Translational
Neuromodeling
Unit

Signal, Noise and Preprocessing*

Methods and Models for fMRI Analysis

October 8th, 2019

Sam Harrison

Translational Neuromodeling Unit

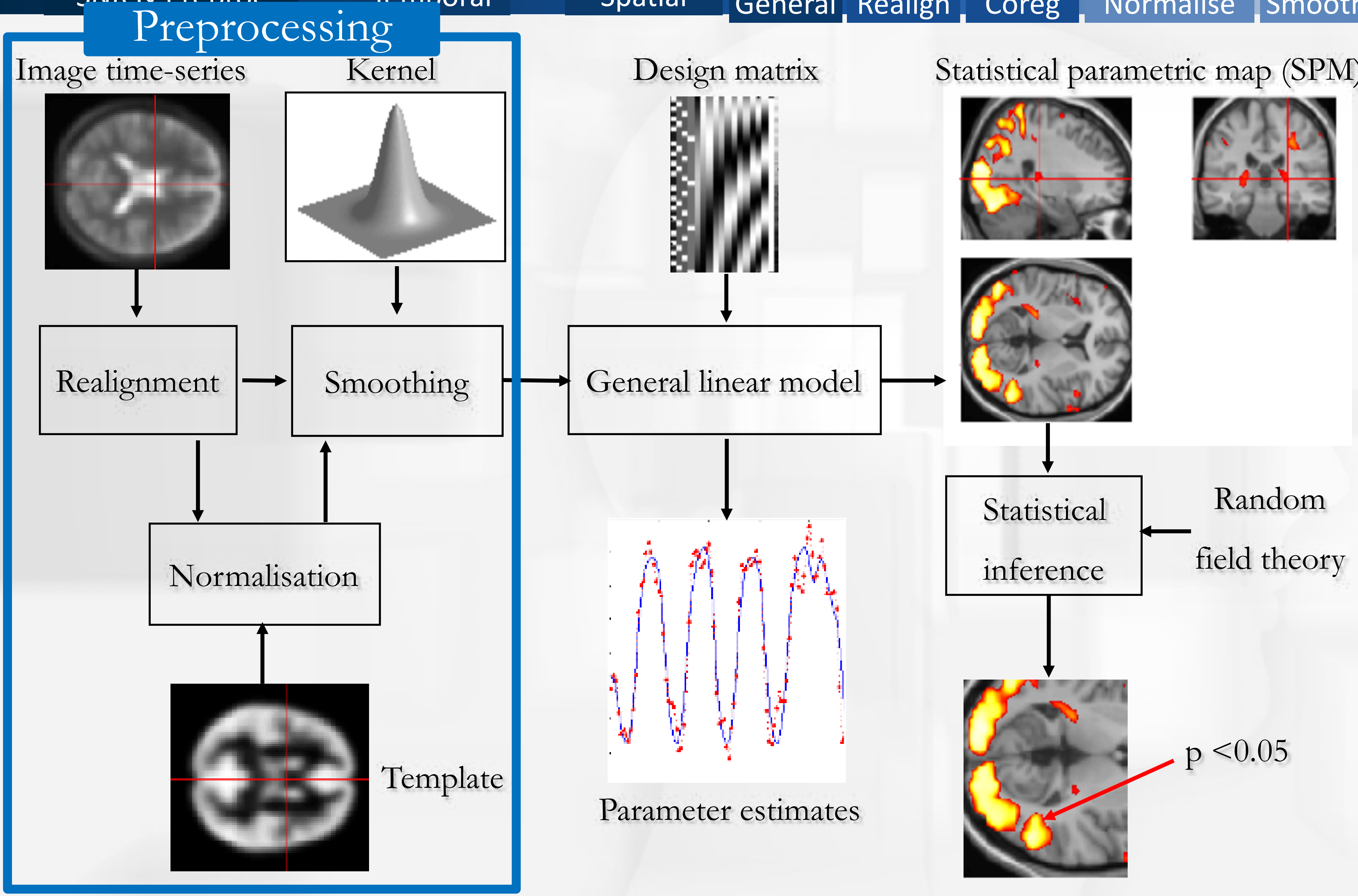
Generous slide support:

<u>Lars Kasper</u>

Guillaume Flandin
Ged Ridgway
Klaas Enno Stephan
John Ashburner

*Huettel et al.

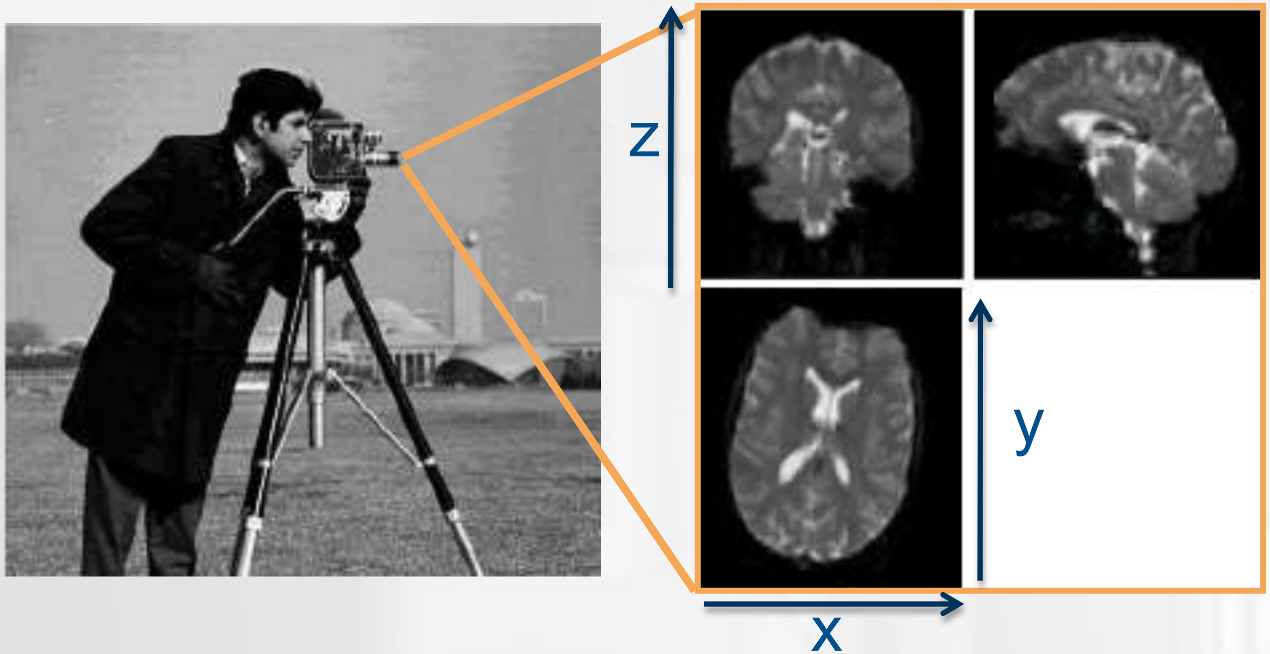
Overview of SPM for fMRI



Preprocessing Aims

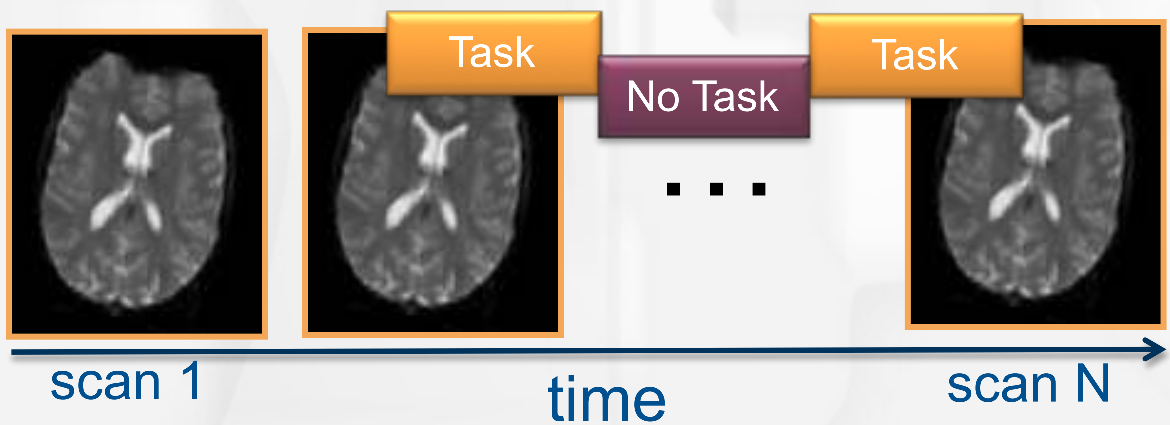
- Broadly speaking, preprocessing does one of three things:
  - Transforms our data so that it is more useful, but without fundamentally changing its properties (e.g. registration).
  - Increases the sensitivity of our analyses, either by boosting signal or removing noise (e.g. motion correction).
  - Adjusts the data such that it fits our modelling assumptions (e.g. smoothing).
- The aim of this lecture is therefore that:
  - You understand *why* the different preprocessing steps are important.
  - You're realistic about what it can, and more importantly can't, do.
  - See e.g. Eklund et al., PNAS, 2016 or Deen & Pelphrey, Nature, 2012 for why it's important to get this right!

fMRI = Acquiring Movies



- …of 3D Blood-Oxygen-Level Dependent (BOLD) contrast images
- typically Echo-Planar Images (EPI)

- Run/session: time-series of images



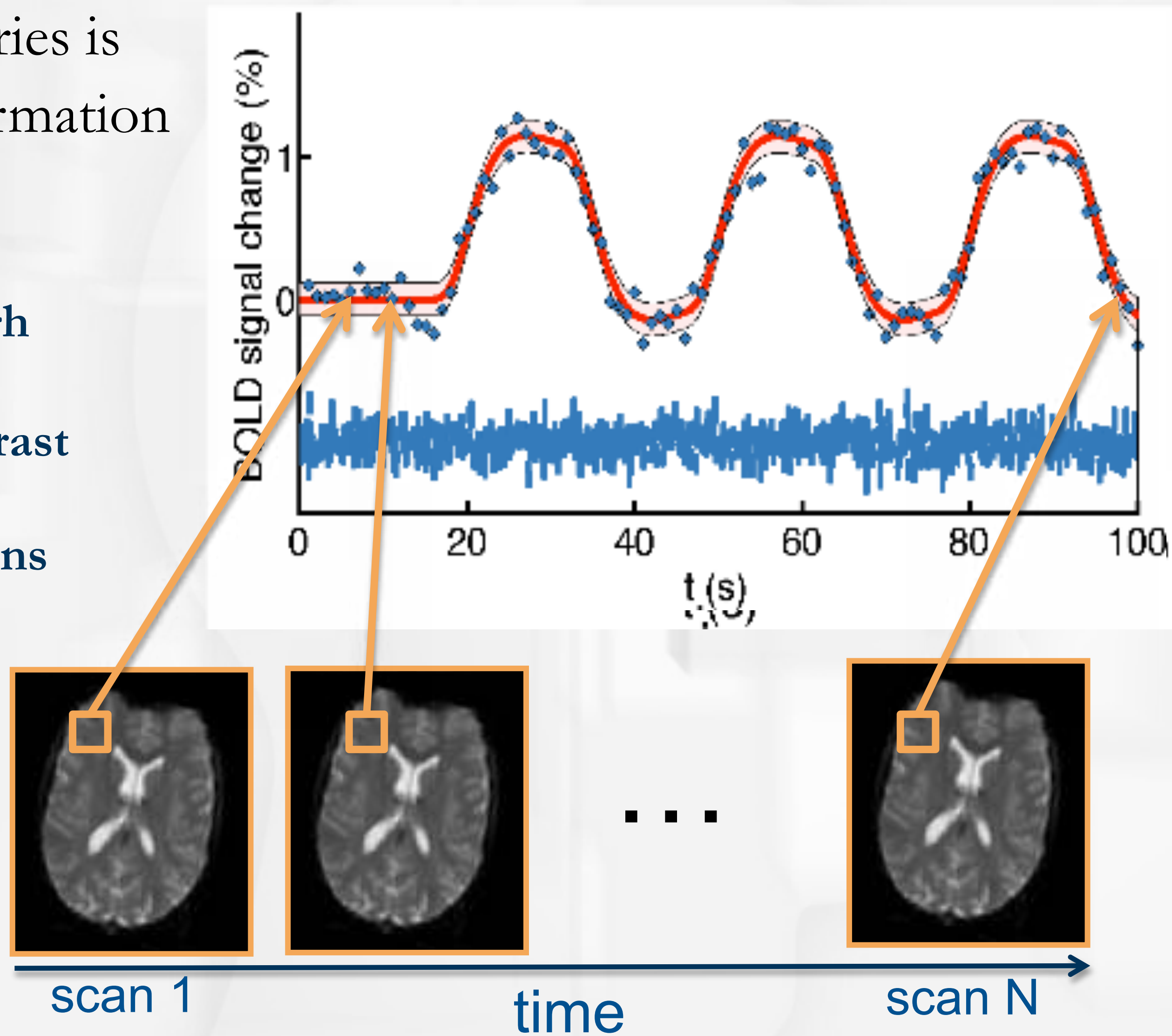
fMRI = Acquiring Movies

- The localized time-series is the fundamental information unit of fMRI

Signal: fluctuation through Blood-Oxygen-Level Dependent (BOLD) contrast

Noise: all other fluctuations

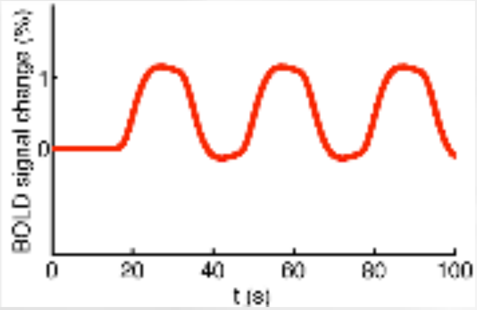
- Run/session: time-series of images

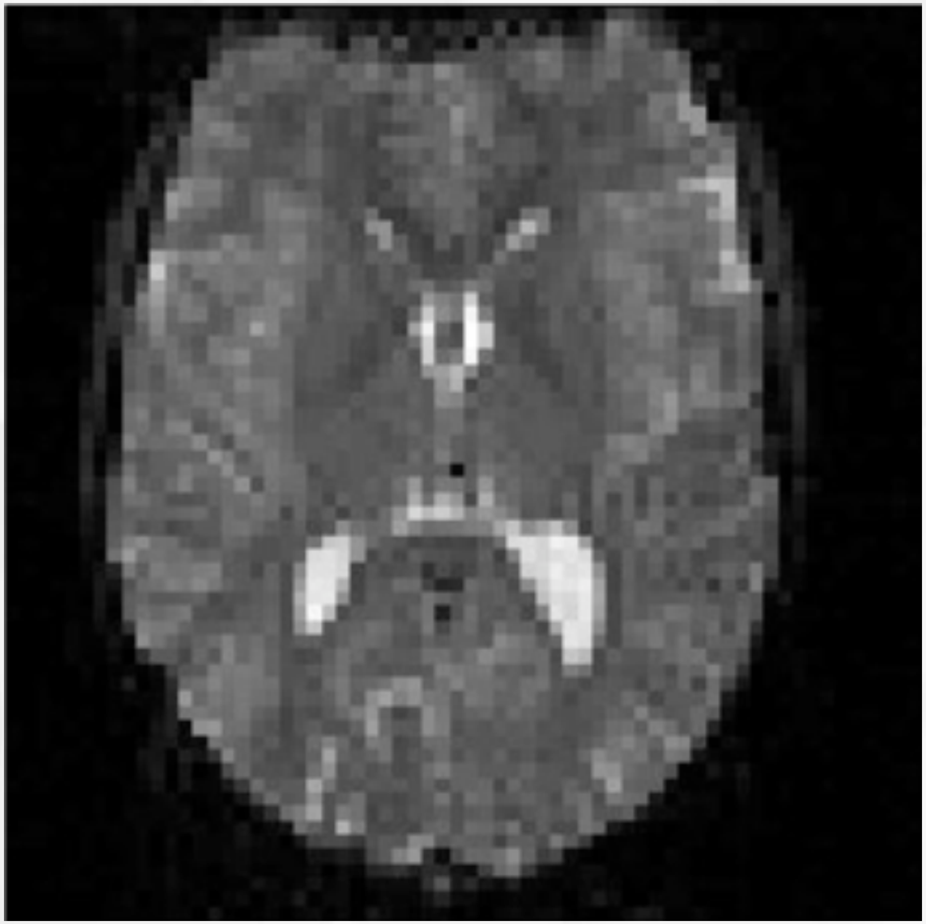


The Golden Rule

Look at your data!!

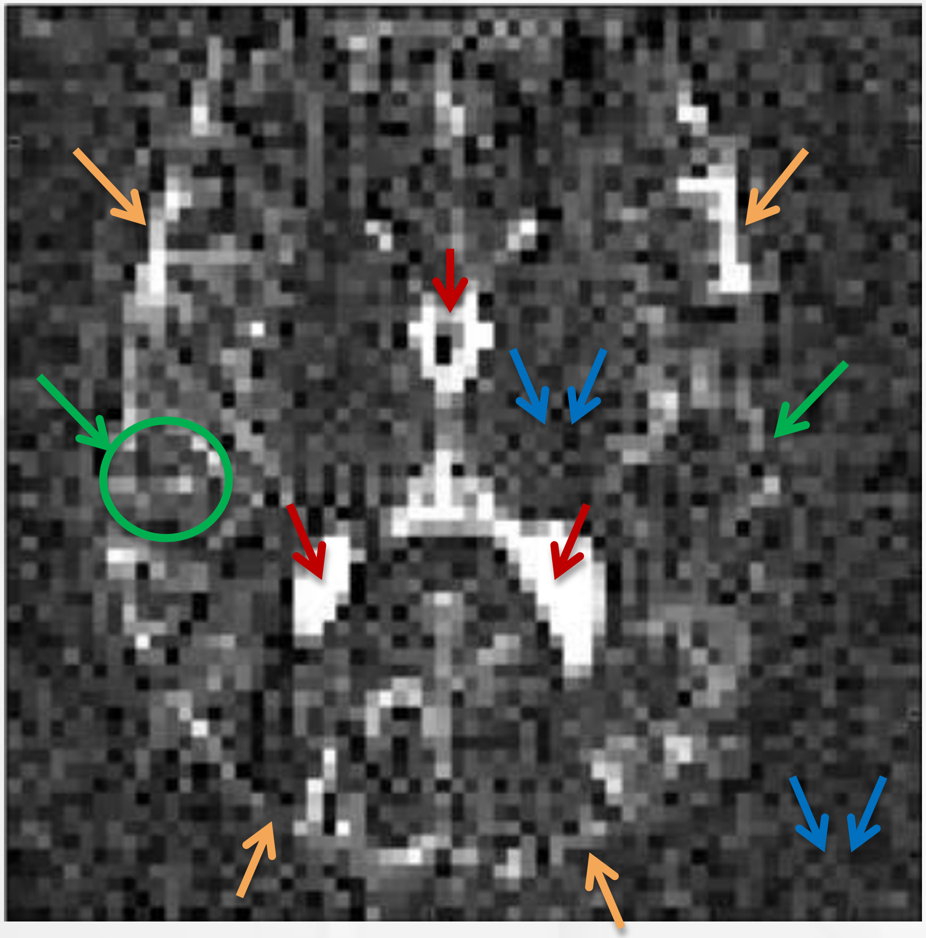
fMRI Movie: An example





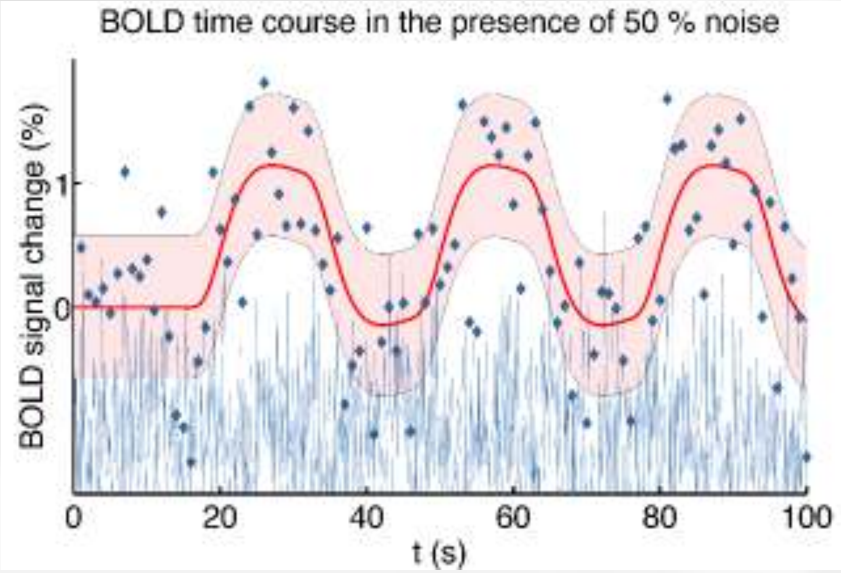
fMRI Movie: Subtract the Mean

- Interested in fluctuations only

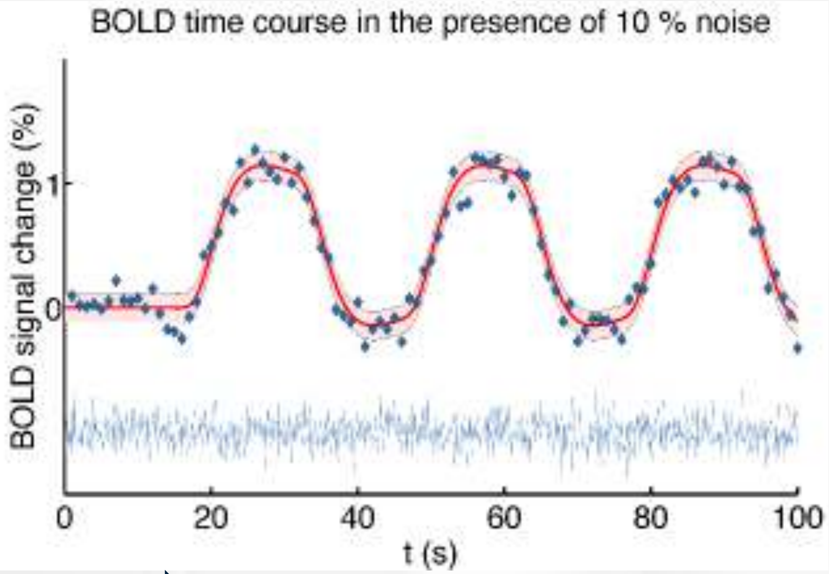


The Goal of Preprocessing

Before



After



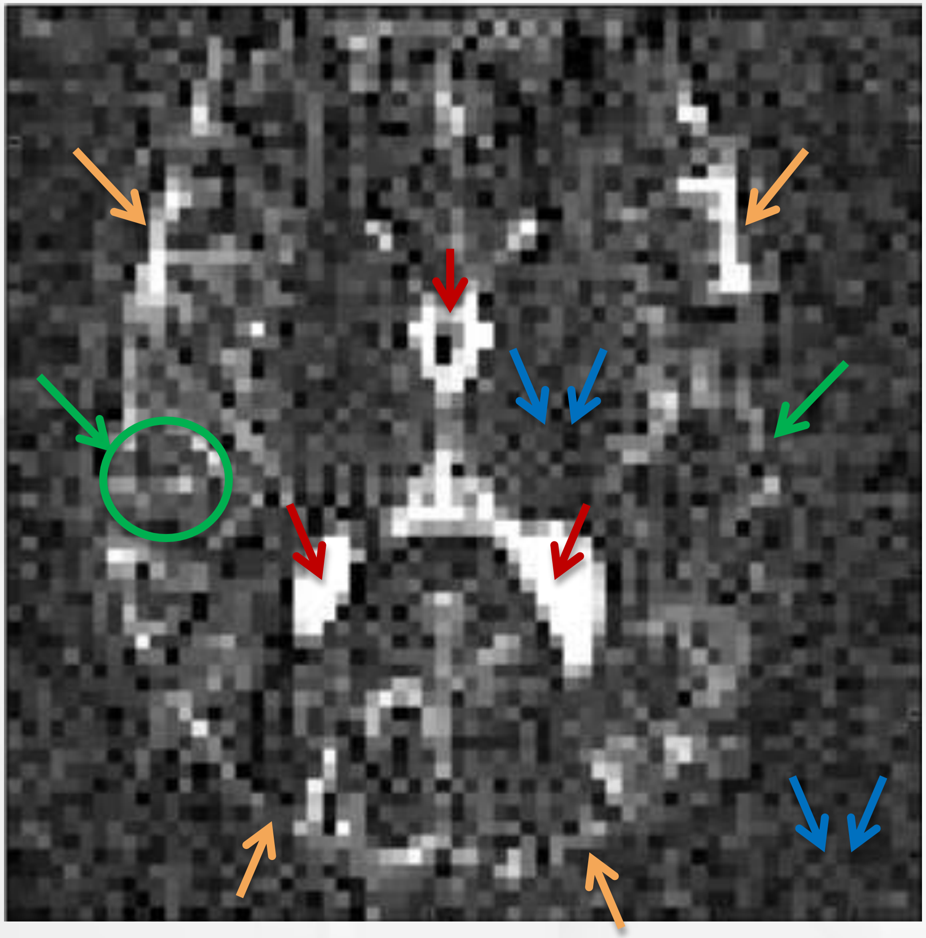
Preprocessing

Sources of Noise in fMRI

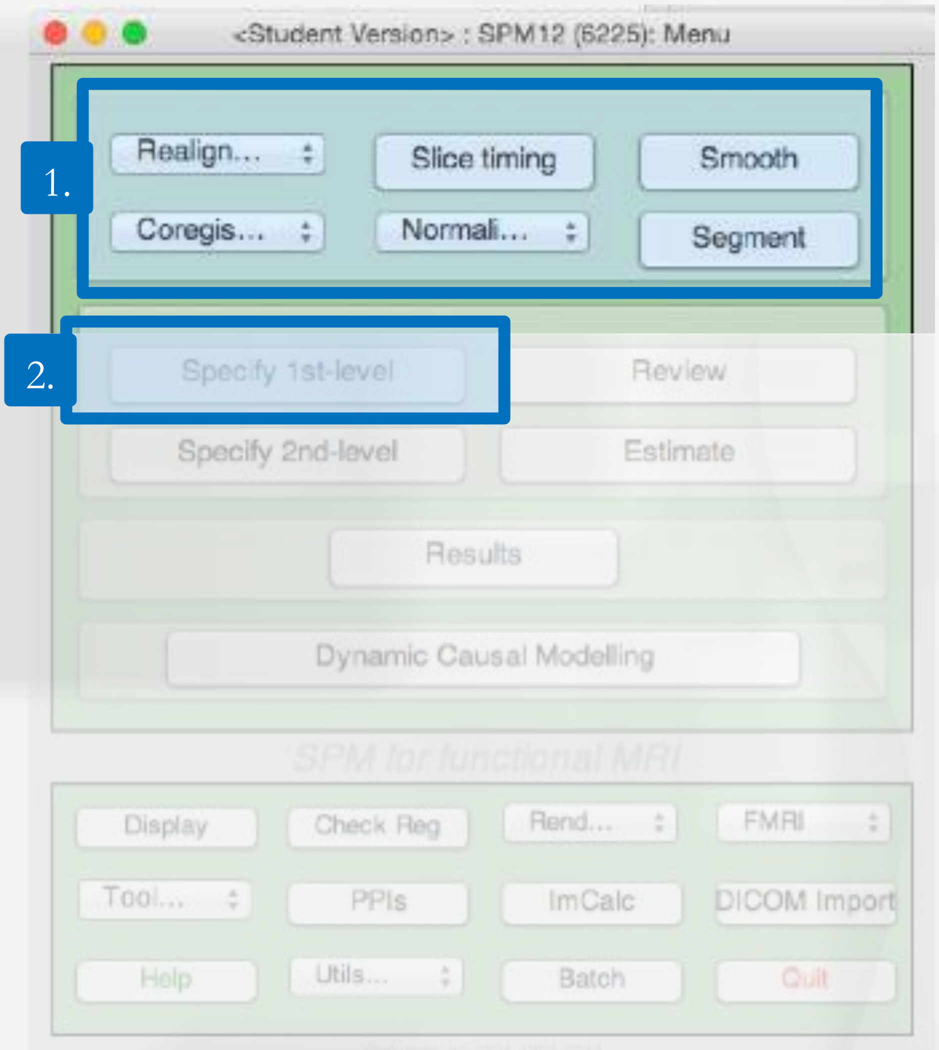
Source	Preprocessing	Method
Acquisition Timing	Temporal Preproc	Slice-Timing
Subject Motion	Spatial Preproc	Realignment
Anatomical Identity	Spatial Preproc	Co-registration
Inter-subject variability	Spatial Preproc	Segmentation
Thermal Noise	Spatial Preproc	Smoothing
Physiological Noise	Noise Modeling	PhysIO Toolbox

fMRI Movie: Noise Sources

- interest in fluctuations only



The SPM Graphical User Interface



- Preprocessing
  - Realignment
  - Slice-Timing Correction
  - Co-registration
  - Unified Segmentation & Normalisation
  - Smoothing. . .
- Noise Modeling
  - Physiological Confound Regressors

Sources of Noise in fMRI

Temporal Preproc

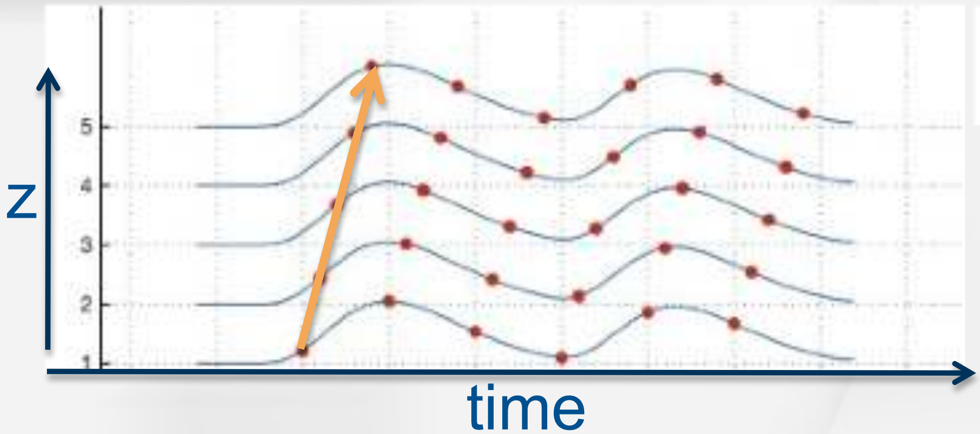
- Acquisition Timing
- Subject Motion
- Anatomical Identity
- Inter-subject variability
- Thermal Noise
- Physiological Noise

- Slice-Timing
- Realignment
- Co-registration
- Segmentation
- Smoothing
- PhysIO Toolbox

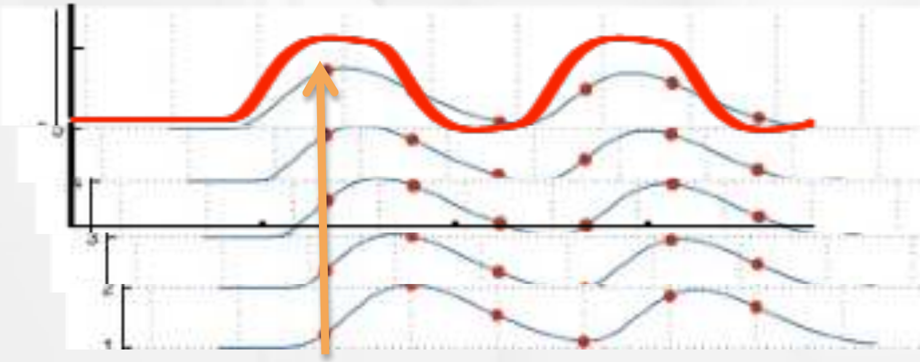
Slice-timing correction (STC)

- Slices of 1 scan volume are not acquired simultaneously (60 ms per slice)
- Creates shifts of up to 1 volume repetition time (TR), so typically several seconds
- This acquisition delay reduces sensitivity for time-locked effects (i.e. a smaller correlation with a temporally fixed model)

True 2D Acquisition

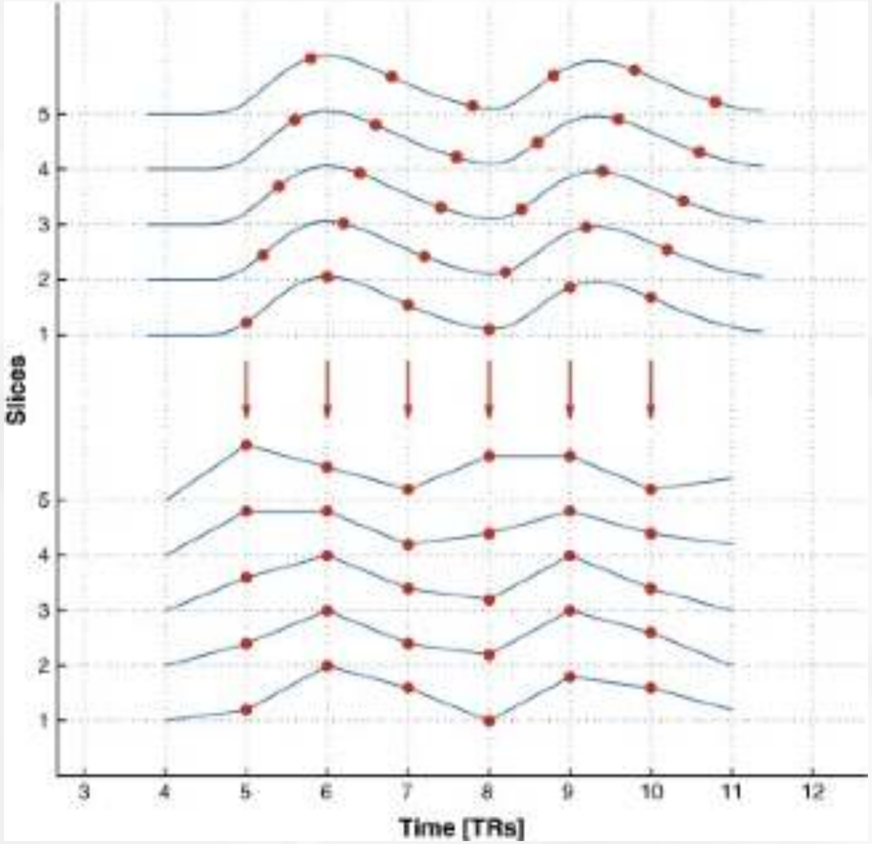


Same-Timepoint Assumption



Slice-timing correction (STC)

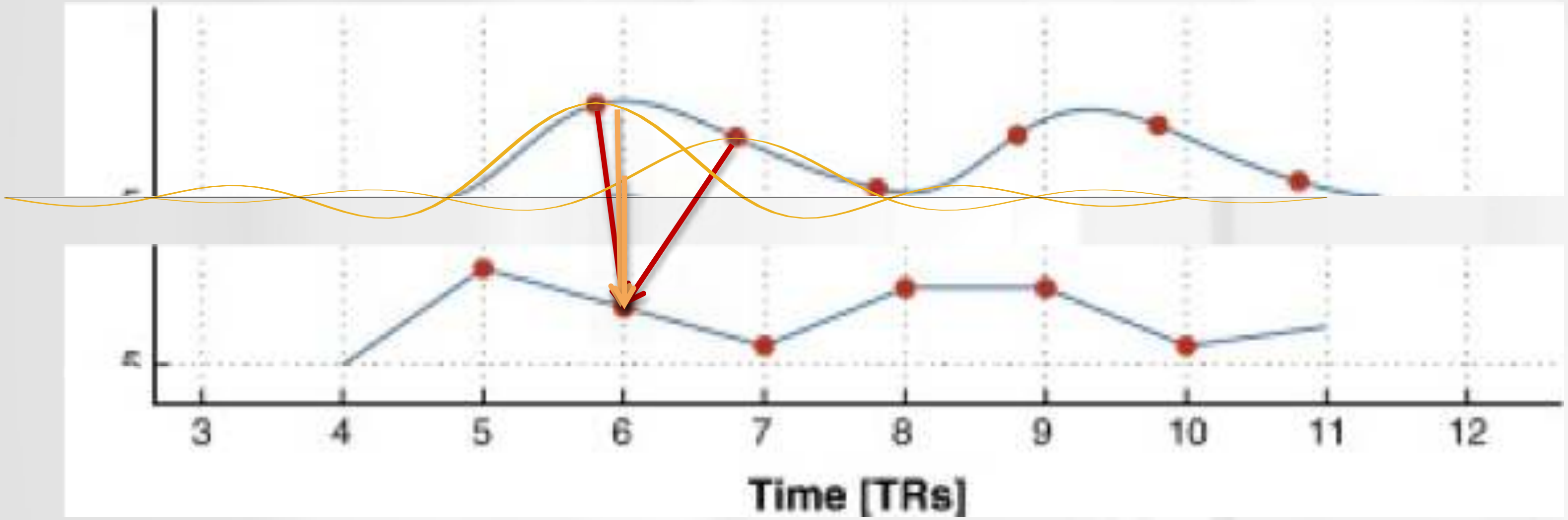
- Slice-timing correction: all voxel time series are aligned to acquisition time of 1 slice
- Missing data is sinc-interpolated (band-limited signal)



Sladky et al, NeuroImage 2011

Interpolation

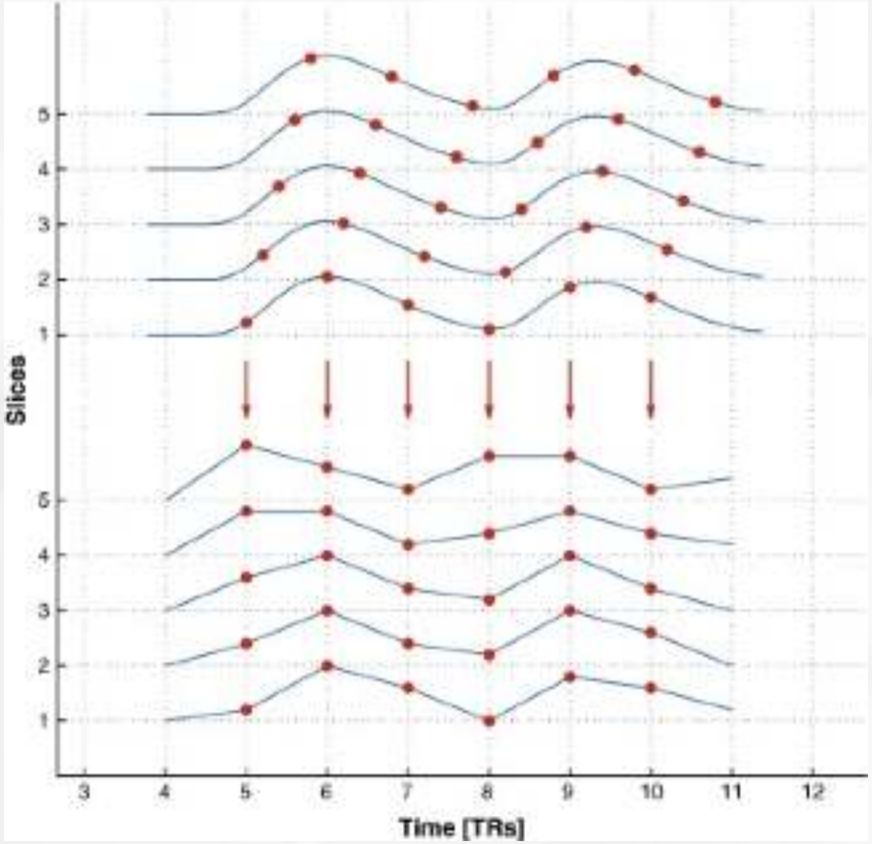
- Interpolation: estimate missing data between existing data via certain regularity assumptions



- Signal at missing point is weighted average of neighbors
- Weighting function = interpolation "kernel"
- Here: assumption of limited frequency range of signal: *sinc*-interpolation

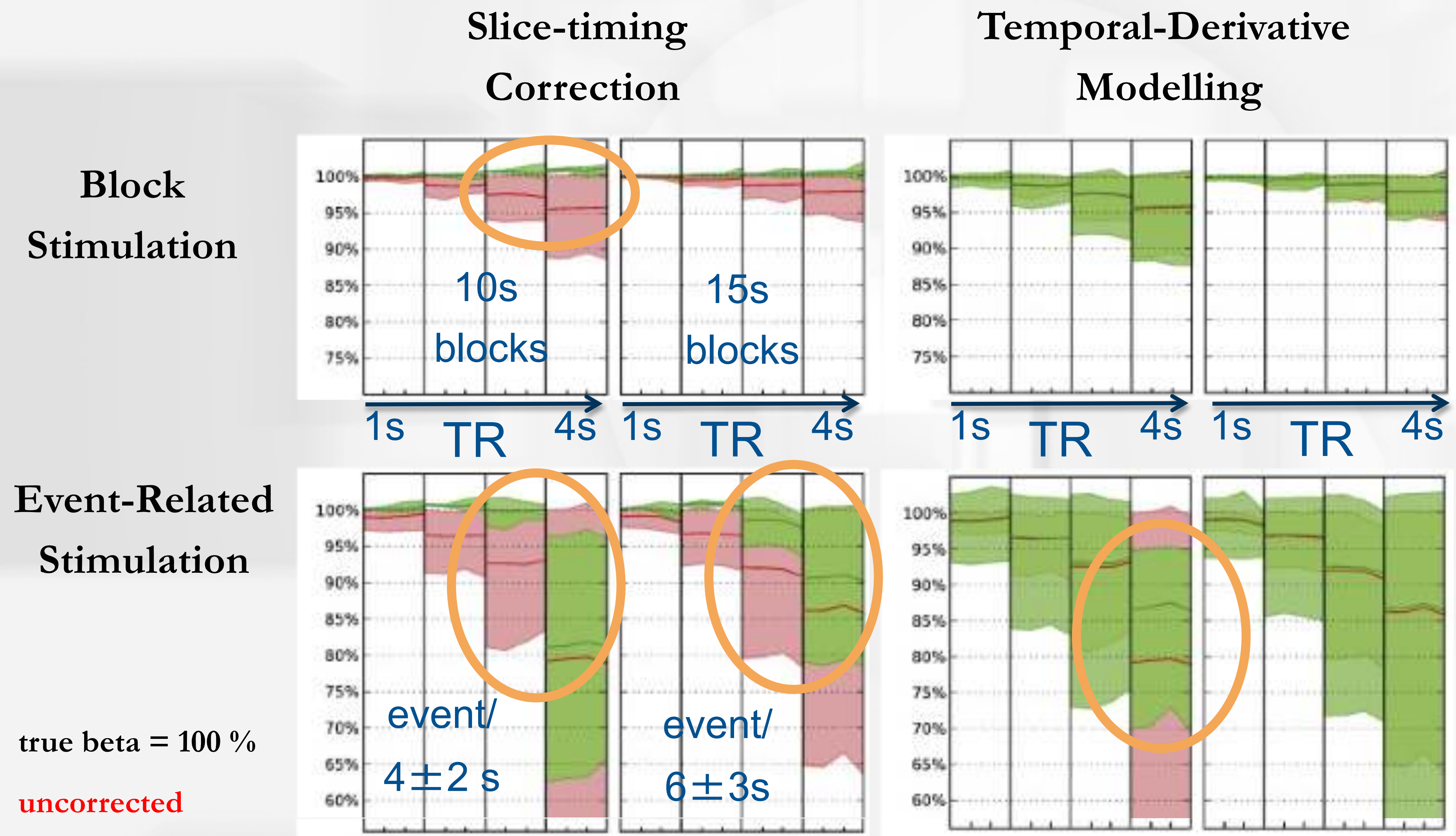
Slice-timing correction (STC)

- Slice-timing correction: all voxel time series are aligned to acquisition time of 1 slice
- Missing data is sinc-interpolated (band-limited signal)
- Before or after realignment?
  - before: dominant through-slice motion
  - after: dominant within-slice motion
- At all?



Sladky et al, NeuroImage 2011

STC Results: Simulation



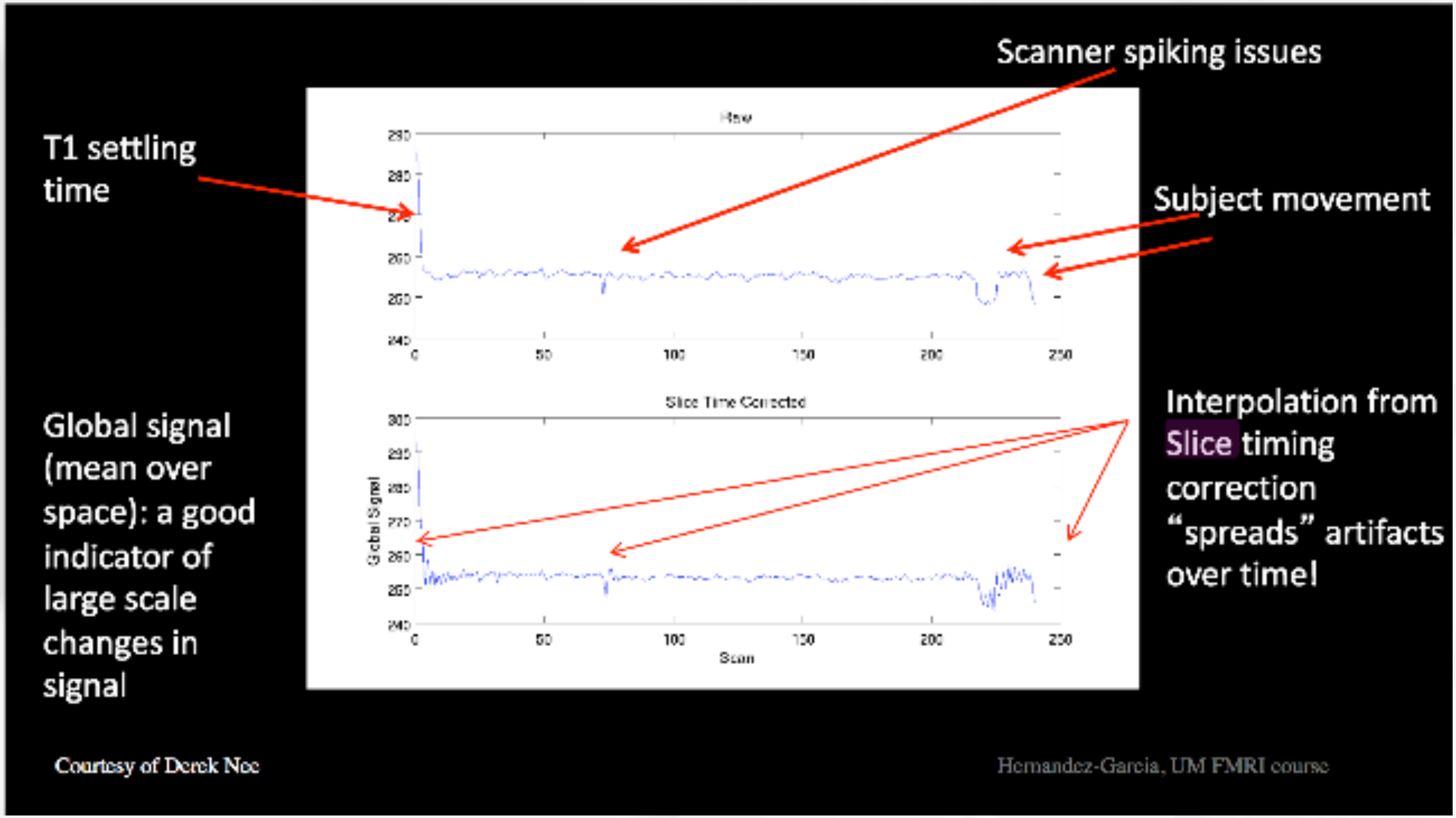
true beta = 100 %

uncorrected

corrected

Sladky et al, NeuroImage 2011

STC Issues: Motion

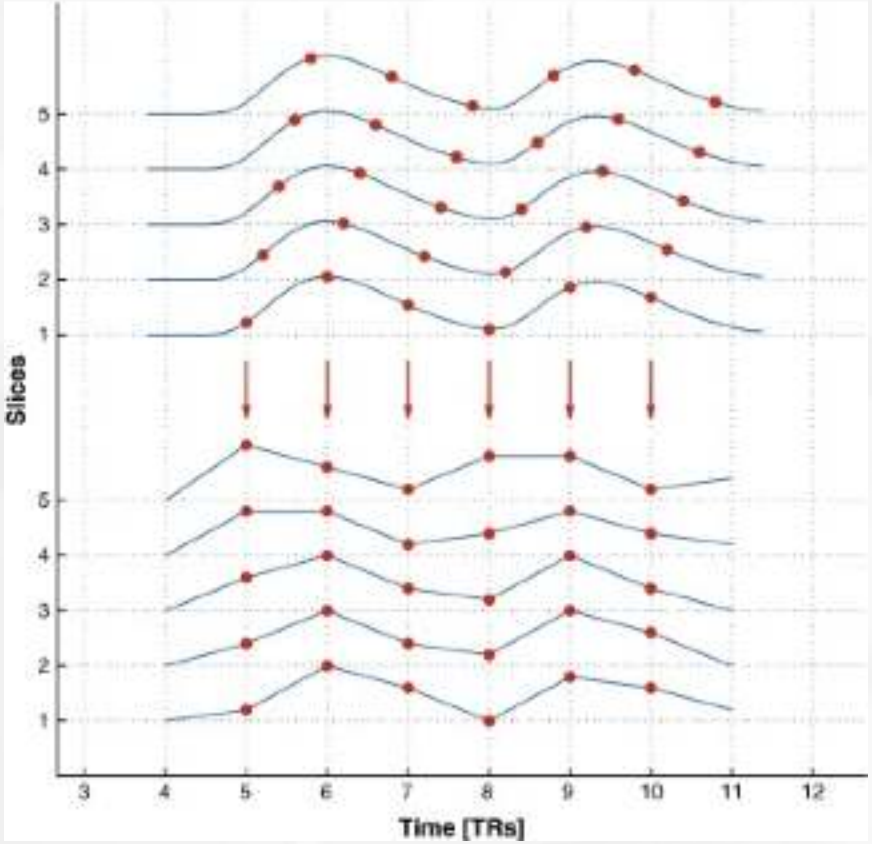


http://imaging.mrc-cbu.cam.ac.uk/imaging/SliceTiming

Power et al., PLoS One, 2017

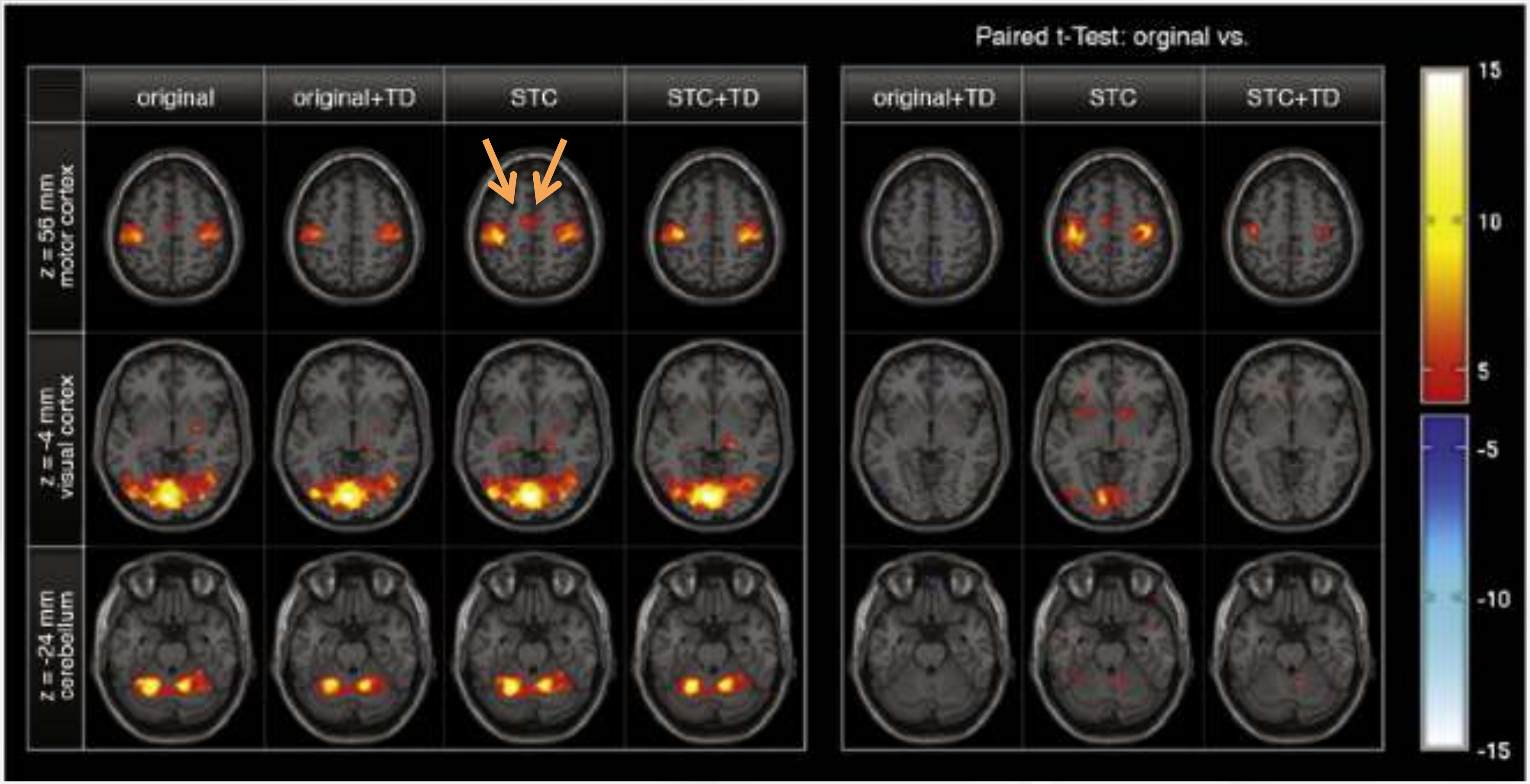
Slice-timing correction (STC)

- Slice-timing correction: all voxel time series are aligned to acquisition time of 1 slice
- Missing data is sinc-interpolated (band-limited signal)
- Before or after realignment?
  - before: dominant through-slice motion
  - after: dominant within-slice motion
- At all?
  - block design: for long TR (3s+) & short blocks (10s) improves estimates > 5 %
  - event-related: for normal TRs (2s+) improves estimates > 5 %



Sladky et al, NeuroImage 2011

STC Results: Experiment



Sladky et al, NeuroImage 2011

Sources of Noise in fMRI

- Acquisition Timing
- Subject Motion
- Anatomical Identity
- Inter-subject variability
- Thermal Noise
- Physiological Noise

Spatial Preproc

Spatial Preproc

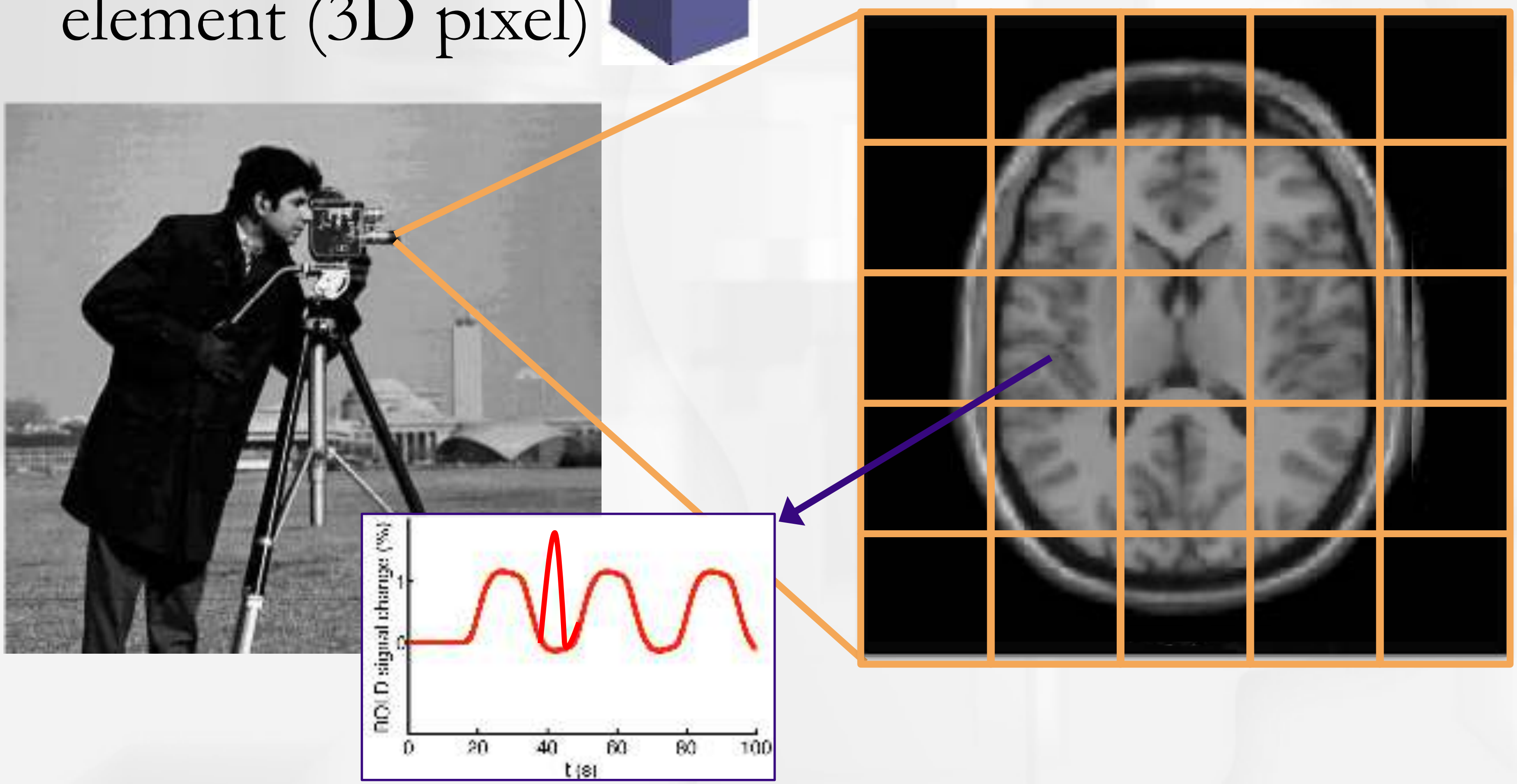
Spatial Preproc

Spatial Preproc

- Slice-Timing
- Realignment
- Co-registration
- Segmentation
- Smoothing
- PhysIO Toolbox

Finite Resolution and Voxel Identity

- voxel = volume element (3D pixel)



Preproc = Correct Voxel Mismatch

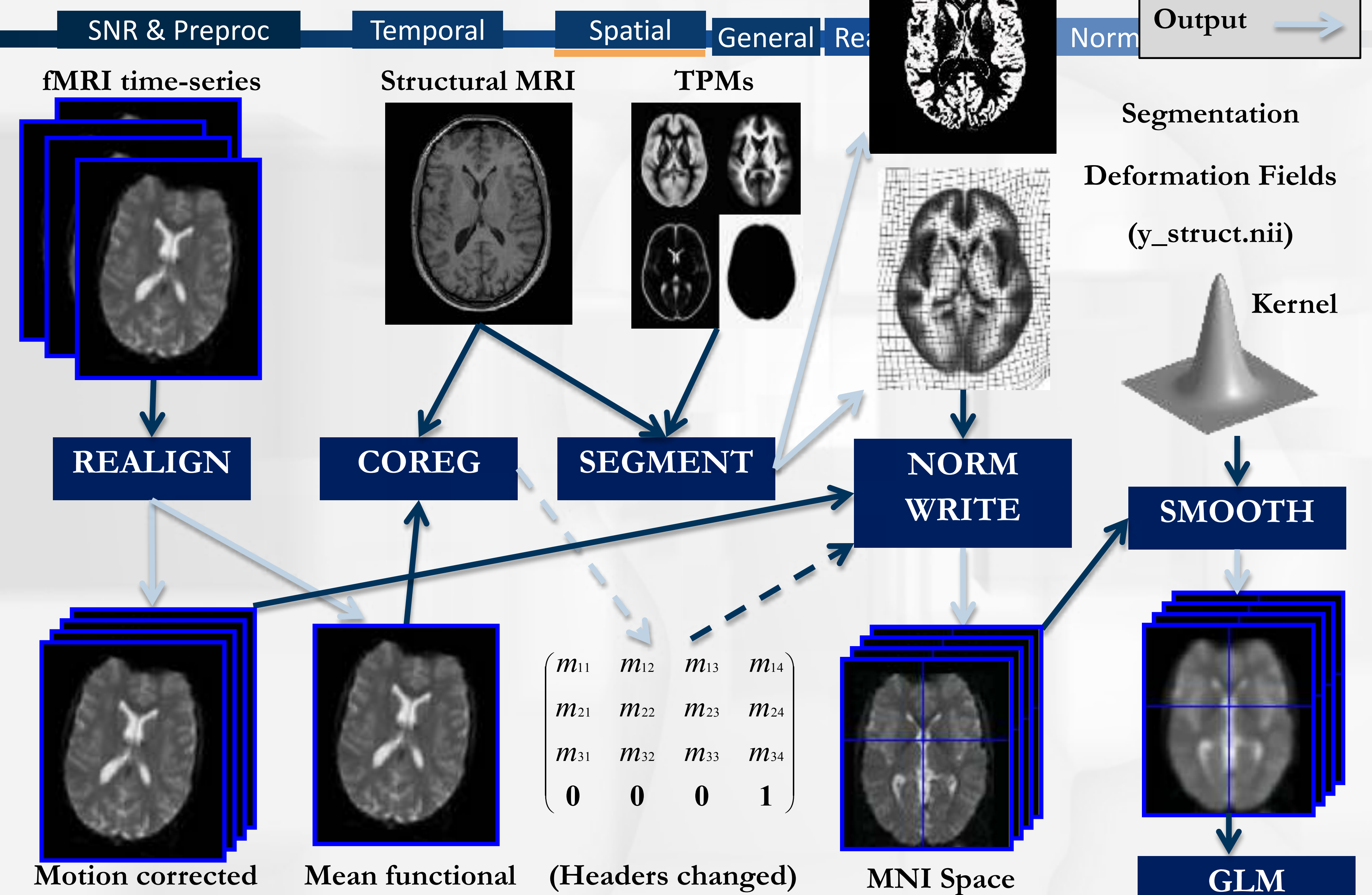
Voxel Mismatch Between

- Functional Scans/Runs
- Functional/Structural Images
- Subjects

Realignment | Inter-Modal Coregistration | Normalisation/ Segmentation

Smoothing

Spatial Preprocessing

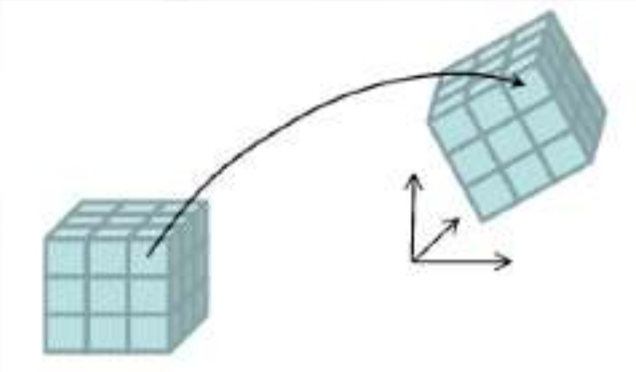


General Remarks: Image Registration

- Realignment, Co-Registration and Normalisation (via Unified Segmentation) are all *image registration methods*
- Goal: manipulate one set of images to arrive in same coordinate system as a reference image
- Key ingredients for image registration
    A. Voxel-to-world mapping
    B. Transformation
    C. Similarity Measure
    D. Optimisation
    E. Interpolation

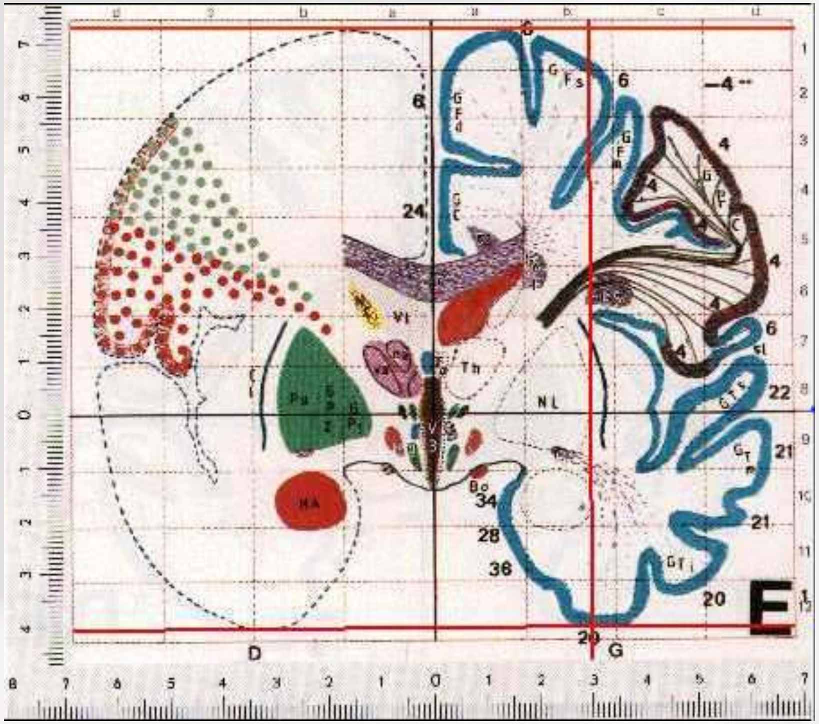
A. Voxel-to-World Mapping

- 3D images are made up of voxels.
- Voxel intensities are stored on disk as lists of numbers.
- Meta-information about the data:
  - image dimensions
    - conversion from list to 3D array
  - "voxel-to-world mapping"
    - Spatial transformation that maps
      - from: data coordinates (voxel column i, row j, slice k)
      - to: a real-world position (x,y,z mm) in a coordinate system e.g.:
      - Scanner coordinates
      - T&T/MNI coordinates

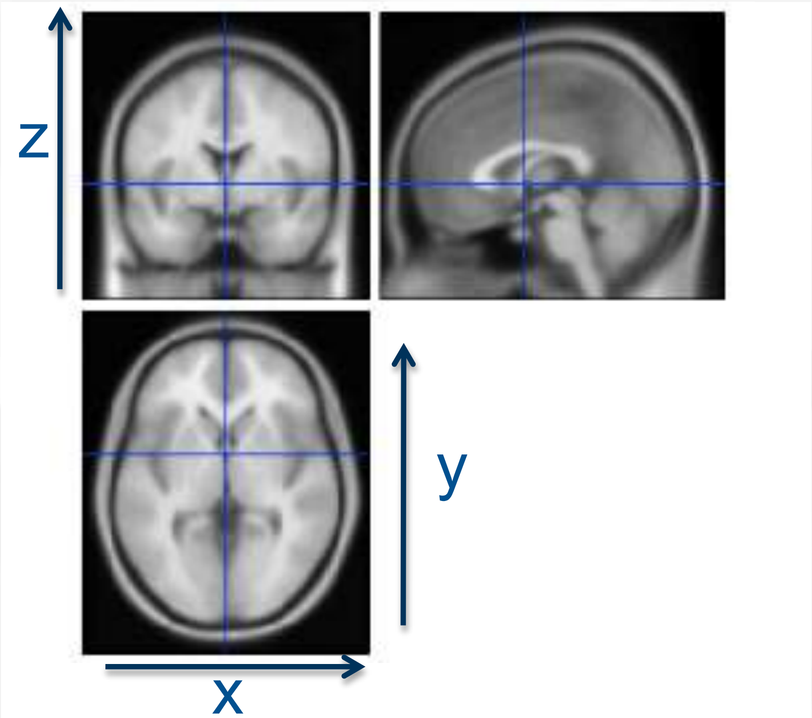


A. Voxel-to-World: Standard Spaces

Talairach Atlas



MNI/ICBM AVG152 Template

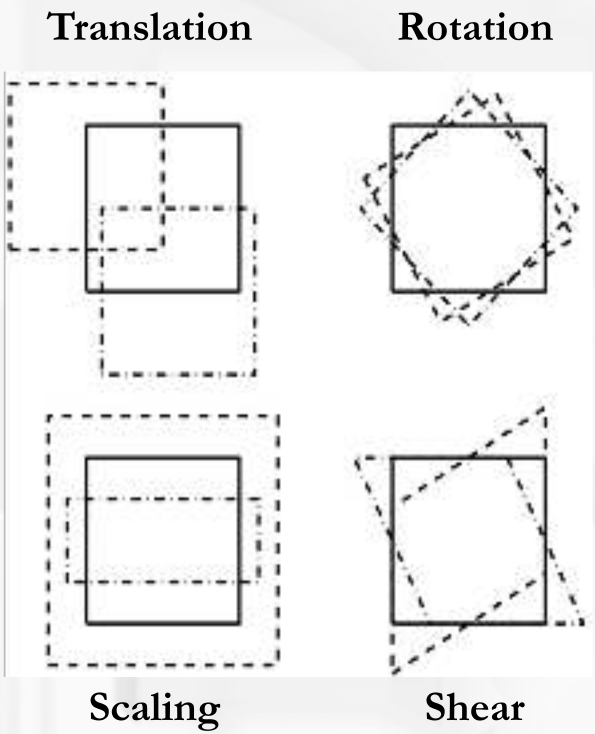


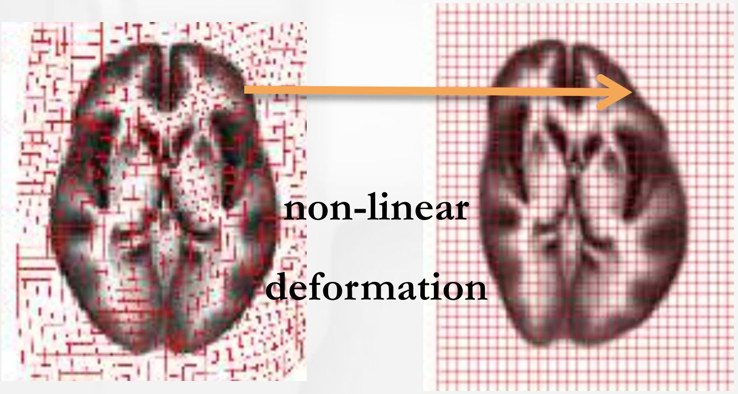
- Definition of coordinate system:
  - Origin (0,0,0): anterior commissure
  - Right = +X; Anterior = +Y; Superior = +Z

- Actual brain dimensions
  - European brains, a bit dilated (bug)

B. Transformations

- Transformations describe the mapping of all image voxels from one coordinate system into another
- Types of transformations
  - rigid body = translation + rotation
  - affine = rigid body + scaling + shear
  - non-linear = any mapping
    - (x,y,z) to new values (x',y', z')
    - described by deformation fields





Spatial Preproc: SPM vocabulary

- SPM uses different names for different modes of image registration
- depending on input images and allowed transformations

Realignment

- Intra-modal image registration
  - e.g. functional images
- rigid body transformations
  - translation/rotation

Co-Registration

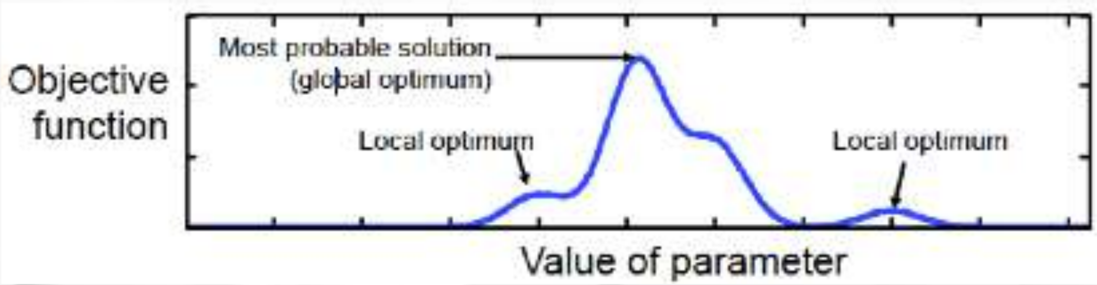
- Inter-modal registration
  - e.g. T1/T2 contrast
  - functional to structural image
- affine transformations
  - rigid body
  - stretching/shearing

Normalisation

- Multi-modal registration
  - e.g. T1 and/or T2
  - structural image(s) to template
- non-linear transformations
  - voxel-wise mapping (deformation fields)

C. Similarity & D. Optimisation

- Similarity measure summarizes resemblance of (transformed) image and reference into 1 number
  - mean-squared difference
  - correlation-coefficient
  - mutual information



- Automatic image registration uses an optimisation algorithm to maximise/minimise an "objective function"
  - Similarity measure is part of objective function
  - Algorithm searches for transformation that maximises similarity of transformed image to reference
  - Also includes constraints on allowed transformations (priors)

Preprocessing Step Categorisation

B. Allowed Transformations

Rigid-Body	Affine	Non-linear	
REALIGN	**COREG**	**SEGMENT**	**NORM WRITE**

C. Similarity Measure

Mean-squared Difference	Mutual Information	Tissue Class Probability

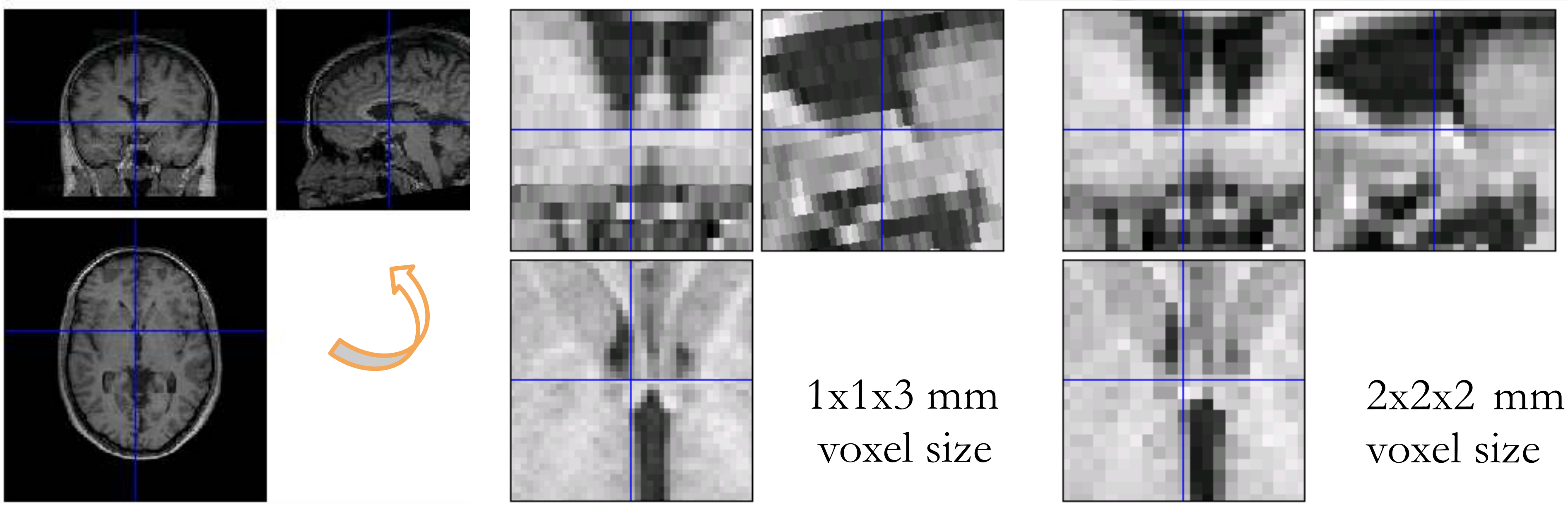
D. Optimisation

Exact Linearized Solution	Conjugate Direction Line Search	Iterated Conditional Modes (EM/Levenberg-Marquardt)

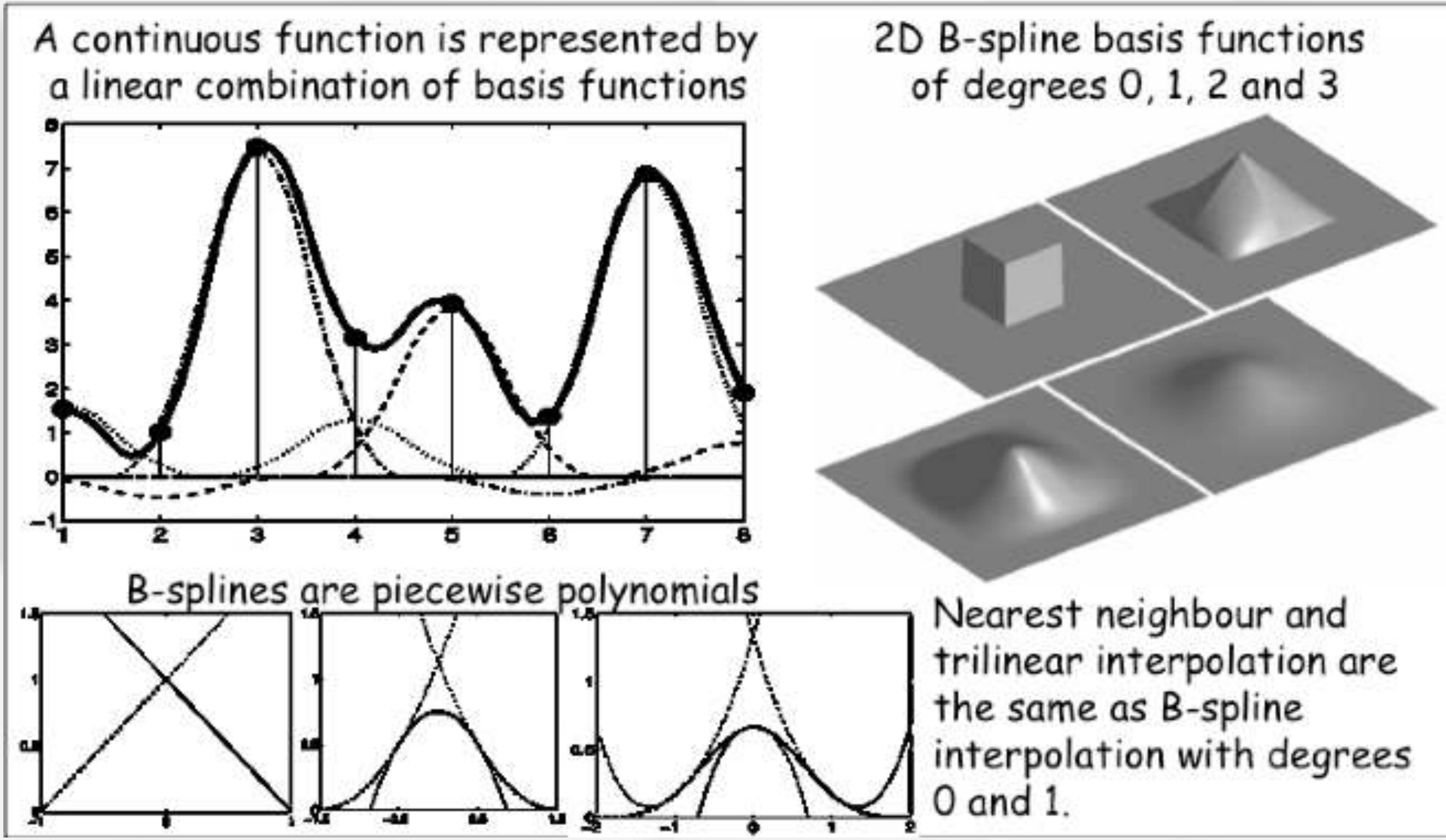
E. Reslicing/Interpolation

- Finally, images have to be saved as voxel intensity list on disk again
- After applying transformation parameters, data is re-sampled onto same grid of voxels as reference image

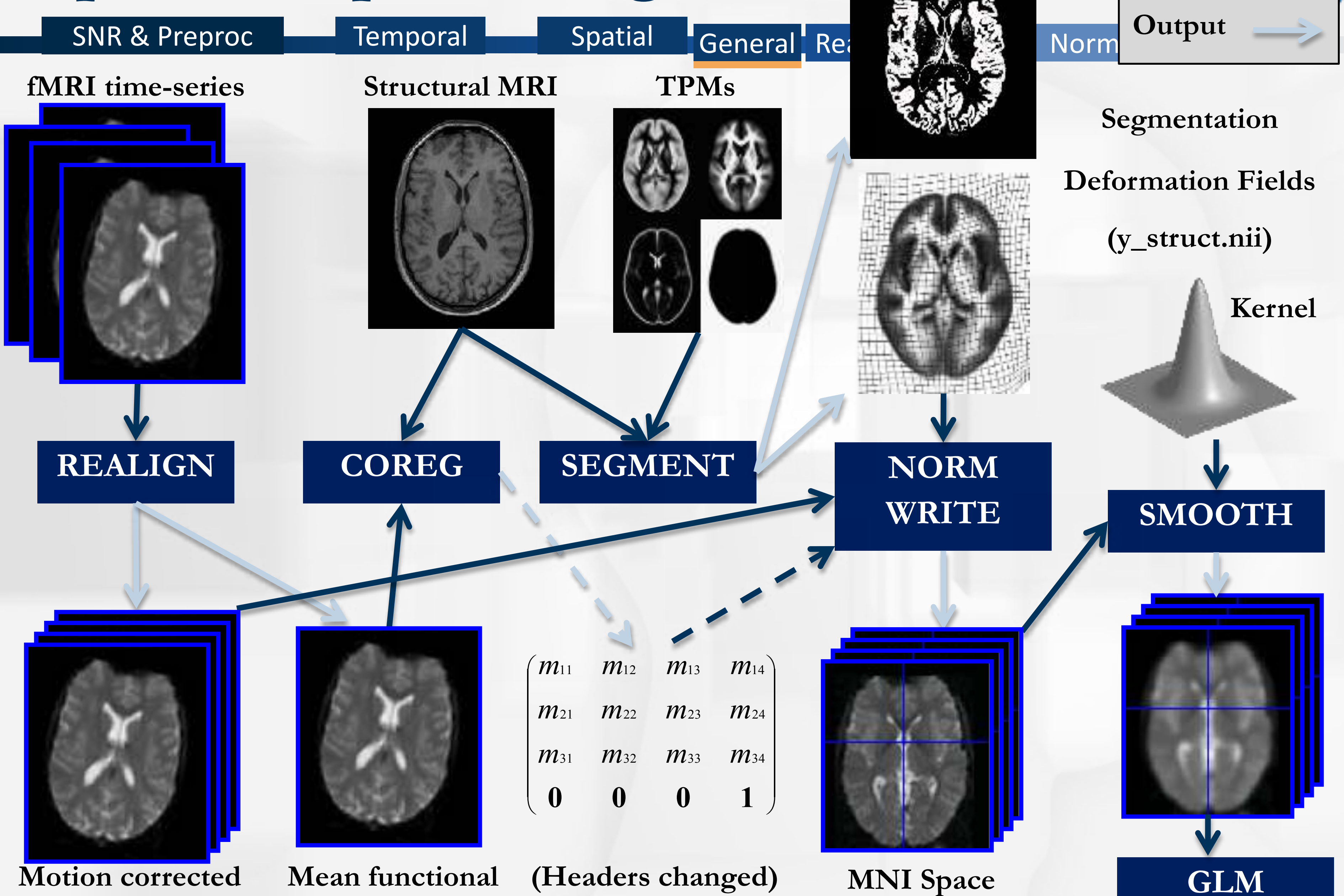
Reoriented **Resliced**



E. B-spline Interpolation

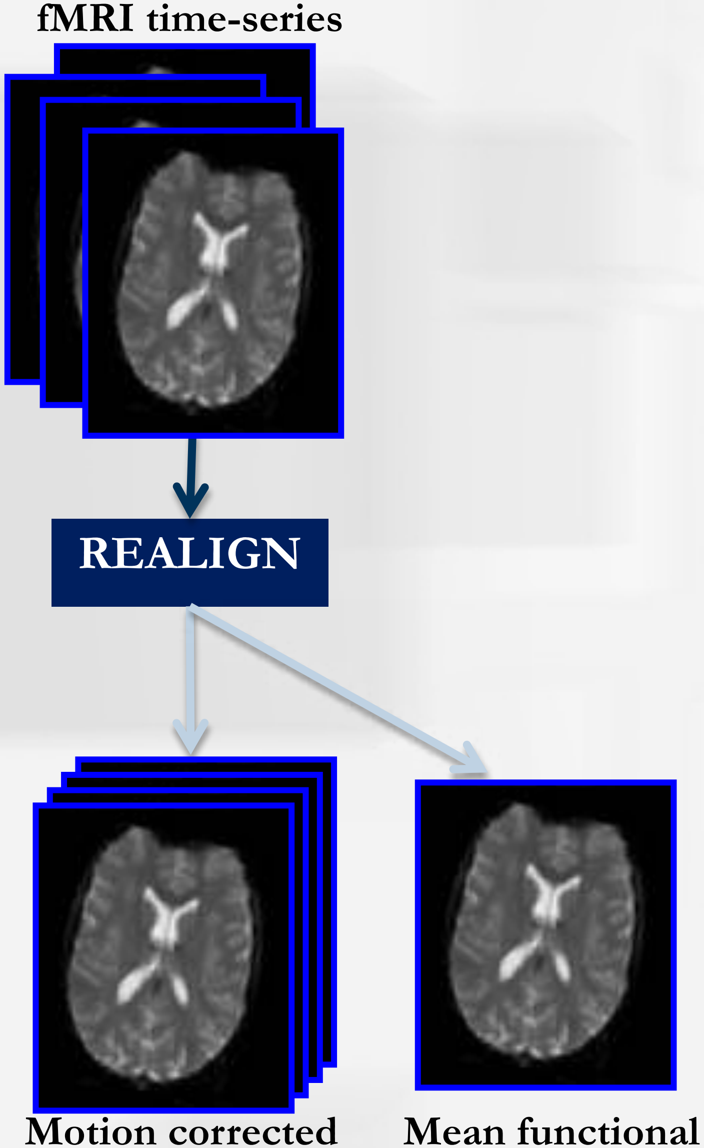


Spatial Preprocessing



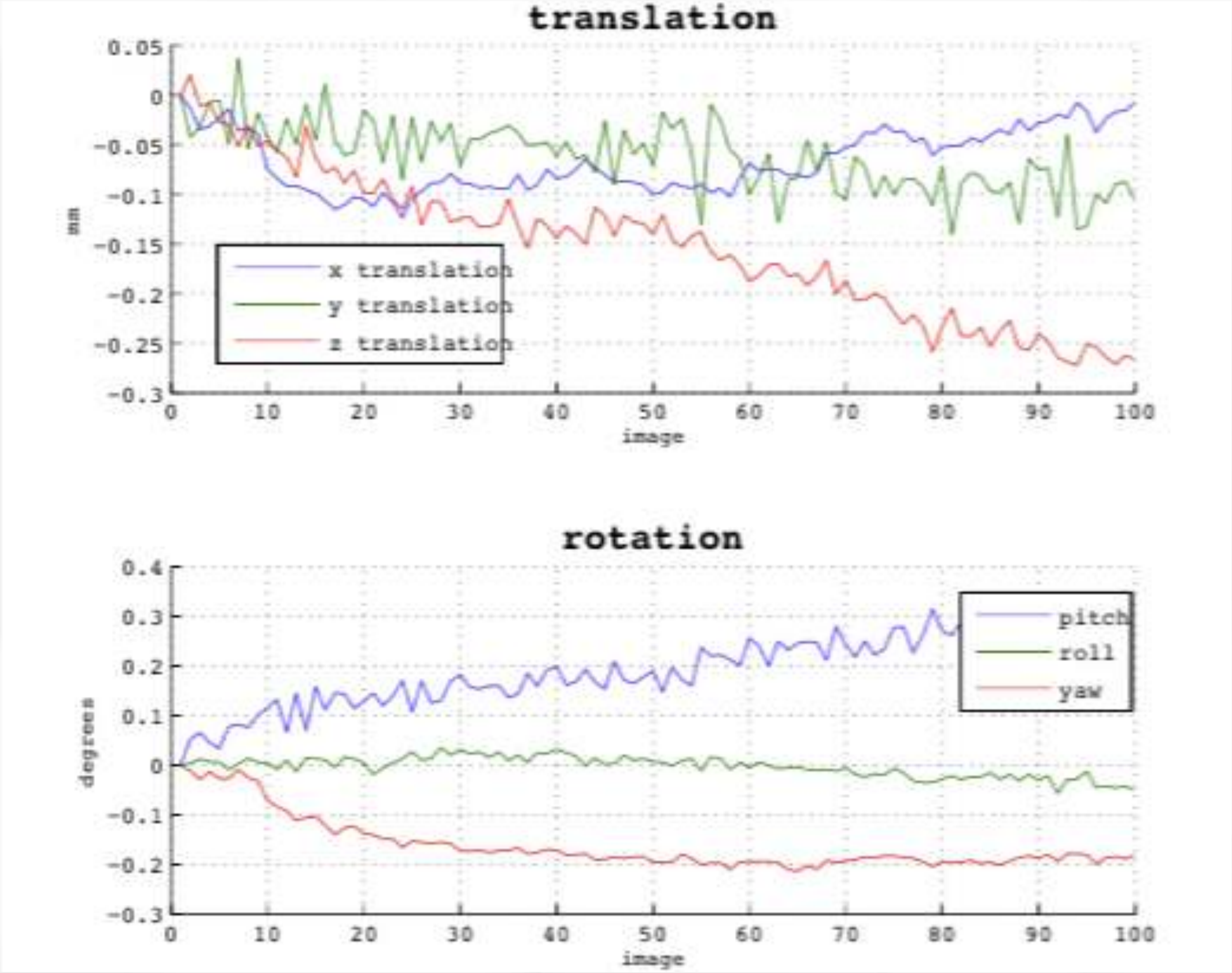
Realignment

SNR & Preproc | Temporal | Spatial | General | Realign | Coreg | Normalise | Smooth

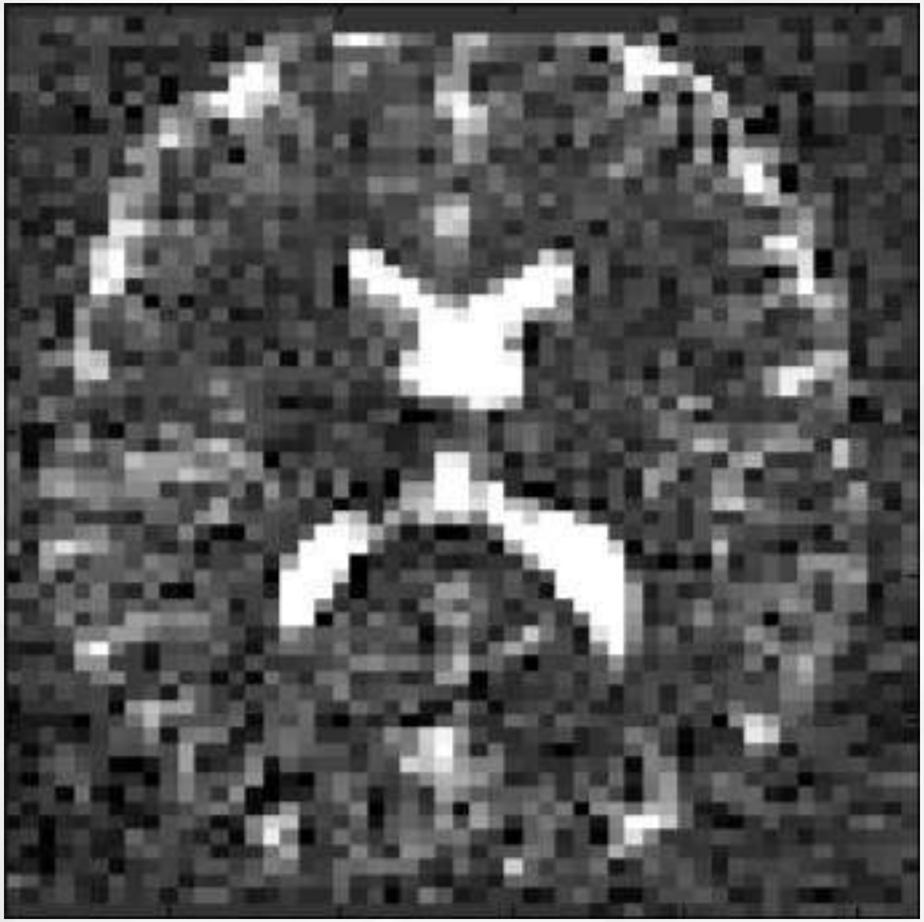


- Aligns all volumes of all runs spatially
- Rigid-body transformation: three translations, three rotations
- Objective function: mean squared error of corresponding voxel intensities
- Voxel correspondence via interpolation

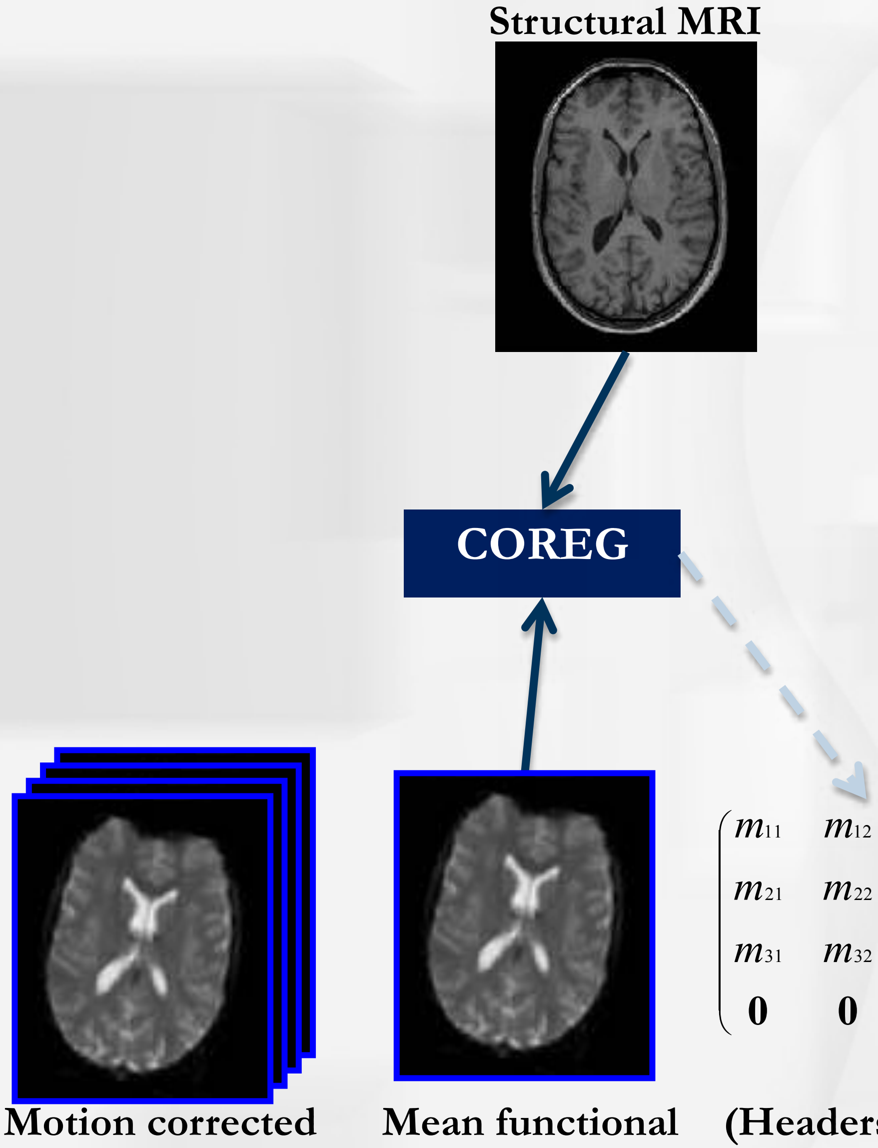
Realignment Output: Parameters



fMRI Run after Realignment



Co-Registration

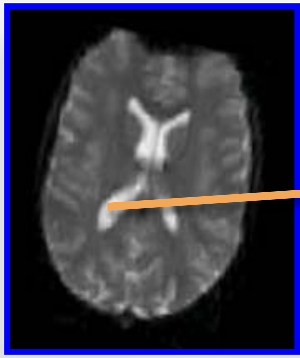


$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

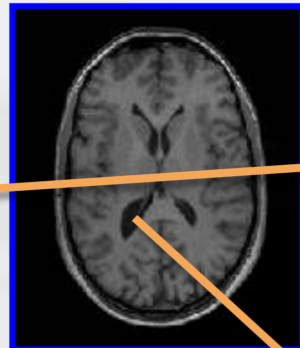
(Headers changed)

- Aligns structural image to mean functional image
- Affine transformation: translations, rotations, scaling, shearing
- Objective function: mutual information (diff. contrast!)
  - Optimisation via Powell's method: conjugate directions, line seach along parameters
- Typically only trafo matrix ("header") changed

Co-Registration: Mutual Information

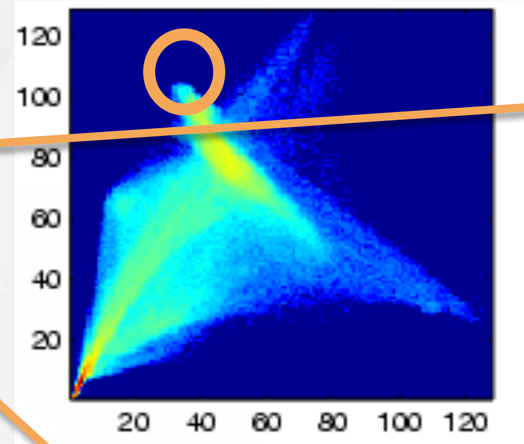


Mean functional

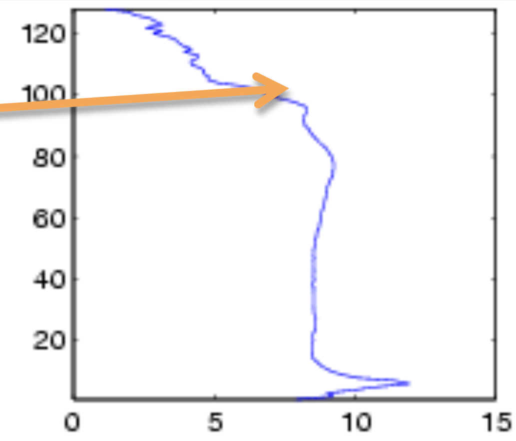


Anatomical MRI

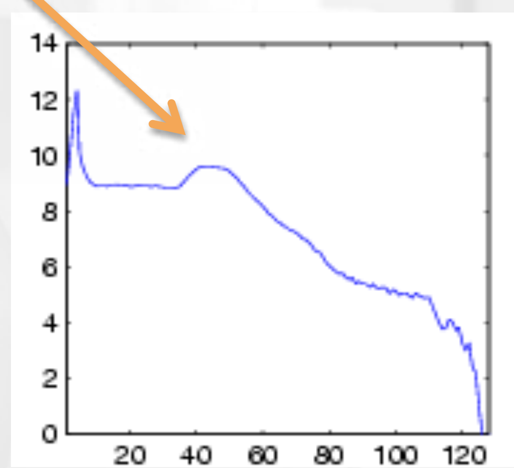
Joint Histogram



Marginal Histogram



- Voxels of same tissue identity have same intensity in an MR-contrast
- In a 2nd MR contrast, intensity might be different, but still the same among all voxels of the same tissue type
- Therefore, aligned voxels in 2 images induce crisp peaks in joint histogram



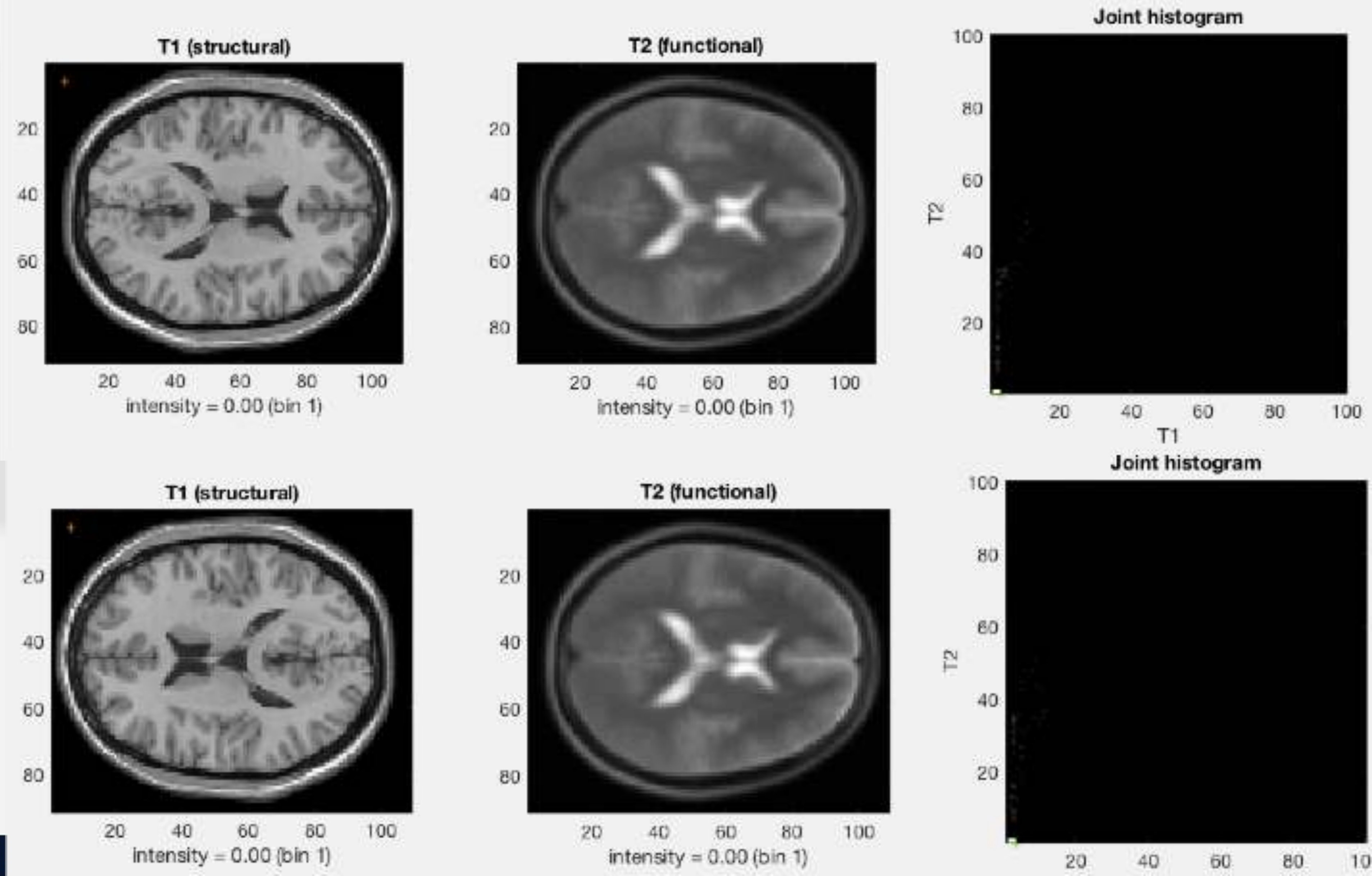
intensity bins structural

intensity bins functional

Joint Histogram: $h(i_f,i_s)$

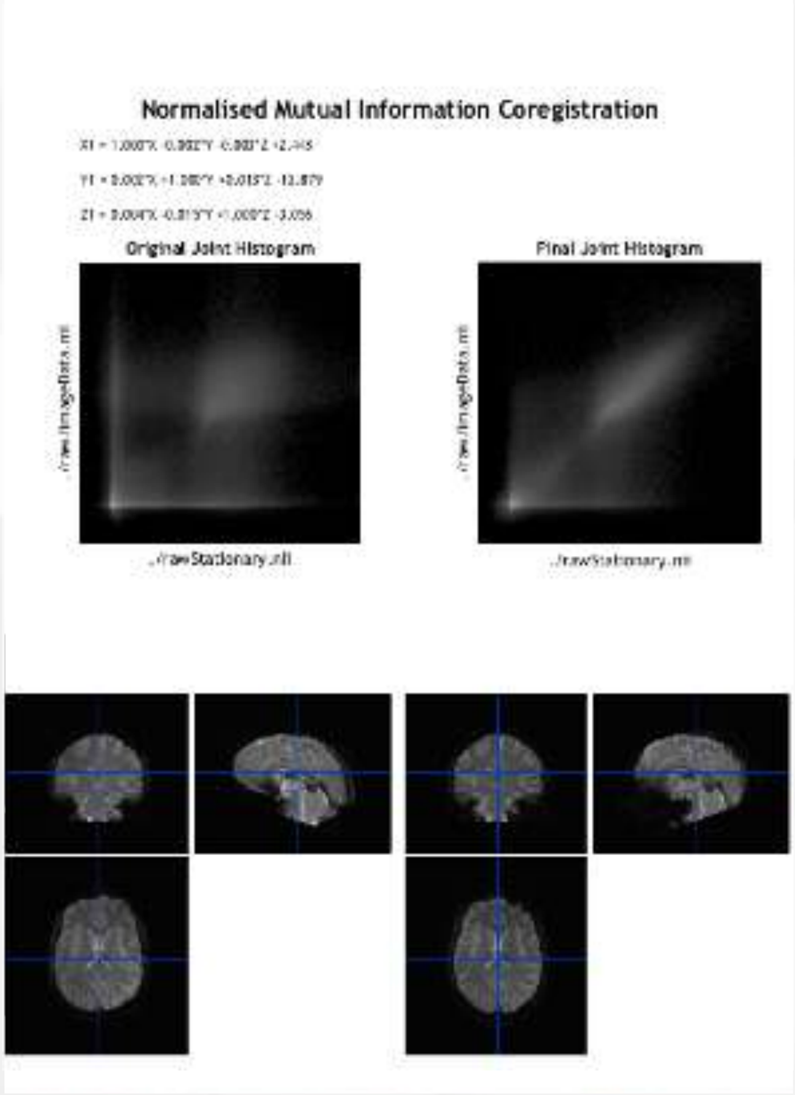
Count of voxels who have intensity $\mathbf{i_f}$ in functional and $\mathbf{i_s}$ in structural image

Co-Registration: Mutual Information



Co-Registration: Output

- Aligned voxels in 2 images induce crisp peaks in joint histogram
- Optimization criterion:
  - Joint histogram: quantify how well voxel intensity in one image predicts the intensity in the other
  - how much shared (=mutual) information
  - Joint histogram: proxy to joint probability distribution



Sources of Noise in fMRI

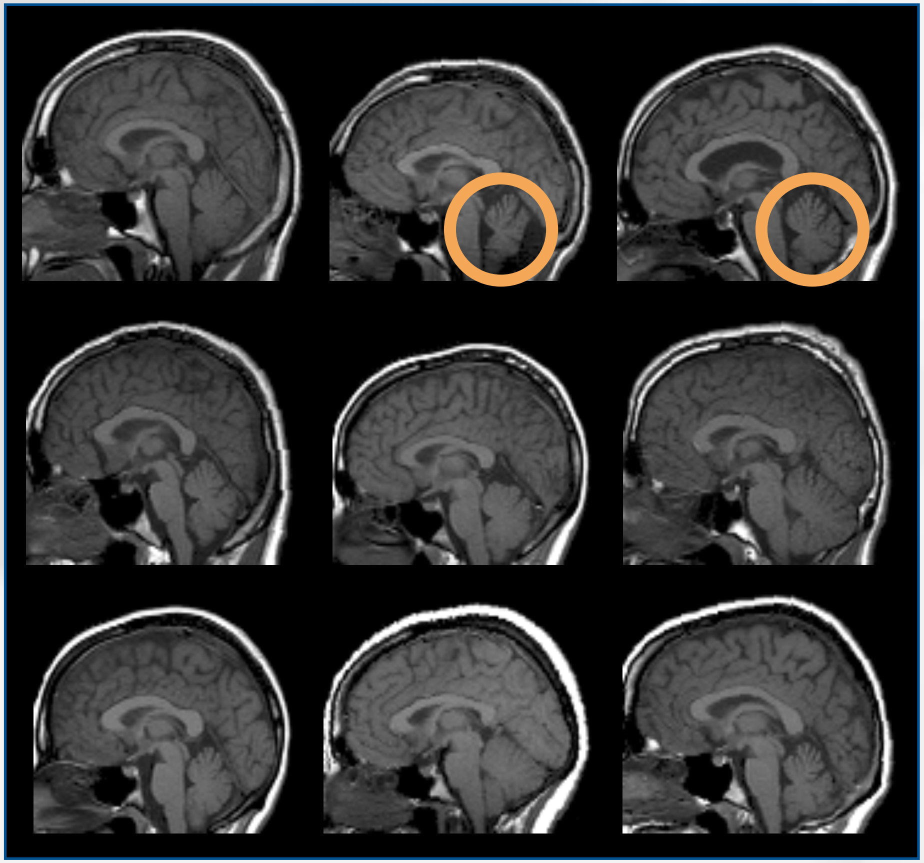
- Acquisition Timing
- Subject Motion
- Anatomical Identity
- Inter-subject variability
- Thermal Noise
- Physiological Noise

Spatial Preproc

- Slice-Timing
- Realignment
- Co-registration
- Segmentation
- Smoothing
- PhysIO Toolbox

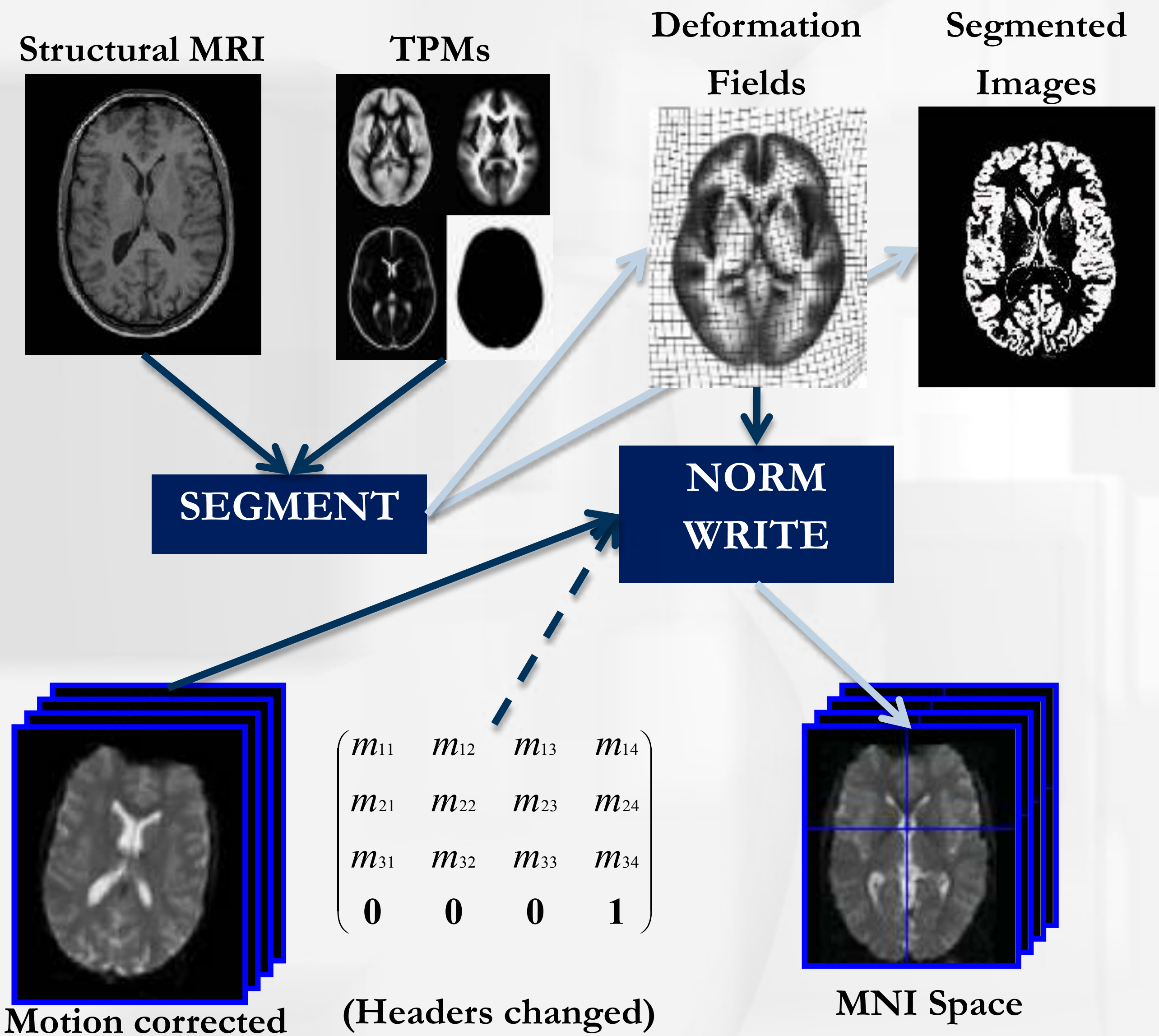
Spatial Normalisation: Reasons

- Inter-Subject Variability



- Inter-Subject Averaging
  - Increase sensitivity with more subjects (fixed-effects)
  - Generalise findings to population as a whole (mixed-effects)
- Ensure Comparability between studies (alignment to standard space)
  - Talairach and Tournoux (T&T) convention using the Montreal Neurological Institute (MNI) space
  - Templates from 152/305 subjects

Unified Segmentation



- Warps structural image to standard space (MNI)
- Non-linear transformation: discrete cosine transforms (~1000)
- Objective function: Bayes probability of voxel intensity

Theory: Segmentation/Normalisation

- Why is normalisation difficult?
  - No simple similarity measure, a lot of possible transformations…
  - Different imaging sequences (contrasts, geometry distortion)
  - Noise, artefacts, partial volume effects
  - Intensity inhomogeneity (bias field)
- **Normalisation** of segmented tissues is more robust and precise than of original image
- Tissue **segmentation** benefits from spatially aligned tissue probability maps (of prior segmentation data)

➡ Motivates a unified model of segmentation/normalisation

Summary of the unified model

- SPM12 implements a generative model of voxel intensity from tissue class probabilities
  - Principled Bayesian probabilistic formulation
  - Gaussian mixture model: segmentation by tissue-class dependent Gaussian intensity distributions
  - Voxel-wise prior mixture proportions given by tissue probability maps
- Deformations of prior tissue probability maps also modelled
  - Non-linear deformations are constrained by regularisation factors
  - Inverse of estimated transformation for TPMs normalises the original image
- Bias field correction is included within the model

Theory: Unified Model Segmentation

$$\mathcal{E} = -\sum_{i=1}^{I} \log\left(\frac{\rho_i(\beta)}{\sum_{k=1}^{K}\gamma_k b_{ik}(\alpha)} \sum_{k=1}^{K} \gamma_k b_{ik}(\alpha)\left(2\pi\sigma_k^2\right)^{-\frac{1}{2}} \times \exp\left(-\frac{\left(\rho_i(\beta)y_i-\mu_k\right)^2}{2\sigma_k^2}\right)\right)$$

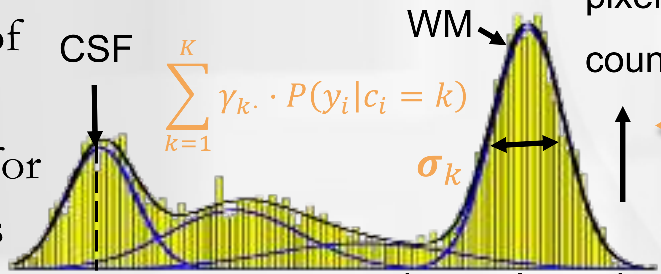
(2005), Neuroimage

- Objective function: log joint probability of all voxel intensities $\boldsymbol{y}$

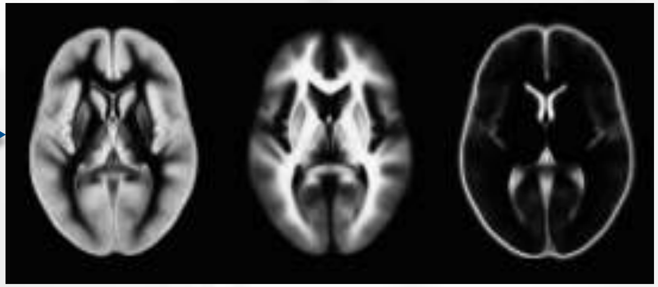
$$\mathcal{E} = \log P(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{b}_{1\ldots K}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

Gaussian Mixture Model

probability of intensity in given voxel for tissue class



Prior: Tissue probability maps

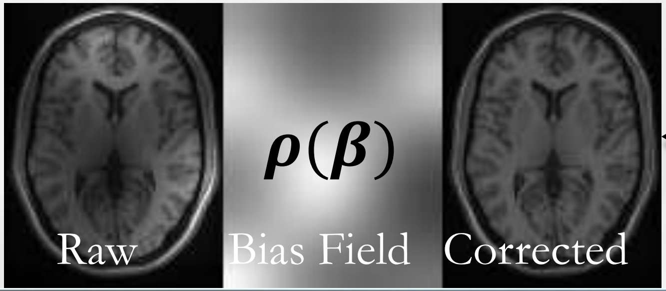


$\boldsymbol{b}_1$ $\boldsymbol{b}_2$ $\boldsymbol{b}_3$

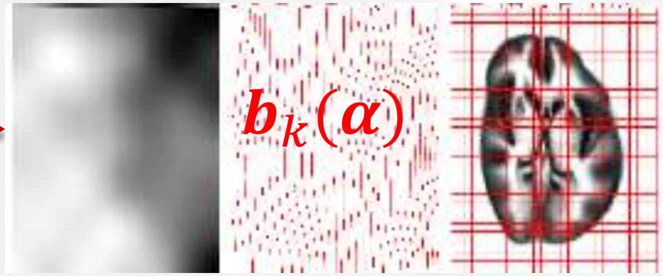
TPMs in MNI space

Bias Field

coil inhomo-geneities



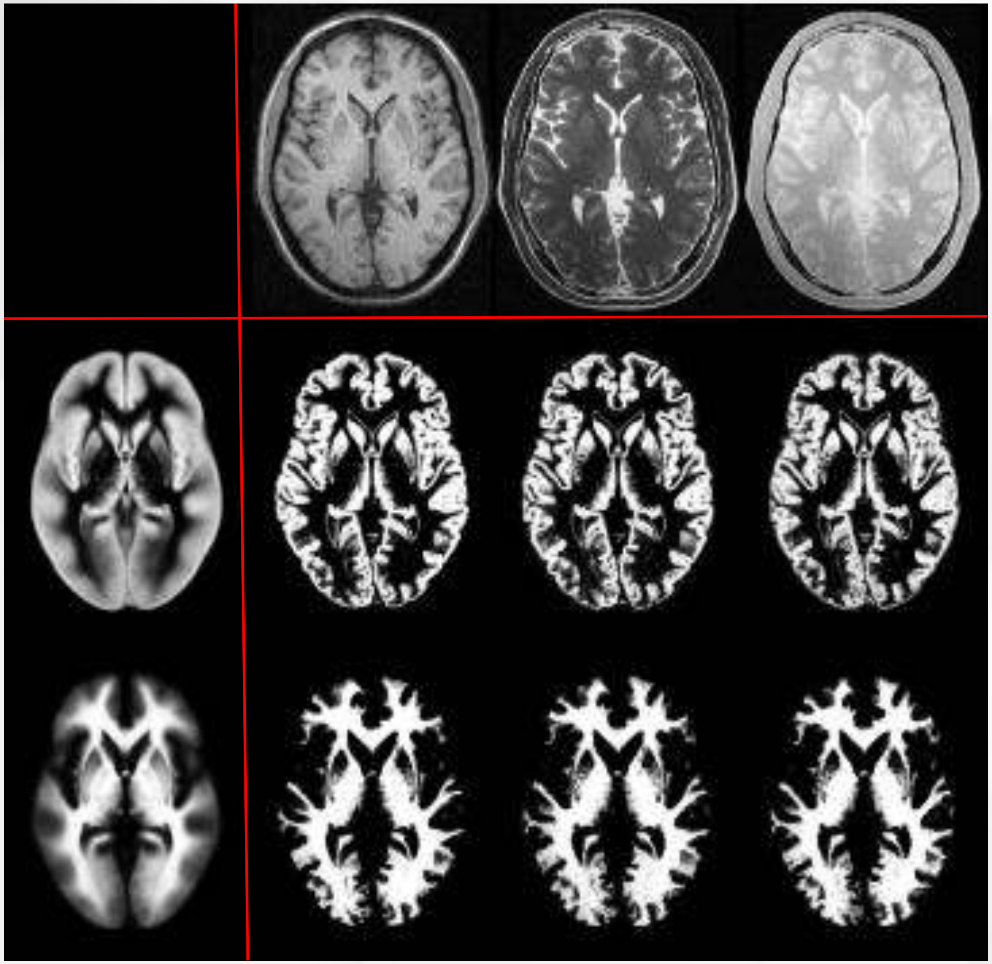
Deformation Fields



~1000 discrete cosine transforms

Segmentation results

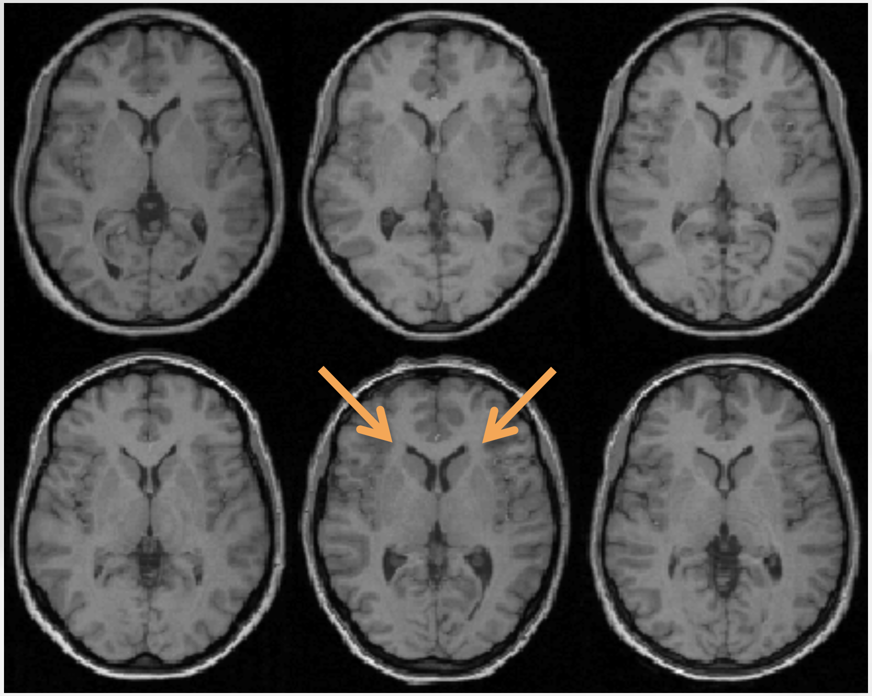
- Segmentation works irrespective of image contrast



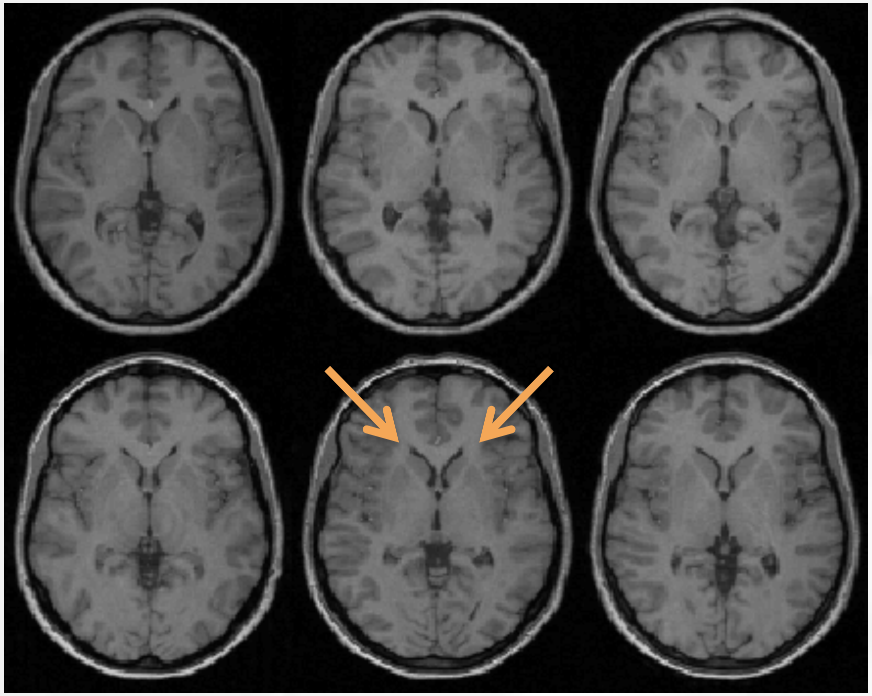
Cocosco, Kollokian, Kwan & Evans. "*BrainWeb: Online Interface to a 3D MRI Simulated Brain Database*". NeuroImage 5(4):S425 (1997)

Benefits of Unified Segmentation

Affine registration



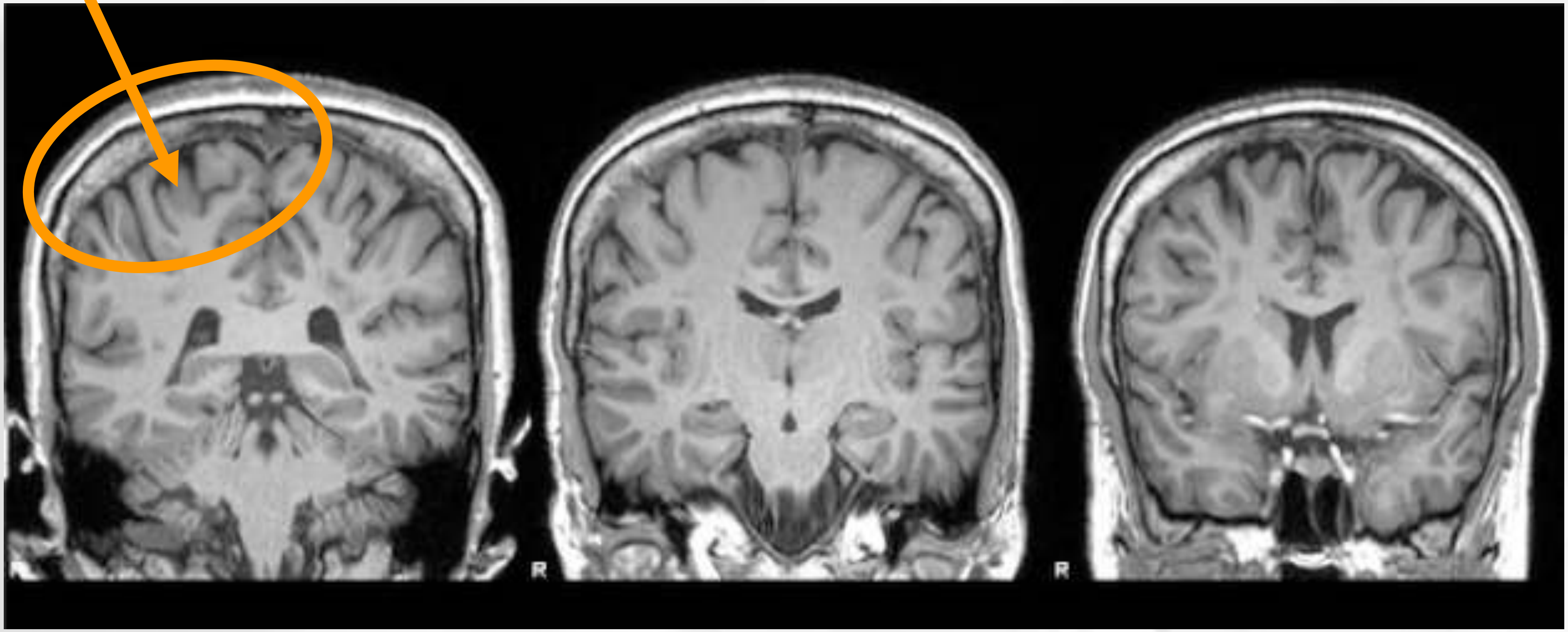
Non-linear registration

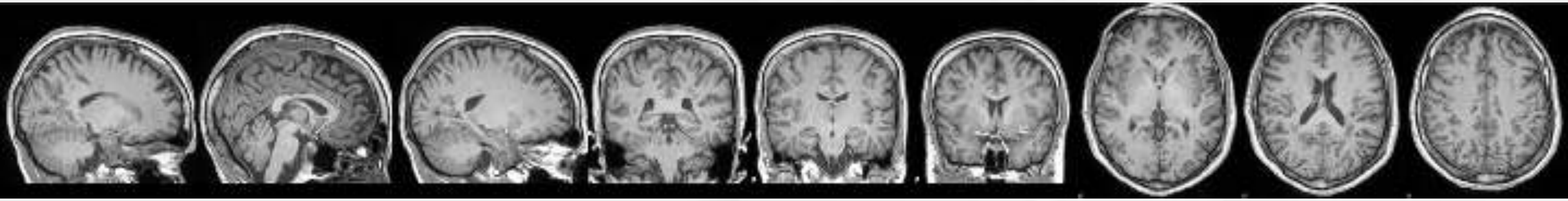


Spatial normalisation – Limitations

- Seek to match functionally homologous regions, but...
  - Challenging high-dimensional optimisation
    - many local optima
  - Different cortices **can** have different folding patterns
  - No exact match between structure and function
    - See e.g. Amiez et al. (2013), PMID:23365257

Spatial normalisation – Limitations





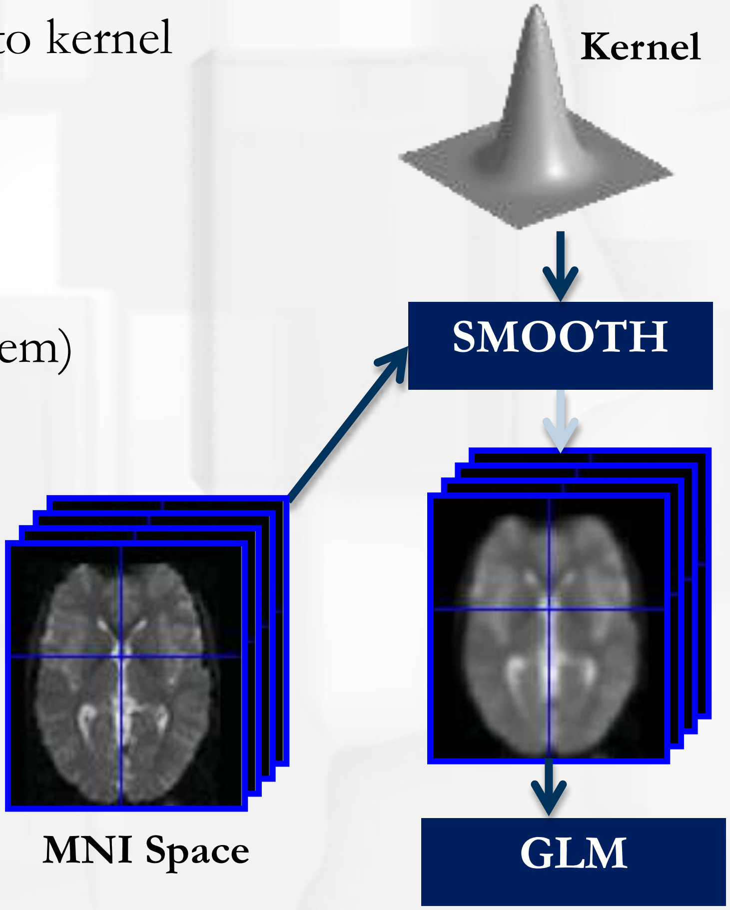
UK Biobank data: eventually, 100,000 subjects and 80 mins!

Spatial normalisation – Limitations

- Seek to match functionally homologous regions, but...
  - Challenging high-dimensional optimisation
    - many local optima
  - Different cortices **can** have different folding patterns
  - No exact match between structure and function
    - See e.g. Amiez et al. (2013), PMID:23365257
- Compromise
  - Correct relatively large-scale variability
  - Smooth over finer-scale residual differences

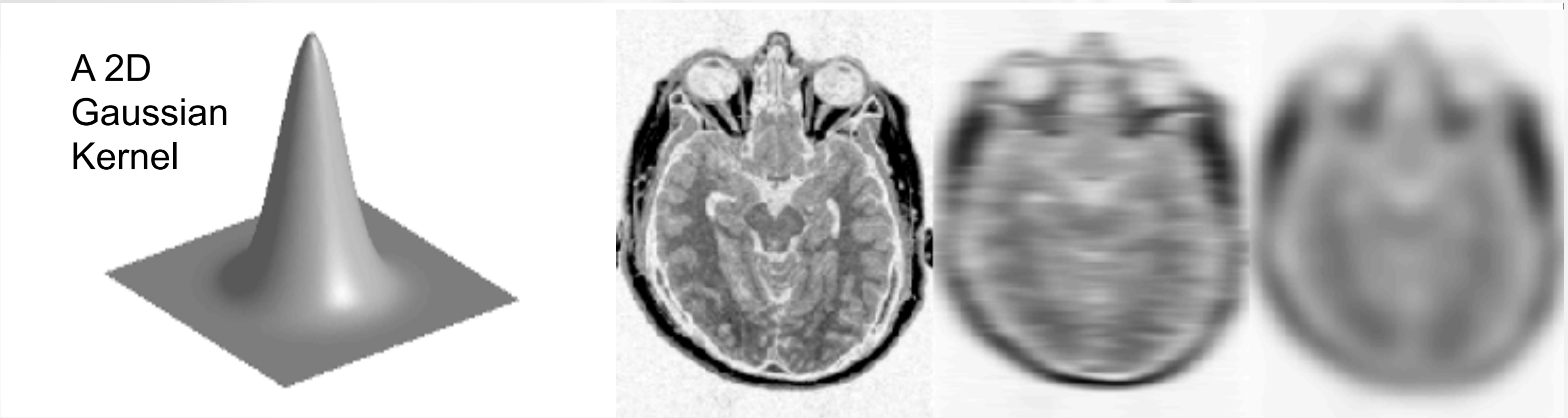
Smoothing – Why blurring the data?

- Intra-subject signal quality
  - Suppresses thermal noise (averaging)
  - Increases sensitivity to effects of similar scale to kernel (matched filter theorem)
- Single-subject statistical analysis
  - Makes data more Gaussian (central limit theorem)
  - Reduces the number of multiple comparisons
- Second-level statistical analysis
  - Improves spatial overlap by blurring anatomical differences



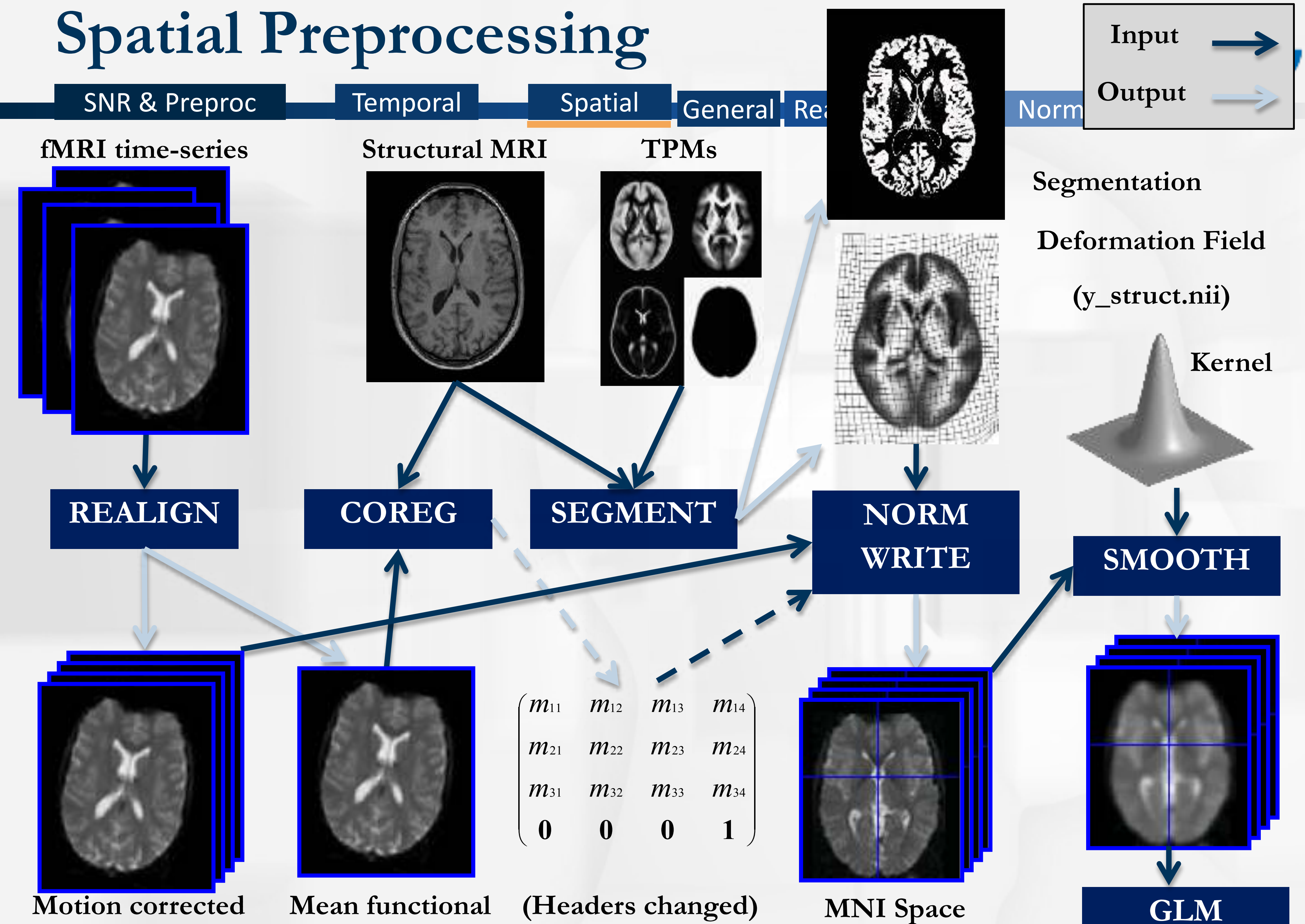
Smoothing – How is it implemented?

- Convolution with a 3D Gaussian kernel, of specified full-width at half-maximum (FWHM) in mm
  - Mathematically equivalent to slice-timing operation or reslicing, but different kernels there (sinc, b-spline)
- Gaussian kernel is separable, and we can smooth 2D data with 2 separate 1D convolutions



fMRI Run after Smoothing

Spatial Preprocessing

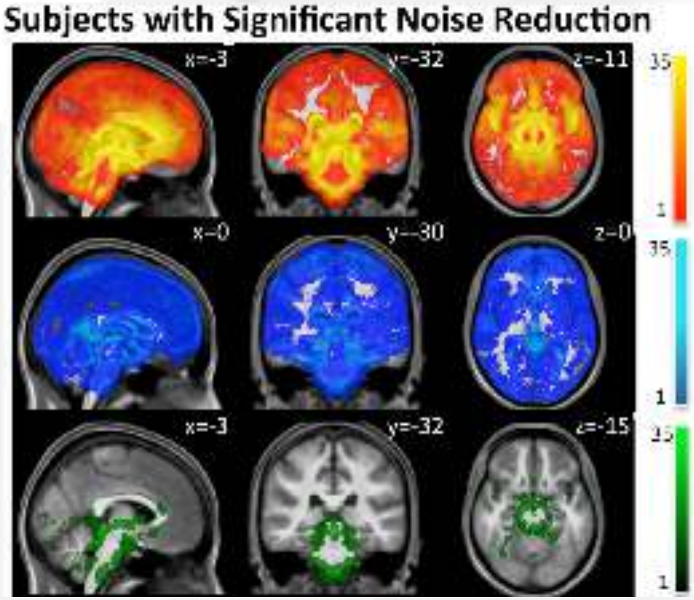


Sources of Noise in fMRI

Source	Preprocessing	Method
Acquisition Timing	Temporal Preproc	Slice-Timing
Subject Motion	Spatial Preproc	Realignment
Anatomical Identity	Spatial Preproc	Co-registration
Inter-subject variability	Spatial Preproc	Segmentation
Thermal Noise	Spatial Preproc	Smoothing
Physiological Noise	Noise Modeling	PhysIO Toolbox

Teaser: PhysIO Noise Modelling

- We can model time series of non-BOLD physiological fluctuations from prior knowledge (locations, dominant frequencies) or peripheral recordings (ECG, breathing belt)
- "Filter" these out via incorporation into general linear model
  - See the next two GLM lectures!
- Result:
  - Cardiac (red), respiratory (blue) physiological time courses, and their interaction (green) contribute severely to remaining non-Gaussian voxel fluctuations



- For more details: See Matthias' talk on **19th November…**

Thank you…

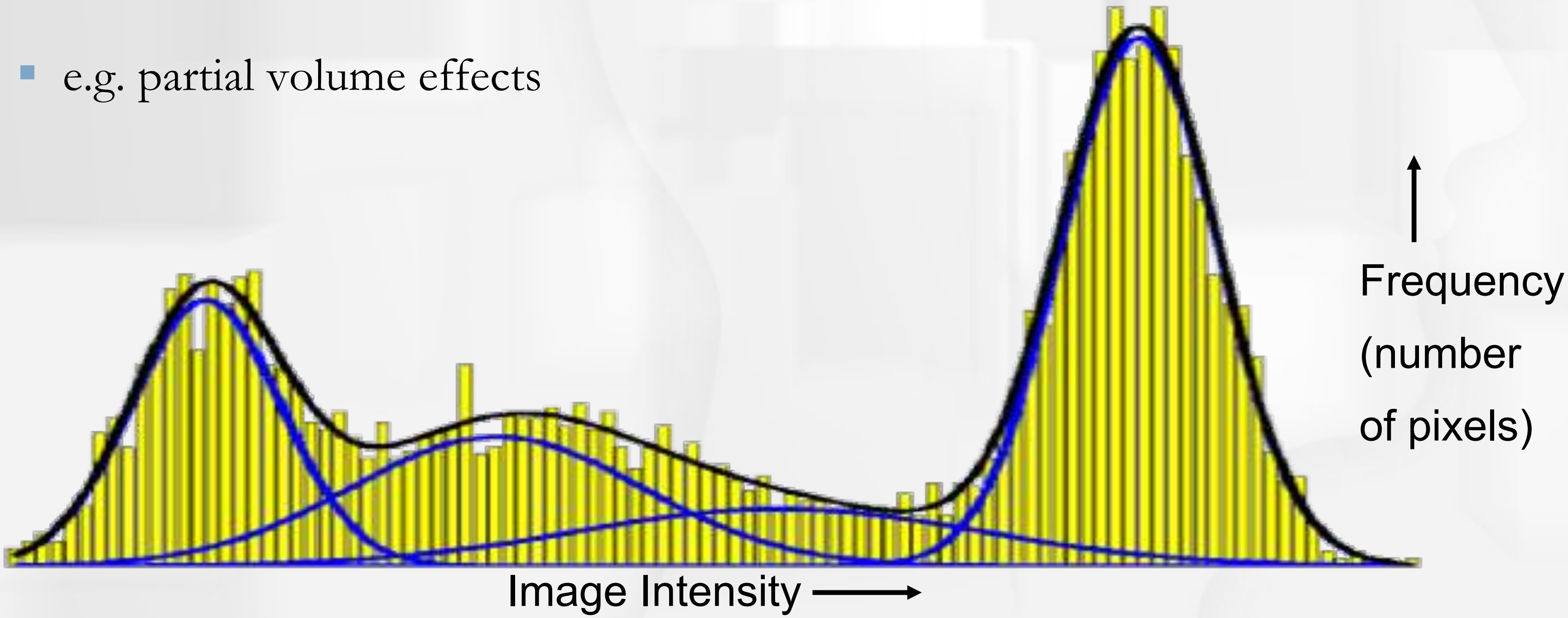
- …and:
  - TNU
  - Lars Kasper
  - Everyone Lars borrowed slides from ☺

Further Reading

- Good Textbook: Karl Friston, J.A., William Penny (Eds.), Statistical Parametric Mapping, Academic Press, London, in particular
  - Ashburner, J., Friston, K., 2007a. Chapter 4 - Rigid Body Registration, pp. 49–62.
  - Ashburner, J., Friston, K., 2007b. Chapter 5 - Non-linear Registration, pp. 63–80.
  - Ashburner, J., Friston, K., 2007c. Chapter 6 - Segmentation, pp. 81–91.
- For mathematical/engineering connoisseurs: (see also extra slides here):
  - Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851. doi:10.1016/j.neuroimage.2005.02.018

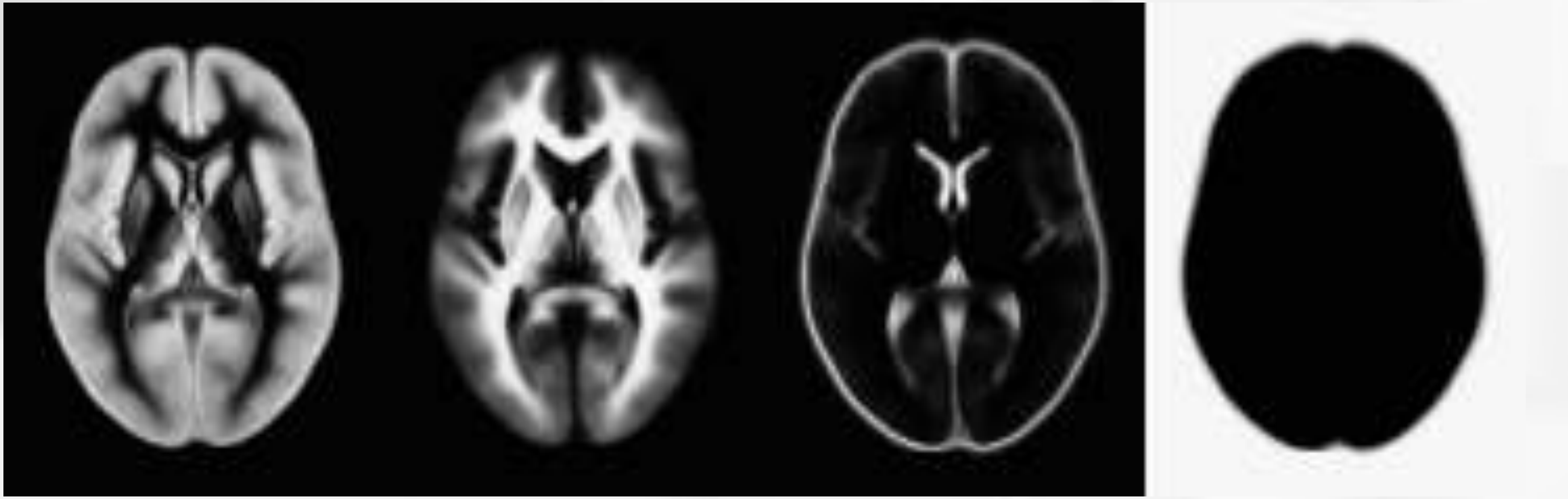
Mixture of Gaussians

- Classification is based on a Mixture of Gaussians model, which represents the intensity probability density by a number of Gaussian distributions.
- Multiple Gaussians per tissue class allow non-Gaussian intensity distributions to be modelled
  - e.g. partial volume effects



Tissue Probability Maps

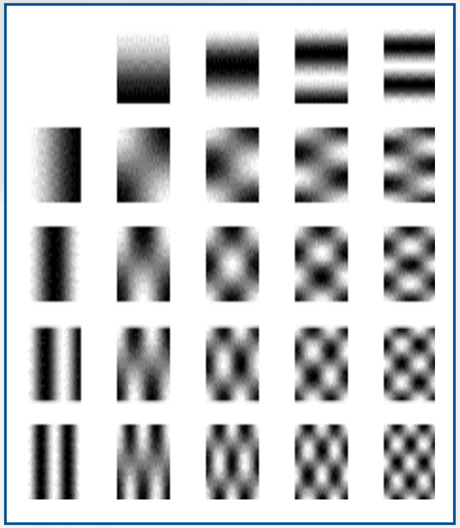
- Tissue probability maps (TPMs) are used as the prior, instead of the proportion of voxels in each class

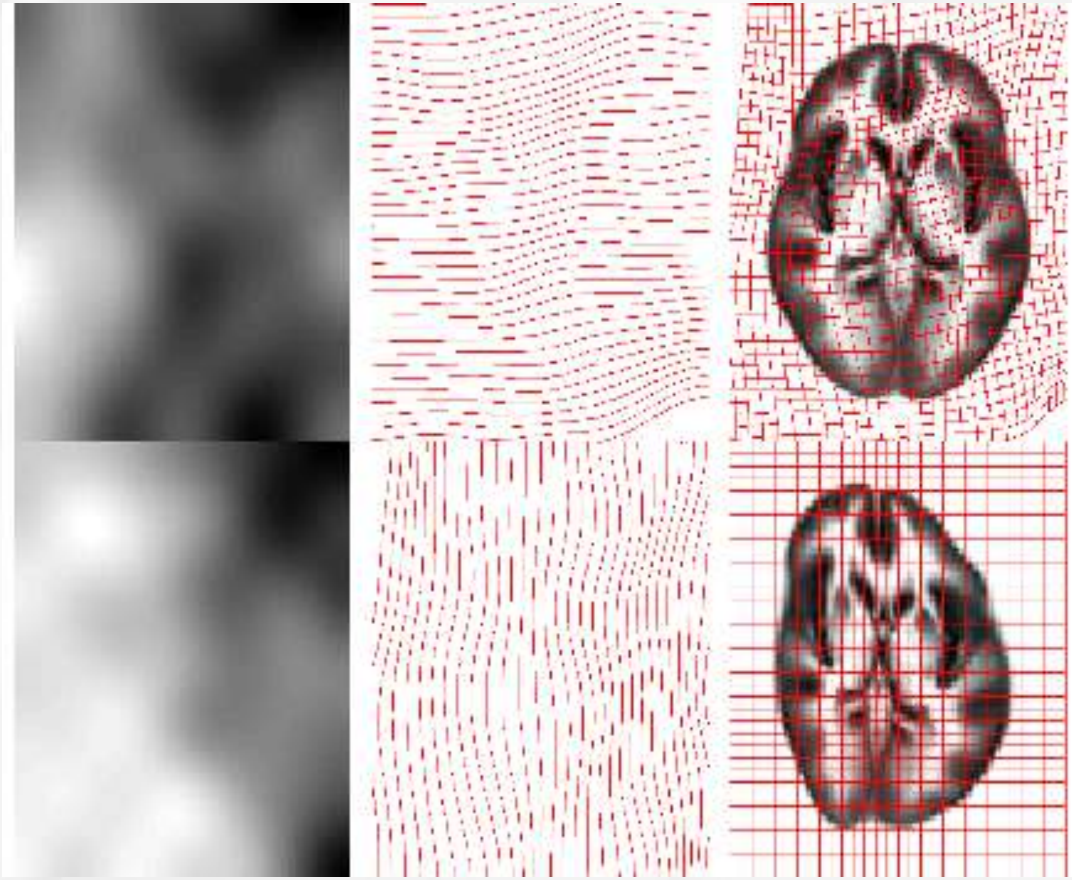


ICBM Tissue Probabilistic Atlases. These tissue probability maps were kindly provided by the **International Consortium for Brain Mapping**

Deforming the Tissue Probability Maps

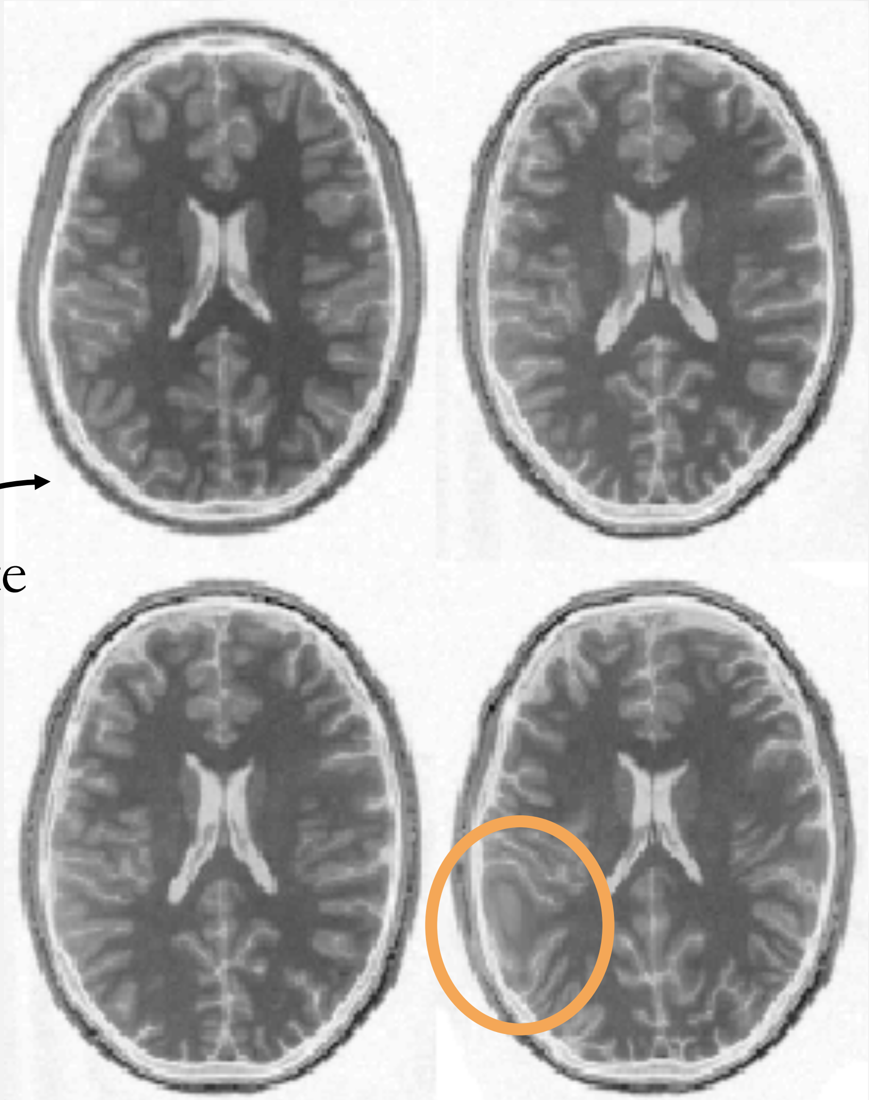
- Tissue probability maps images are warped to match the subject
- The inverse transform warps to the TPMs





Why regularisation? – Overfitting

- Regularisation constrains deformations to realistic range (implemented as priors)



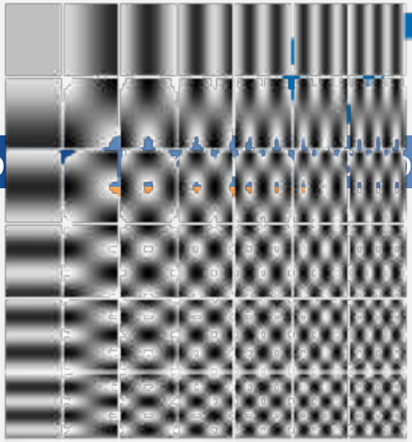
Affine registration (error = 472.1)

Template image

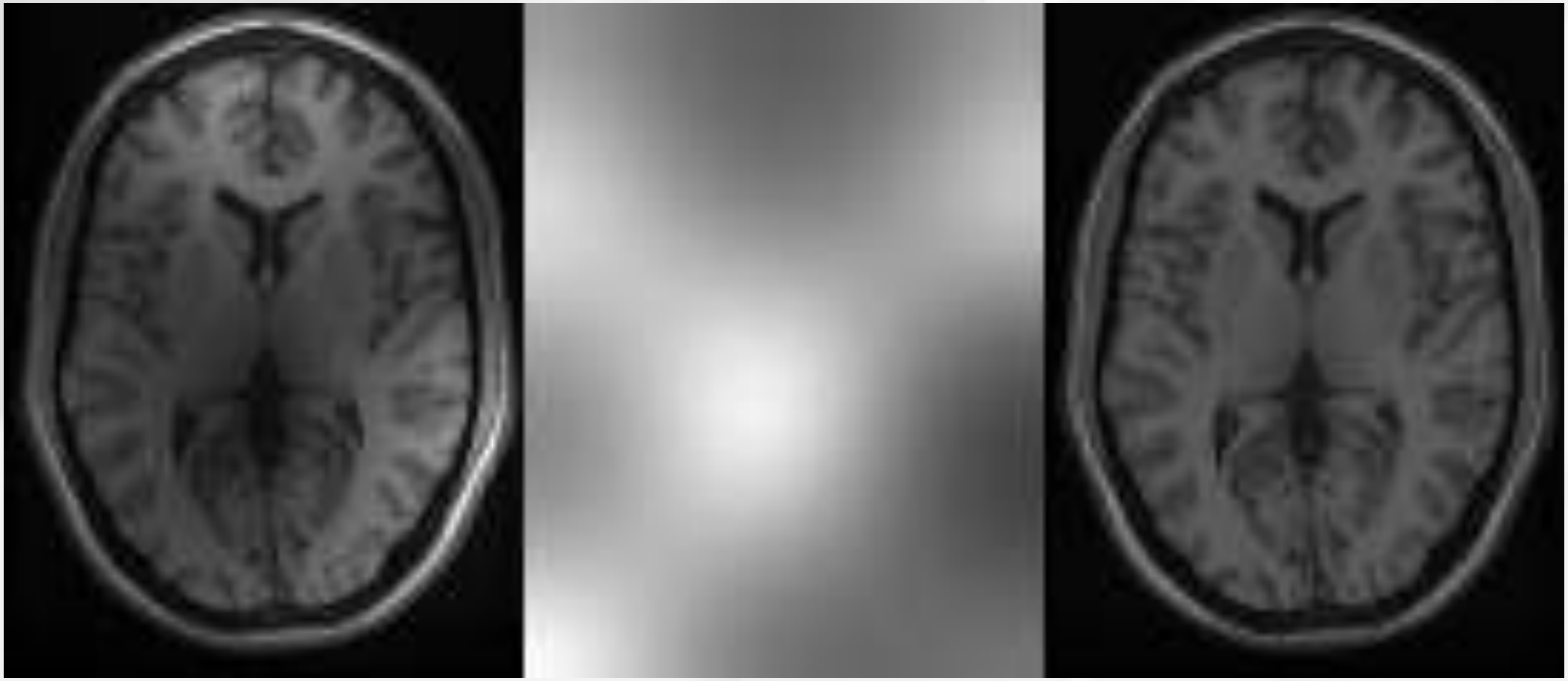
Non-linear registration using regularisation (error = 302.7)

Non-linear registration without regularisation (error = 287.3)

Modelling inhomogeneity



- A multiplicative bias field is modelled as a linear combination of basis functions.



Corrupted image **Bias Field** **Corrected image**

Unified segmentation: The maths

- Mixture of Gaussians: probability of voxel i having intensity y_i, given it is from a specific cluster k (e.g. tissue class gray matter)

$$P(y_i|c_i = k, \mu_k, \sigma_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (1)$$

  - Prior probability of voxel's tissue class (e.g. voxel proportion) γ_k

$$P(c_i = k|\gamma_k) = \gamma_k$$

  - Joint Probability: $P(y_i, c_i = k|\mu_k, \sigma_k, \gamma_k) = P(y_i|c_i = k, \mu_k, \sigma_k) P(c_i = k|\gamma_k)$
  - Marginal probability of voxel intensity:

$$P(y_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \sum_{k=1}^{K} P(y_i, c_i = k|\mu_k, \sigma_k, \gamma_k)$$

  - Joint probability all voxels' intensity:

$$P(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \prod_{i=1}^{I} P(y_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \prod_{i=1}^{I}\left(\sum_{k=1}^{K} \frac{\gamma_k}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2}\right)\right) \quad (5)$$

US Maths: Bias Field

- Implemented by adjusting the Means and Variances of the Gaussians on a pixel-by-pixel basis by a function smoothly varying in space, $\rho_i(\boldsymbol{\beta})$:
  - $\mu_k \mapsto \frac{\mu_k}{\rho_i(\boldsymbol{\beta})},\ \sigma_k^2 \mapsto \left(\frac{\sigma_k}{\rho_i(\boldsymbol{\beta})}\right)^2$
  - ρ_i is the exponential of a linear combination of low frequency basis functions
  - Parameters to be estimated: vector $\boldsymbol{\beta}$
- intensity probability conditioned on cluster identity:

$$P(y_i|c_i = k, \mu_k, \sigma_k, \beta)$$
$$= \frac{1}{\left(2\pi(\sigma_k/\rho_i(\beta))^2\right)^{\frac{1}{2}}} \exp\left(-\frac{(y_i - \mu_k/\rho_i(\beta))^2}{2(\sigma_k/\rho_i(\beta))^2}\right)$$
$$= \rho_i(\beta)\frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \exp\left(-\frac{(\rho_i(\beta)y_i - \mu_k)^2}{2\sigma_k^2}\right)$$

US Maths: Spatial Priors by TPMs

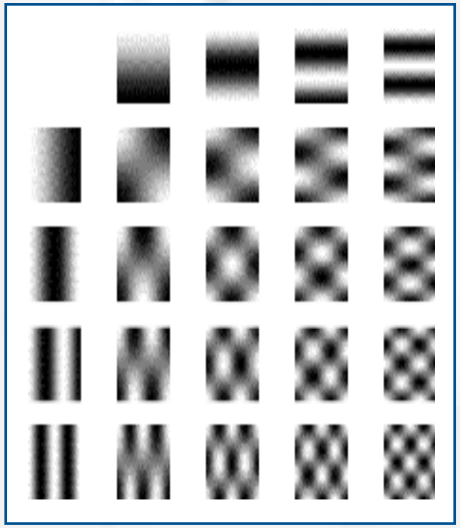
- Replacing stationary mixing proportions γ_k by voxel-dependent proportions which are informed by the prior tissue probabilities b_{ik} for this voxel i and different tissue types k

- $\gamma_k \mapsto \gamma_k(i) = \gamma_k \cdot \frac{b_{ik}}{\sum_{j=1}^{K} \gamma_j b_{ij}}$

- Note: K can be larger than the number of tissue classes, since each class can be reflected by a mixture of Gaussians, e.g. 3 Gaussians for gray matter (to allow for non-Gaussian distributions per tissue class)
  - E.g. partial volume effects

US Maths: Deformation Fields

- Deformation (and thereby normalisation) is implemented by allowing the prior TPMs (which are in MNI-space) to be spatially transformed by a parameterised mapping
  - $\mathrm{b}_{ik} \mapsto \mathrm{b}_{ik}(\alpha) \Rightarrow P(c_i = k|\gamma,\alpha) = \frac{\gamma_k b_{ik}(\alpha)}{\sum_{j=0}^{K} \gamma_j b_{ij}(\alpha)}$
  - Parameter vector to be estimated: $\boldsymbol{\alpha}$
  - about 1000 discrete cosine transforms



US Maths: Regularisation

SNR & Preproc | Temporal | Spatial | General | Realign | Coreg | Normalise | Smooth

- Linear Regularisation of Bias Field and Deformation Field Estimates
  - By including prior distributions for α and β as zero-mean multivariate Gaussians
  - Covariance: $\alpha^T C_\alpha \alpha = bending\ energy; \rho(\beta) = \exp(K_{70mm} * N(0,\beta))$
- Thus, the final objective function to be maximised is the log-joint probability of intensity, bias and deformation field parameters:

$$P(\mathrm{y}, \beta, \alpha | \gamma, \mu, \sigma^2) = P(\mathrm{y} | \beta, \alpha, \gamma, \mu, \sigma^2) P(\beta) P(\alpha)$$

- Equivalently, the negative free energy is minimised:

$$\mathcal{F} = -\log P(\mathrm{y}, \beta, \alpha | \gamma, \mu, \sigma^2) = \mathcal{E} - \log P(\beta) - \log P(\alpha)$$

$$\mathcal{E} = -\sum_{i=1}^{I} \log\left(\frac{\rho_i(\beta)}{\sum_{k=1}^{K} \gamma_k b_{ik}(\alpha)} \sum_{k=1}^{K} \gamma_k b_{ik}(\alpha) (2\pi\sigma_k^2)^{-\frac{1}{2}} \times \exp\left(-\frac{(\rho_i(\beta) y_i - \mu_k)^2}{2\sigma_k^2}\right)\right)$$