MRI Data Processing

Course on signal and image processing (FFYS7086)

Harri Merisaari

haanme@utu.fi

Department of Diagnostic Radiology, Turku Brain and Mind Center, University of Turku

Adjunct Professor in Medical Image Analysis

MRC

TBMC Turku Brain and Mind Center

UNIVERSITY OF TURKU

Learning goals

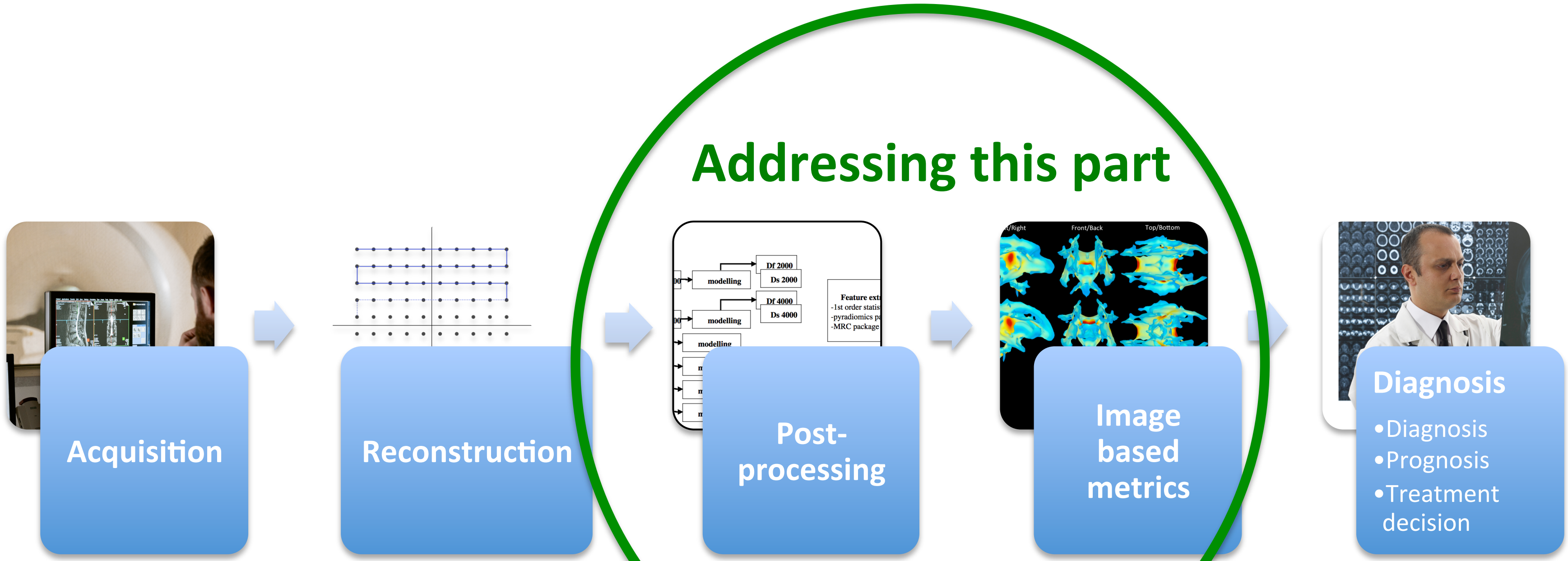
Basic understanding in:

- How MRI data is organized
- Goals of MR image processing
- Concepts of common MRI data processing steps

In zoom session, please leave questions/comments to the chat, I will answer them at the end of the talk

- Tool(s) useful for each particular processing step, are marked with ***blue bold font***

What is post-processing and why it is important?



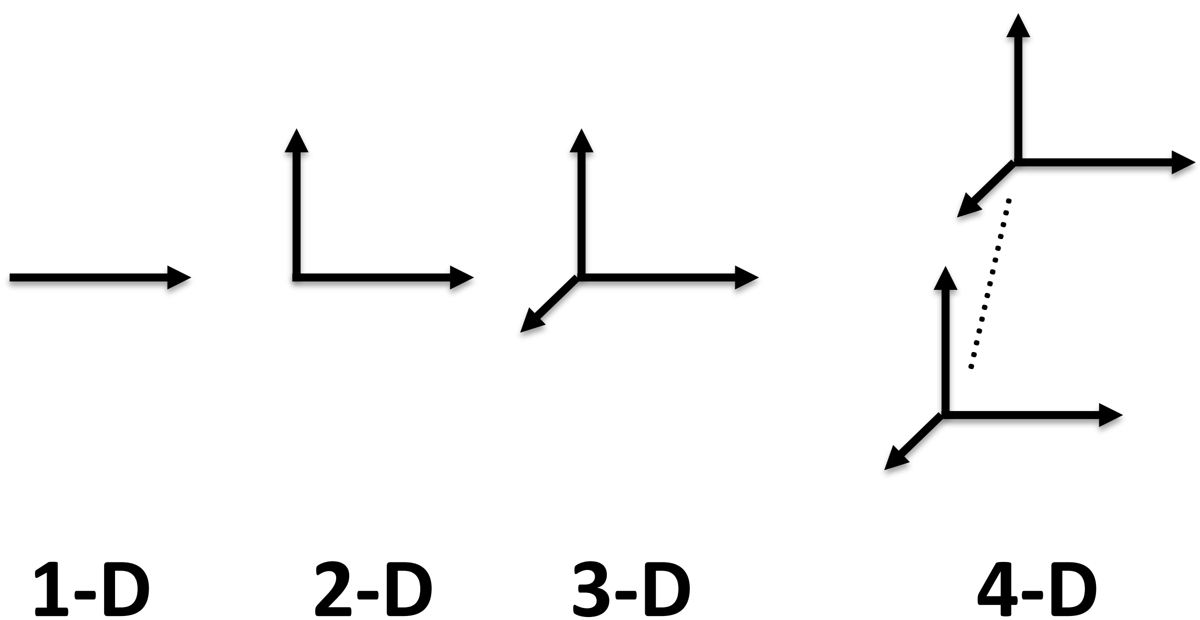
Goals of post-processing:

- Assure data quality
- Salvage otherwise unusable or dubious data
- Apply model fitting to obtain their specific parameters
- Compose information not available with mere manual inspection

Outline

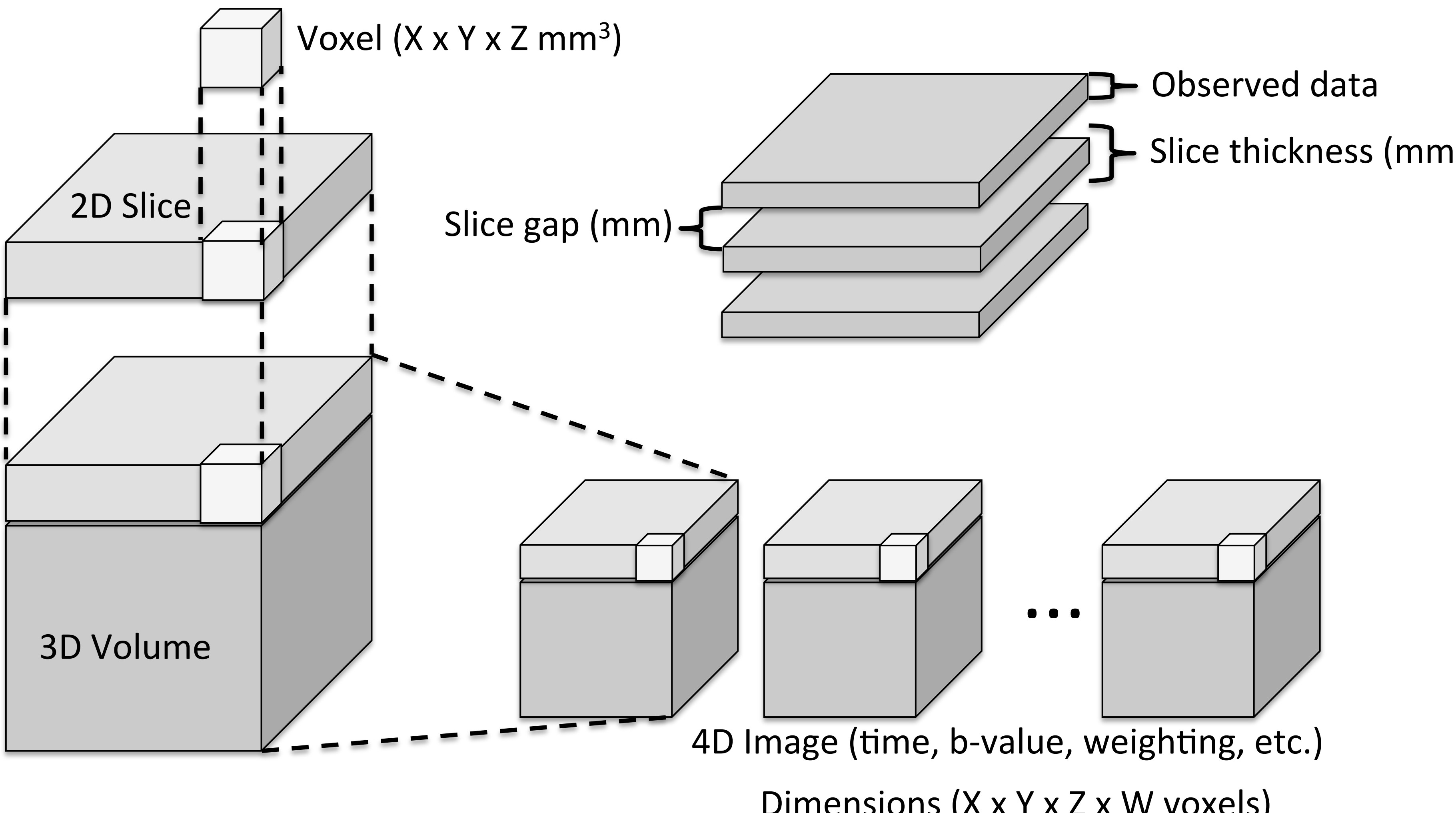
1 Data organization
2 MR Data modalities
3 Image formats
4 Quality Assurance
5 Image processing steps
6 Radiomics
7 Artificial Intelligence

1 Data organization



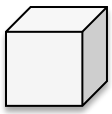
Data organization: Components

Medical imaging data is multi-dimensional signal



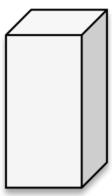
Data: Continued

Isotropic voxel



Voxel (X x Y x Z mm^3)

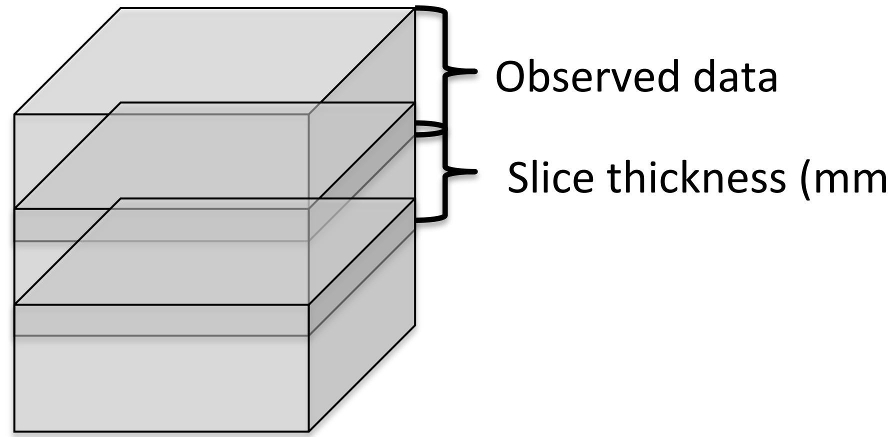
Anisotropic voxel



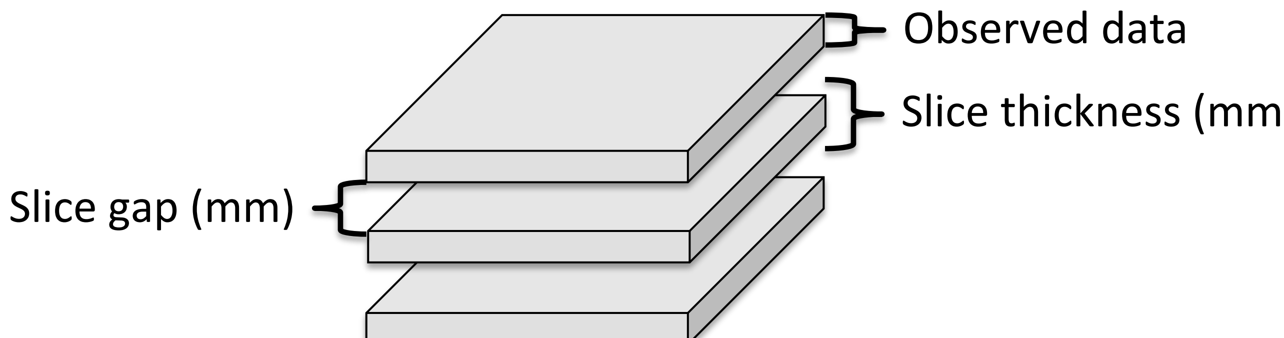
Voxel (X x Y x Z mm^3)

- Voxel shape and sampling is important when analysis considers information in 3D
- Slice gap and anisotropic voxel cause extra data loss when image is rotated

Oversampled acquisition

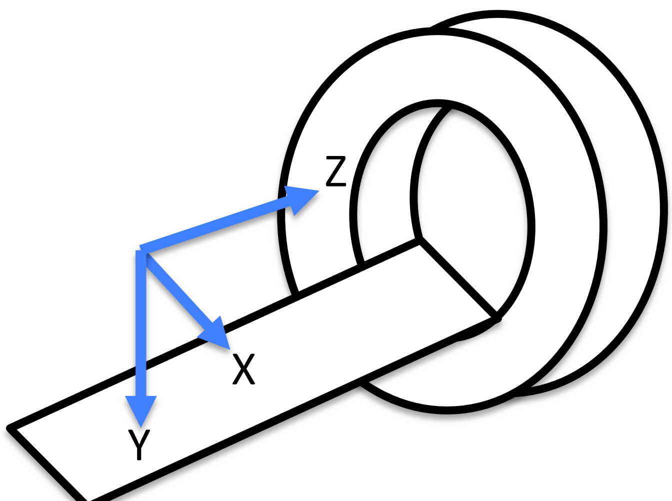


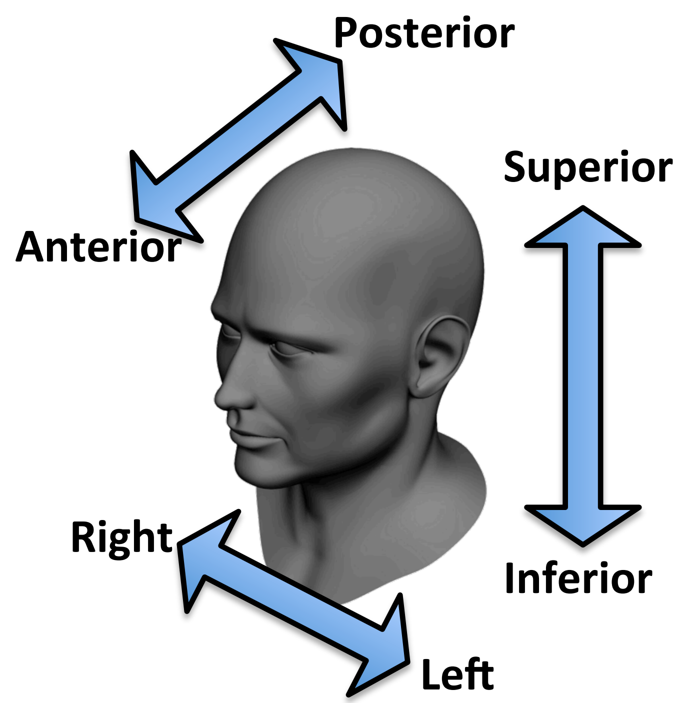
Undersampled acquisition



Data: Orientation

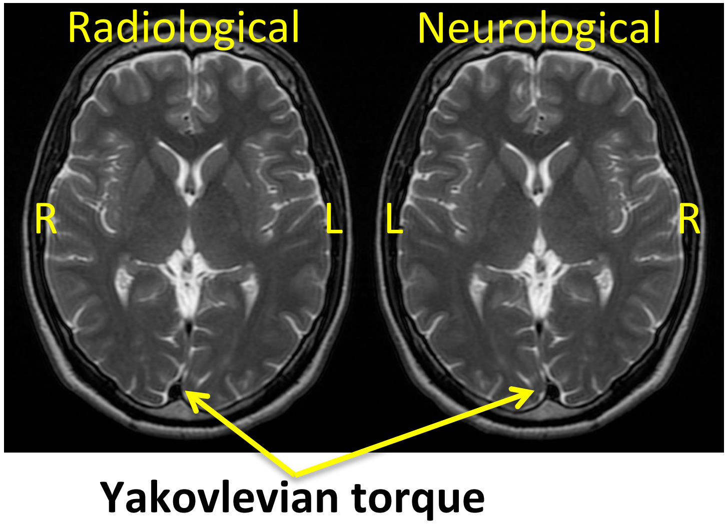
- Image data is collected in **transaxial/coronal/sagittal** slices in respective to the scanner
- In addition, subject may be positioned in various orientations into the scanner, depending on the acquisition procedure





Data: Orientation

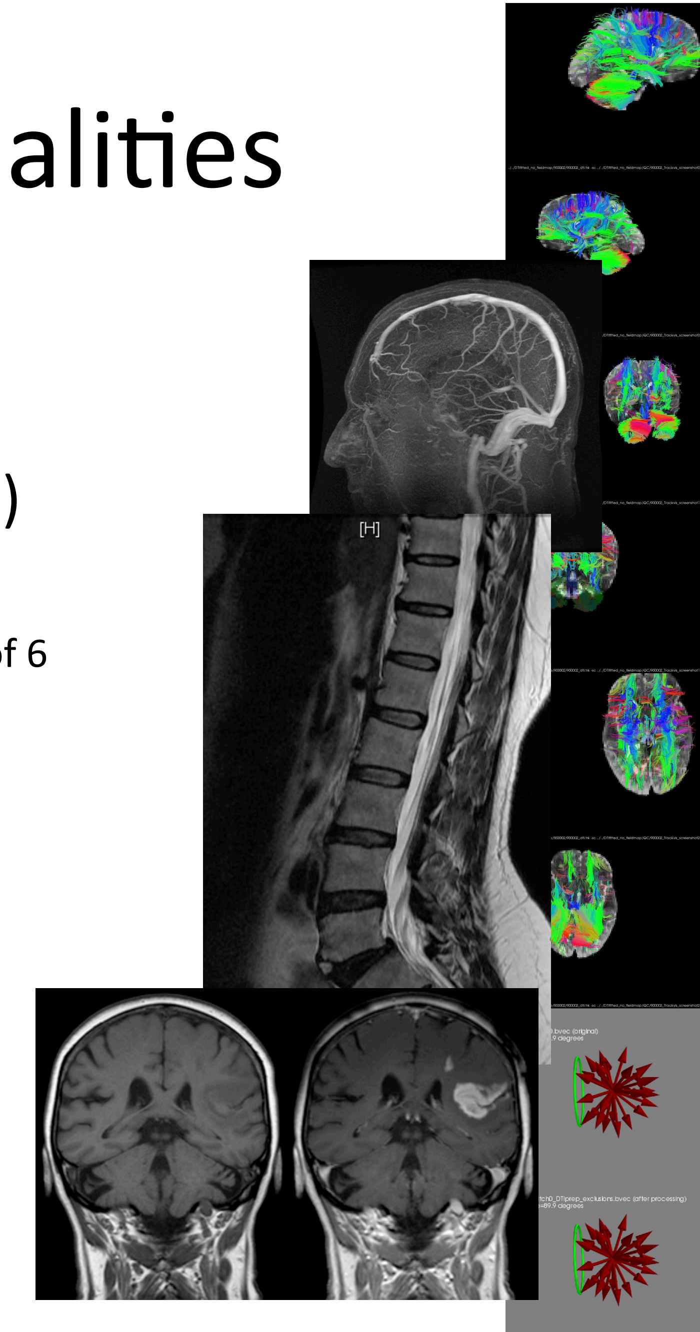
- Visualization conventions:
  - Radiological convention: Subject right is on the left
  - Neurological convention: Subject right is on the right
- How do I know which way is up/left/front in the image?
  - localize anatomical landmarks (e g Yakovlevian torque, heart etc.)
  - consult data (e g DICOM tags (0018,5100), (0020,0020), (0020,0037))
  - localize stereotaxic marker (e g fish oil capsule)
  - consult person responsible for acquisition



2 MR Data modalities

MR Image modalities

- **T1W** (3D), T1 relaxation time
  - gadolidium contrast agent (paramagnetic)
- **T2W** (3D), T2 relaxation time
- **DWI**, Diffusion Weighted Imaging (4D)
- **DTI**, Diffusion Tensor Imaging (4D)
  - DWI with diffusion encoding directions, minimum of 6
- Others
  - fMRI (BOLD signal)
  - FLAIR (FLuid Attenuation Inversion Recovery)
  - Contrast enhanced images (injected contrast agent enhances signal, 3D/4D)
- <u>**Non-invasive** (apart from contrast agent)</u>
  - Comparison to X-ray, CT, PET, SPECT no radiation
  - Magnetic field may limit use



3 Image formats

Group - Element	Description	Type	Length	Value
0002 0000	Group 0002 Length	UL	4	200
0002 0001	File Meta Information Version	OB	2	(binary data)
0002 0002	Media Storage SOP Class UID	UI	26	1.2.840.10008.5.1.4.1.1.7
0002 0003	Media Storage SOP Instance UID	UI	54	1.2.826.0.1.3680043.2.1208.34948545565201041914916296
0002 0010	Transfer Syntax UID	UI	18	1.2.840.10008.1.2
0002 0012	Implementation Class UID	UI	18	1.2.804.114118.3
0002 0013	Implementation Version Name	SH	6	
0002 0016	Source Application Entity Title	AE	18	
0002 0100	Private Information Creator UID	UI	0	(empty)
0002 0102	Private Information	OB	0	(binary data)
0008 0005	Specific Character Set	CS	12	ISO_IR 100
0008 0008	Image Type	CS	24	ORIGINAL\PRIMARY\OTHER
0008 0012	Instance Creation Date	DA	10	20100419
0008 0013	Instance Creation Time	TM	14	140916.000000
0008 0016	SOP Class UID	UI	26	1.2.840.10008.5.1.4.1.1.7
0008 0018	SOP Instance UID	UI	54	1.2.826.0.1.3680043.2.1208.34948545565201041914916296
0008 0020	Study Date	DA	10	20100419
0008 0021	Series Date	DA	10	20100419
0008 0030	Study Time	TM	14	140916.000000
0008 0031	Series Time	TM	14	140916.000000
0008 0032	Acquisition Time	TM	14	140916.000000
0008 0033	Content Time	TM	14	140916.000000
0008 0050	Accession Number	SH	2	
0008 0060	Modality	CS	4	OT
0008 0064	Conversion Type	CS	4	WSD
0008 0070	Manufacturer	LO	4	IIS
0008 0080	Institution Name	LO	2	
0008 0090	Referring Physician's Name	PN	2	

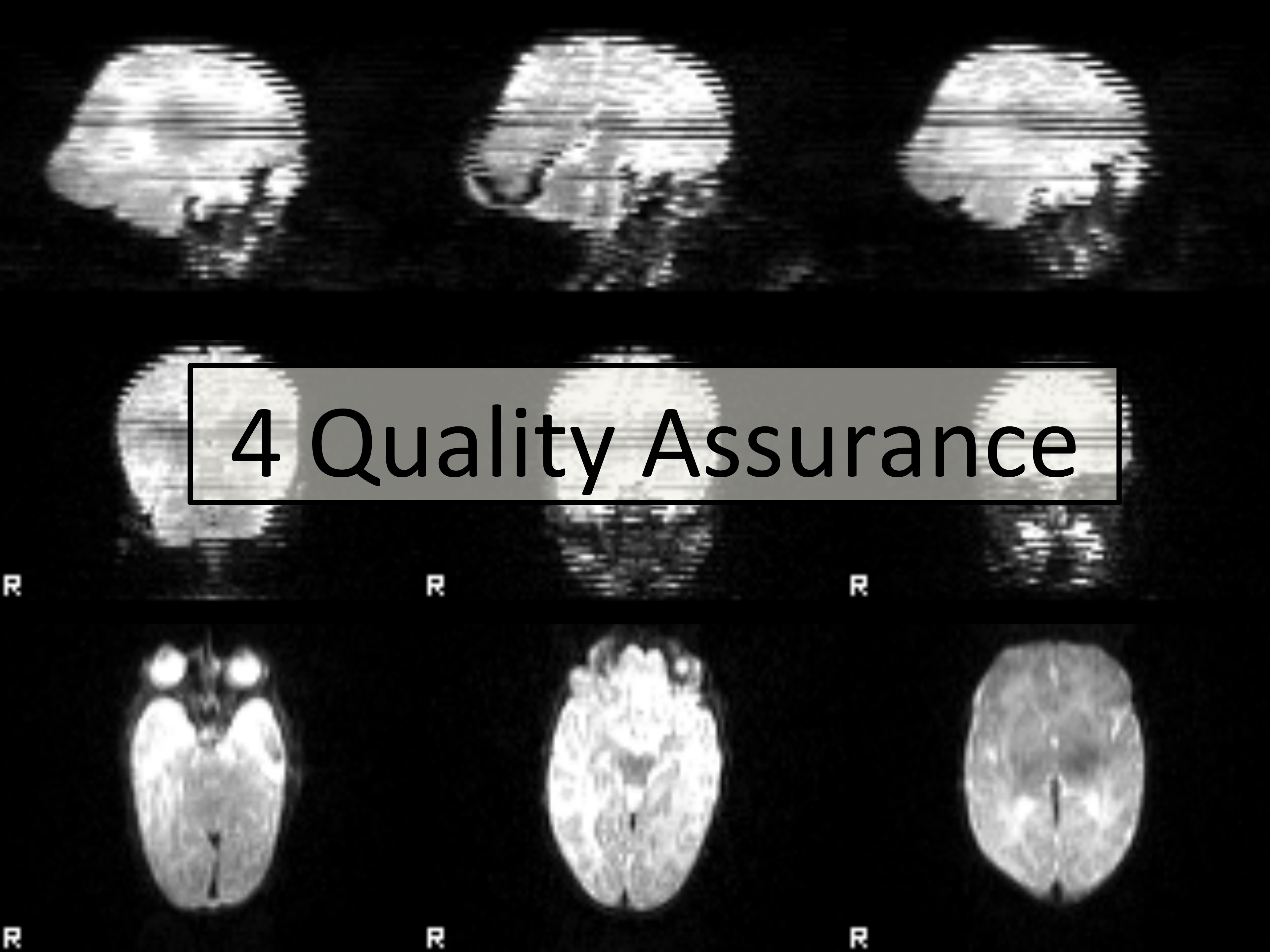
Image data formats

- DICOM (Digital Imaging and Communications in Medicine, ***pydicom, matlab***)
  - Most common image format in hospitals
  - Used in both storing and transferring image data
  - Data stored is commonly as one file/slice
  - Original image data acquired at the MR scanner is usually DICOM, possible that some data loss occur in conversions
- Nifti (***nibabel, matlab***, all main visualization tools support this)
  - Convenient for image processing
  - One file per 3D/4D image
- Others: mnc (***MINC toolkit***), vtk (***visualization toolkit***), nrrd, mhd (***Slicer3D***)

Summary

- Most important things to understand before analysis:
  - What is the modality
  - Isotropic or anisotropic voxel, spatial resolution
  - Which way is subject <u>left/right</u>, anterior/posterior

4 Quality Assurance



Quality Assurance

- Quality Assurance composes of automatic or semi-automatic methods to verify that the data is valid **<u>for particular use</u>**
  - While target region ultimately determine if the data is acceptable, artefacts outside it may indicate that the target region may be compromized
- Image Quality consists of
  - Spatial resolution
  - Image Contrast
  - Signal to Noise Ratio (SNR)
  - Artefacts
    - Motion
    - Susceptibility
    - Inhomogeneity
    - Other
    - Please see https://radiopaedia.org/articles/mri-artifacts-1 for more exhaustive list

Quality Assurance

- Quality Assurance methods
  - Visual inspection for artefacts (by one or more persons)
  - Sanity checks for data integrity (data organization, reference region)
  - Noise profile
  - Artefact detection (automatic/manual)
- Things to consider in addition to image quality
  - Fit for particular use (e g inclinical application)
  - Repeatability (needs specific measurements, with test-retest setting)
  - Reproducibility (needs specific measurements, e g with phantom)
- MRI scan is a compromise between scan time and quality
  - In practice we need to compromise
- Image quality is not good enough, then what?
  - Some problems can be addressed afterwards, others not

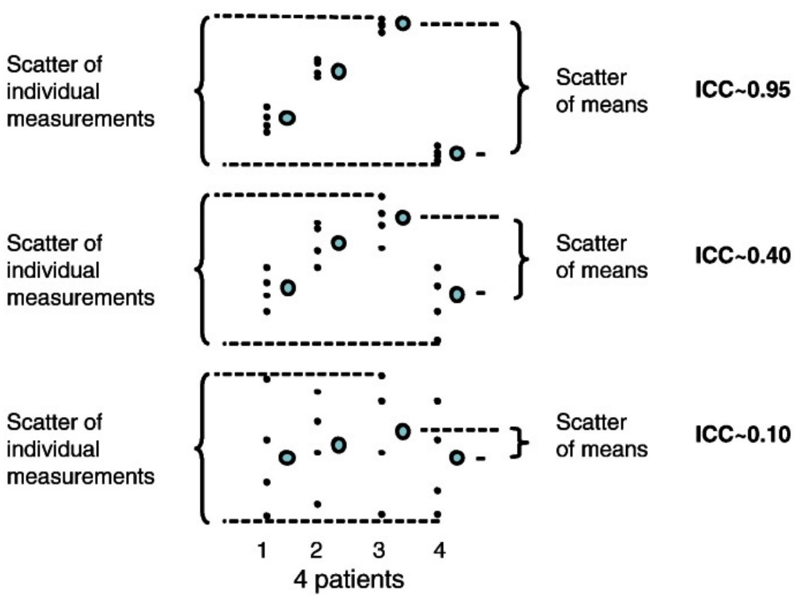
Quality Assurance:Tools

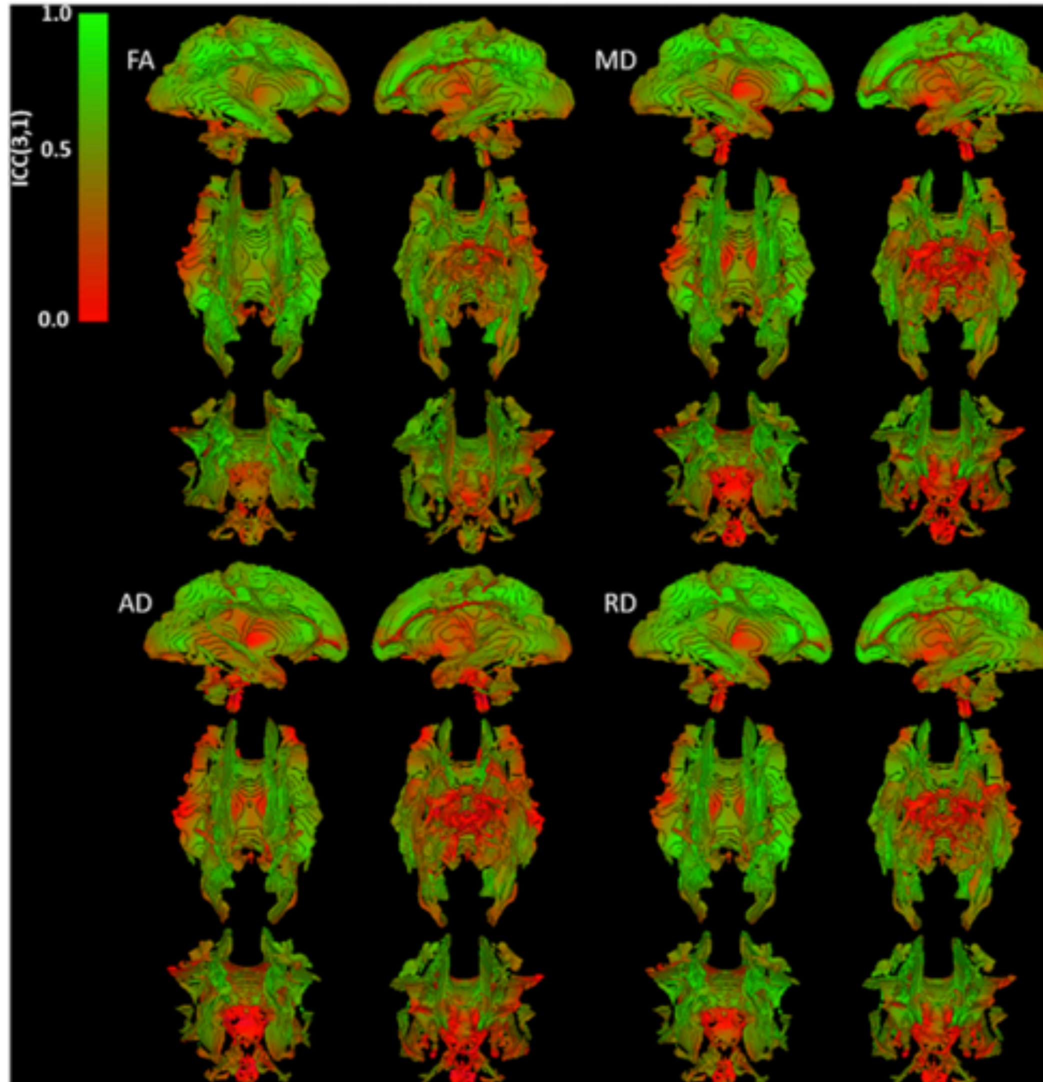
- Vendor (Siemens, Philips, GE, etc) specific visualization at workstation (DICOM, 3D/4D)
  - limited scalability
  - fit for particular use
- ***fsleyes*** (fsl toolbox's visualization tool Nifti, 3D)
  - part of fsl tools (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki), designed mainly for brain
  - fusion imaging, support for DTI
- ***eddyQC***(fsl quality control tool for DTI)
  - creates quality control report for DTI data
- ***MRIcron*** (Nifti, 3D)
- ***Slicer3D*** (DICOM, Nifti, 3D/4D)
  - most extensive list of features
  - learning curve more steep
- ***ITK-SNAP*** (DICOM, Nifti, 3D/4D)
  - light-weight viewer opening almost any image
- ***DTIprep*** (nrrd)
  - DTI specific quality assurance tools
- ***DSIstudio***
  - DTI specific visualization, contains tracktography pipeline
- ***Carimas*** (developed at Turku PET Centre)
  - PET specific tool can be used with MRI data as well

In addition to QC of individual image: image quality at group level

Intraclass Correlation coefficient (ICC)

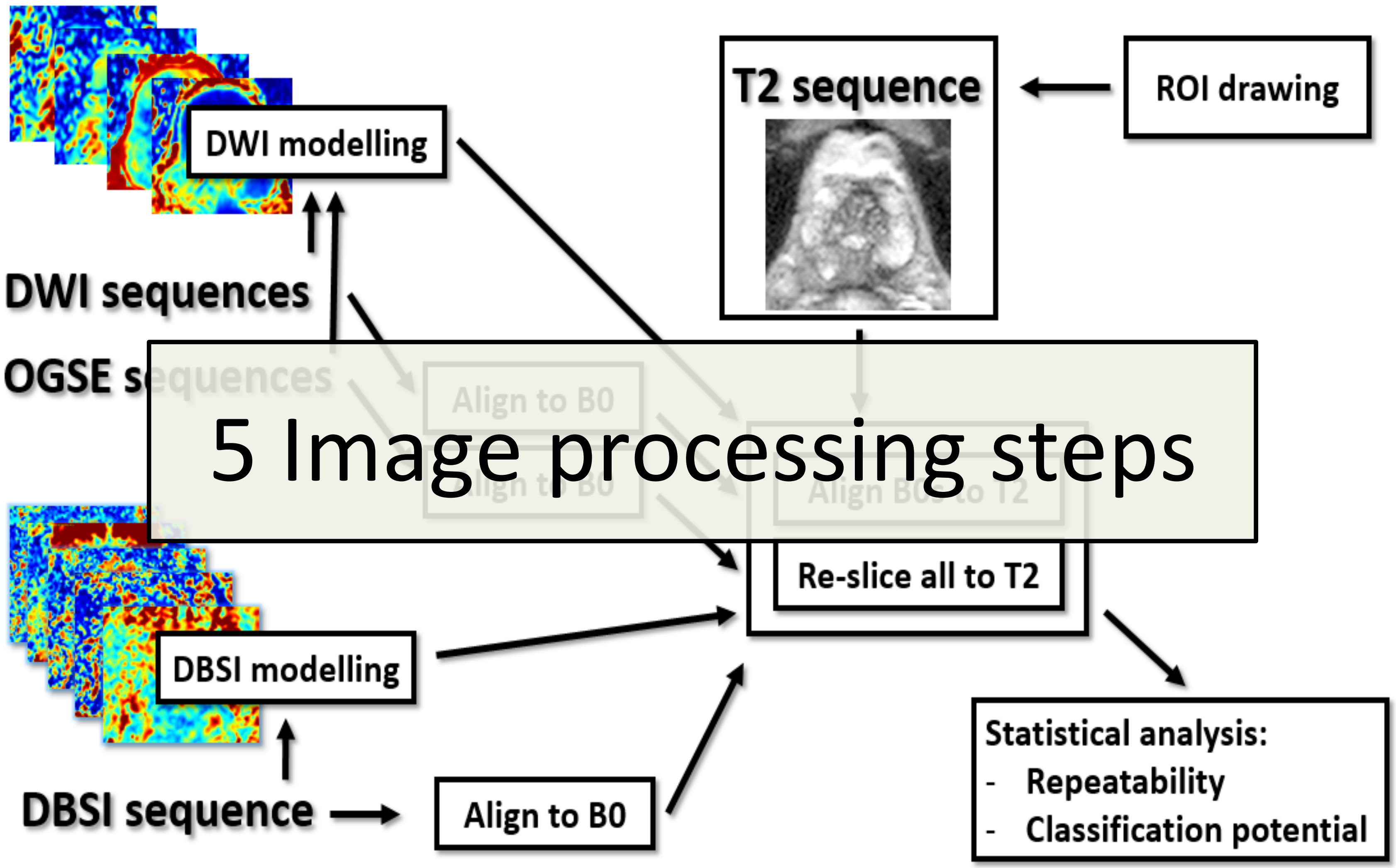
Contrast between measurement variance and difference between inividuals





Summary

- Early detection of problems is less costly
  - In best scenario, the acquisition can be repeated
  - In worst scenario, diagnosis quality is compromised
- When doing quality control, things to look for:
  - Missing data
  - Artefacts
    - Correct if possible
  - Signal-to-Noise ratio
  - Is a data quality problem occurring at random?
    - e g patients may express more motion artefact than healthy > measurement is about artefact rather than actual target tissue



5 Image processing steps
DWI modelling
DWI sequences
OGSE sequences
Align to B0
T2 sequence
ROI drawing
Re-slice all to T2
DBSI modelling
DBSI sequence
Align to B0
Statistical analysis:
- Repeatability
- Classification potential

Where to process the data

- In practise, large data/processing is run on servers
- Standardized image visualization may need specific monitors
- The trend is towards OS independent processing
- Patient anonymity!

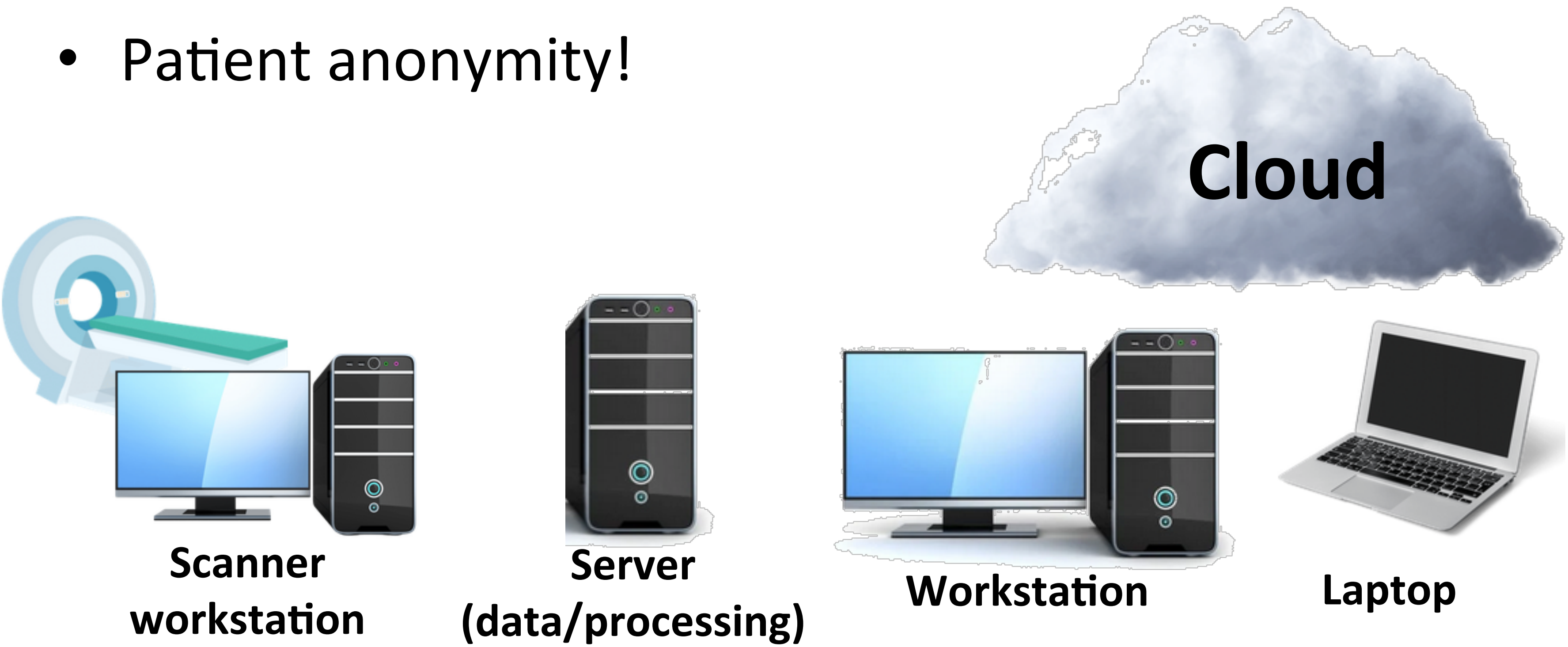


Image processing steps

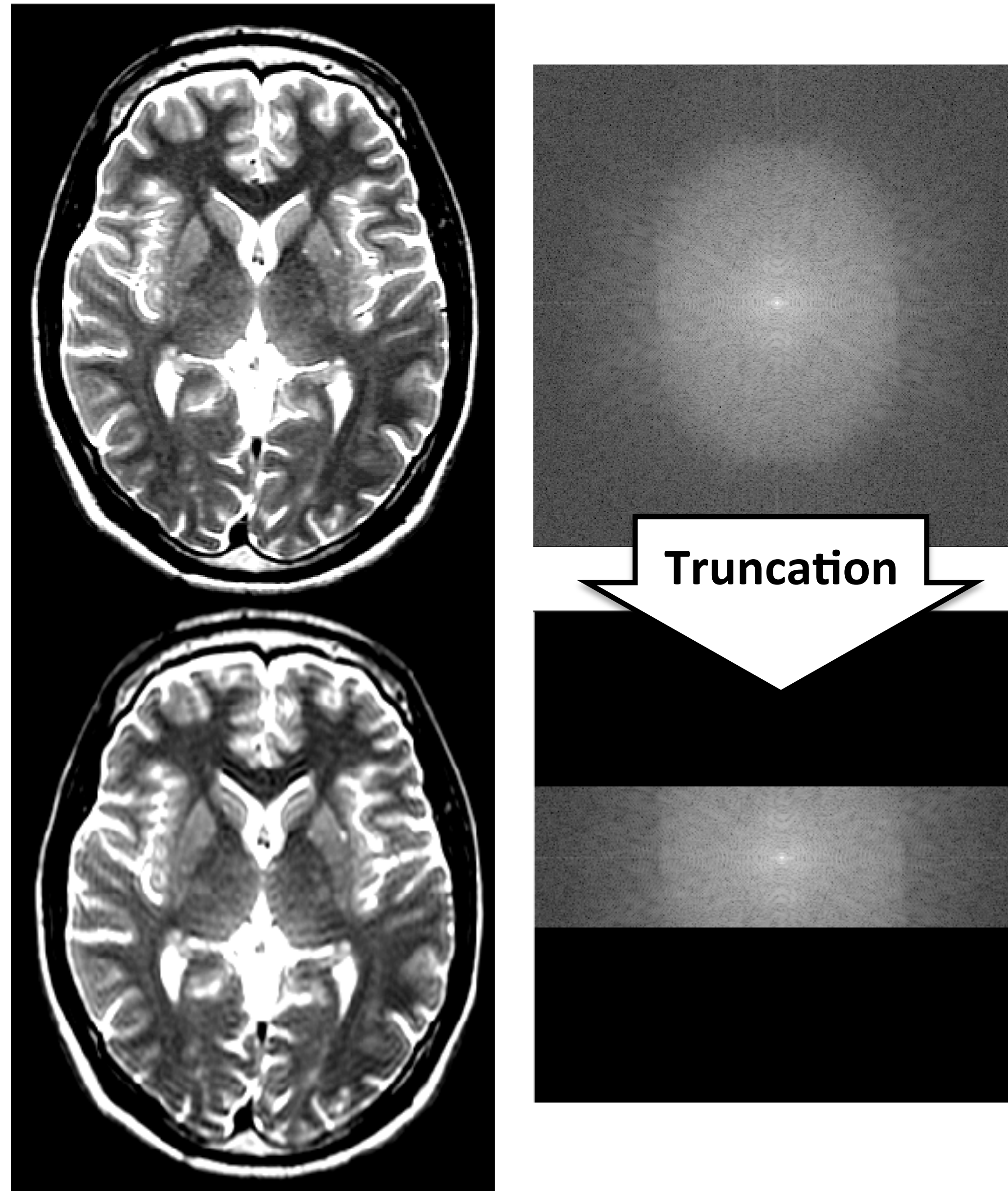
- Artefact removal
- Signal modelling
- Region of Interest (ROI) placement
- Radiomics
- Computation of diagnostic/prognostic scores

Artefact removal

Gibbs Ringing

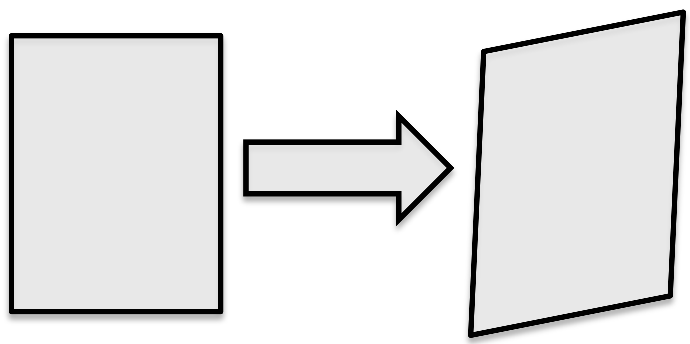
- Abruptly truncating signal in *k*-space introduces “ringing” to the image
- To fix: mrxtrix ***degibbs***
  – Removes ringing artefact if present

native slice fft space



Eddy Current correction

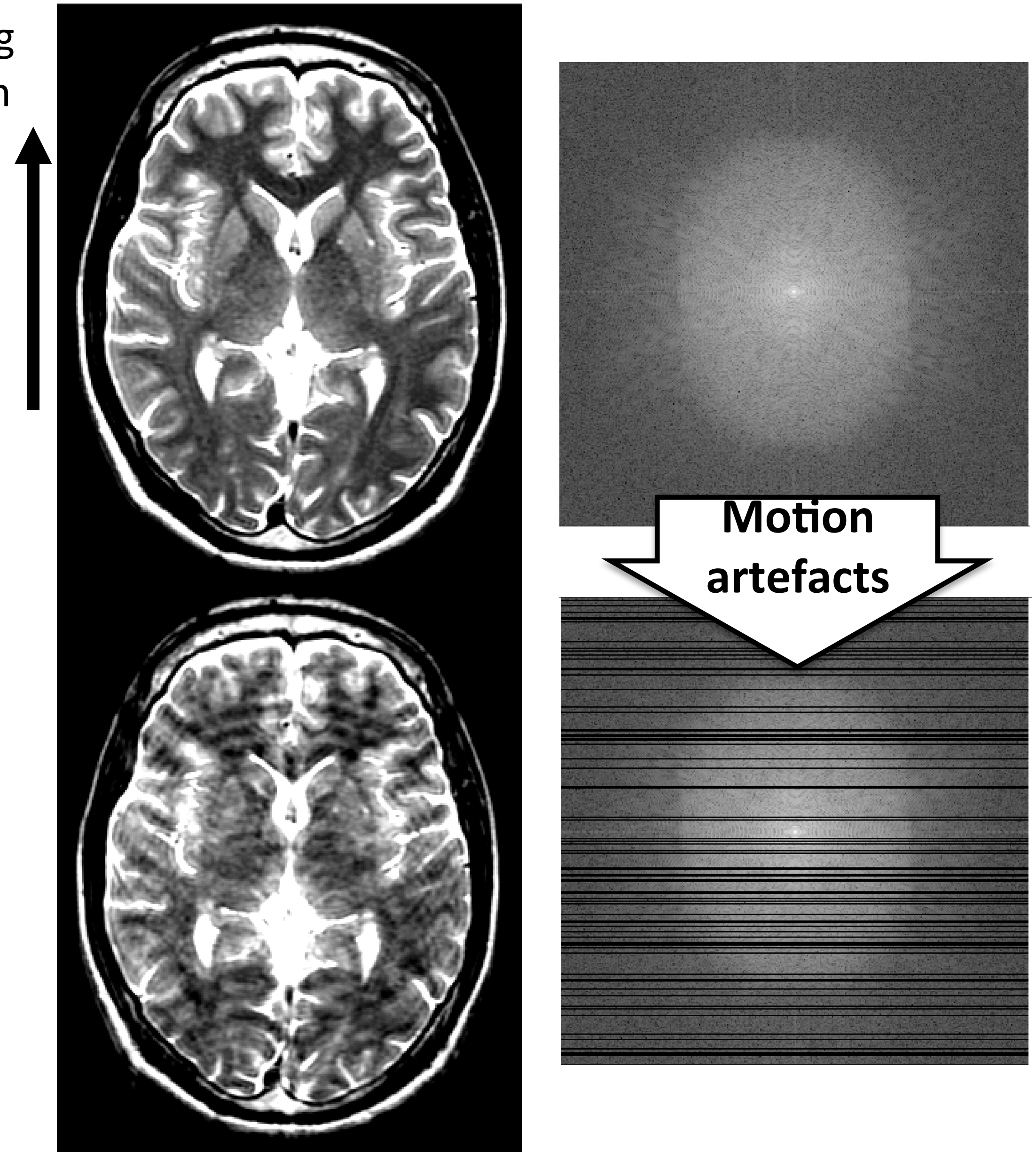
- **Eddy currents** result from gradient magnetic fields
  - Eddy currents generate their own magnetic fields, which distort the spatial and temporal performance of the overall desired magnetic field.



Eddy current distortion in DTI

Motion artefact

- Occurs **within** (see right) **and between slices and volumes**
- Tools (between volumes motino correction): fsl ***eddy, mcflirt***, ***ANTs***

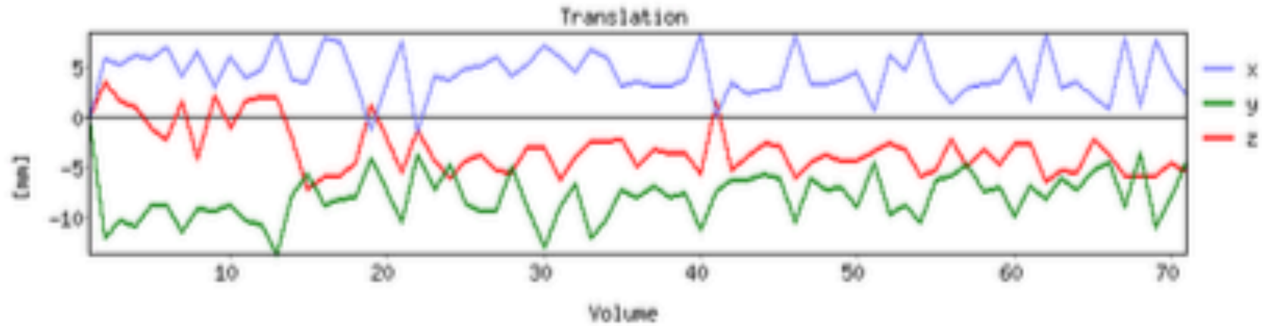


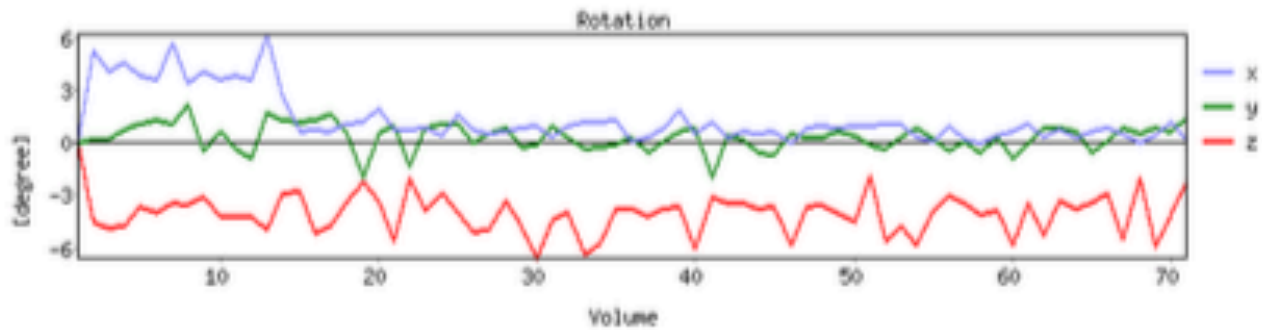
Motion correction

- Approaches for assessing motion
  - Motion suppression by means of acquisition setting (positioning, pillows etc)
  - Protocols/sequences optimized to address motion (PROPELLER, navigator images, oversampling etc.)
  - **Post-processing**
  - If possible, estimate occurred motion, and use as covariate
- With multi-volume MRI data, eddy current distortion is corrected at the same time with motion

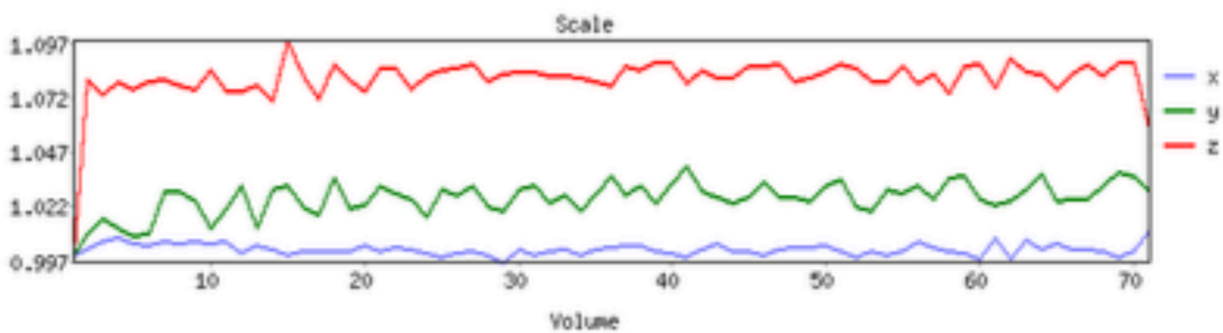
Motion correction example

Estimated motion
Mean dislocation : 9.976763 [mm]
Mean translation: (x y z)=(4.531128, -7.783876, -3.033722) [mm]
Mean rotation: (x y z)=(1.405032, 0.362363, -3.988247) [degree]





Estimated distortion
Mean scale: (x y z)=(1.002559, 1.026221, 1.081162)
Mean skew: (x y z)=(0.029569, -0.001893, -0.016641)



N/2 Ghosting artefact

- Shifted trajectory is sum of 2 shifted undersampled trajectories (or single undersampled trajectory)
- Causes aliasing (“N/2 ghosting”)
- To fix: measure shifts with reference scan, shift back in reconstruction

Undersampled Fourier space

Undersampled Fourier space

Acquired Fourier space

Effect of 1 voxel shift

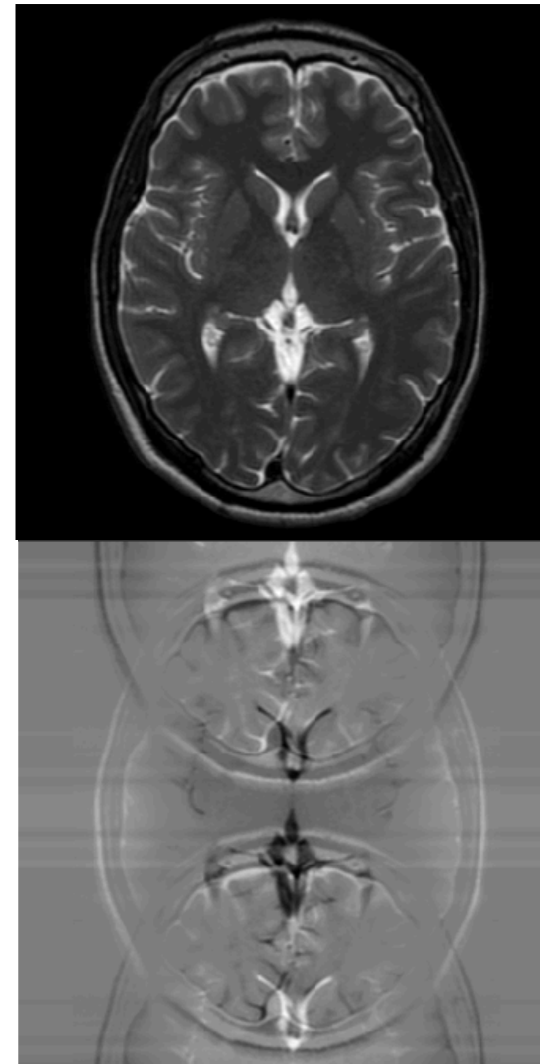
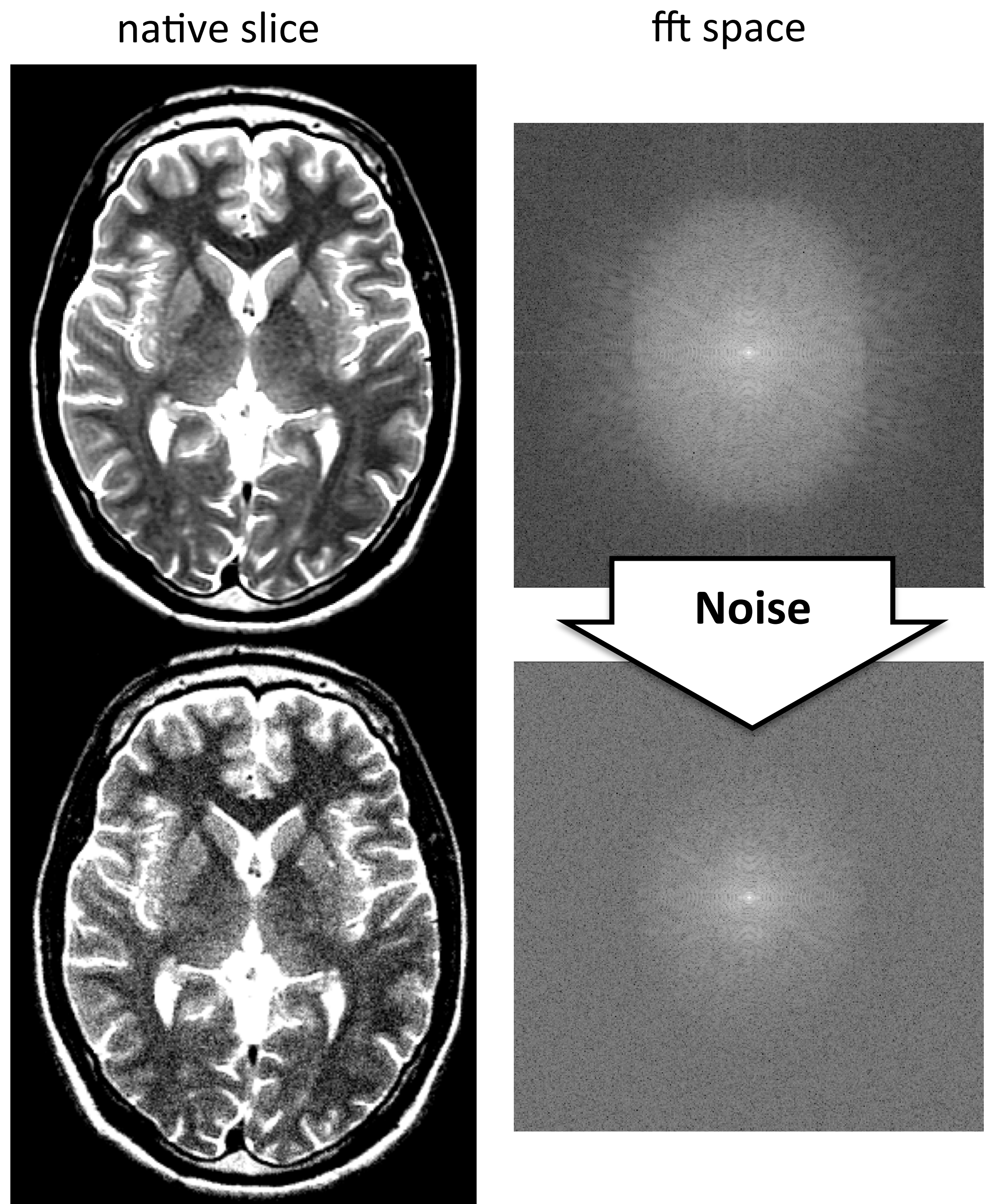
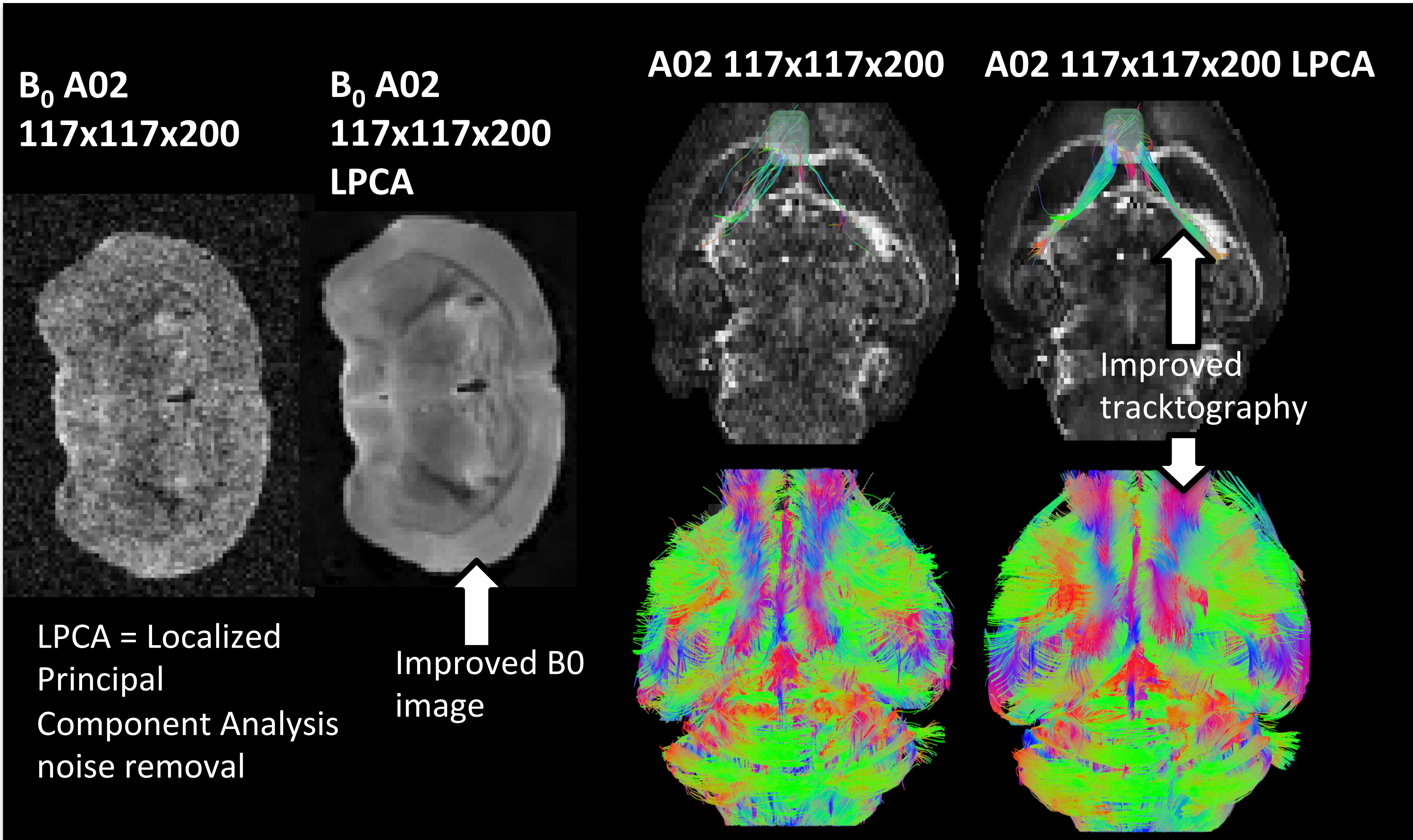


Image Noise

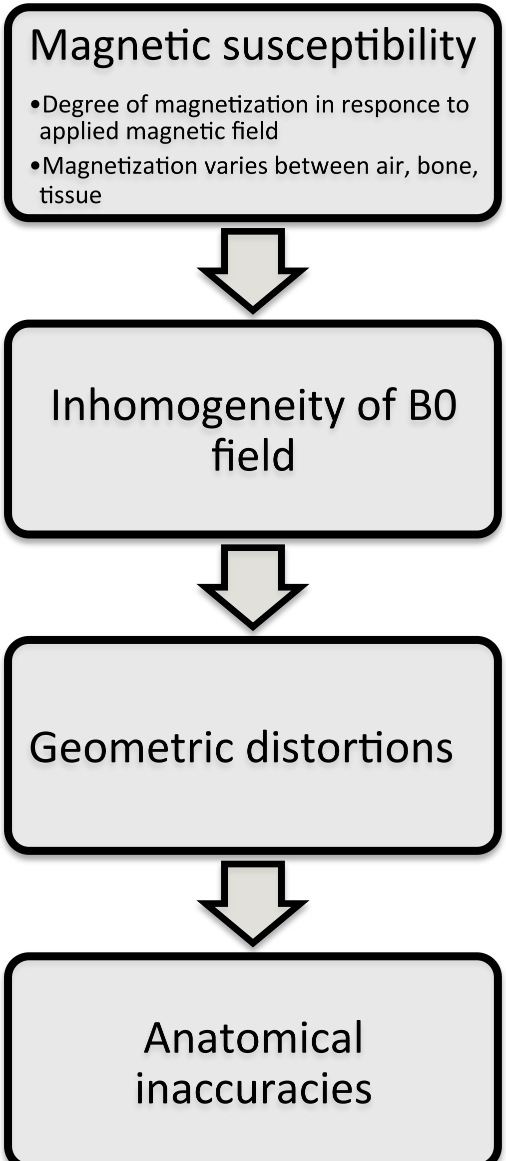
- Noise arises from target tissue and scanner
- Noise is additive in phase and magnitude data, theoretically resulting Rician noise in reconstructed image
- To fix: mrtrix ***dwidenoise***
  - Aims to remove noise according to Marchenko-Pastur theorem for image data components, in separation signal from noise



DTI Noise removal example: Pre-clinical murine with PCA

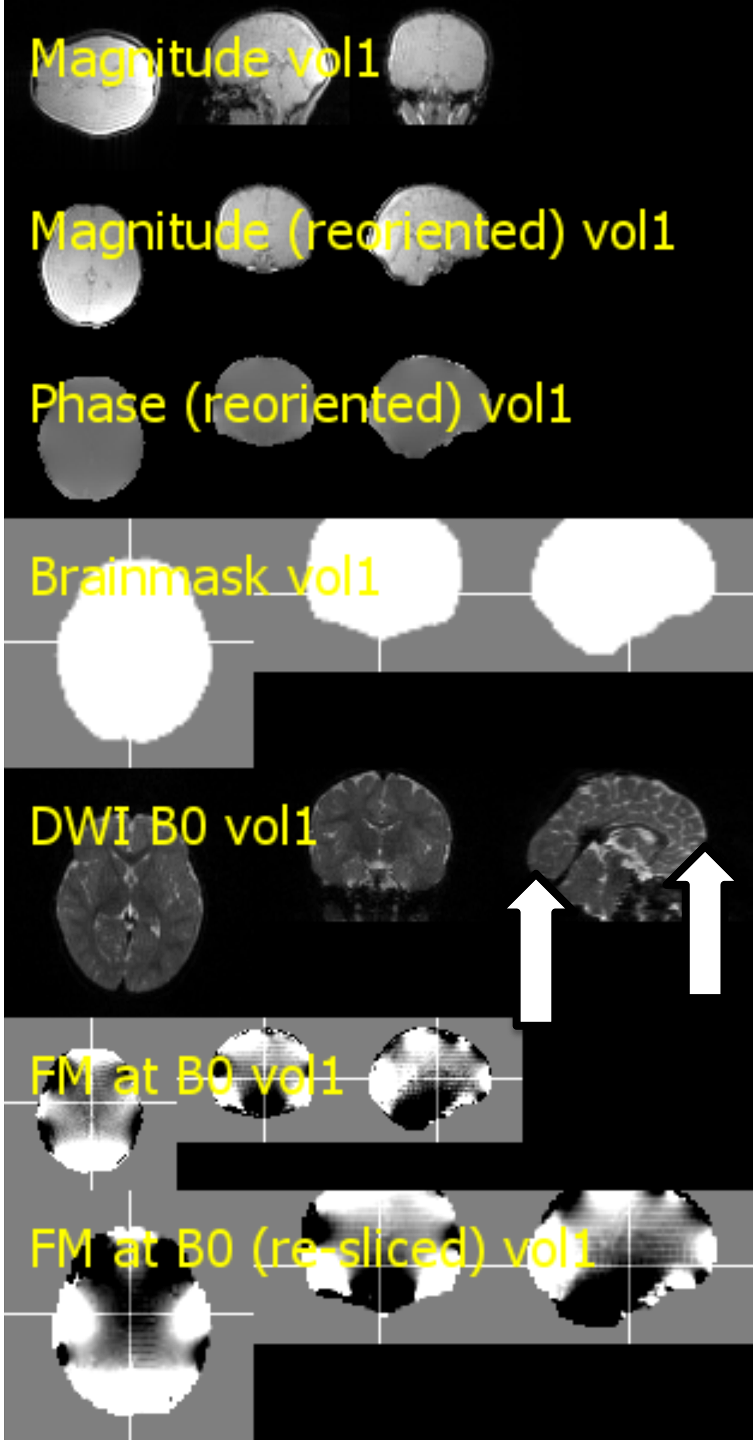


Distortion correction

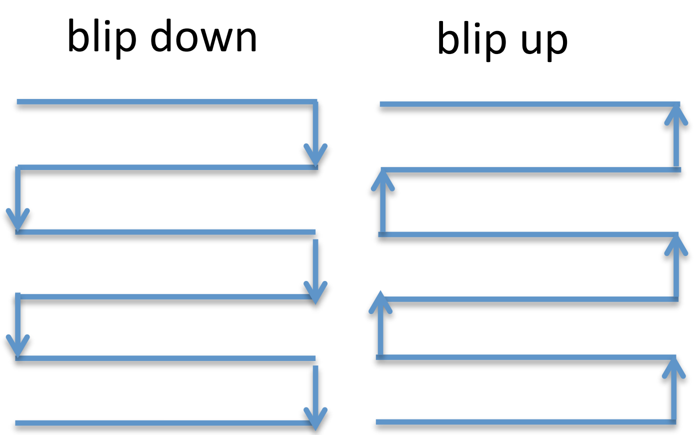


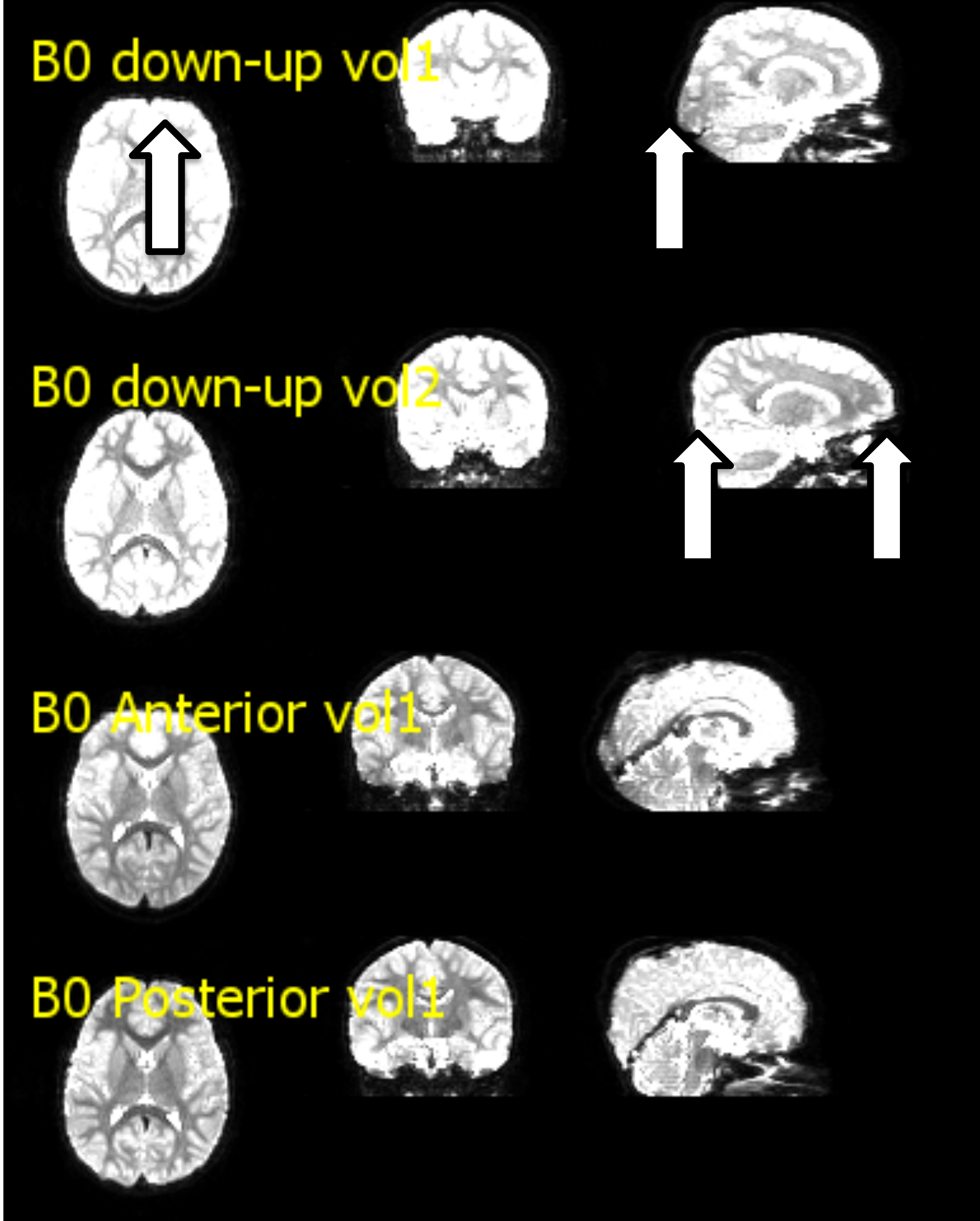
- Magnetization processes at a different rate than expected, and reconstruction places the signal at the wrong location
- Correction options
  - 1) Specifically acquired field maps
  - **2) Acquire opposite blip direction image, fsl *topup***
  - 3) Estimate distortion using anatomical reference e g T1W image ***SynB0*** tool

Distortion correction example: Separately acquired field map



Distortion correction example: TOPUP

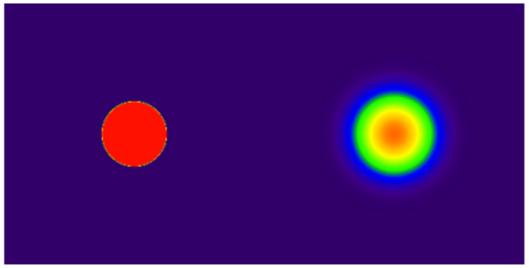


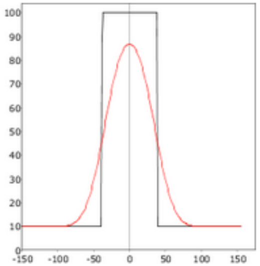


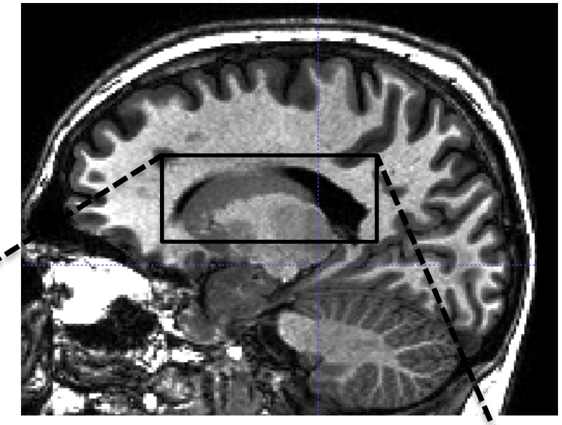
Partial Volume Effect

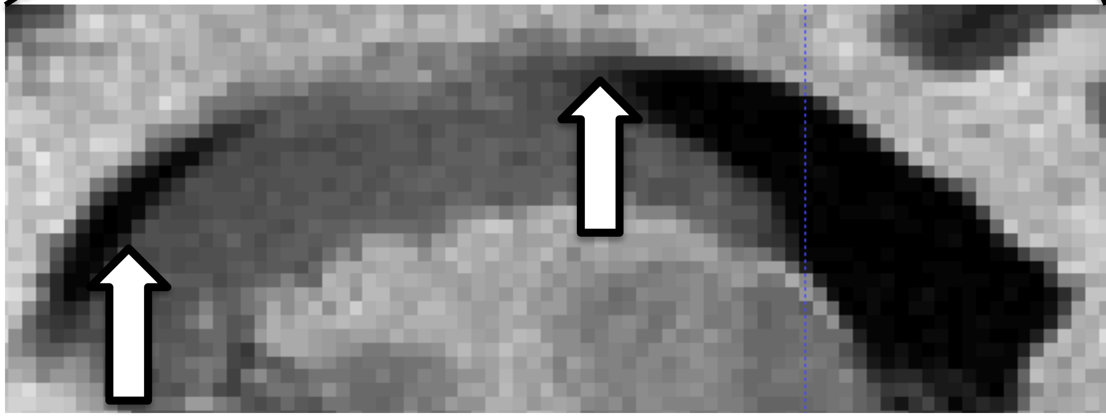
- Correct for Partial Volume Effect (PVE)
  - Potential spill-over from e g cerebrospinal fluid to juxtaposed region
  - Tools ***fslmaths***

    python ***skimage***









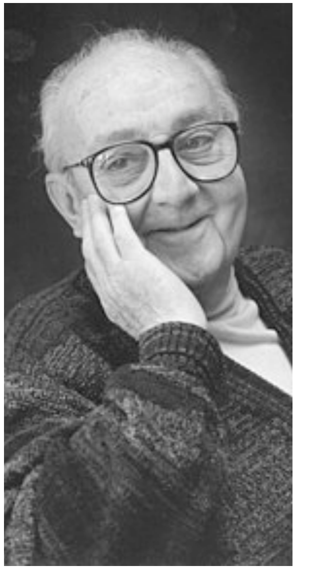
Summary

- Many of the artefacts in the image can be avoided by better optimization of acquisition
- Motion, susceptibility and noise may need to be addressed with post-processing

Signal Modelling

At this point we assume data is all good, now doing the actual image processing

"All models are wrong, but some are useful"
- George Box, statistician



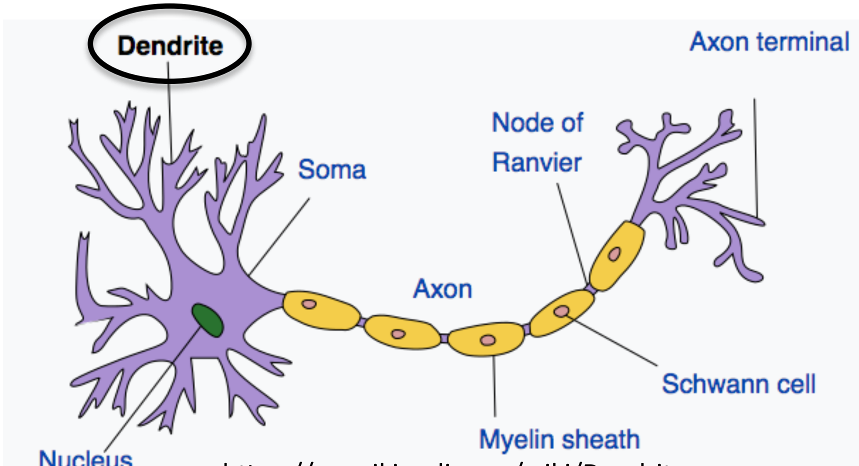
Signal Modelling

- Modelling is usually done for each ***individual voxel*** in the image, although sometimes for ***ROI mean values***
- Parameter estimates in medical imaging can be used as ***biomarkers*** - objective measure capturing of what is happening in a cell or an organism at a given moment
- Model parameters have a level of ***unspecificity****:* Parameter value may have multiple different physiological interpretations at cell or organism level
- In medicine, ultimate criteria for good parameter is usefulness in ***clinical context***

Modelling T1W/T2W

- T1W
  - T1 times depend strongly upon magnetic field strength 1.5T, 3T, 7T etc.
  - In brain, T1 have support to detect changes in water content, myelin, fiber orientation, iron
- T1W/T2W ratio
  - can be used a measure for myelin, axon/dendrite density, iron content ***matlab***, ***python***
- T1W: Voxel Based Morphometry (VBM) ***SPM***
  - gray matter 'density'

Mechelli A, Price CJ, Friston KJ, Ashburner J. Voxel-based morphometry of the human brain: methods and applications. Current Medical Imaging. 2005 Jun 1;1(2):105-13.



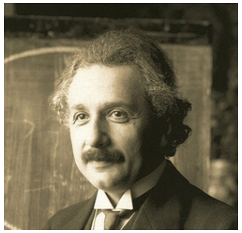
https://en.wikipedia.org/wiki/Dendrite

Modelling DWI

https://en.wikipedia.org/wiki/Diffusion

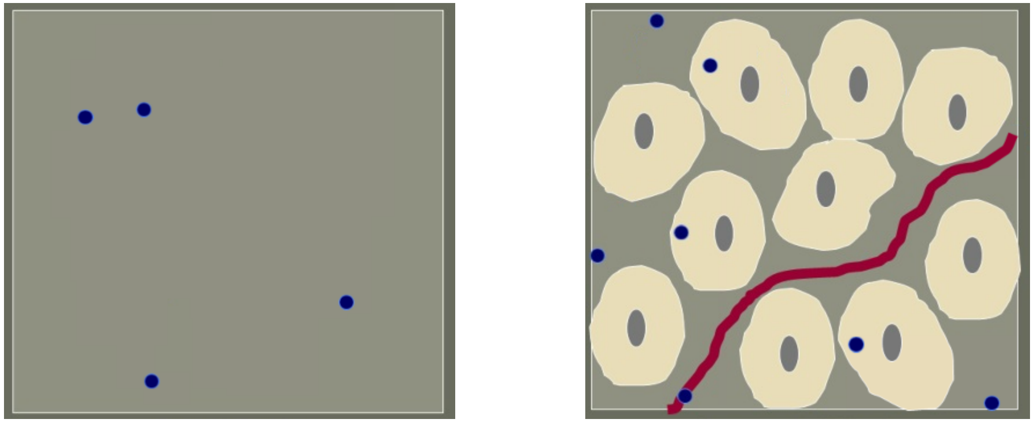
Brown, 1827: Continuous and spontaneous random motion of pollen grains suspended in water, with microscope

Einstein, 1905: Brownian motion, i. e. random motion of particles in a fluid



Tools: C code ***dlib***
www.github.com/haanme,
matlab, ***Carimas, python dipy*** etc.

Modelling DWI



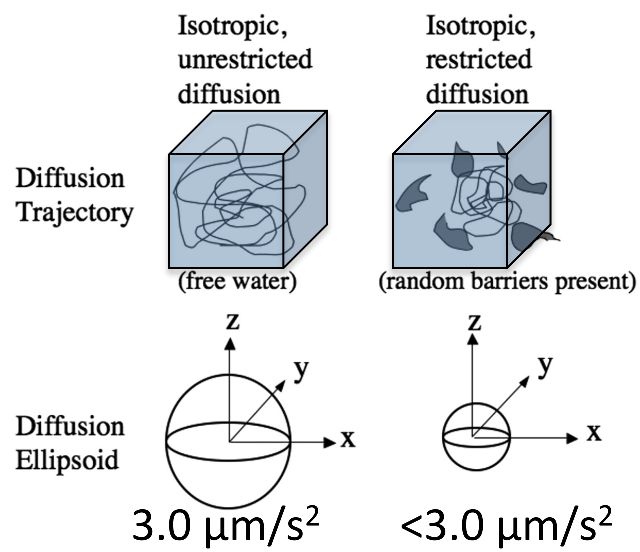
Free diffusion

Restricted/Hindered diffusion

- Most common DWI measure is Apparent Diffusion Coefficient (ADC):
  - –Microcapillary perfusion
  - –Fluid homogeneity
  - –Macromolecules
  - –Cellular **density**
  - –Cell membranes integrity
  - –Microstructural organisation

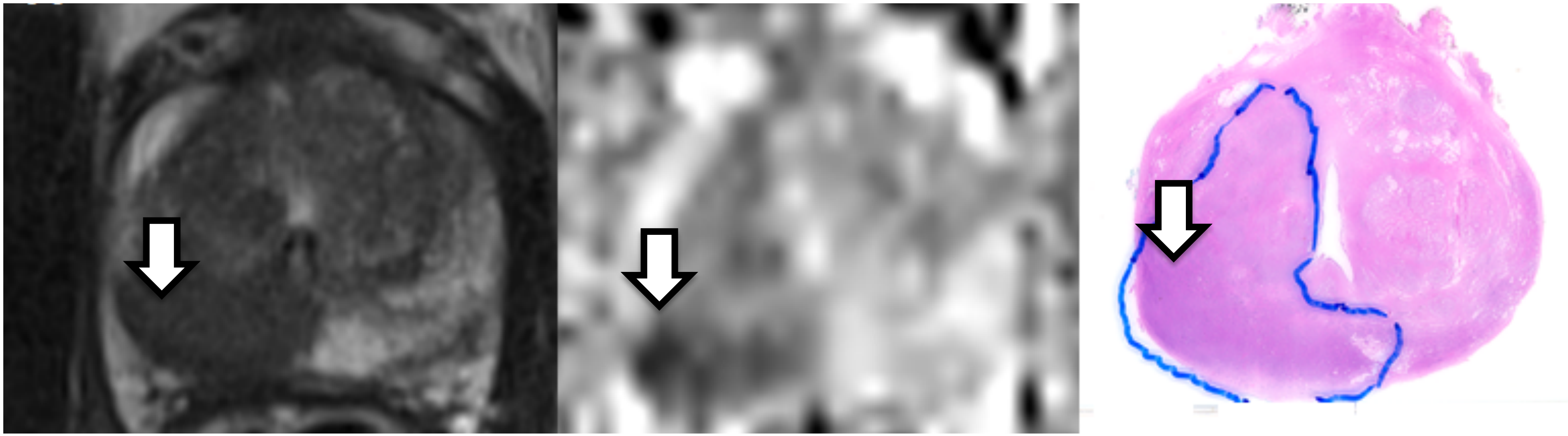
Physiological example values of ADC

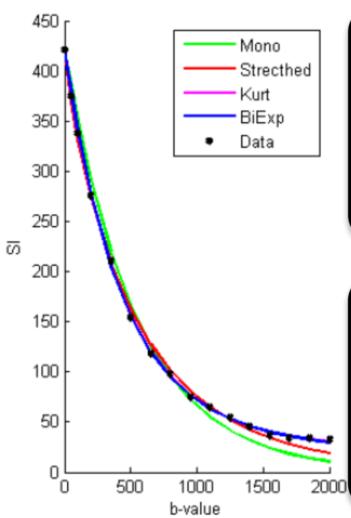
- 0 µm/s² - No diffusion
- 3.0 µm/s² - Free water
  (e g water in a glass)



Apparent Diffusion Coefficient example: Prostate Cancer

T2W **ADC** **Histology**





More signal loss at higher b-values → **More diffusion of water particles** → **Less dense tissue**

Less signal loss at higher b-values → **Less diffusion of water particles** → **More dense tissue** → **Tumor**

Modelling DWI: Monoexponential

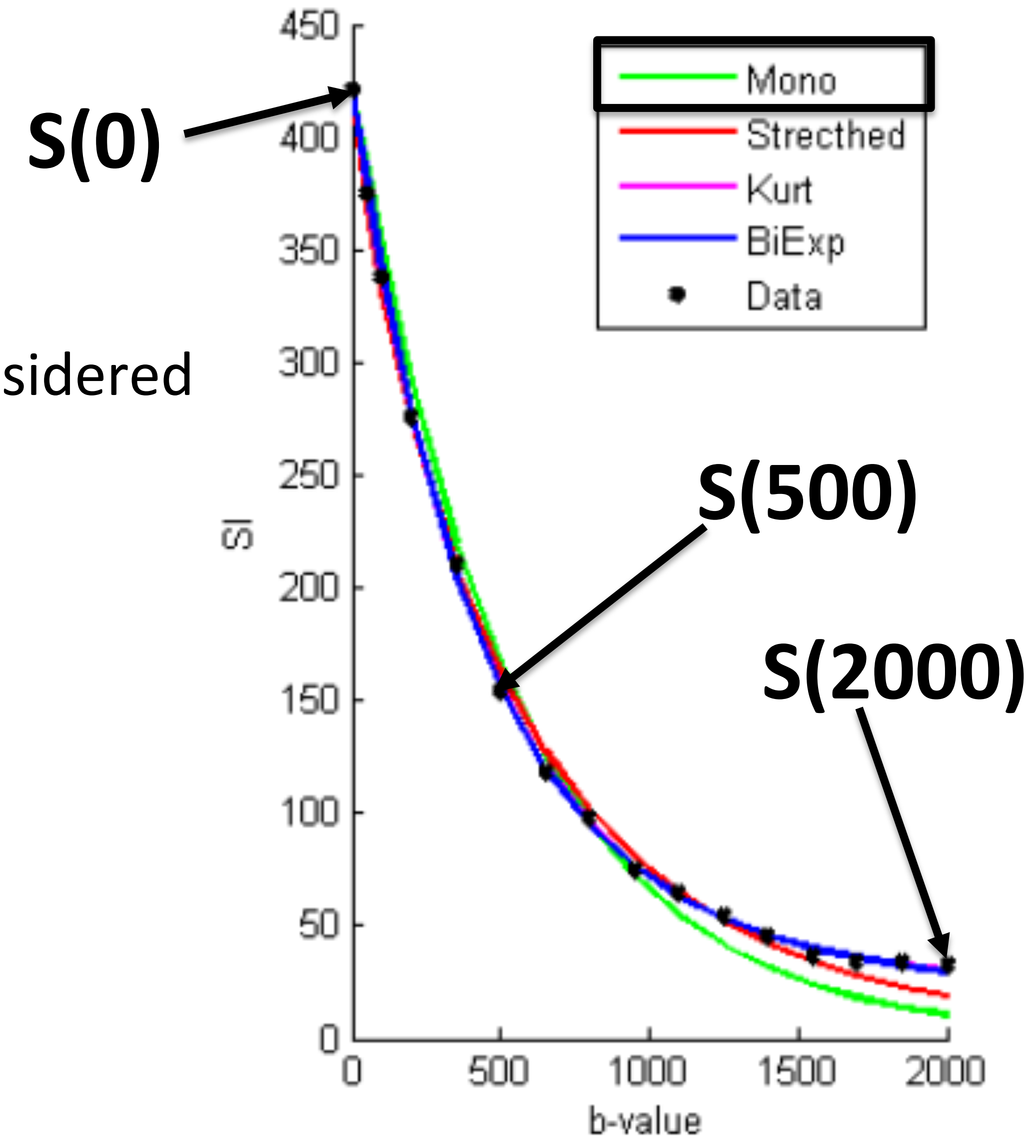
$$S(b) = S_0 e^{-bADC_{mono}}$$

Monoexponential:

- Single exponential decay
- ADC can be generally considered to reflect tissue density:

↑ less dense tissue

↓ more dense tissue



Modelling DWI: Other models

Stretched exponential

$$S(b) = S_0 e^{-(bADC_{streched})^{\alpha}}$$

Kurtosis model

$$S(b) = S_0 e^{(-bADC_{kurtosis} + \frac{1}{6} b^2 ADC_{kurtosis}^2 K)}$$

Intra-voxel Incoherent Motion (IVIM)

Biexponential

$$S(b) = S_0((1-f)e^{-bD_s} + fe^{-bD_f})$$

Modelling: Stretched exponential

$$S(b) = S_0 e^{-(bADC_{streched})^{\alpha}}$$

Stretched exponential:

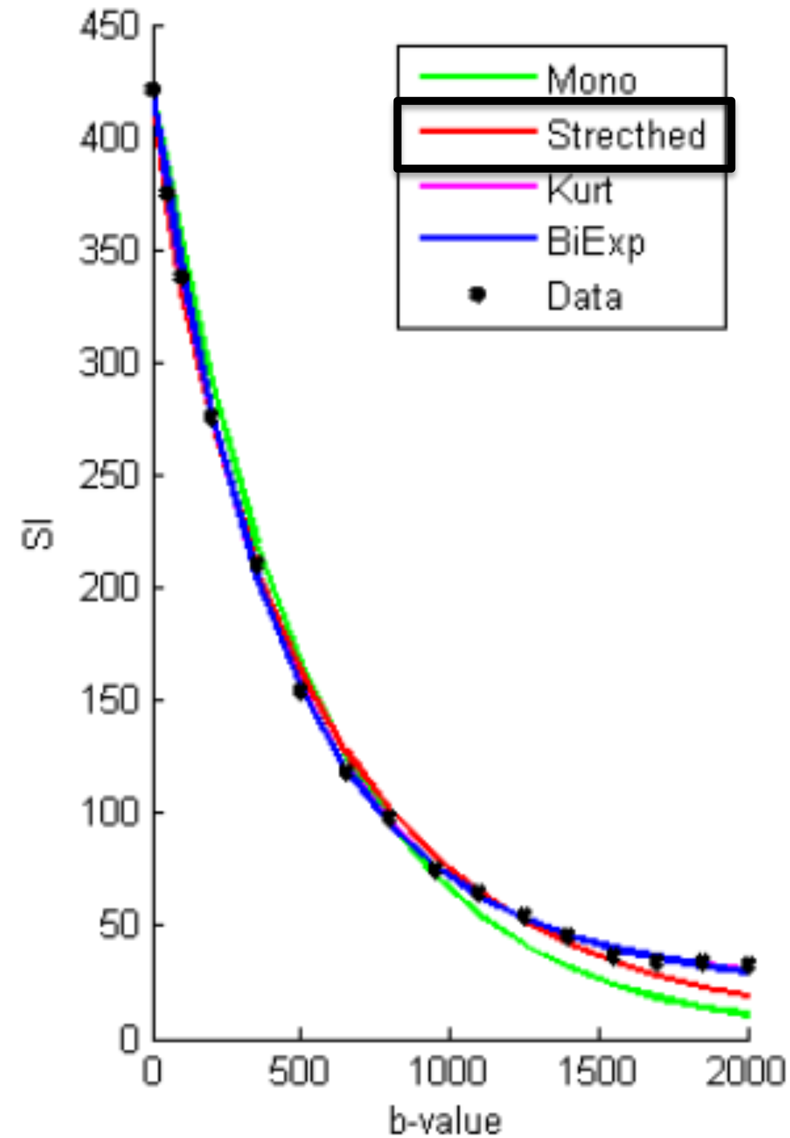
- Heterogeneity index α reflects composition of multiple exponentials

- α < 1 for more than

one exponential present

↑ more homogenous tissue

↓ more heterogenous tissue



Modelling DWI: Kurtosis model

$$S(b) = S_0 e^{(-bADC_{kurtosis} + \frac{1}{6} b^2 ADC_{kurtosis}^2 K)}$$

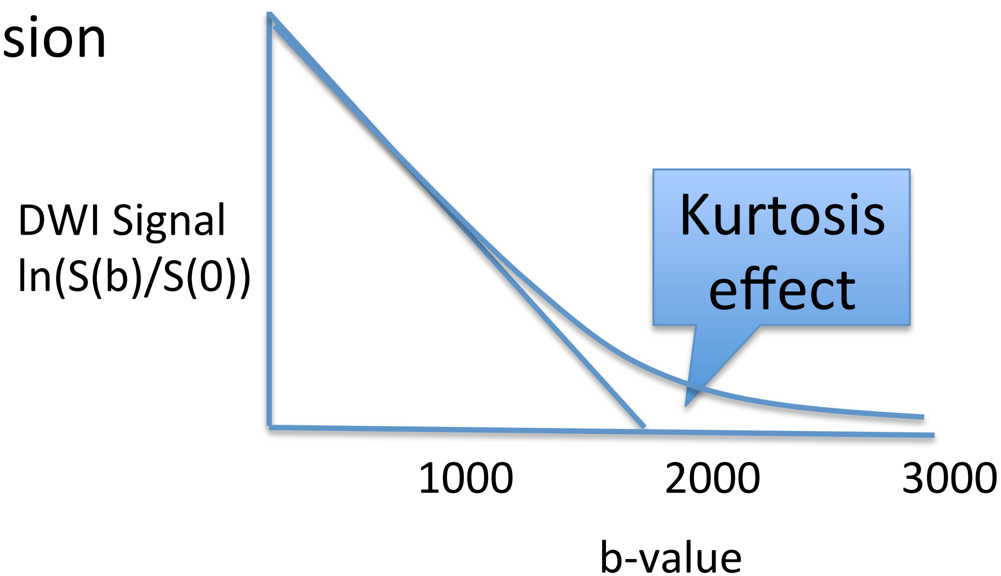
Kurtosis:

- K reflects deviation from Gaussian shape, physically associated with structure

- Based on Taylor series expansion of signal

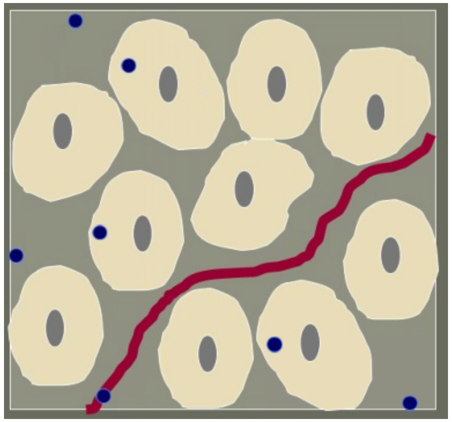
↑ more structures in tissue

↓ less structure



Modelling DWI: Bi-exponential

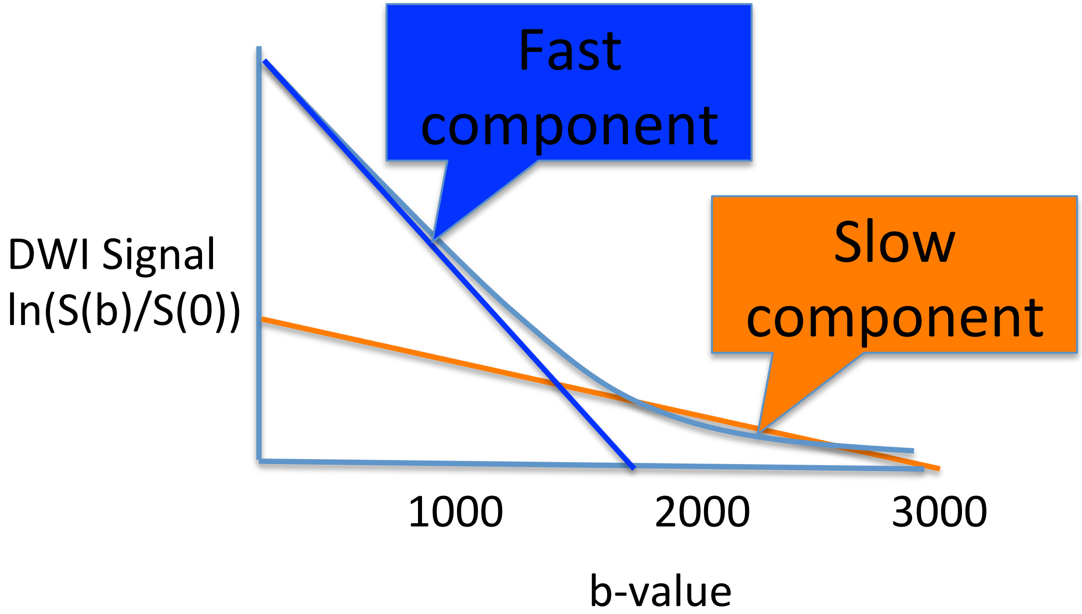
$$S(b) = S_0((1 - f)e^{-bD_s} + fe^{-bD_f})$$



Biexponential:

- Df and Ds reflect fast and slow exponential decays

- Interpretation depends on used b-values, and the tissue under study



Modelling IVIM
(IntraVoxel Incoherent Motion)

$$S(b) = S_0\left[\left(1-F_p\right)e^{-ADC_D\ b} + F_p e^{-b(ADC_P\ +\ ADC_D)}\right]$$

Le Bihan D, Radiology 1988

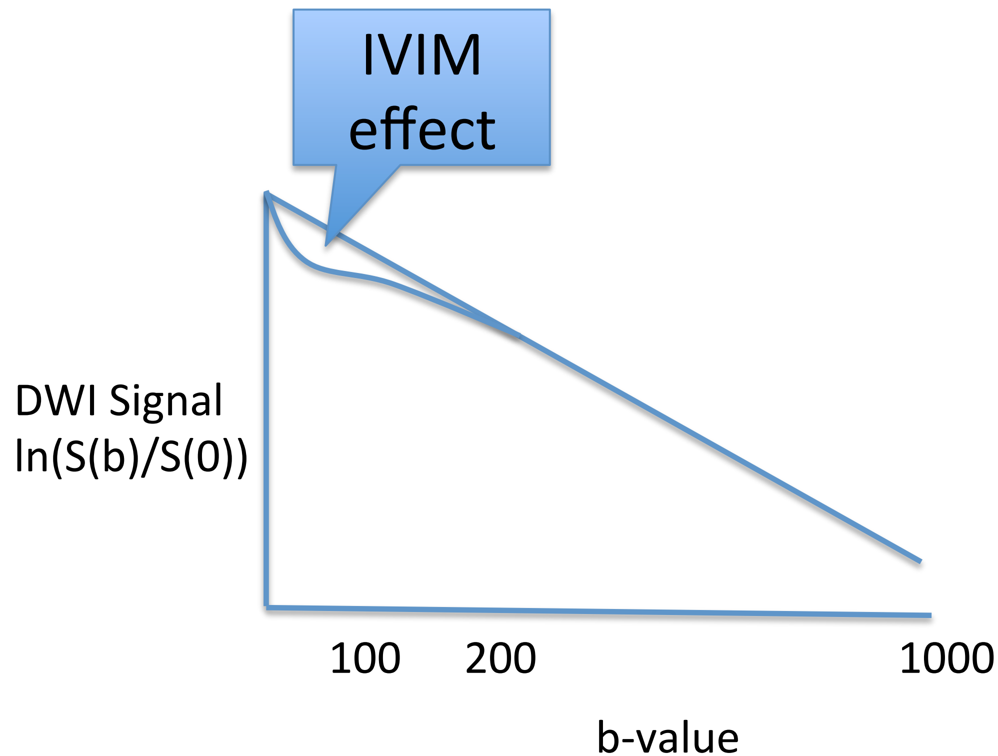
IVIM:

- Mathematically the same a bi-exponential model

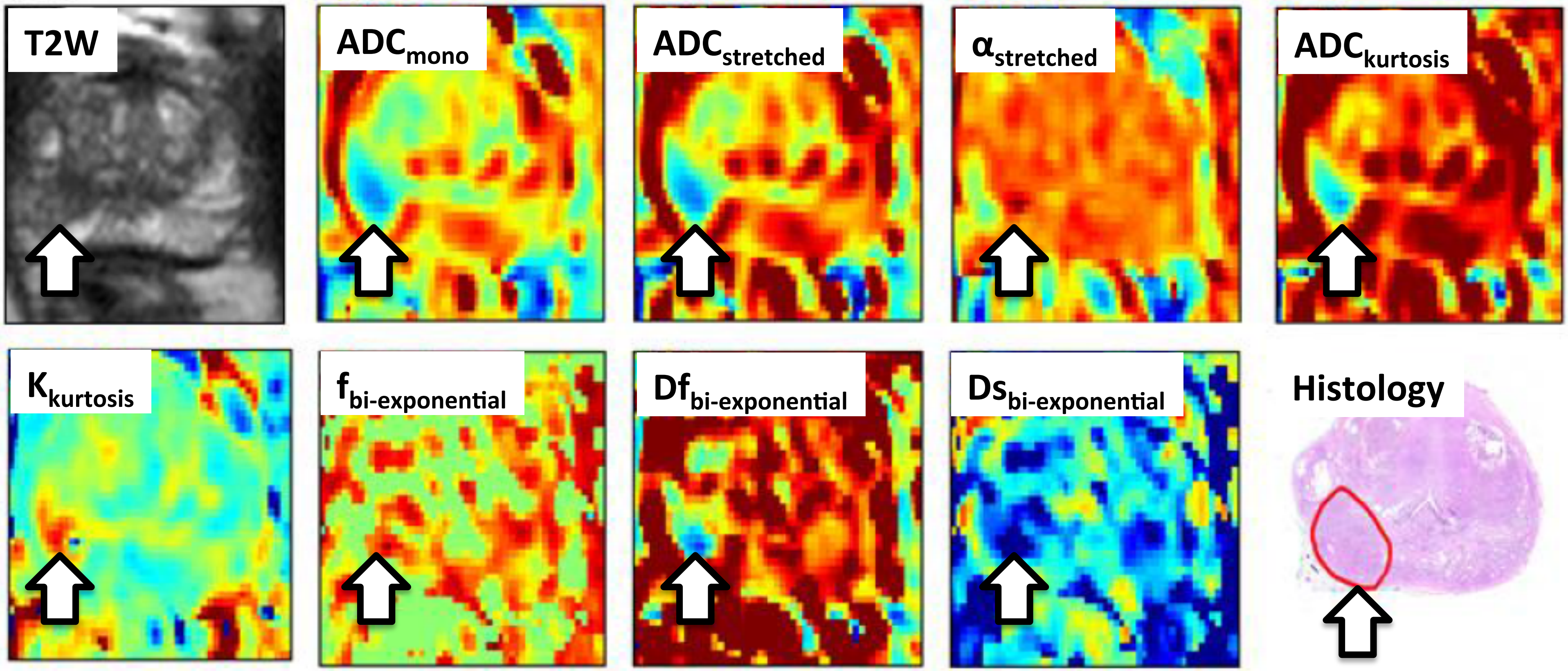
- Difference arises from the acquired b-value samples at range of IVIM effect

b-value < 200 s/mm^2

- D* (pseudodiffusion) reflects blood perfusion



Modelling DWI: Example of prostate cancer parameter maps



DWI models reveal different contrast with tumour and within tumour, with data from the same DWI acquisition

Modelling DTI

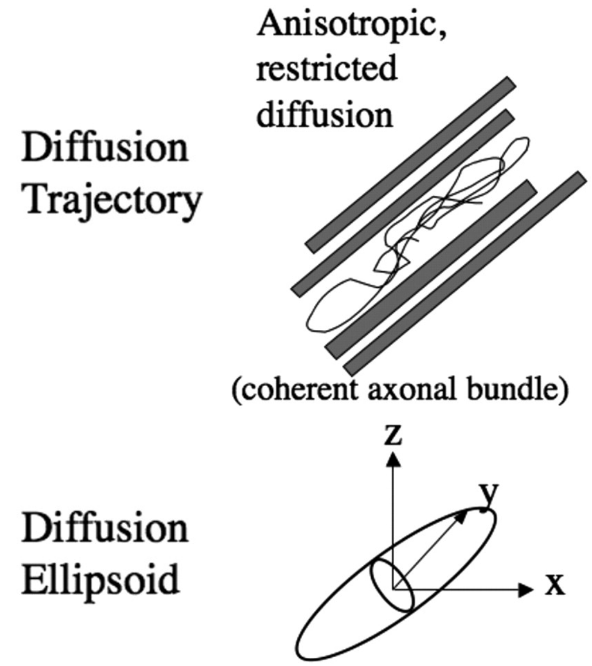
- DTI model assumes one 3D ellipsoid instead of sphere
- Convention is to use DTI in brain white matter (WM)
- Tools: fsl ***dtifit***, ***DSIstudio, mrtrix, diffusion toolkit***
- DTI scalar interpretations:

Fractional Anisotropy (**FA**):
↑high integrity of fibers ↓low integrity

Mean Diffusivity (**MD**):
↑ less dense tissue ↓ more dense tissue

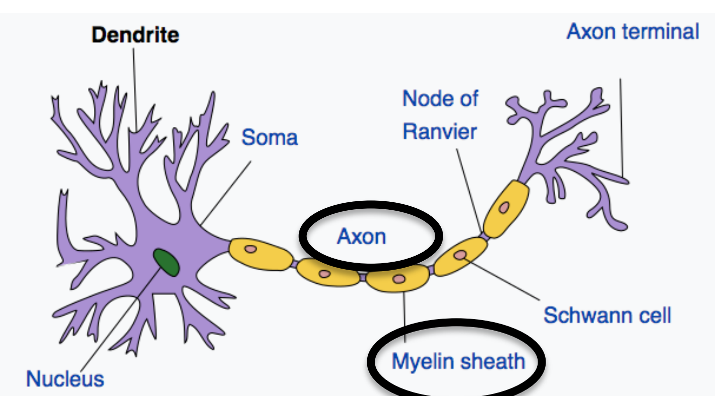
Axial Diffusivity (**AD**):
↑ healthy fibers ↓ axonal injury

Radial Diffusivity (**RD**):
↑ myelin loss ↓ more myelin



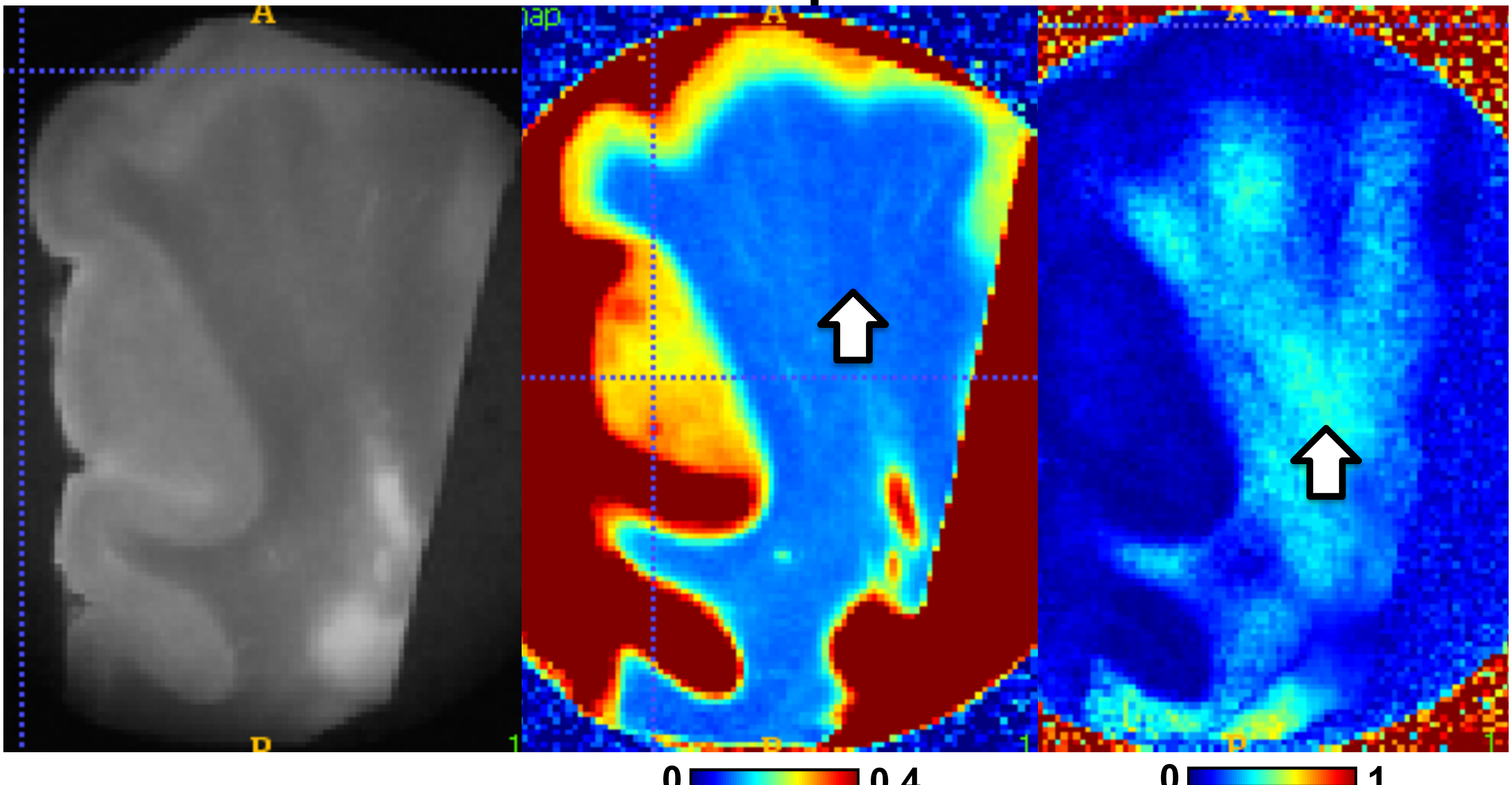
Diffusion Tensor

$$\begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$$



https://en.wikipedia.org/wiki/Dendrite

Modelling DTI: Example brain sample

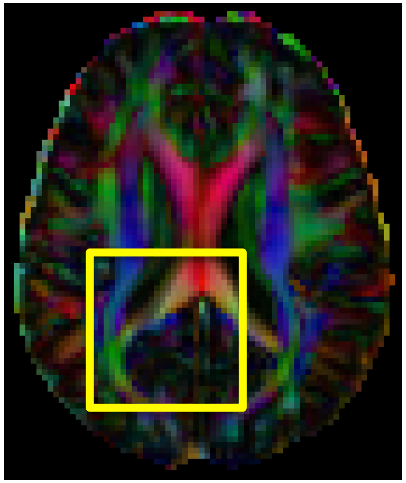


B0 **Mean Diffusivity** **Fractional Anisotropy**

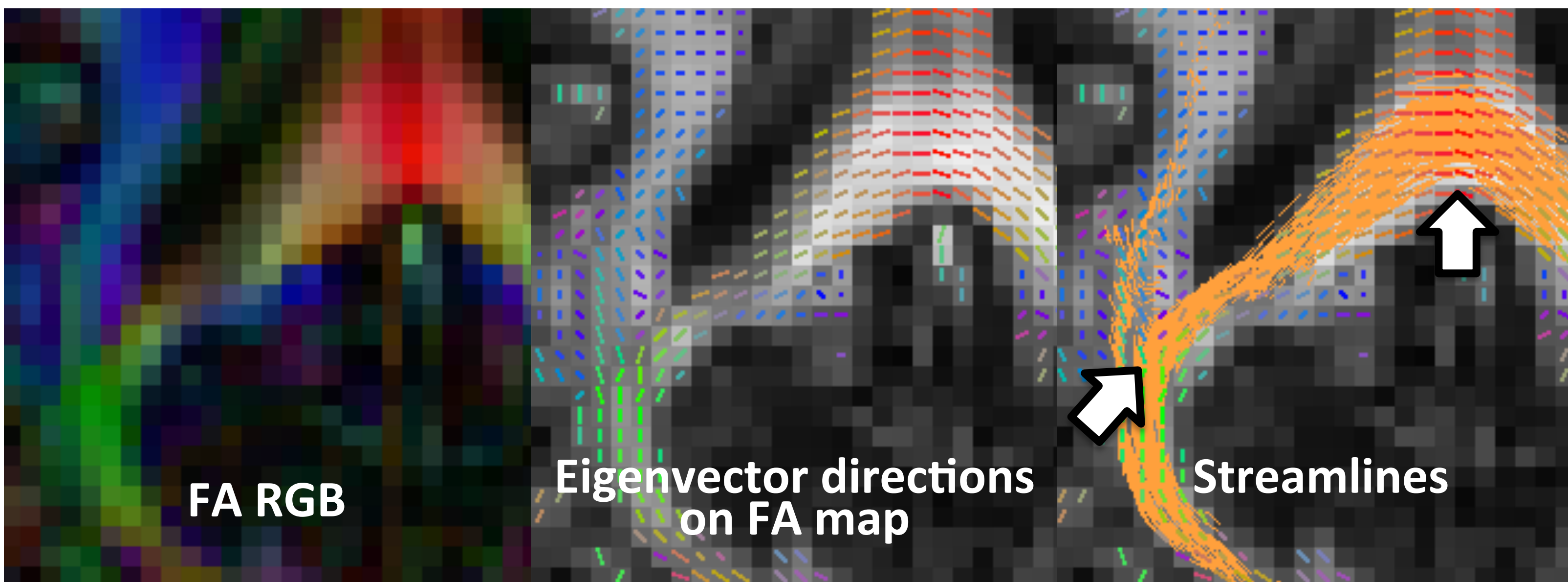
Mean Diffusivity shows dense white matter (blue, middle)

Fractional Anisotropy map highlighting white matter integrity (bright blue, right)

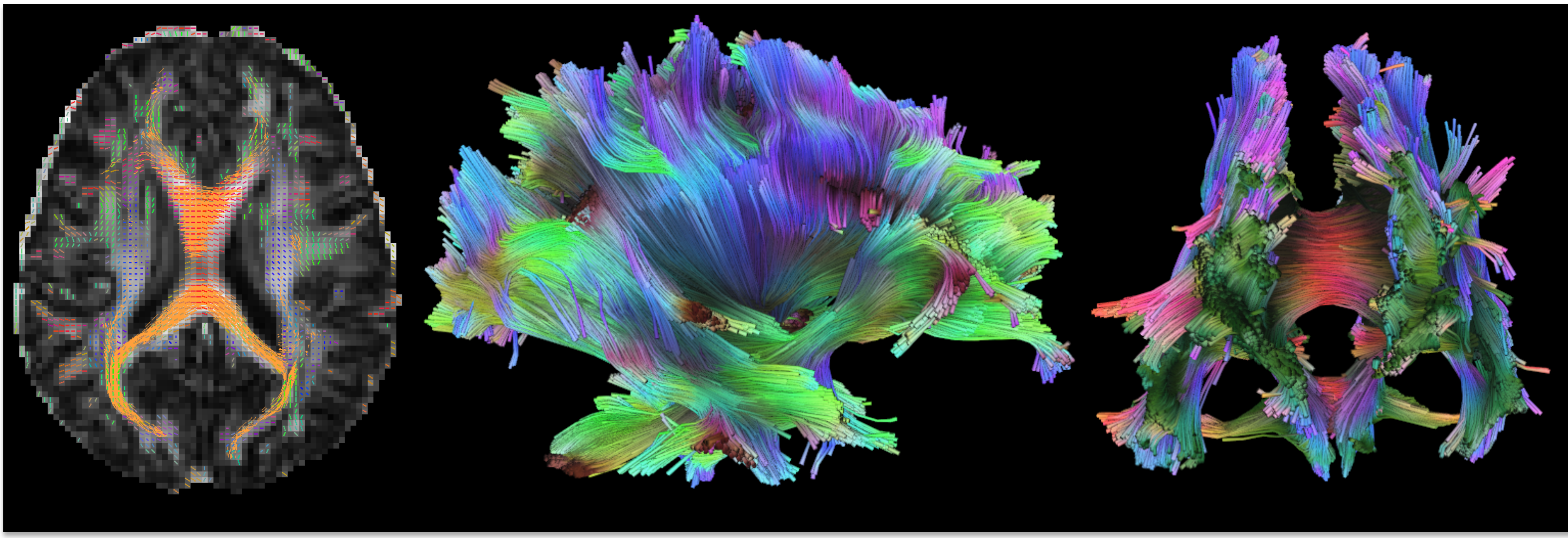
DTI Tracktography



- **Streamline** algorithm (repeat to cover seed region, e g 10k times):
  1. Start from a location in source region
  2. Move along voxels along their major eigenvector
  3. End when stop region is reached
  4. Repeat from step 1.
- **Deterministic** tracktography: One streamline per seed location ***DSIstudio, trackvis***
- **Probabilistic** tracktography: Distribution of all streamlines from all seed voxels fsl ***bedpostx, probtrackx***



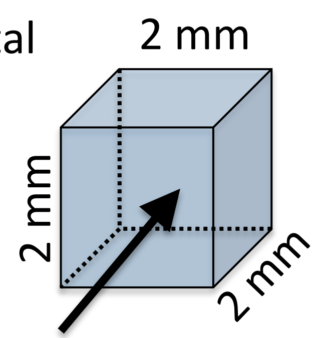
DTI Tracktography: Example of Deterministic tracktography



Eigenvectors on FA map

Tracktogram sagittal view

Tracktogram coronal view

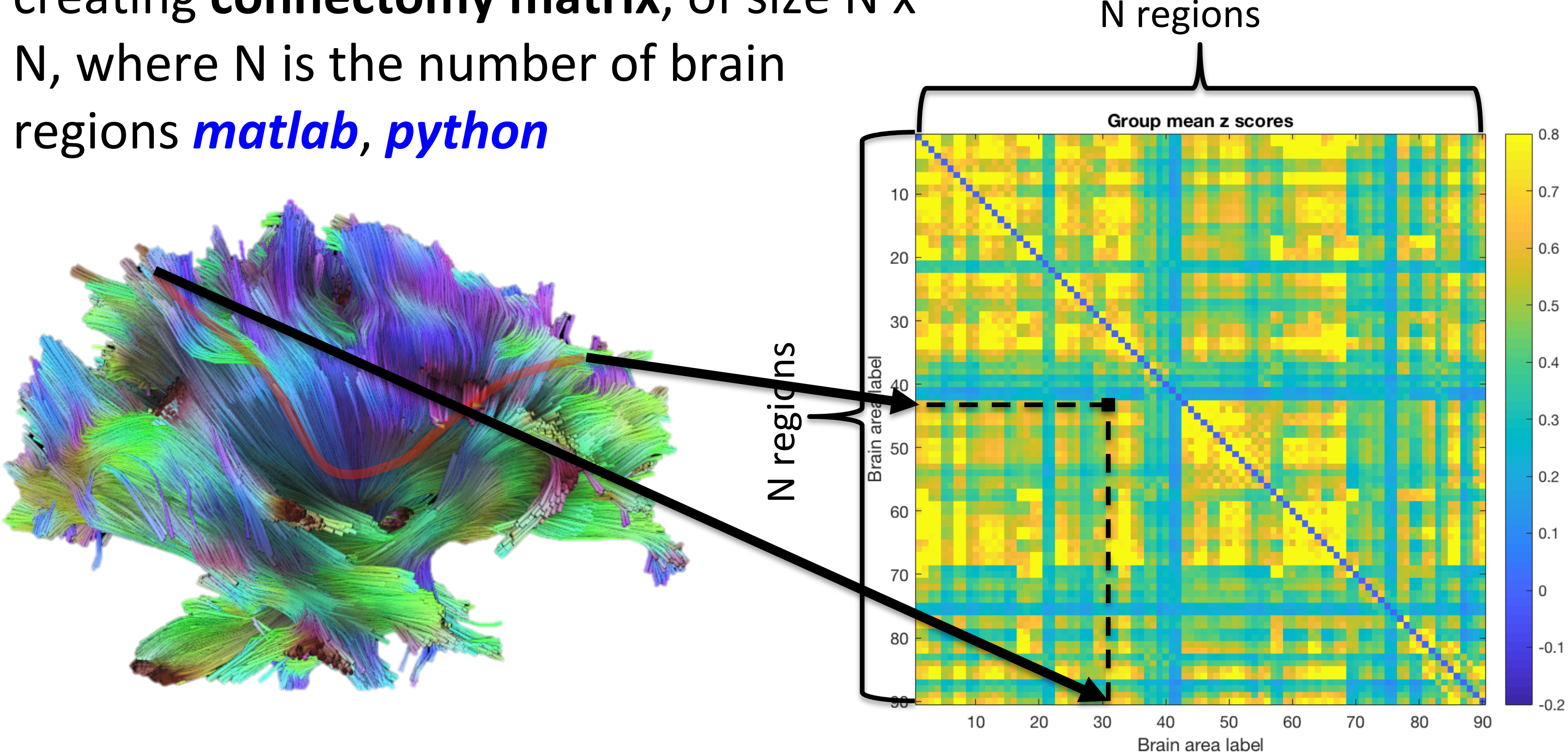


Voxel of size 2x2x2 mm^3 can contain approximately from ~0.5 million axons to >5 million axons!

Walhovd, K.B., et al., 2014. Neuroscience, 276, pp.2-13.

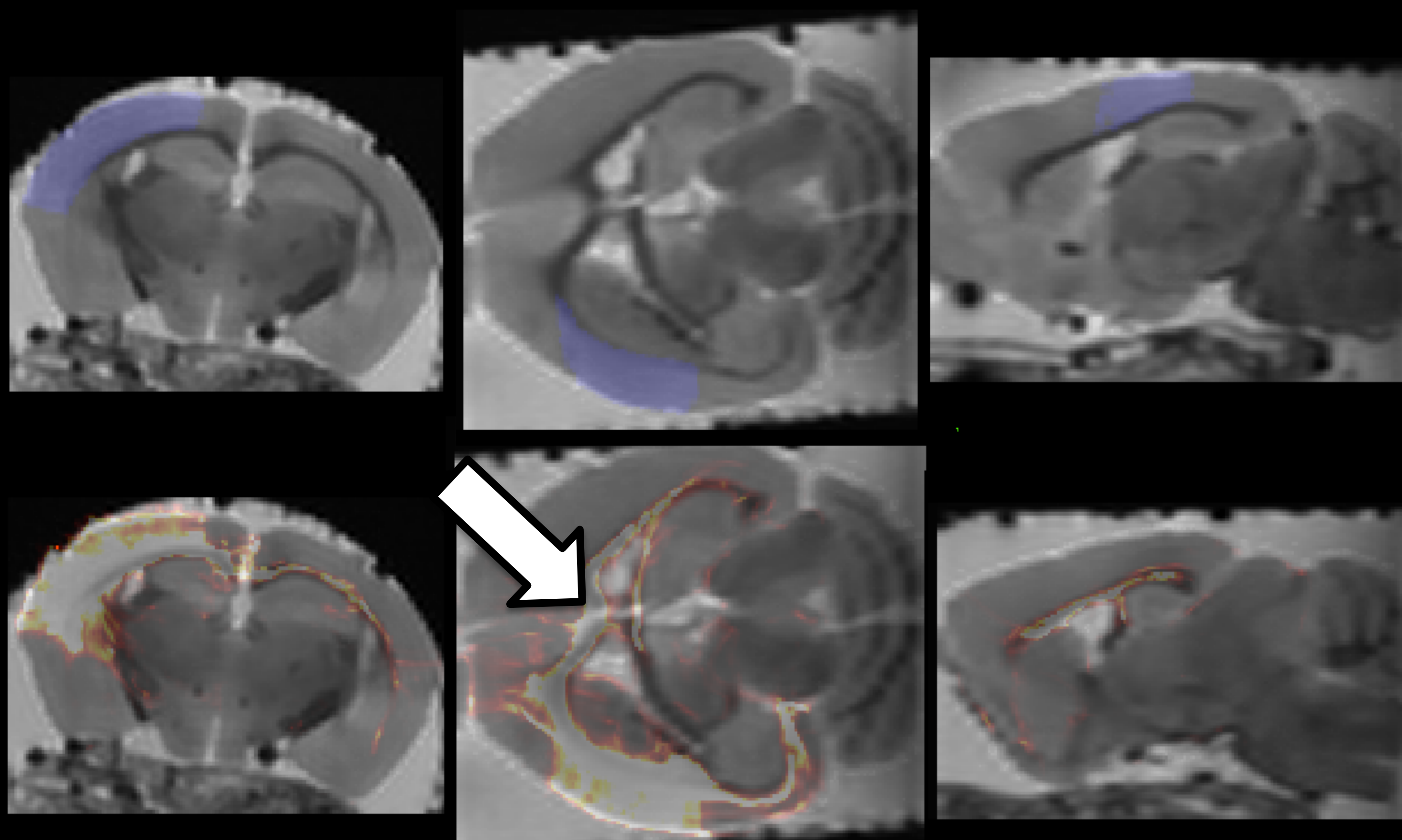
DTI Tracktography: Connectomy matrix

Tracktography can be executed e g between pairs of gray matter regions, creating **connectomy matrix**, of size N x N, where N is the number of brain regions ***matlab***, ***python***



DTI Tracktography: Example of Probabilistic tracktography

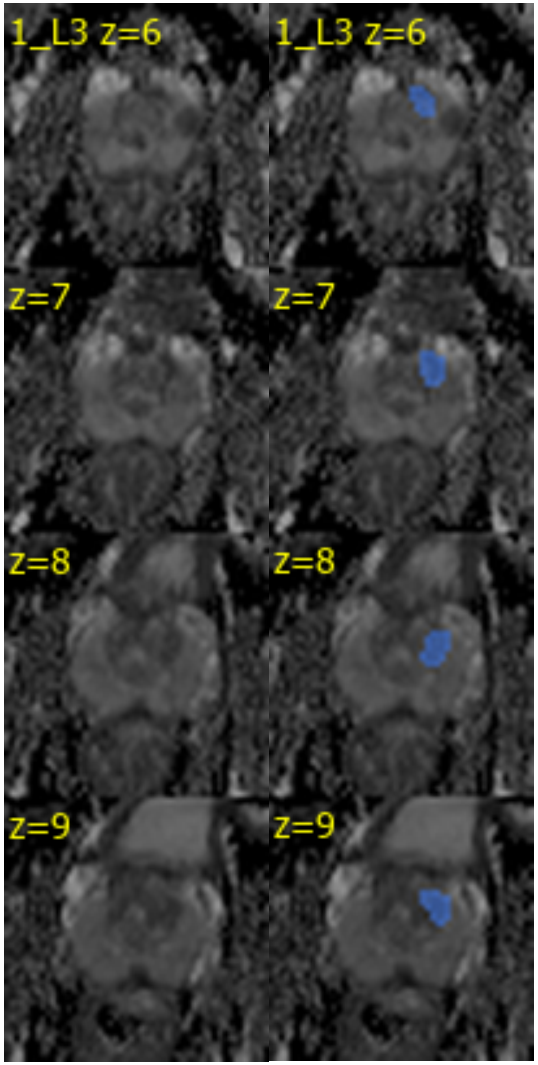
Murine TBI model right somatosensory cortex as seed region, arrow points corpus callosum where seed region connects to the opposite side



Region of Interest (ROI) placement

Region of Interest (ROI) placement

- Manual delineation
  - laborous
  - intra/inter-reader variability
- Automatic delineation
  - Atlas-based ***ANTs***, ***elastix***, ***SPM***, fsl ***fnirt, Freesurfer***:
    - Procedure: 1) align image to template 2) bring atlas to individual space 3) may require manual edits/QA
  - Surface segmentation ***Freesurfer***



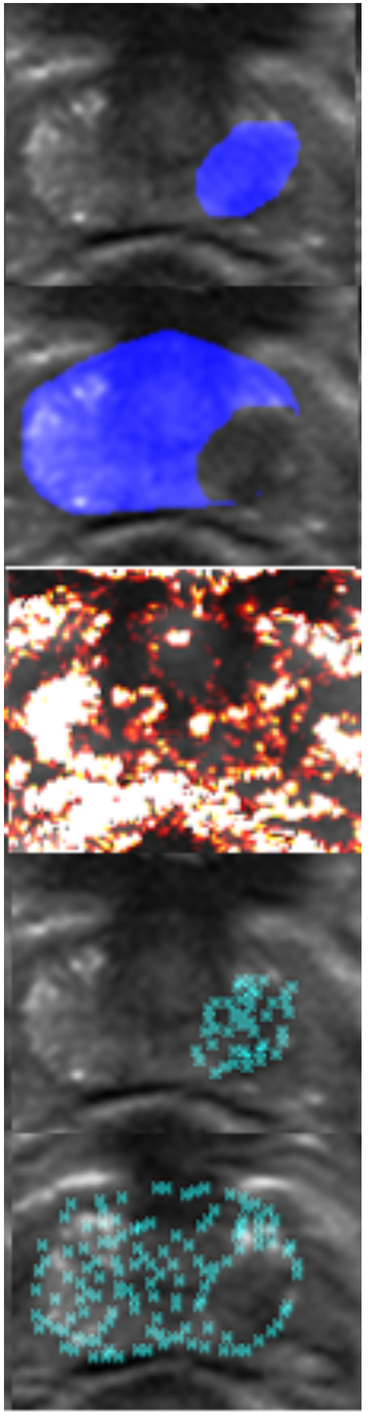
6 Radiomics

A radiomic feature is an extracted measure value from radiology data

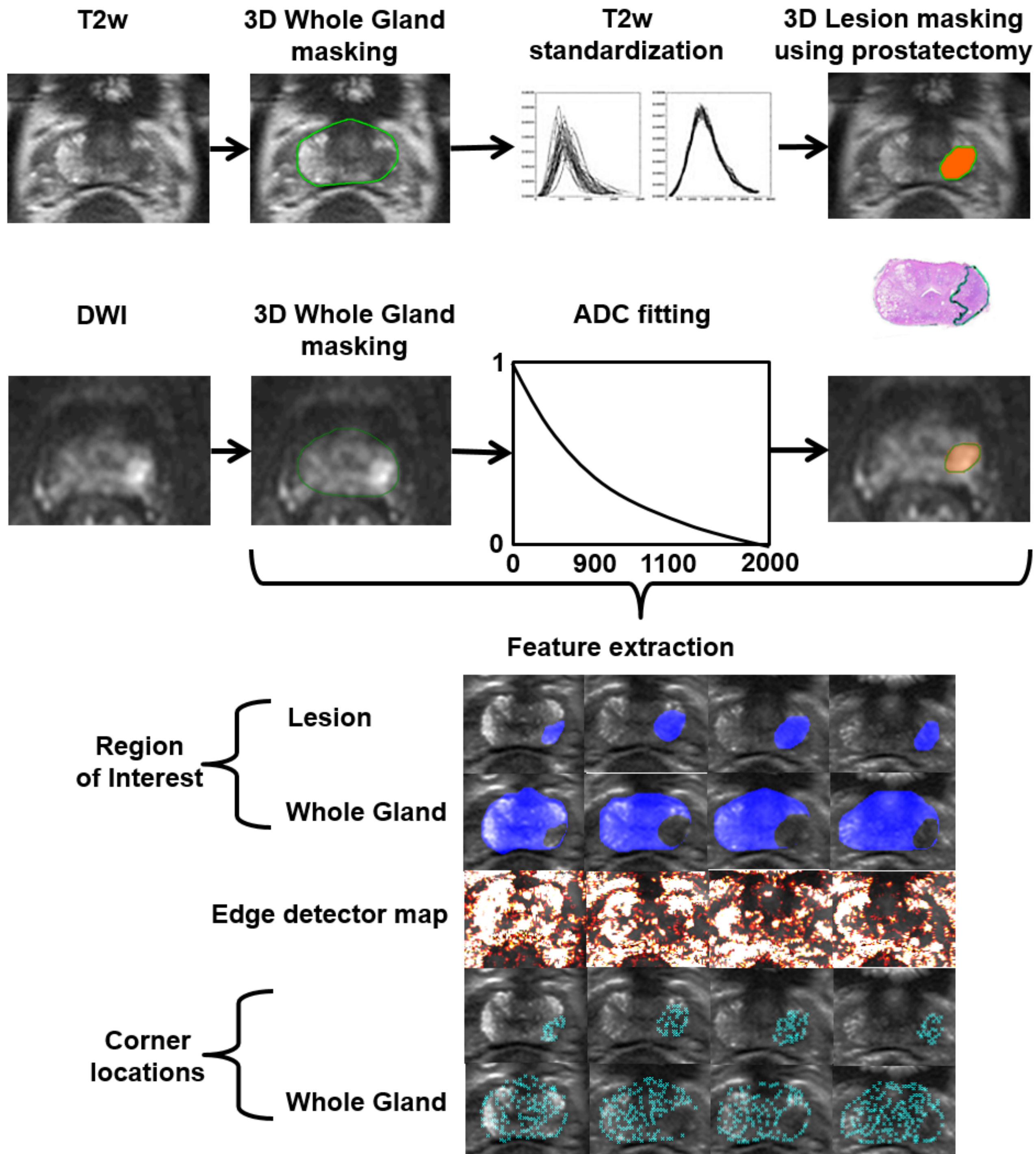
Field of study about radiomic features is ***Radiomics***

Radiomics

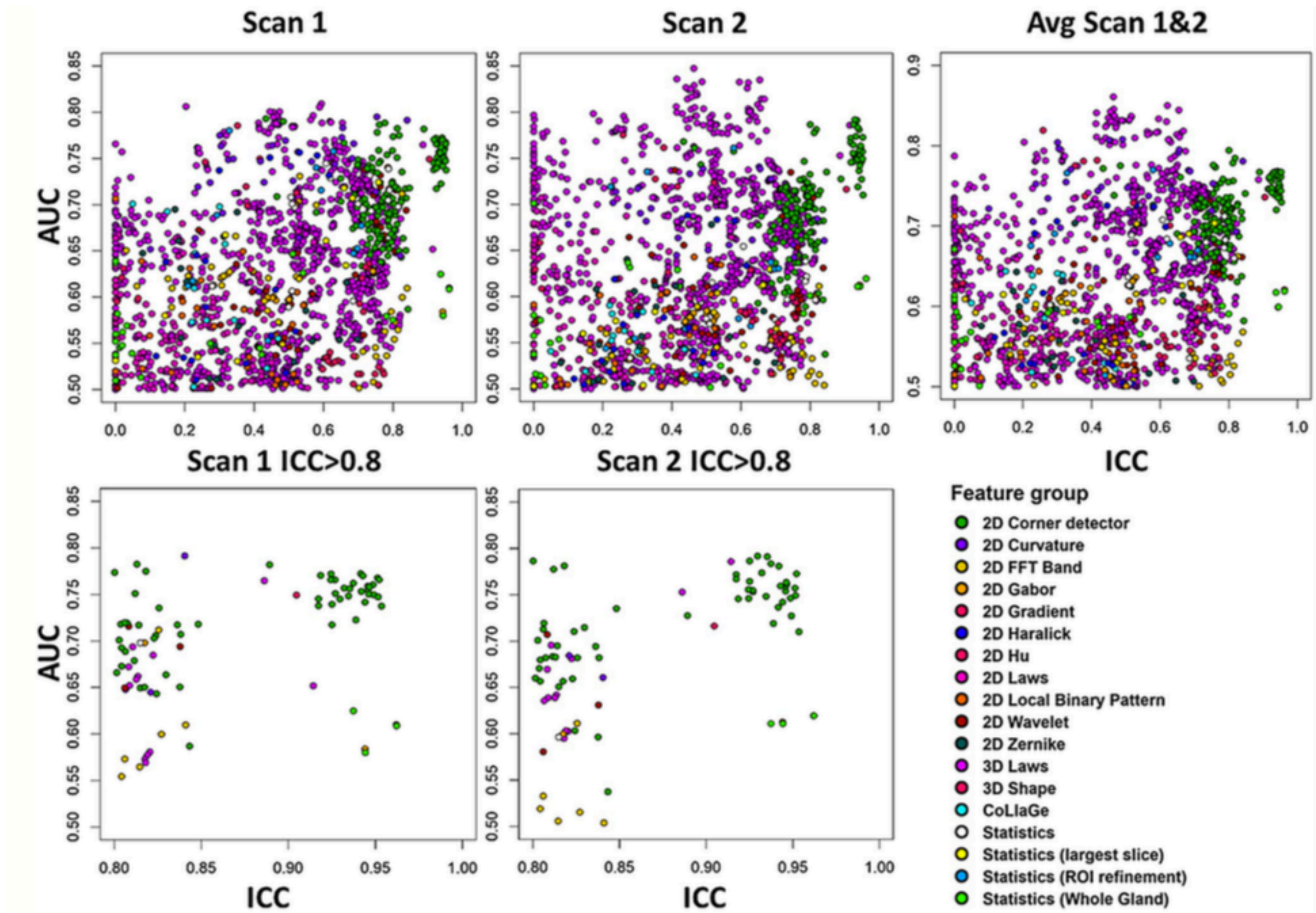
- Conventional analysis method is to take mean intensity (or median) in the ROI
- Radiomic feature extraction involves calculating other metrics from the voxel intensities inside ROI
  - Statistical descriptors (e g skewness, kurtosis percentiles)
  - Texture features (i e features describing textures)
  - Shape features (e g surface curvature, sphericity)
  - **Automatic features** (deep learning framework) or **Hand-crafted features** (by design)
- Performance in particular may depend on
  - Spatial resolution
  - Signal to noise ratio
  - Overall intensity level
- Field of study about radiomic features is called ***Radiomics***
- Tools: ***pyradiomics, MRC tools, pytorch, tensorflow***



Radiomic feature extraction example: Prostate cancer



Radiomics at group level: Repeatability vs clinical performance



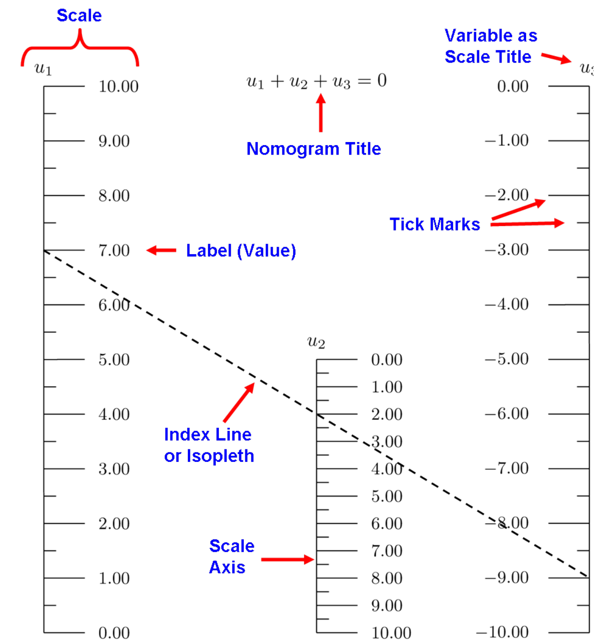
ICC(3,1) vs AUC for significant protate cancer classification with ADC, optimal features are in top right corner

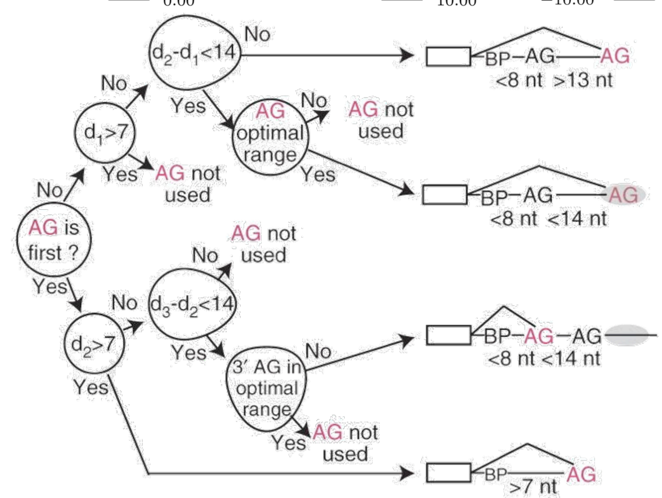
Radiomics: Feature selection

- Feature selection:
  - Selection of most relevant features relevant to the given problem
  - Crucial part of machine learning
  - Techniques
    - MRMR (Maximum Relevance Minimum Redundancy) algorithm
    - Univariate analysis
- Tools: R ***mRMRe,*** python ***scikit-learn****, pearson/spearman correlation between features, t-test/wilcoxon rank sum test for target groups, repeatability of features*

Computation of diagnostic/ prognostic scores: Models

- **Radiomic metrics** can be used **directly** as biomarkers
- **Nomogram** be used to combine multiple variables into one score
- **Decision trees** give detailed information about the logic of how classification is made
- **Regression models**
- **Neural networks** can be considered as large functions





7 Artificial Intelligence

At this point all measurements are done, now giving output for clinical use

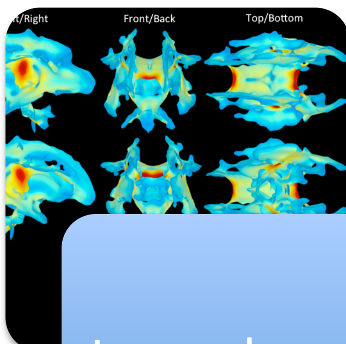
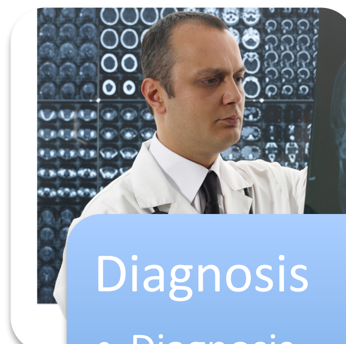


Image based metrics

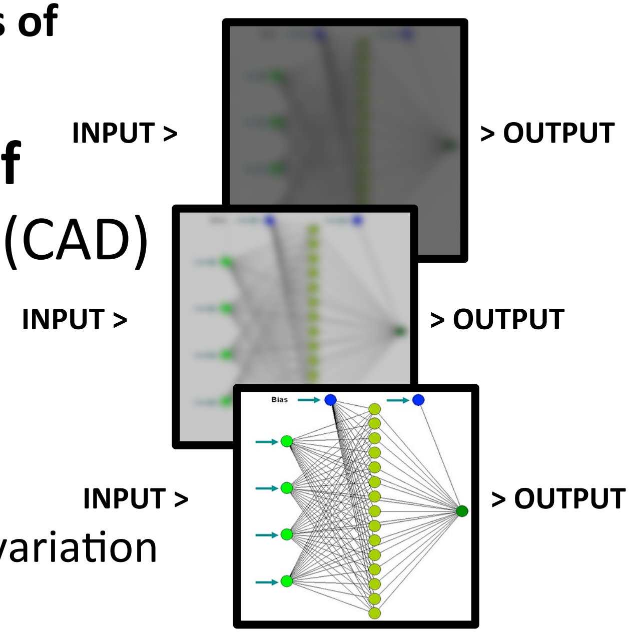


Diagnosis

- Diagnosis
- Prognosis
- Treatment decision

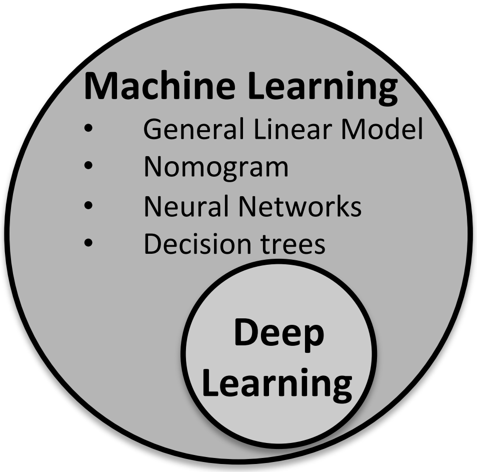
Artificial Intelligence

- Artificial Intelligence (AI) is used to tackle problem of large data
  - It is hard for human to benefit from all **large amounts of multidimensional data**
- AI can be used as a tool **assisting analysis of radiologist** in Computer Assisted Diagnosis (CAD)
  - CAD for region of interest
  - CAD for direct diagnosis/prognosis
- Pros & Cons of AI
  - Pro: AI is repeatable: no intra-reader or inter-reader variation with the same input
  - Con: ***black-box nature***
    - We do not know exactly why the AI tool gives certain specific answer – only that the answer is based on training data



Artificial Intelligence

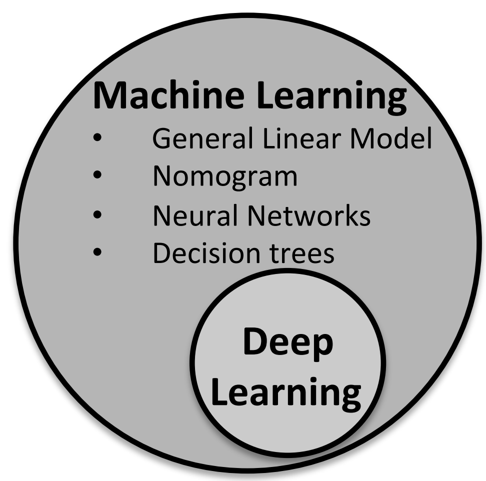
- **Machine learning** is generally used to train classifiers
  - Makes use of both radiomic features and clinical variables
  - Tools: R packages of: ***stats*** (GLM), ***rms*** (nomogram), ***neuralnet*** (ANN), ***rpart*** (decision trees), python ***scikit-learn***
- **Deep learning** is specific type of Machine Learning, where neural networks have many internal layers, making the 'deep' architecture
  - Input is either 2D (image slice) or 3D/4D (image volume, multiple channels)
  - Require high amount of data and processing power
  - Tools: python ***pytorch***, ***tensorflow***, require CUDA



Machine Learning

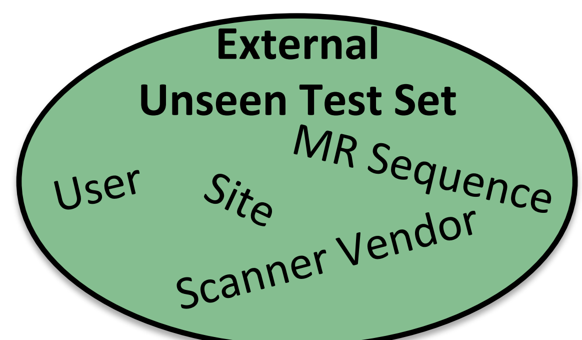
- Steps for doing machine learning:
  1. Preprocessing of data to uniform format
  2. Feature Extraction
  3. Data split
  4. Training/Evaluation

  5. (Final test of fixed developed models)



Training AI

- Data is typically divided into training and testing set, used to verify that the trained model captures correct signal and is not **overfitting** to data
- Separate test data is created by dividing the available data into portions
  - **Stratification**: ensure same proportion of groups are represented in both training and testing
- **Testing set** is sometimes called '**unseen**', '**external**', or '**hold out**'
  - Not used in any method development/adjustments
- For designing training strategy and developing models use **training set**

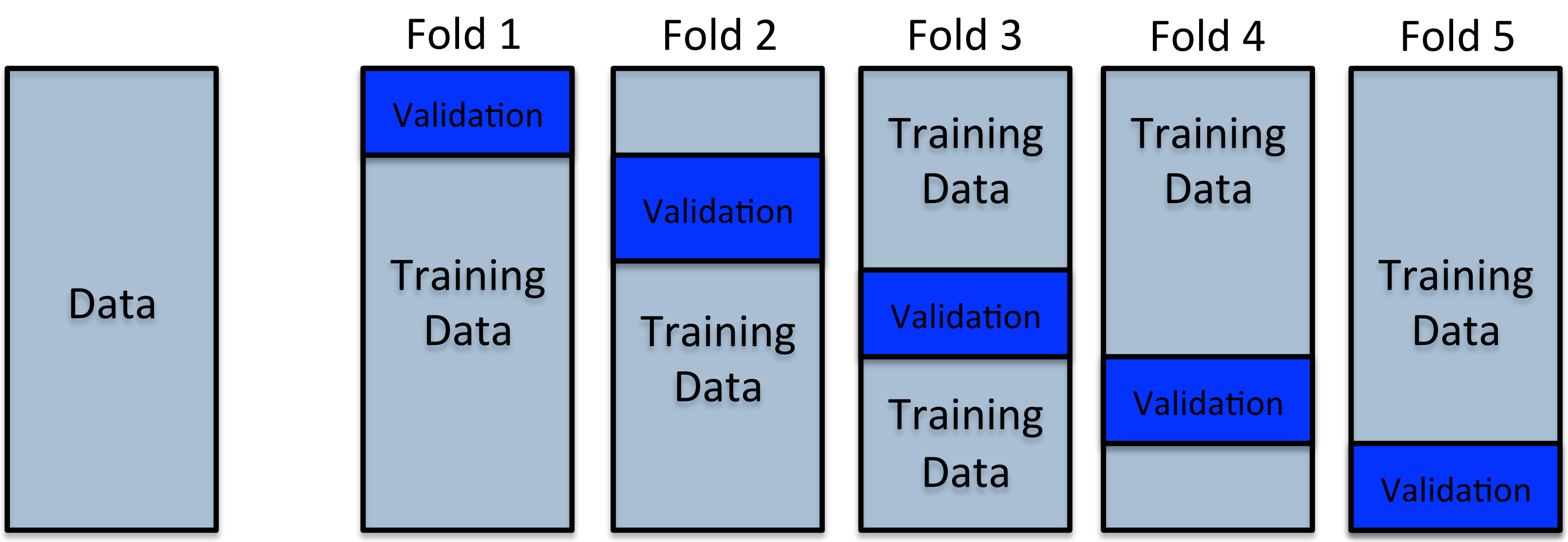


Cross-validation

- Typical N-fold cross-correlation analysis:
  - Data is split to folds
  - Each fold is used as validation set in turn to create estimate for performance
  - Running cross-validation multiple times causes multiple trials problem: likelyhood that we get good performance score by chance increases
- Cross-validation is used to obtain ***Unbiased estimation of generalization performance***

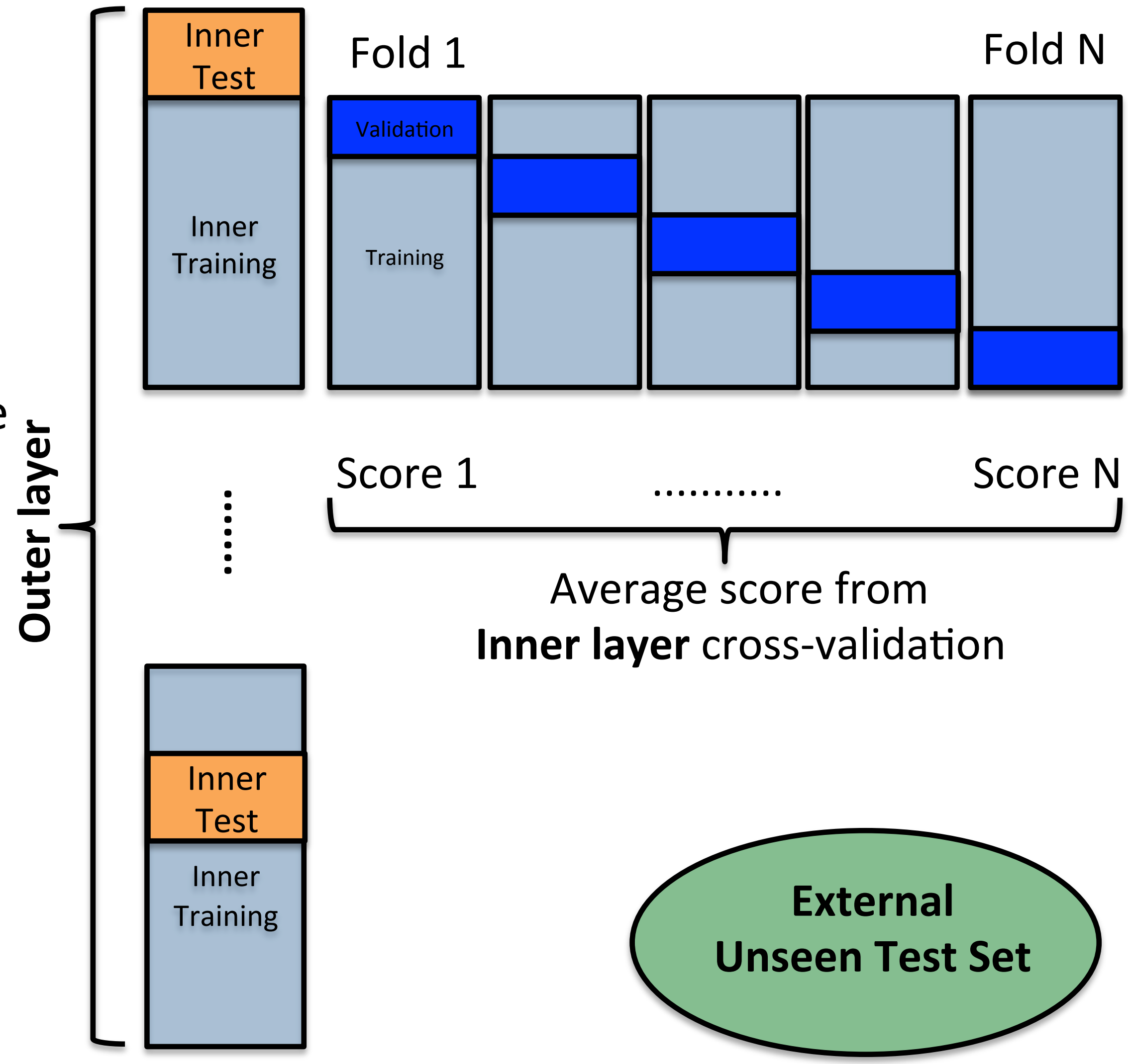
≠
in contrast to

Actual model for inference

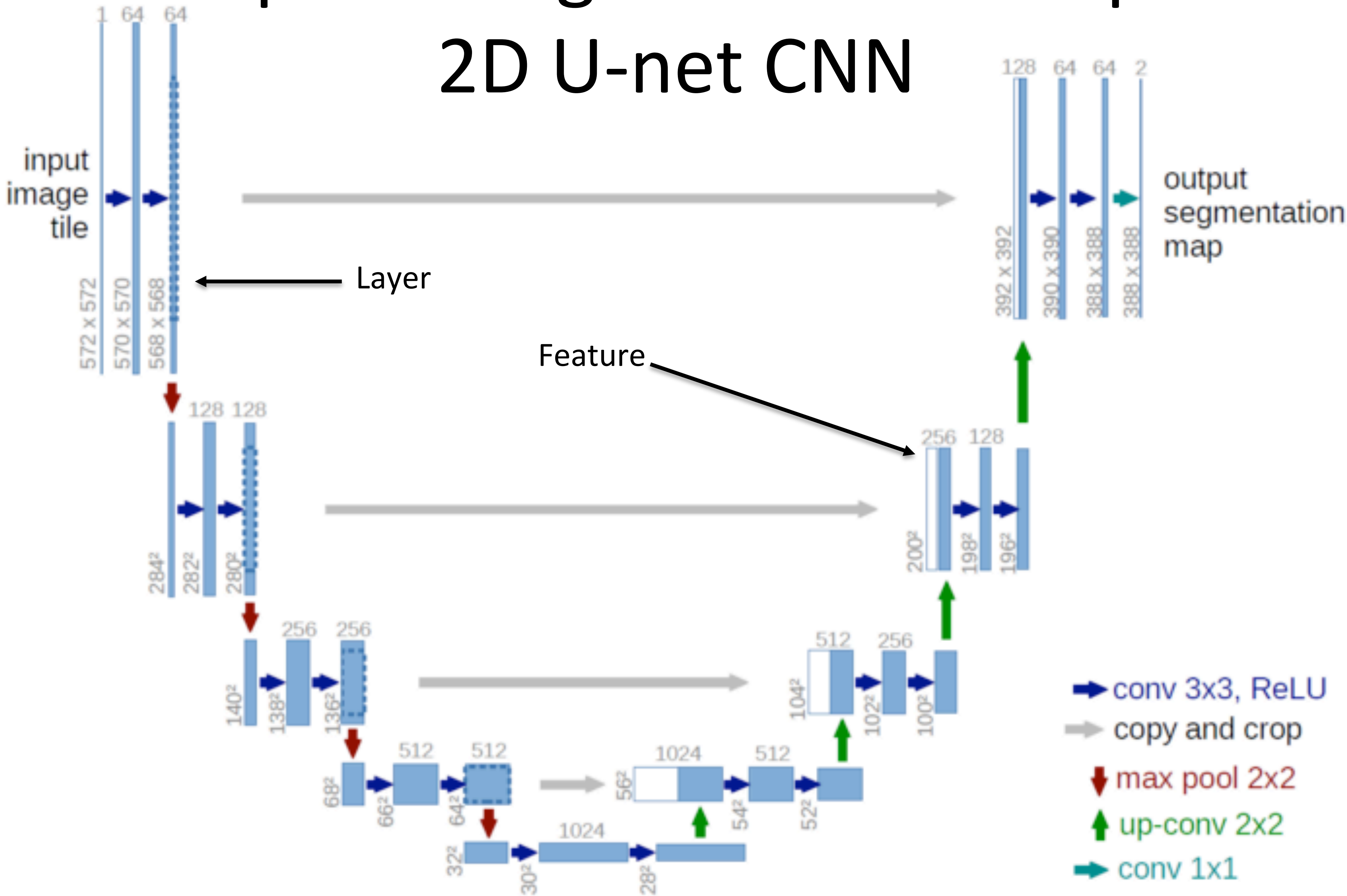


Nested Cross-validation

- Nested cross-validation (double cross-validation):
  - Cross-validation is repeated multiple times
  - Each repetition average score is calculated (as in cross-validation) with independent Test set (Inner Test)
  - **Inner layer** is used multiple times to choose model, hyperparameters, etc.
  - **Outer layer** gives unbiased estimator of generalization performance for the *whole process*

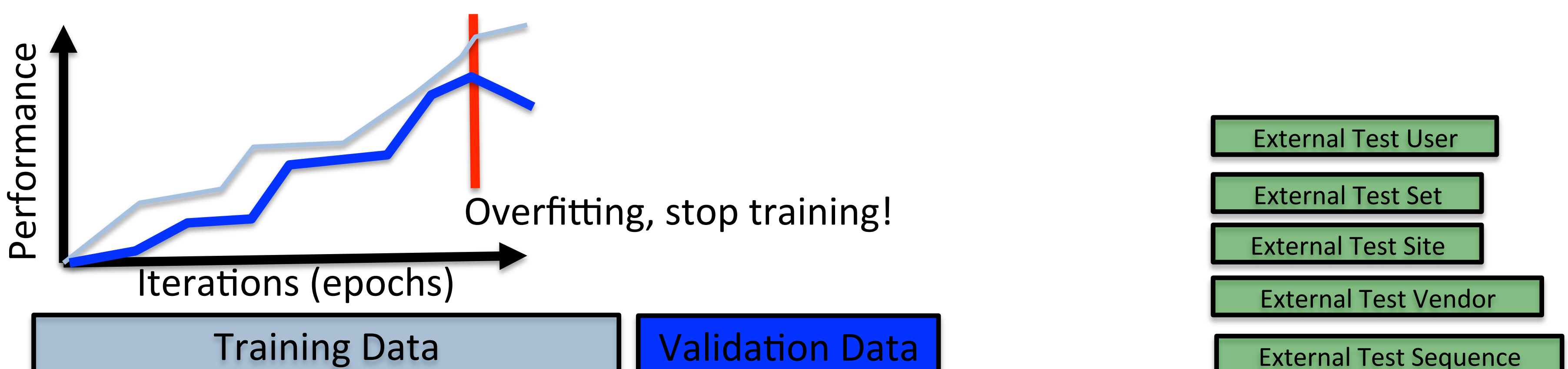


Deep learning network example: 2D U-net CNN



Deep learning

- Convention is to use one validation set since execution of evaluations is costly, even with GPU
- **Validation set** is used to see how well the model performs in external data
- When validation performance starts to drop while training performance increases, model started to **overfit to the data**
- **Ending criteria** is also part of the model development, so external test set is needed to verify the performance of trained model



Machine Learning in Medicine

- Prevalence of a condition/disease may be small
- Usually only one sample per case
- Proper Ground Truth –data not always available

Thank you!

Any questions relating to MRI data processing?