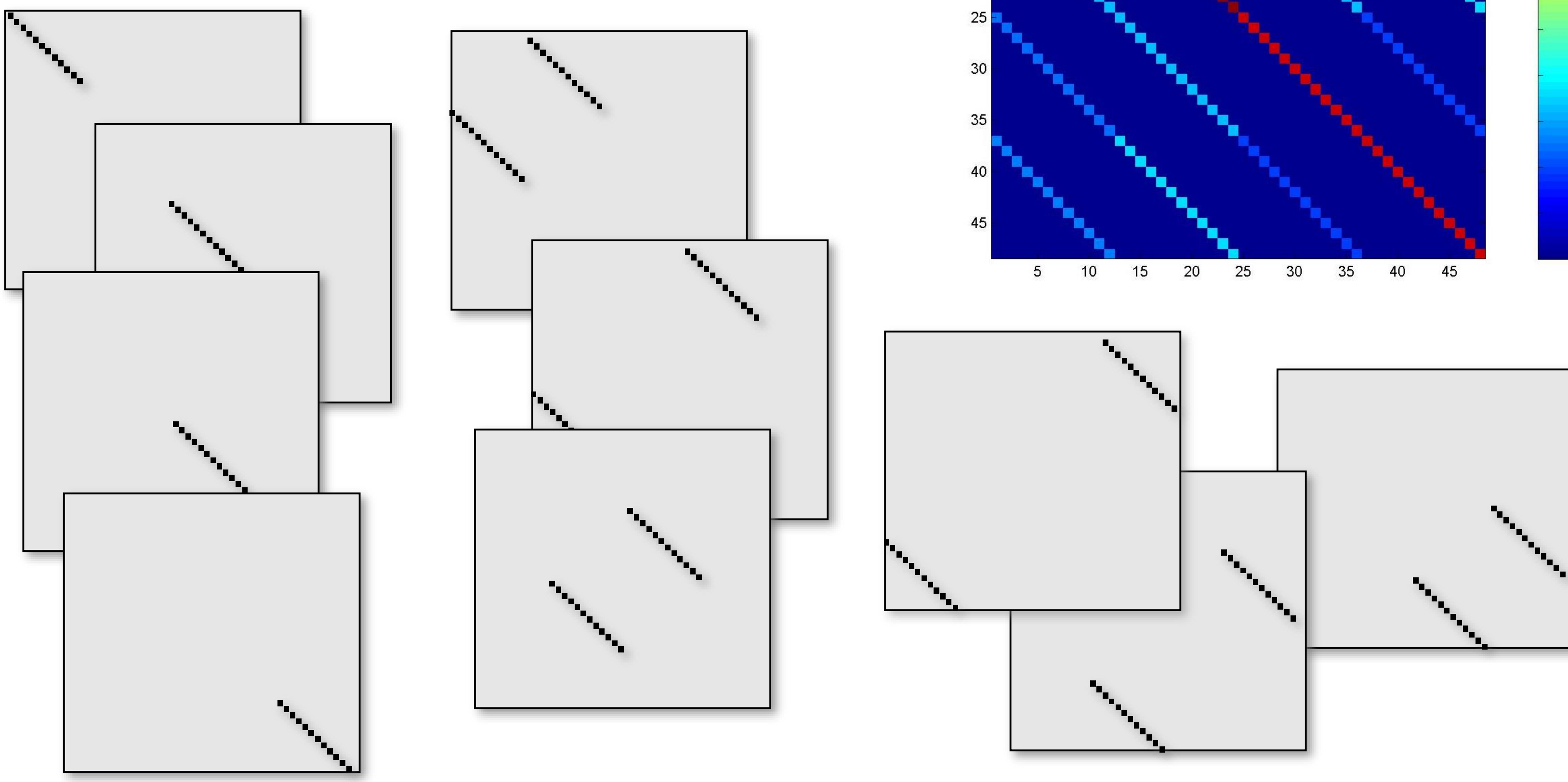
Modelling 2nd level non-sphericity

Errors are not independent
and not identical (4 conditions)

residuals covariance matrix



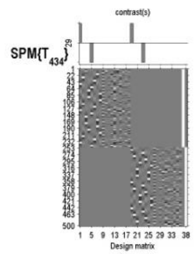
Q_k's:



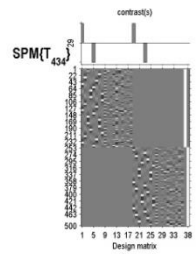
Modelling 2nd level non-sphericity

1: motion **2: sounds** **3: motion** **4: sounds**

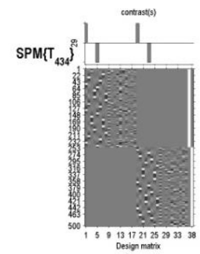
1st level



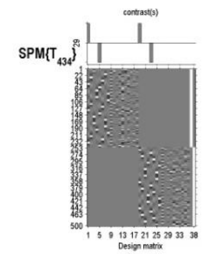
? =



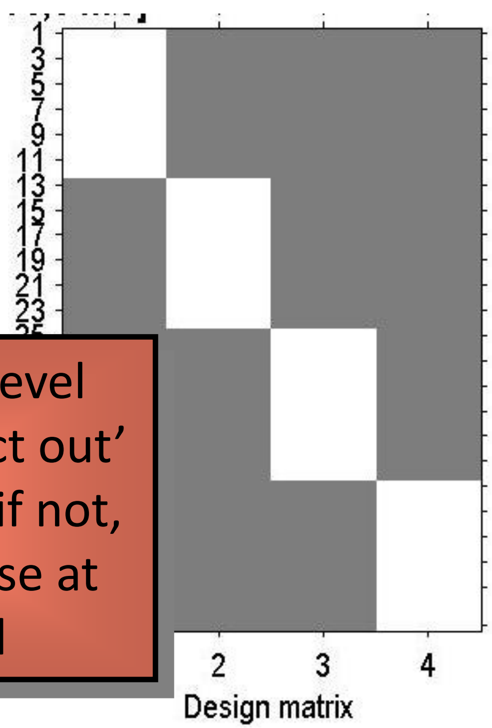
? =



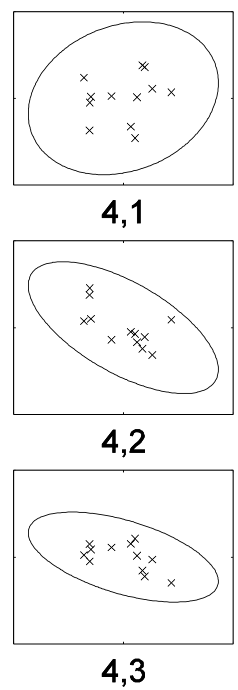
? =



2nd level



N.B. These 1st level contrasts 'subtract out' subject effects – if not, must model these at the 2nd level



Block design study

Repeated measures ANOVA model

Which regions are sensitive to semantic content of words across 4 conditions?

Noppeney et al.

Modelling 2nd level non-sphericity

Assumptions

- As at 1st level, needed for cross-voxel pooling: homogenous across 'active' voxels
- Within subject covariance still homogenous
- HF plus pooled variance at 2nd level

Advantages

- Fast relative to 'full' mixed-effects procedures
- May be more sensitive
- Flexibility of possible 2nd level models

Summary

fMRI models need to take account of

- Hierarchical nature of data
- Multiple sources of variability at each level

Estimation & correction for resulting nonsphericity

- Some assumptions
- If correct, optimise estimation & inference
- SPM enables very flexible 2nd level models

2-stage GLM

Single subject

Each has an independently acquired set of data
These are modelled separately
Models account for **within subjects variability**
Parameter estimates apply to individual subjects

1st level

Single subject **contrasts of parameter estimates** taken forward to 2nd level as (spm_con*.img) **'con images'**

Group/s of subjects

To make population inferences, 2nd level models account for **between subjects variability**
Parameter estimates apply to group effect/s

2nd level

Statistics compare **contrasts of 2nd level parameter estimates to 2nd level error**

References

- Simple group fMRI modeling and inference. Mumford & Nichols, *Neuroimage*, 2009
- *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2007.
- *Generalisability, Random Effects & Population Inference*. Holmes & Friston, NeuroImage, 1999.
- *Classical and Bayesian inference in neuroimaging: theory*. Friston et al., NeuroImage, 2002.
- *Classical and Bayesian inference in neuroimaging: variance component estimation in fMRI*. Friston et al., NeuroImage, 2002.
- Mixed-effects and fMRI studies. Friston, Stephan, Lund, Morcom, Kiebel *Neuroimage, 2005*
- *Flexible factorial tutorial* by Glascher and Gitelman, 2008, copy at: www.sbirc.ed.ac.uk/cyril/cp_fmri.html