



Gaussian Processes models in developmental neuroimaging

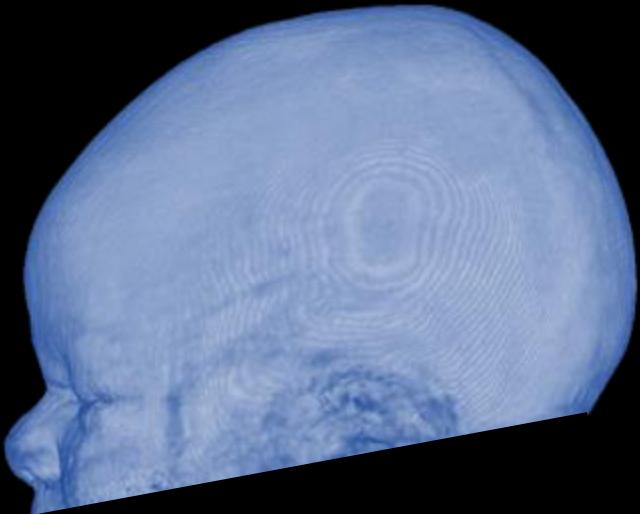
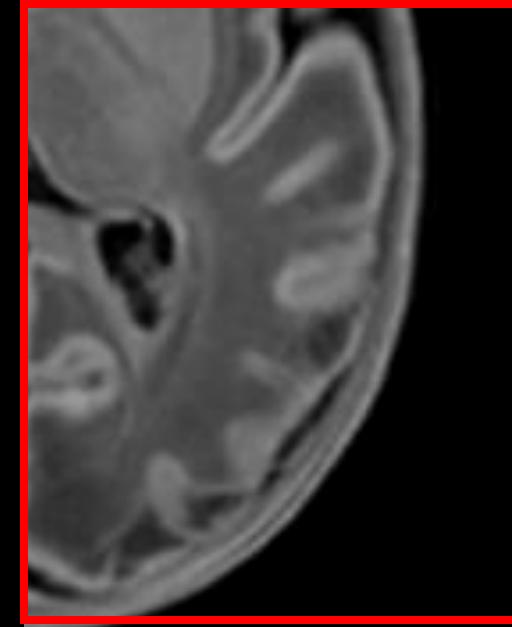
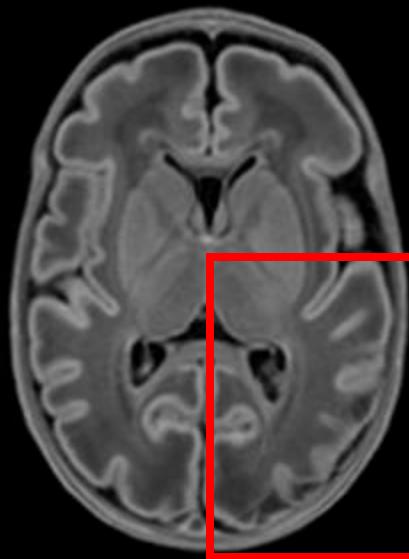
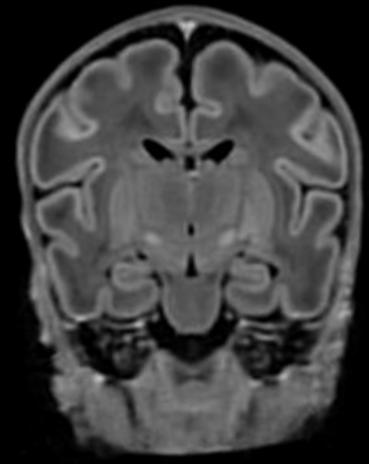
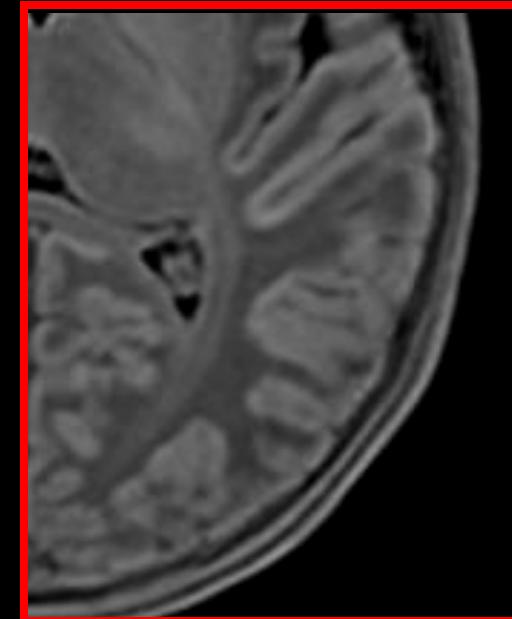
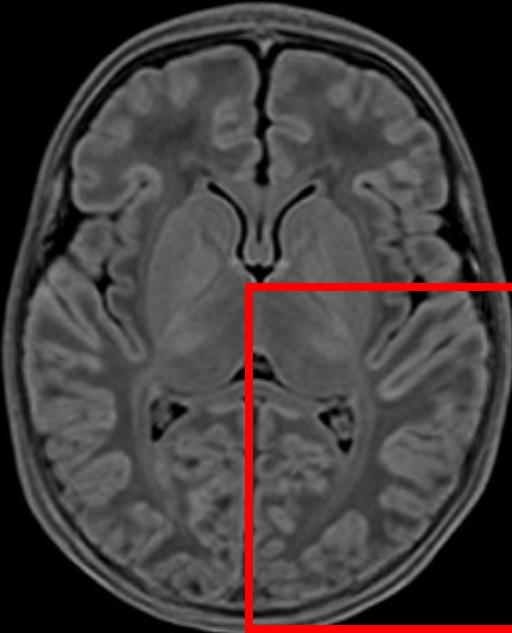
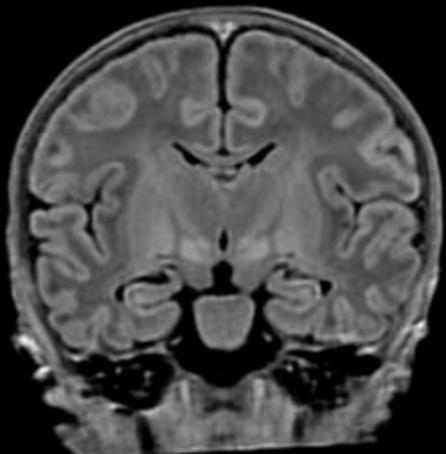
Jonathan O'Muircheartaigh

Summary

- I want to be able to detect if a single (cross-sectional) data point is different than expected over the neonatal / childhood period
- Given that the population average and standard deviation is changing with age in both **intensity** and **shape (deformation in 3 dimensions)**
- I probably don't know the real relationships between age and data
- I want to be able to interpret resulting maps
- Data is very noisy, probably lots missing

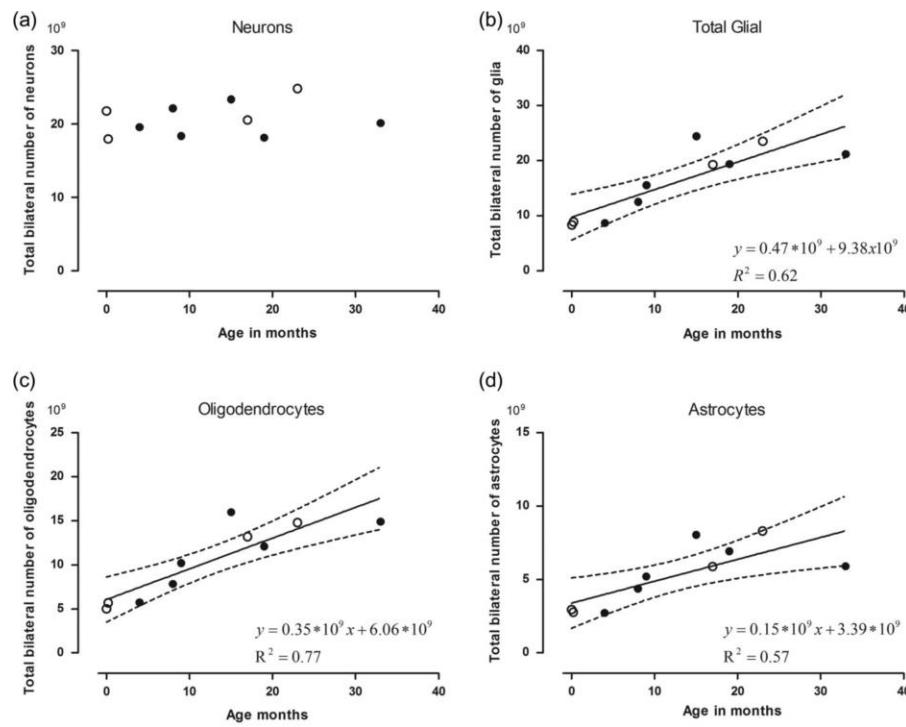
39 Weeks

29 Weeks



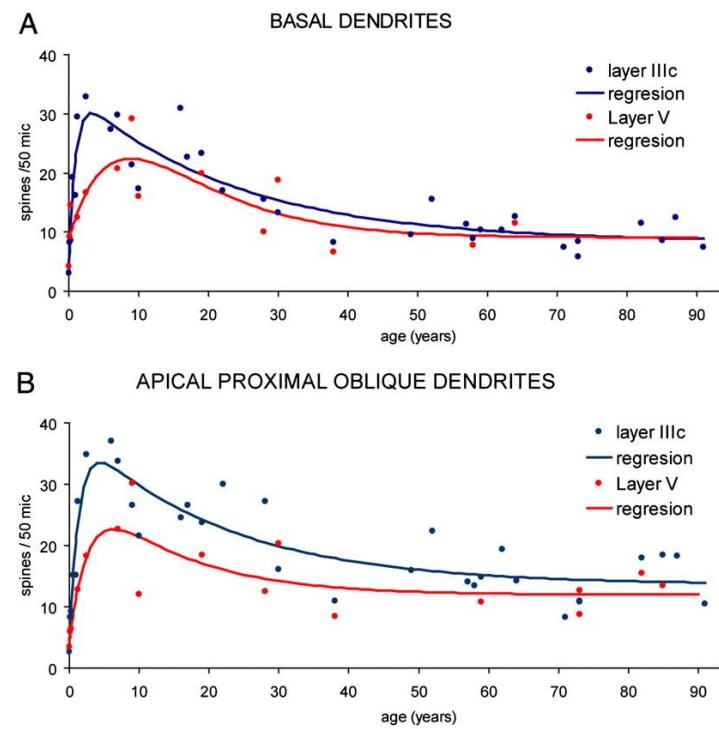
What does these changes reflect? (probably)

Brain Cell Count

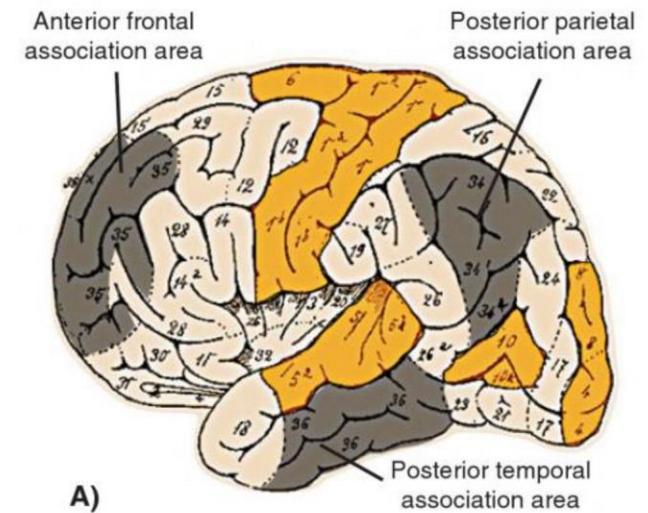


(Kjær et al, 2016)

Synaptic Density

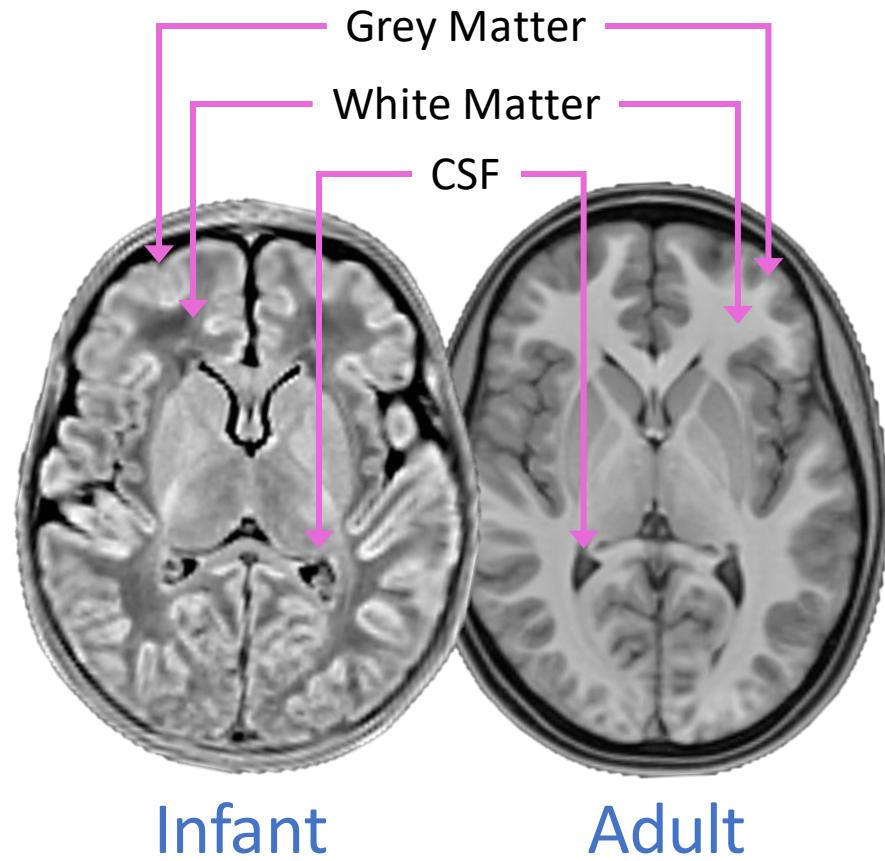


(Petanjek et al 2011)

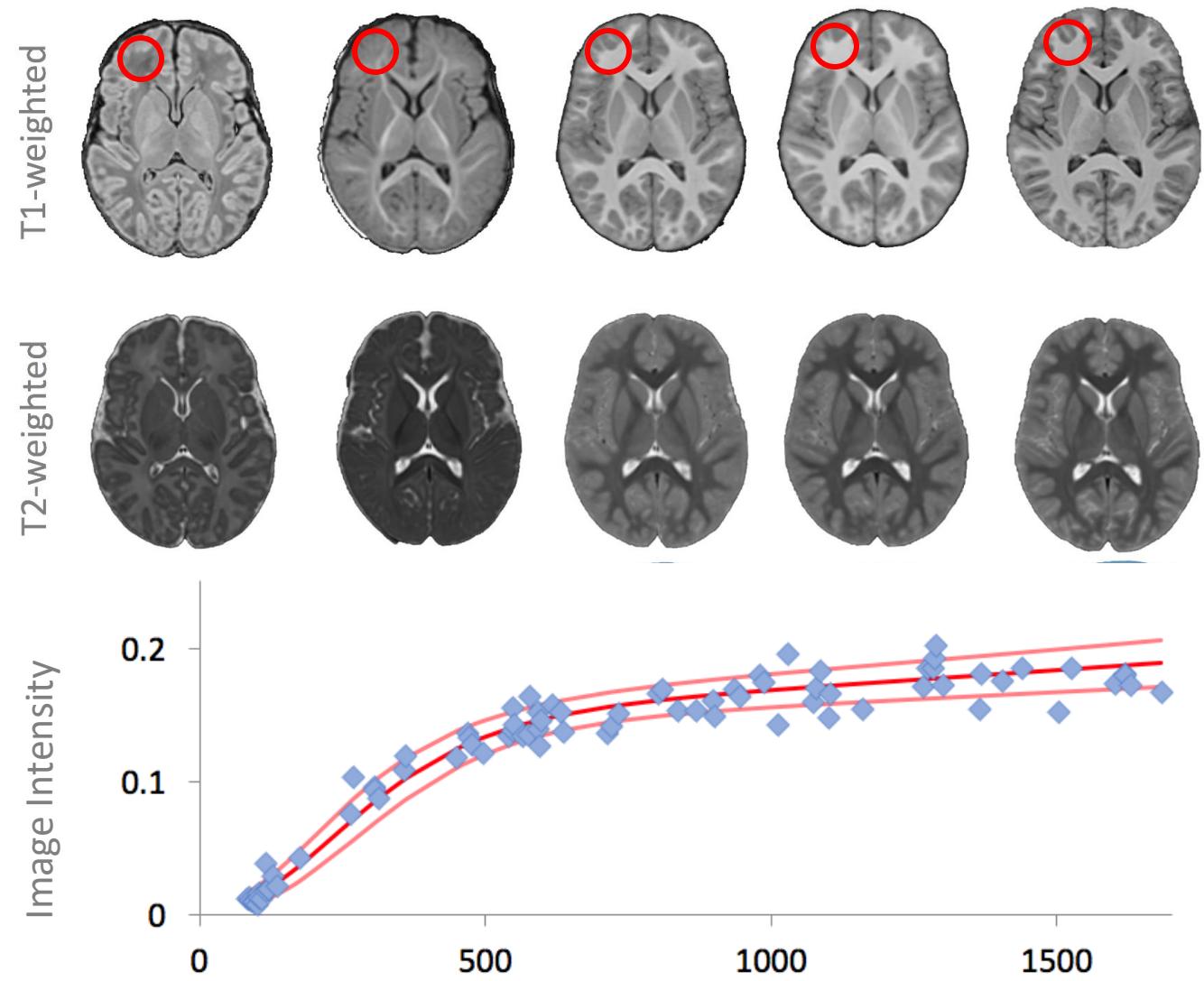


Flechsig (1909)

What does this look like in an image?

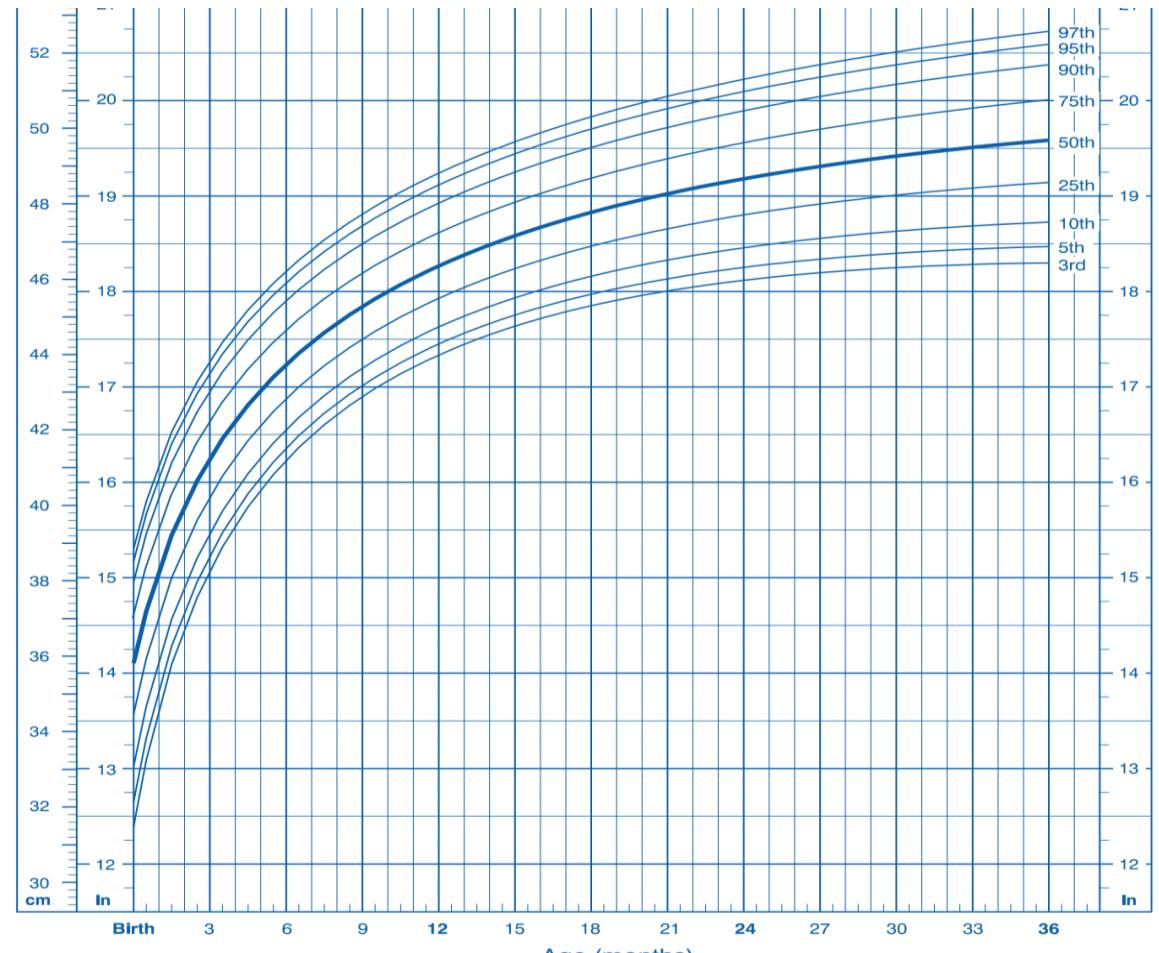


Dean et al 2014



What do I want from a model?

- Ideally something interpretable →
- A curve such as this took 1000s of samples
- Parametric (e.g. sigmoid, quadratic, gompertz) are tricky to fit over larger scales – only describe discrete periods / places in the lifespan.
- Non-parametric are tricky to optimise & can be difficult to interpret



Published May 30, 2000.

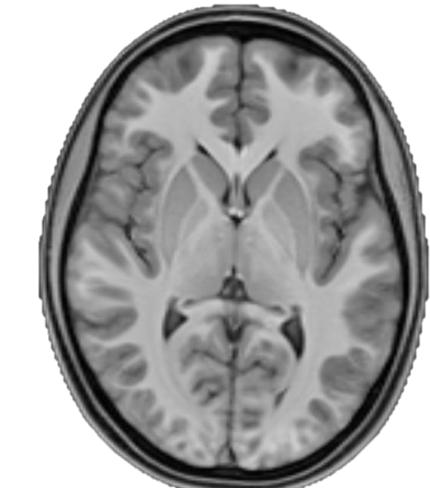
SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).



SAFER • HEALTHIER • PEOPLE™

Starting with something parametric....

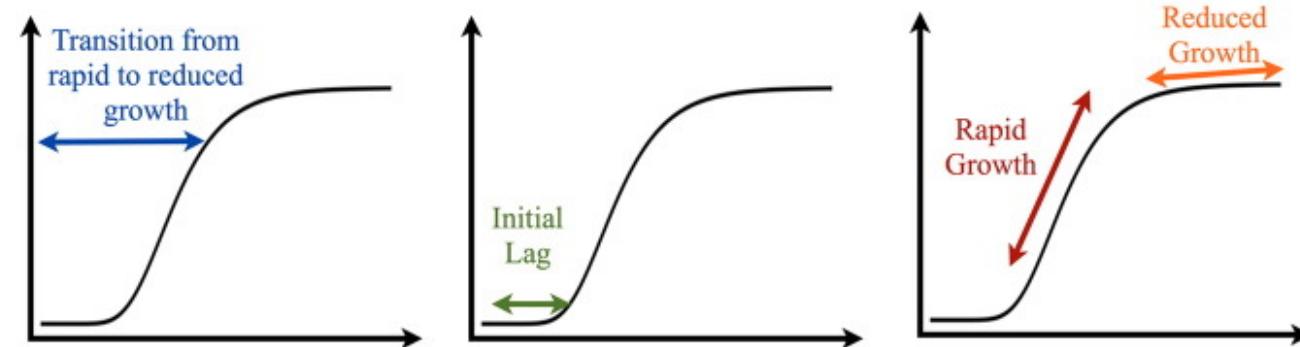
- Brown Baby Imaging Study
- >600 children aged 3months to 5 years scanned during natural sleep
- Longitudinal scans every 6-12 months
- Primary imaging sequence multi-component relaxation mapping (mcDESPOT) – similar contrast to quantitative T1 but with a model to account for free water (Deoni et al 2012)
- Primary aim to parametrically map white matter development



Parametric Models using qMRI

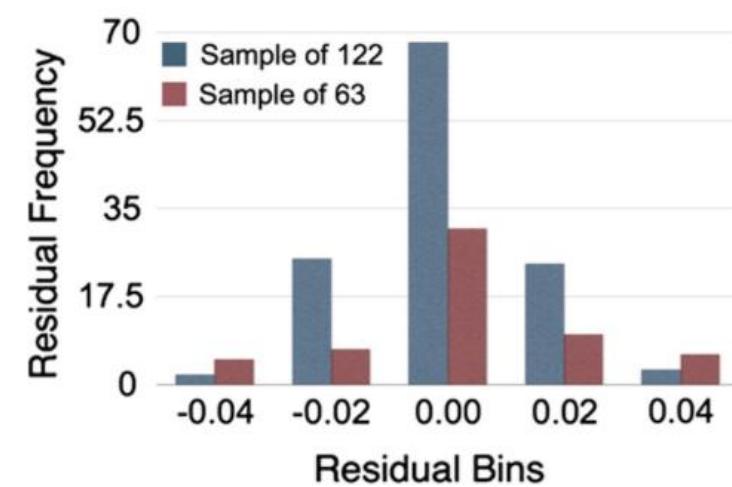
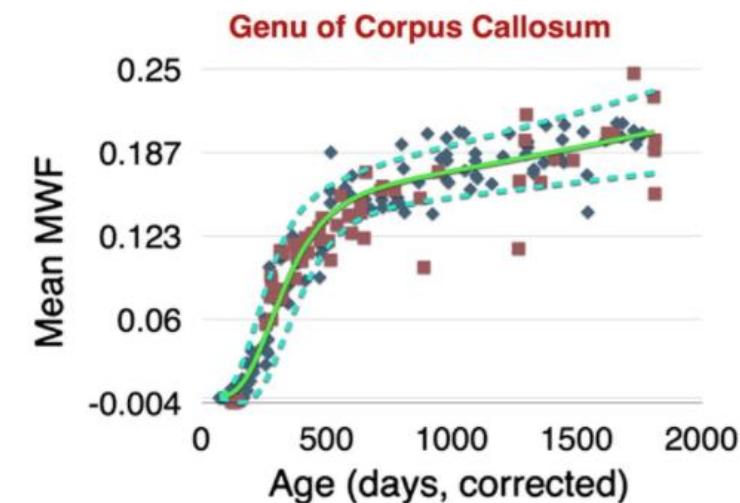
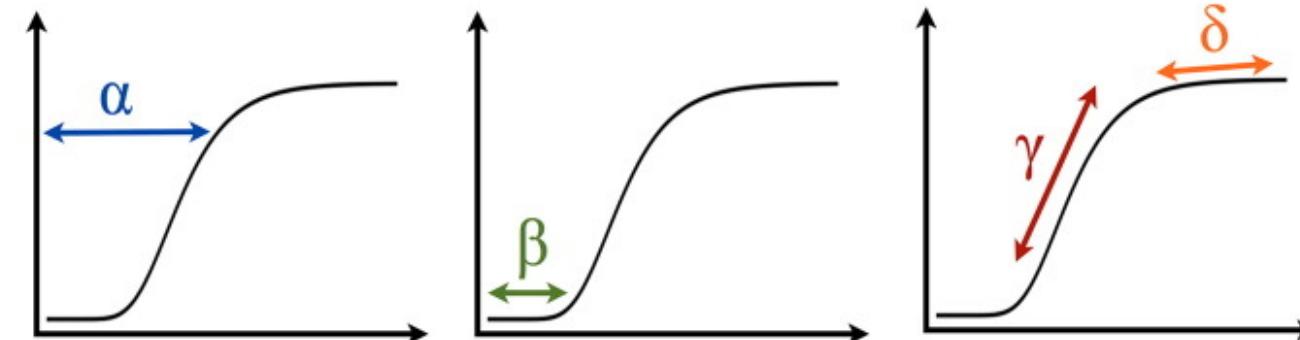
A

Characteristics of Sigmoid Function

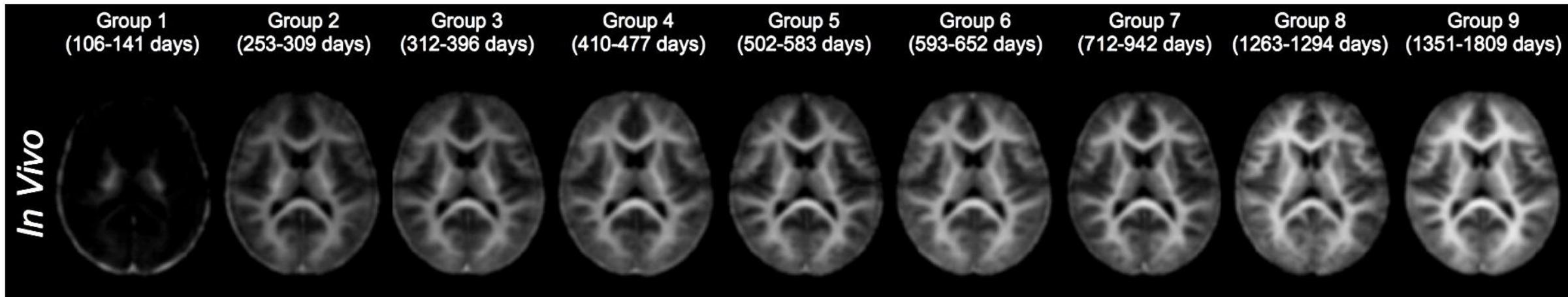


B

Modified Gompertz Function: $MWF(\text{age}) = \alpha e^{-e^{\beta-\gamma*\text{age}}} + \delta * \text{age}$



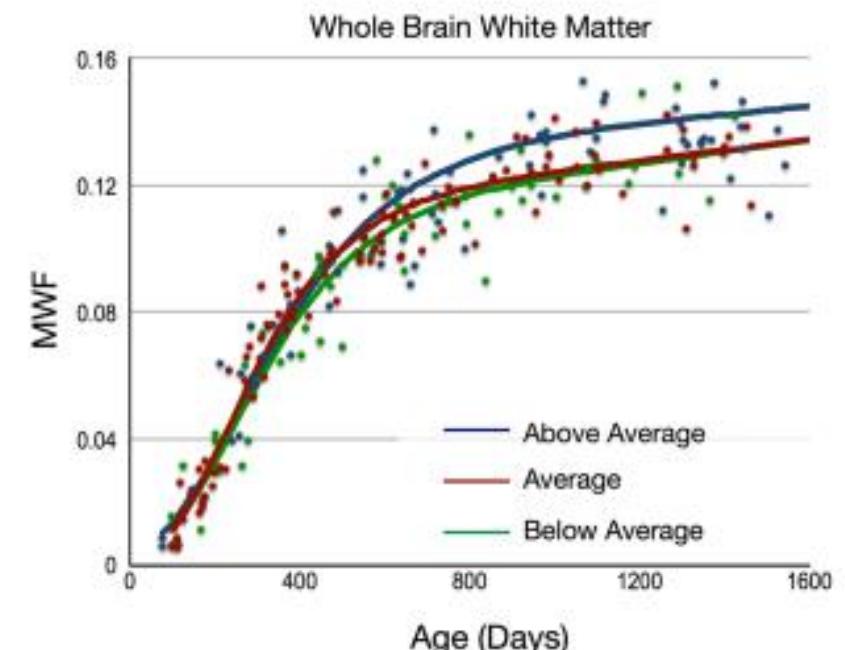
Parametric Models using qMRI



- Dean et al (2014) using a quantitative MRI sequence (mcDESPOT) to measure tissue content over development
- Fitting a four parameter Gompertz model (voxelwise) to describe developmental trajectories (see Sadeghi et al, 2013, for diffusion)
- Fits the data very well in this age group (also bone, muscle etc) but loses some of the goodness in older / younger age groups

Once you have a model?

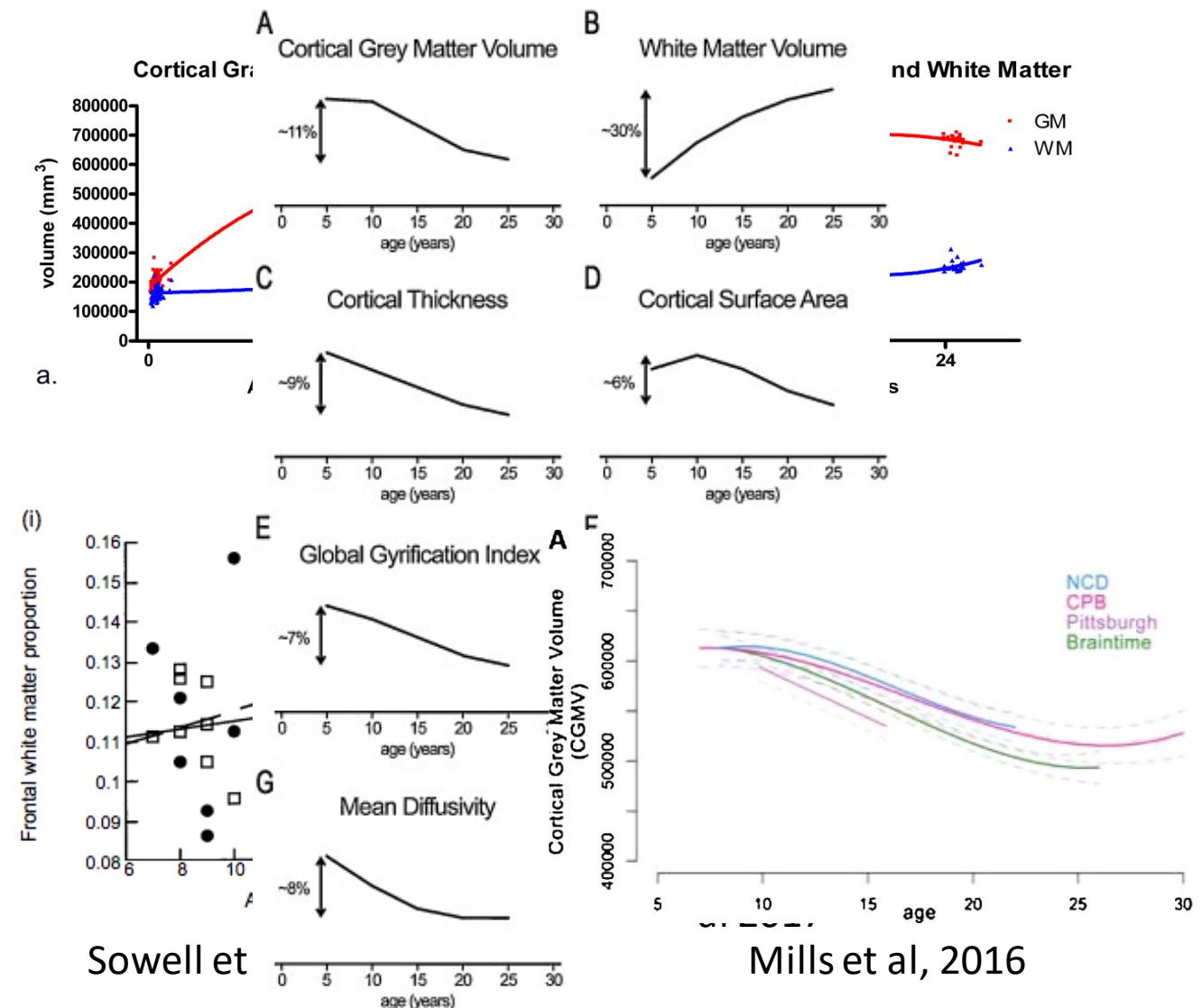
- **Every time you include age as a covariate you are implicitly saying you know the relationship is linear (or log or whatever)
- Instead calculate deviations from a model – individual differences (Z) at different age points - Cross-sectional – percentiles from bootstrapping
- Can calculate statistics on the parameters themselves (e.g. does the variability of growth rates between subjects relate to neurodevelopment) - Longitudinal



Deoni et al 2015

Downsides to parametric models

- What curves do we pick?
- How many parameters?
- Over what age range?
- A line will best fit an arbitrarily small age range in most instances
- What parameters to fit to which parameters?
- What if you don't know?



Modelling when you're comfortable that you've no idea the model shape should be...

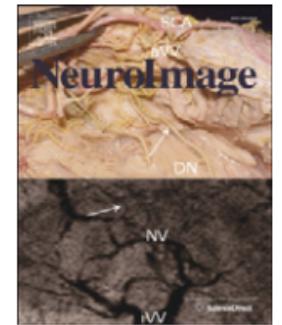
NeuroImage 97 (2014) 333–348



Contents lists available at [ScienceDirect](#)

NeuroImage

journal homepage: www.elsevier.com/locate/ynimng



Individualized Gaussian process-based prediction and detection of local
and global gray matter abnormalities in elderly subjects



G. Ziegler ^{a,b,*}, G.R. Ridgway ^a, R. Dahnke ^b, C. Gaser ^{b,c}, for The Alzheimer's Disease Neuroimaging Initiative¹

^a Wellcome Trust Center for Neuroimaging, Institute of Neurology, London, UK

^b Department of Psychiatry, Jena University Hospital, Jena, Germany

^c Department of Neurology, Jena University Hospital, Jena, Germany

Gaussian Process Regression

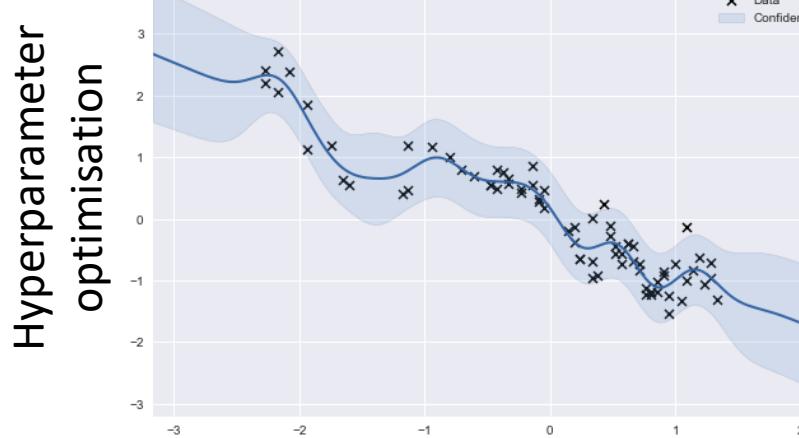
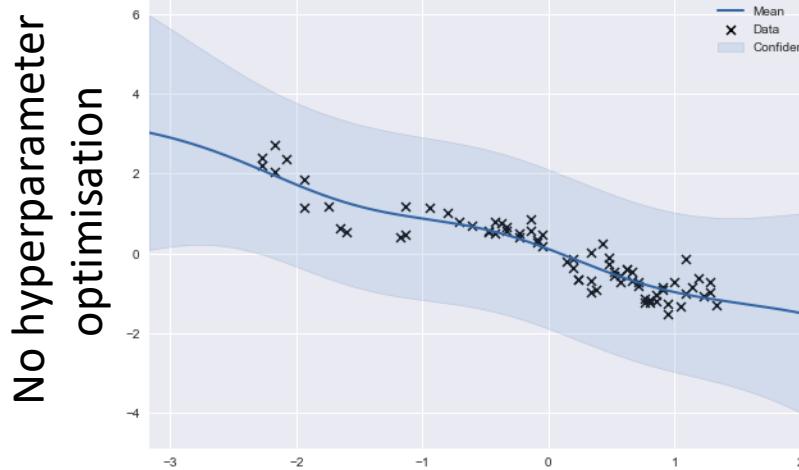
- A form of Bayesian “non-parametric” regression
- Actually very parametric – a distribution of possible functions (constrained by the relationship defined in the covariance kernel), given the data
- ~Linear regression if a linear covariance kernel is used
- ~Non-linear if a non-linear covariance kernel is used (like here)

Gaussian Process Regression

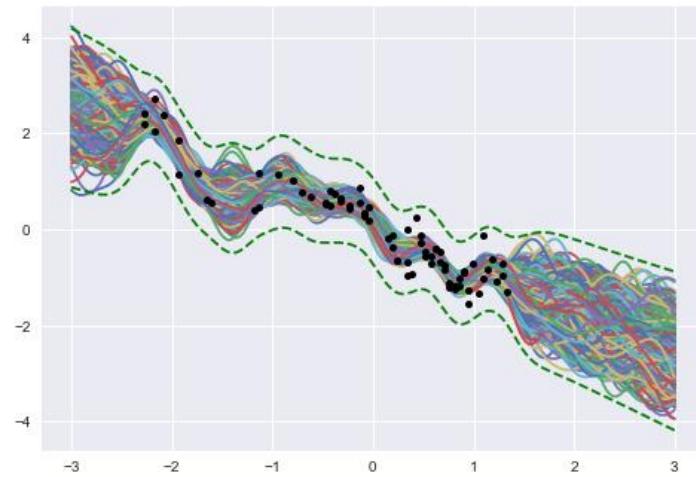
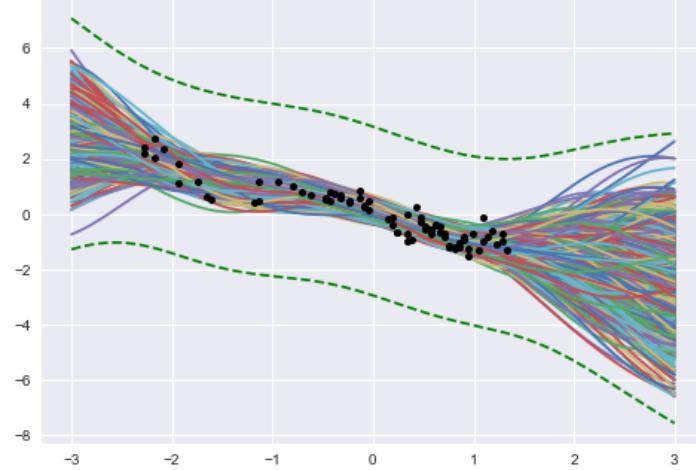
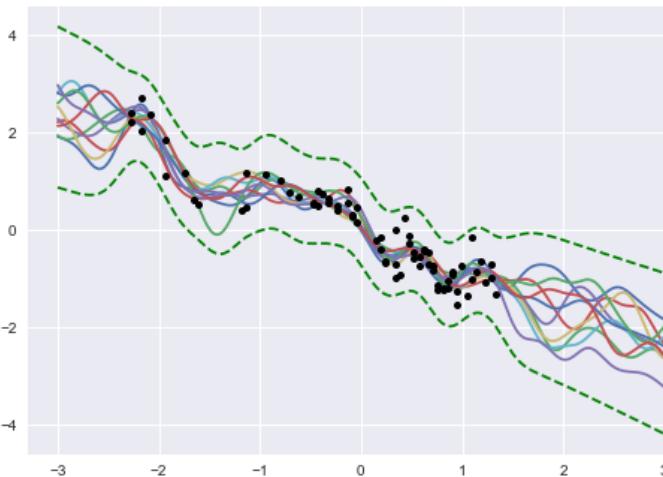
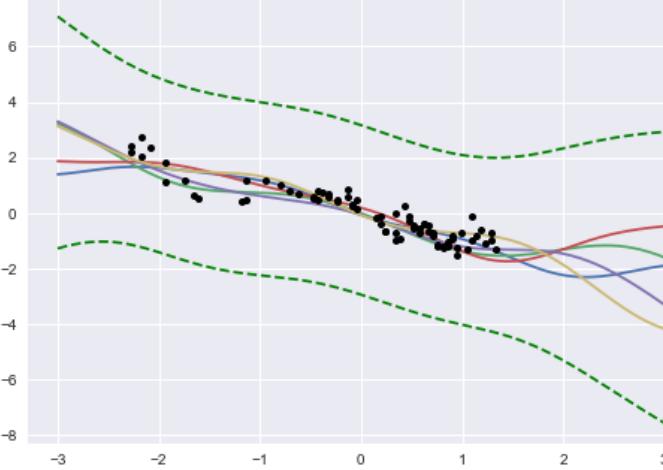
- For a squared exponential kernel (smooth curves – datapoints are locally related)
- Just as in parametric non-linear regression, needs some initial estimate of **memory / wiggliness** of functions (length scale) BUT doesn't fit a single function – provides a distribution
- Hyperparameters can be optimised analytically using log marginal likelihood (not through cross validation)
- Importantly provides a form on prediction interval
- A bit similar to bootstrapping in the sense that you are using the range of possible functions to provide your CI / PI

Distributions of functions (fake data)

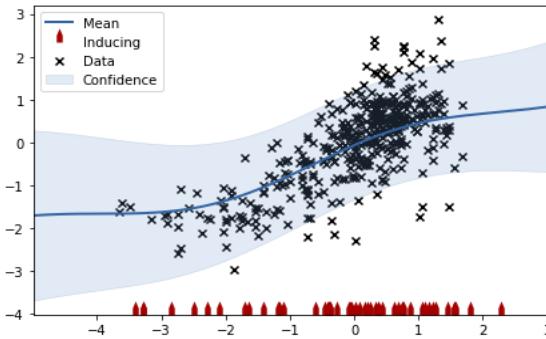
Linear + SE kernel



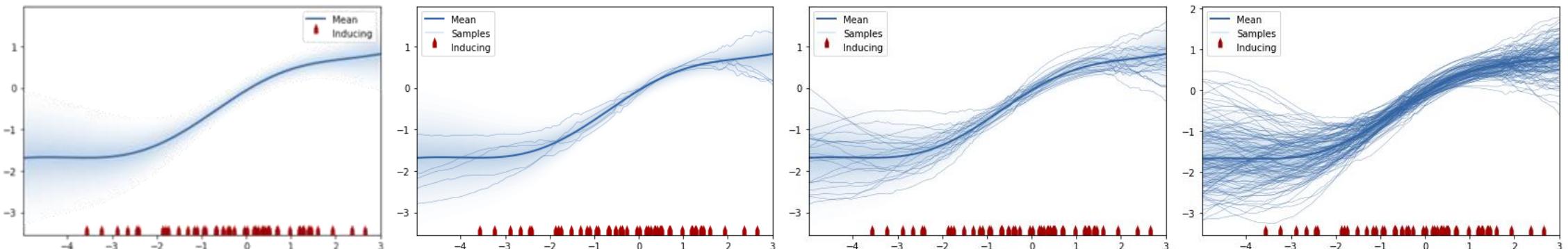
Samples from the posterior



Distributions of functions (real data)

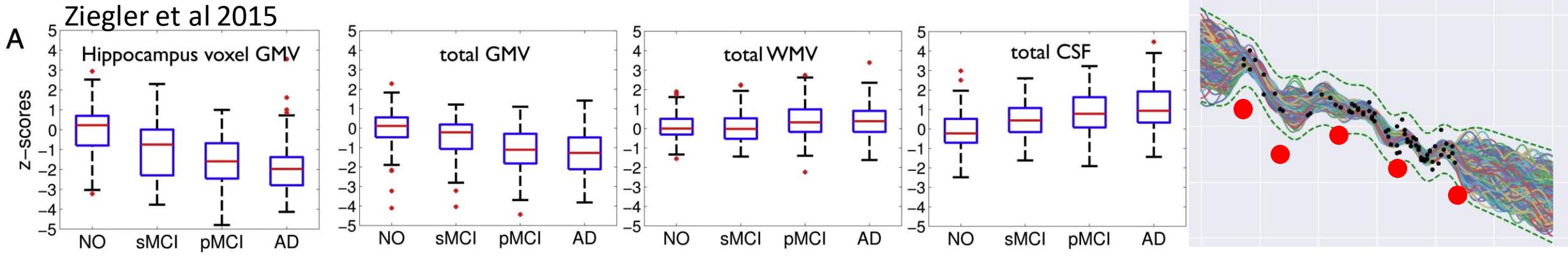


- Where there are observations, you have fairly high certainty of where a new datapoint will lie, where there aren't, you don't
- May seem obvious but hey.



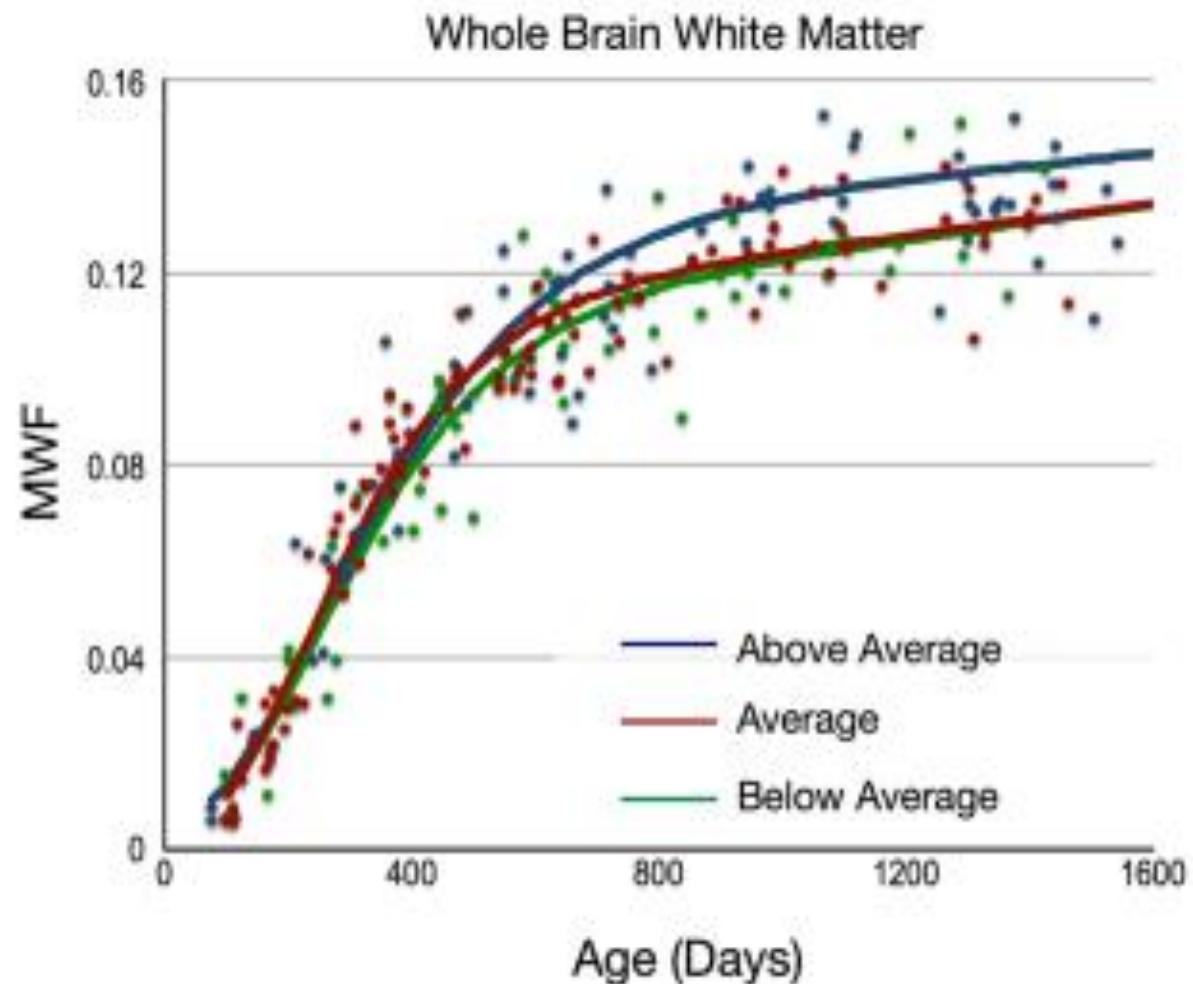
Use in MRI?

- Classification problems (Marquand et al, 2014, Ecker et al 2017)
- Feed in large multi-variable data (e.g. MRI volume) to predict categorical outcome (e.g. disease state) or continuous (brain age etc)
- Switching this round isn't a problem and is more standard for GPs
- Quantify deviations from the model (e.g. Z score as in Ziegler, 2014) – just like residuals from a regression (~IQ, head circumference etc)



Back to the Brown Imaging Study

- Dean et al used a parametric Gompertz model generally had a pretty good fit to observed MRI data in infants and children
- Fitting a GP model (separately at every voxel) with a SE kernel shows reduced MAE and has a better qualitative fit in certain areas compared to parametric (I picked a really obvious bit)



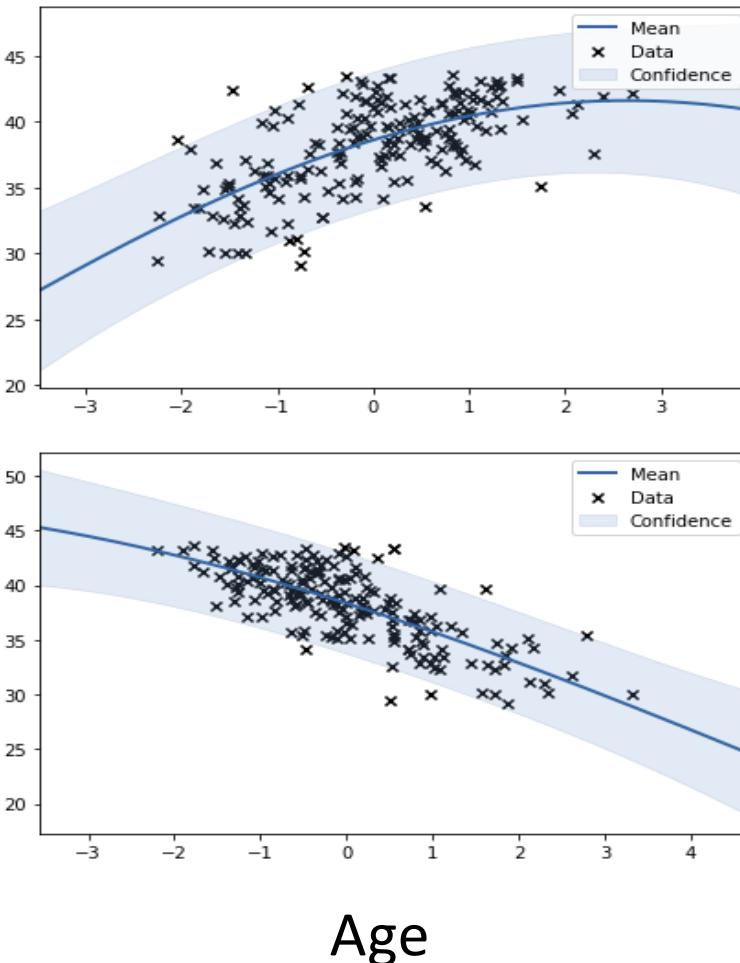
What about multiple correlated output

- The method can be extended to multiple output (vector-valued outputs) and indeed function outputs (for longitudinal data)
- Multiple outputs in imaging is pretty attractive as contrasts are (a) correlated and (b) noisy
- Multi-output GPs can address both problems
- Here using the intrinsic coregionalization approach in GPs (reviewed in Álvarez et al 2011)
- You fit one set of hyperparameters but with a further co-regionalisation matrix that encodes similarity between outputs and (through Kronecker product) a scaling of the hyperparameters per output

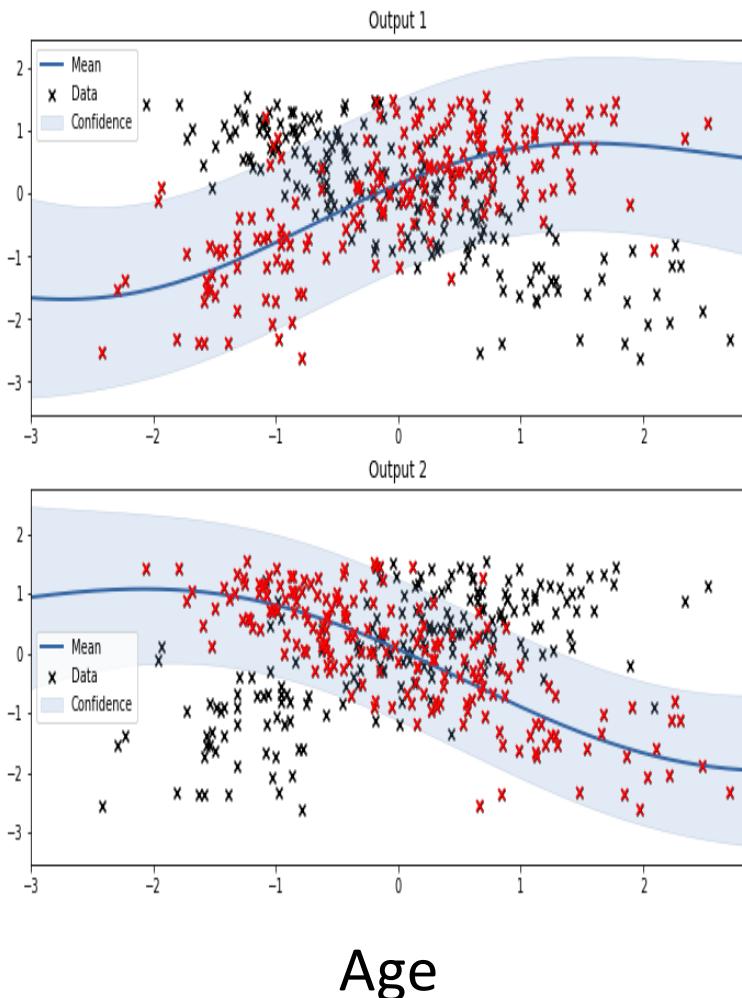
Multiple outputs (e.g. vector valued output)

Single Modality Training

T1w



Joint Training



The values of T1 and T2 voxels are (largely) correlated.

The joint information (encoded by a covariance matrix over outputs) can act to regularise estimates of one where data is noisy (most MRI / paediatric data)

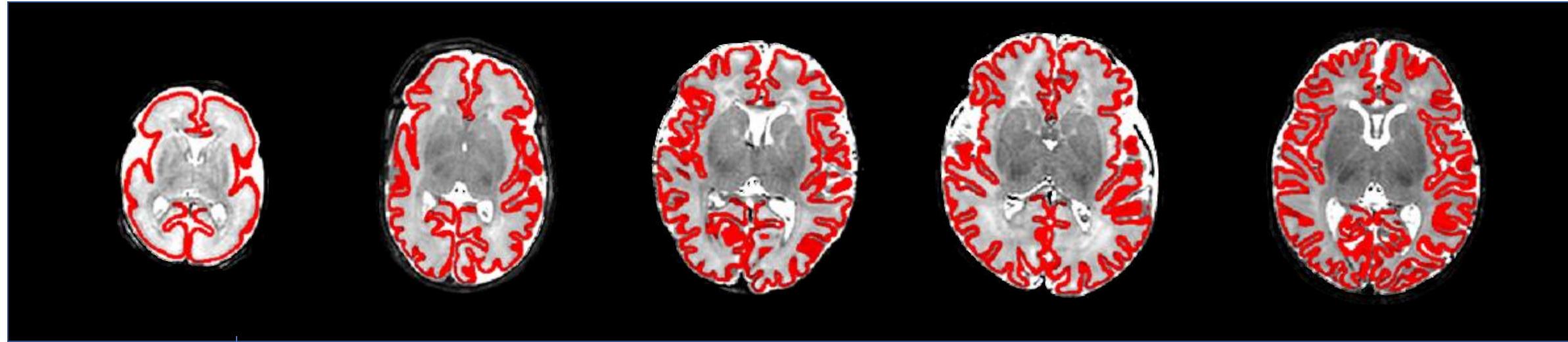
If you add the image to template warps, you can also model shape

The data here



- Developing Human Connectome Dataset
- 424 neonates including preterm children scanned close to birth
- Post-menstrual age-range of 28-44 weeks
- Lots of data available (already released for 40 babies, 500 more later this year!)
- Here we're focussing on T1w and T2w volumes, processed using the dHCP structural pipeline** (Makropoulos et al, 2018)
- Multi-shot acquisitions + super-resolution = 0.5mm iso voxels

How does this work?



Raw

Register all T2w to
initial template
using ANTS (Syn,
Avants et al 2008)

Median Age Template



Fit a GP regression model
(squared exponential
kernel) at every voxel in
“standard” space

What model to fit?

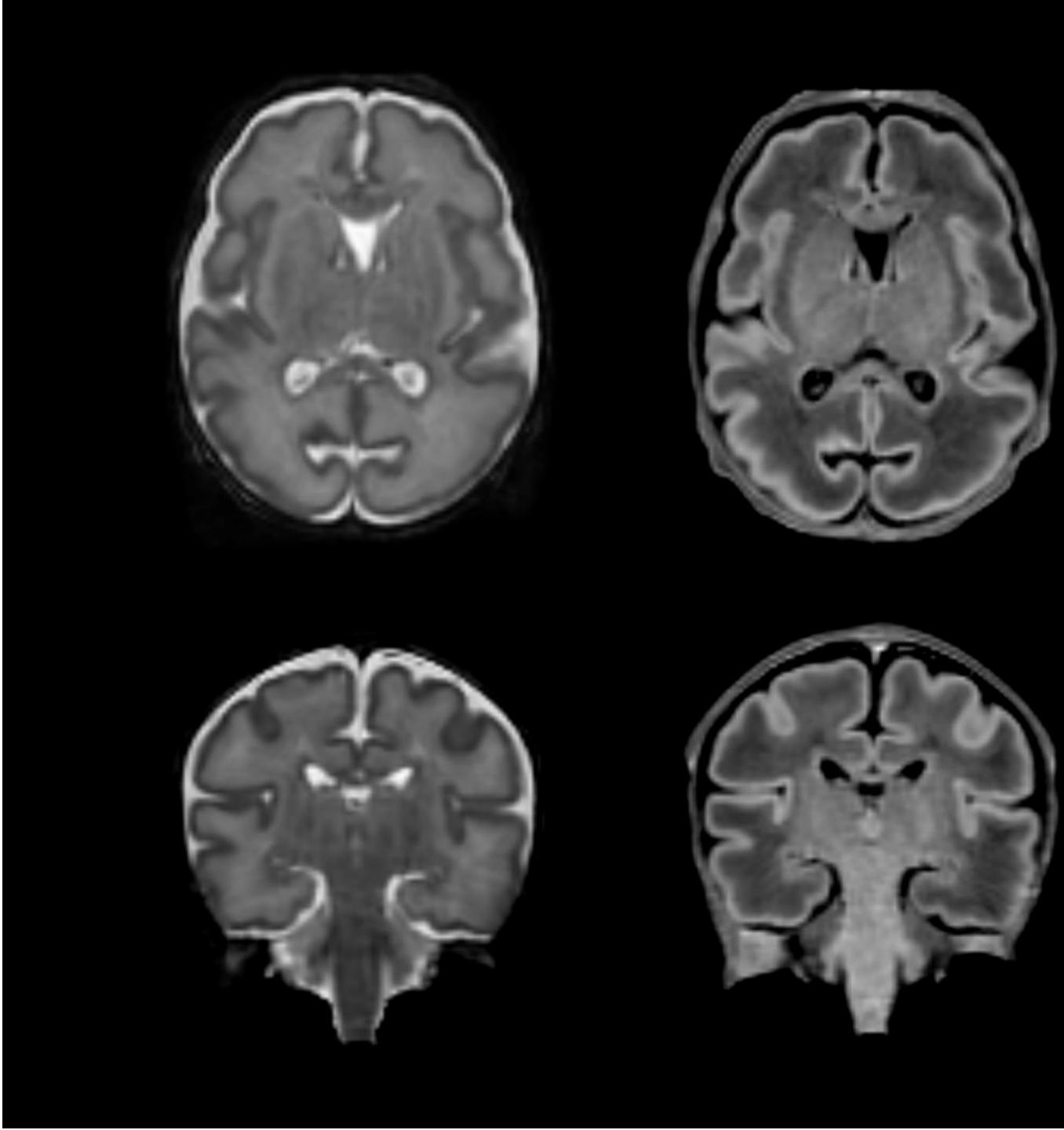
- This is similar to standard regression/GLM – though with multiple output variables, and less assumption of model relationship
- Here, I've used (1) age at scan (2) sex and (3) age at birth
- Given this model, what should a voxel look like in unseen data:
- Generate predicted T1w / T2w volumes for unseen data **AND** their non-linear warp from native (well, affine) space to a standard space
- This is a multi-output regression (intrinsic coregionalized regression) implemented in GPy (Álvarez et al, 2012)
- Unique noise term per output, matrix encoding information sharing

a Raw to
Initial template

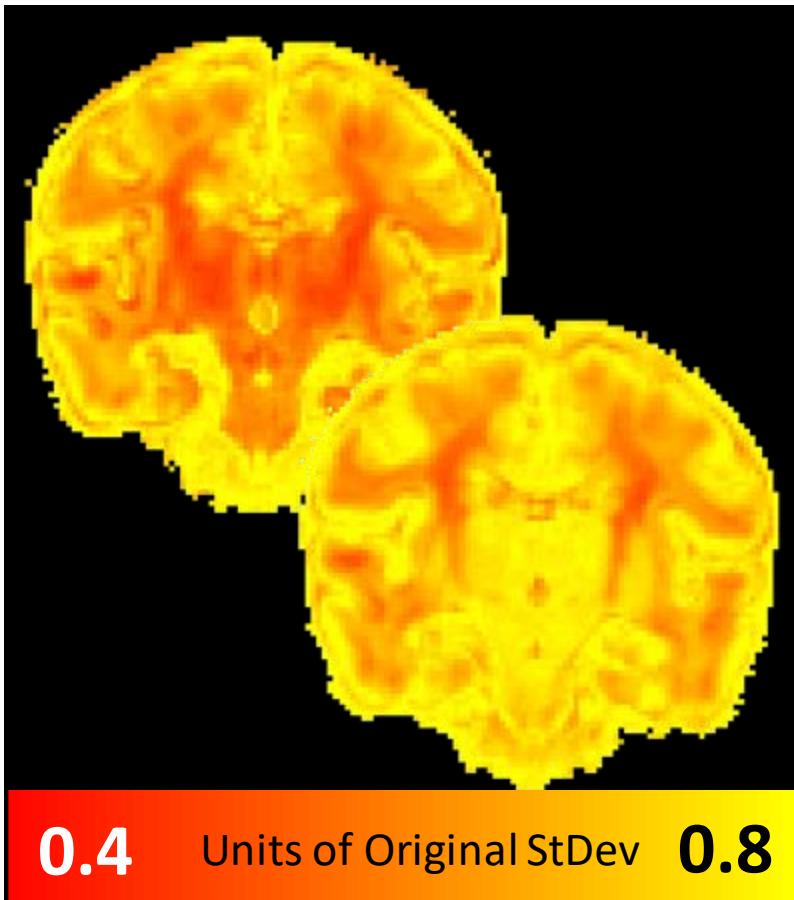
T1w and T2w post
registration

T2w always good
(clinical read), T1w
more variable

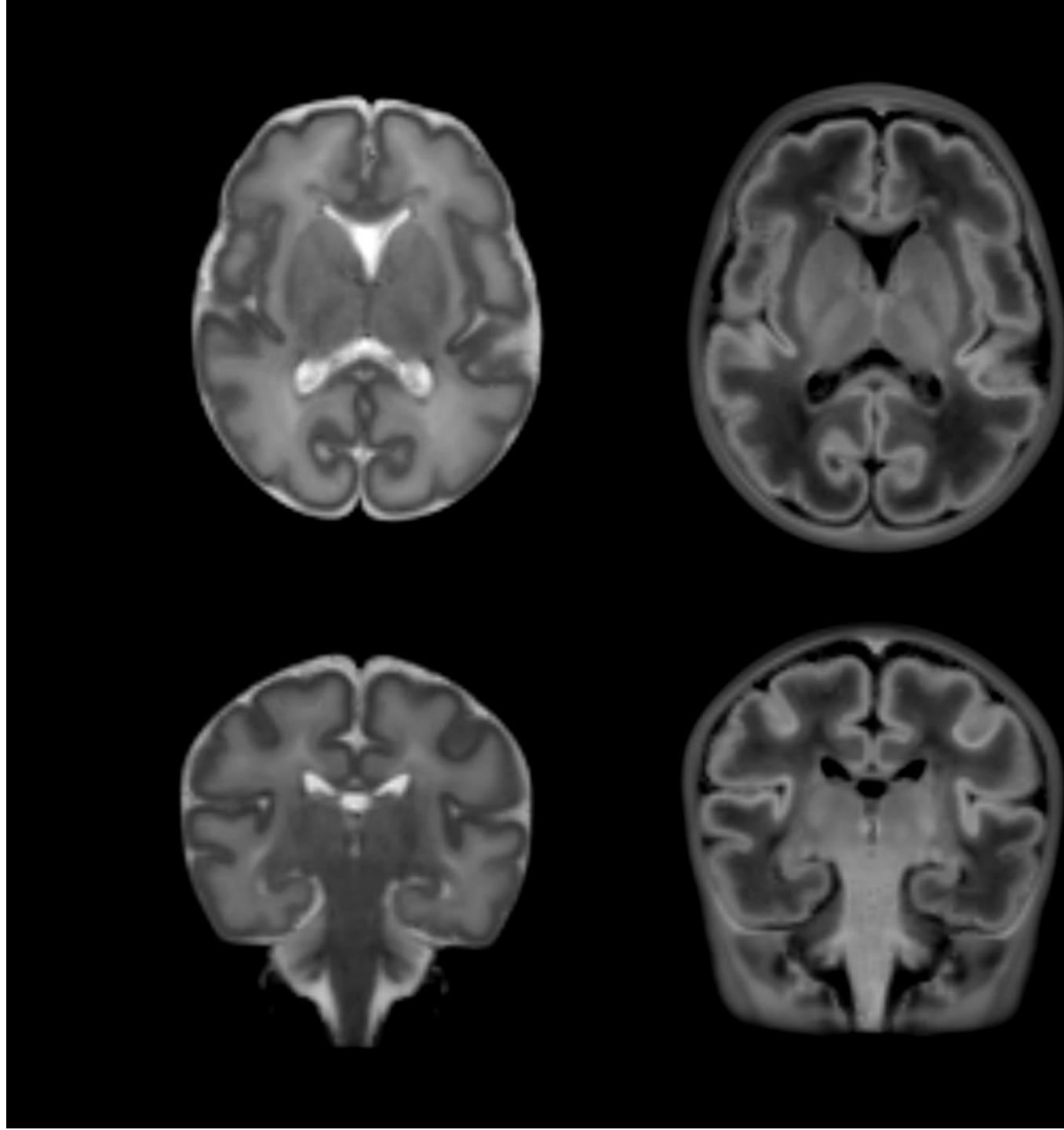
30 weeks – 44 weeks PMA



b Run GPR - guess
what the image
should looks like



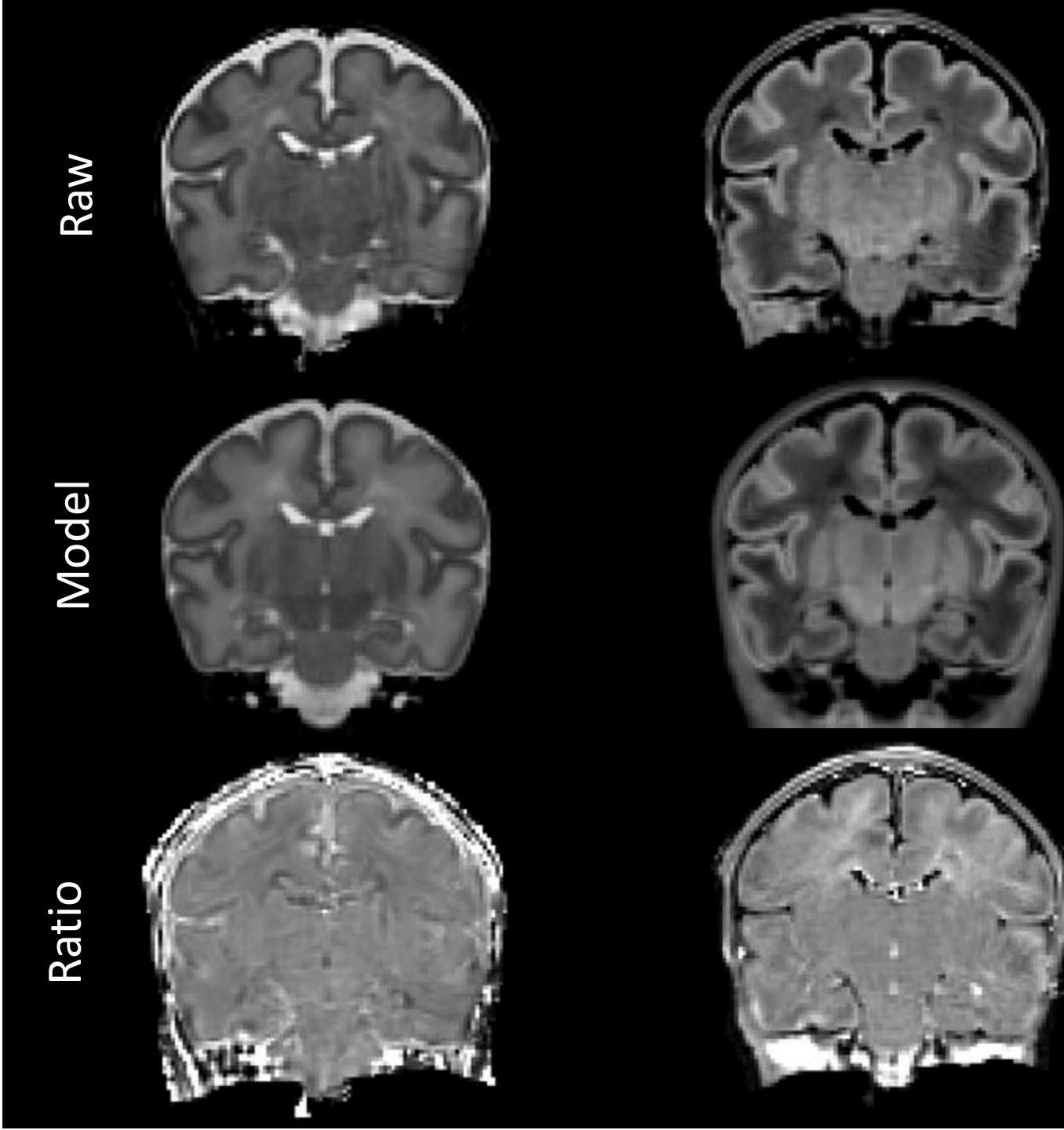
30 weeks – 44 weeks PMA



c Does it get the contrast right
(don't worry about absolute values...)

Here we have the full contrast range of the weighted images, no priors beyond an initial template

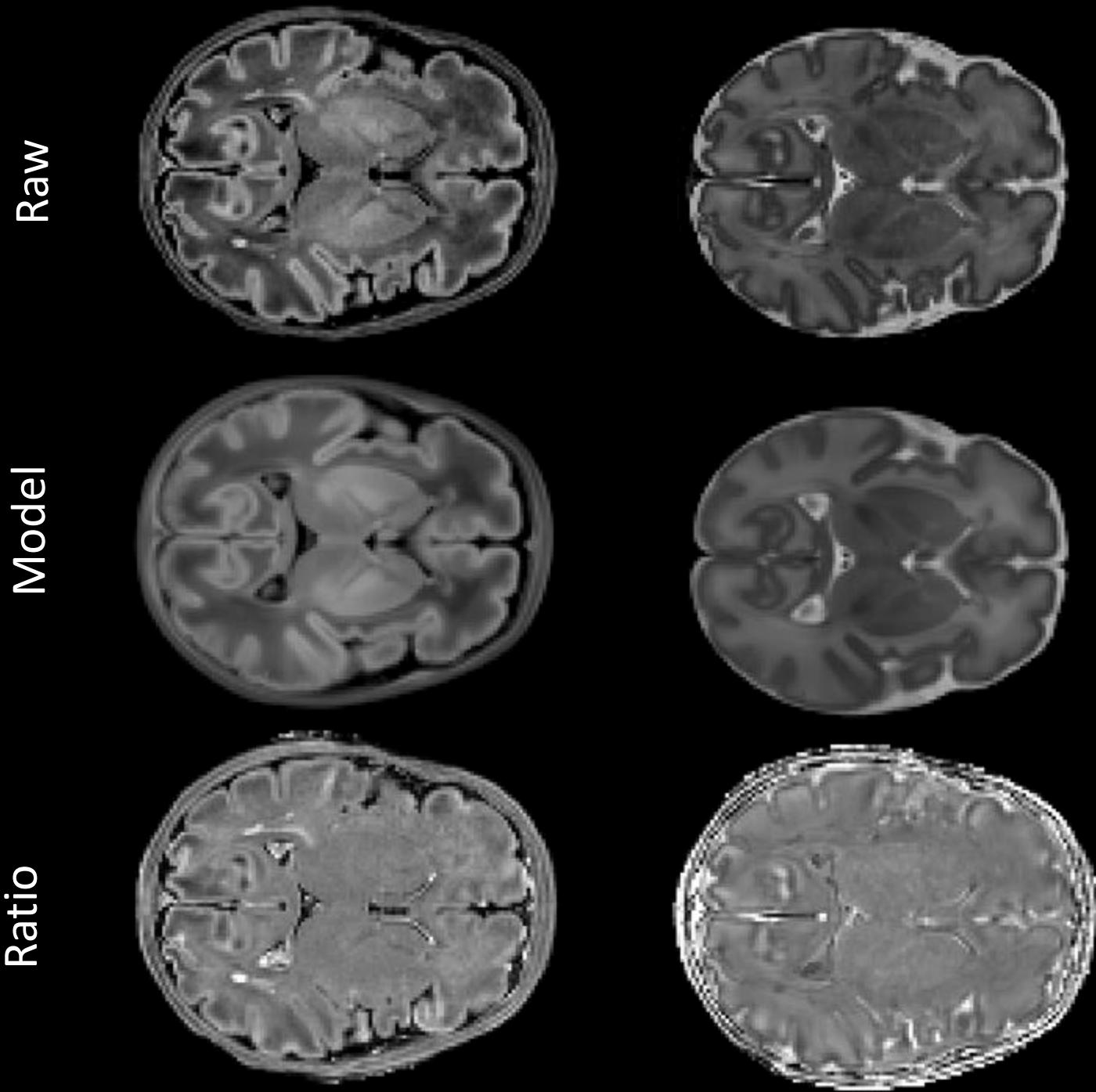
30 weeks – 44 weeks PMA



c Does it get the contrast right?

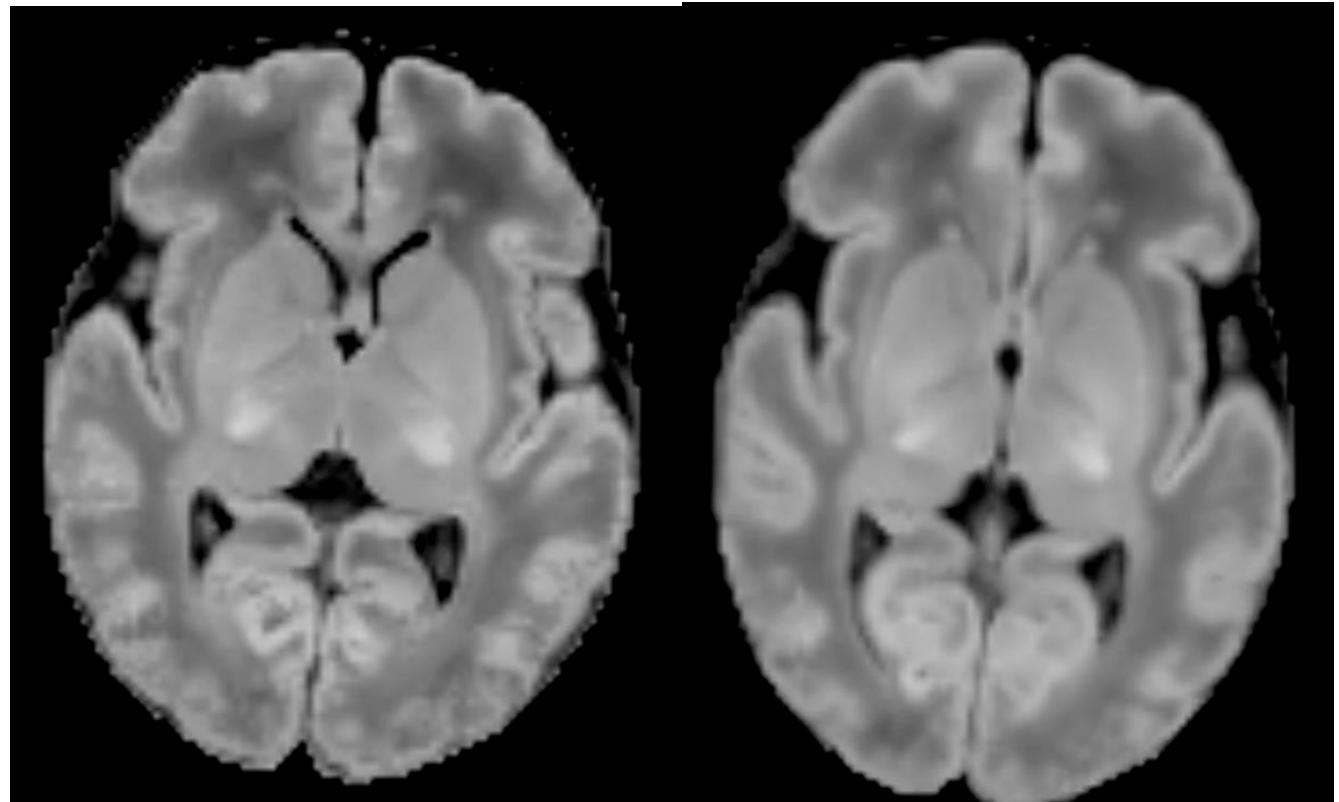
What about
incidental
findings?
Abnormal tissue?

30 weeks – 44 weeks PMA



Tracking Development?

- We can predict things like shape and intensity in out of sample data (e.g. visualise the model)
 - Here an estimate of:
 - 41 week old male
- Born at 28 weeks –
40 weeks in two day steps
- Reflects known shape changes

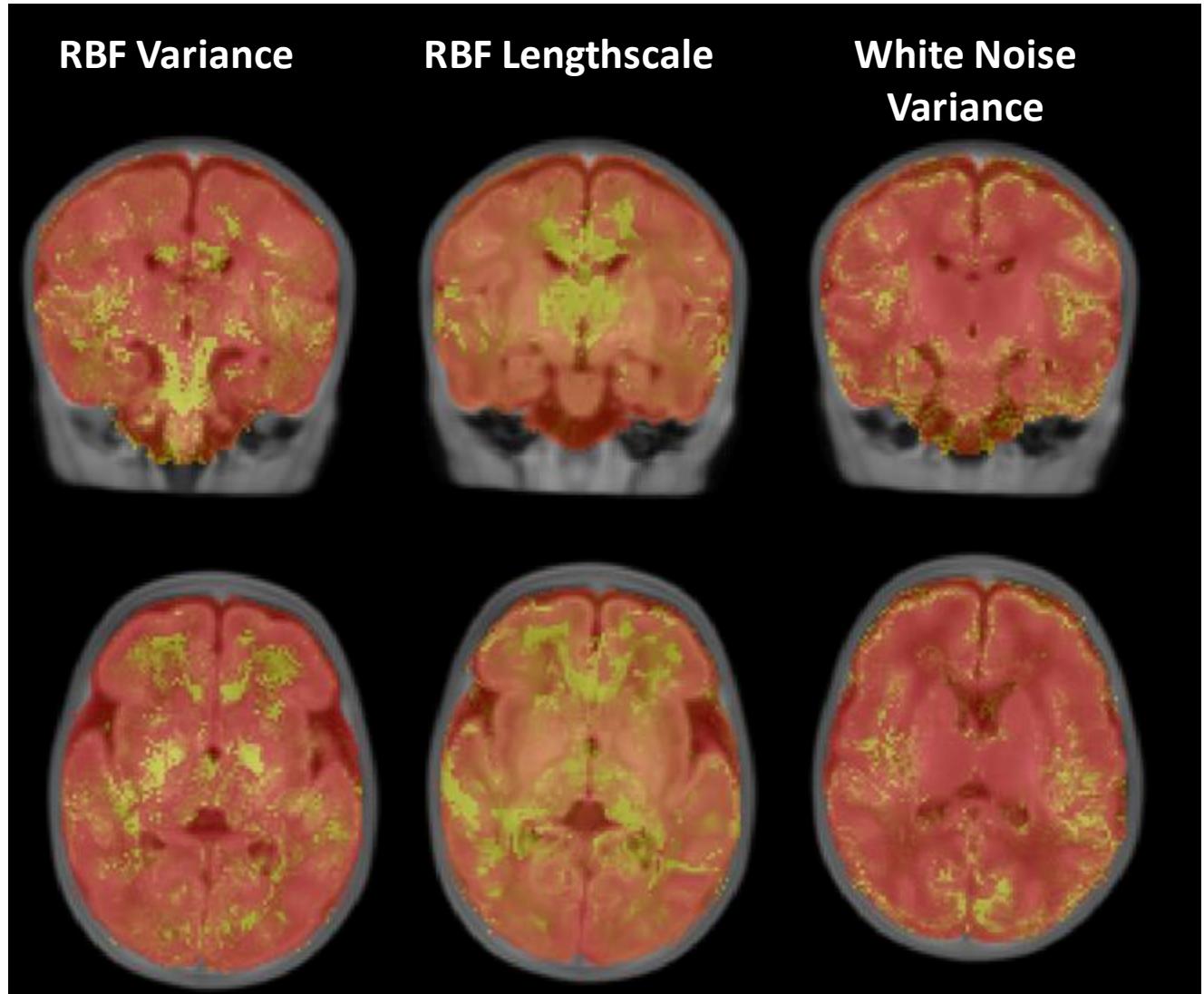


GPR Parameter Maps

Lengthscale is probably the interesting one here (wiggles – how much memory)

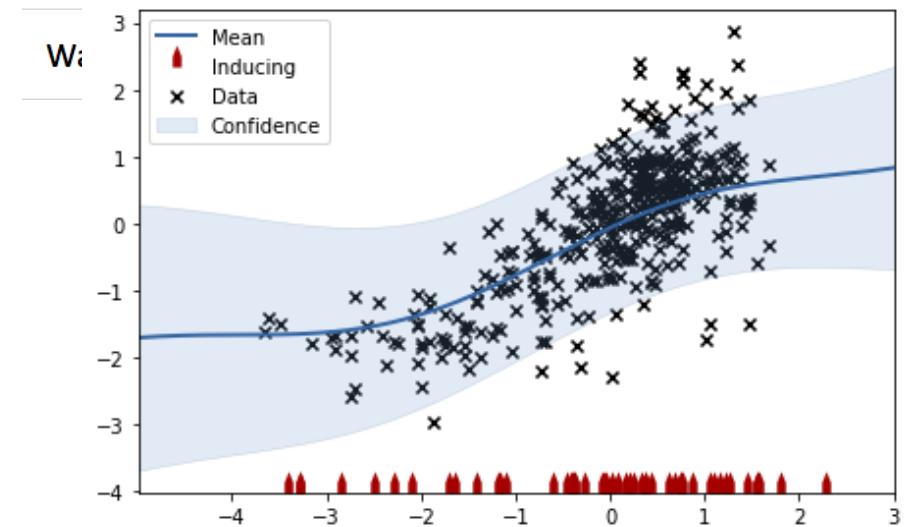
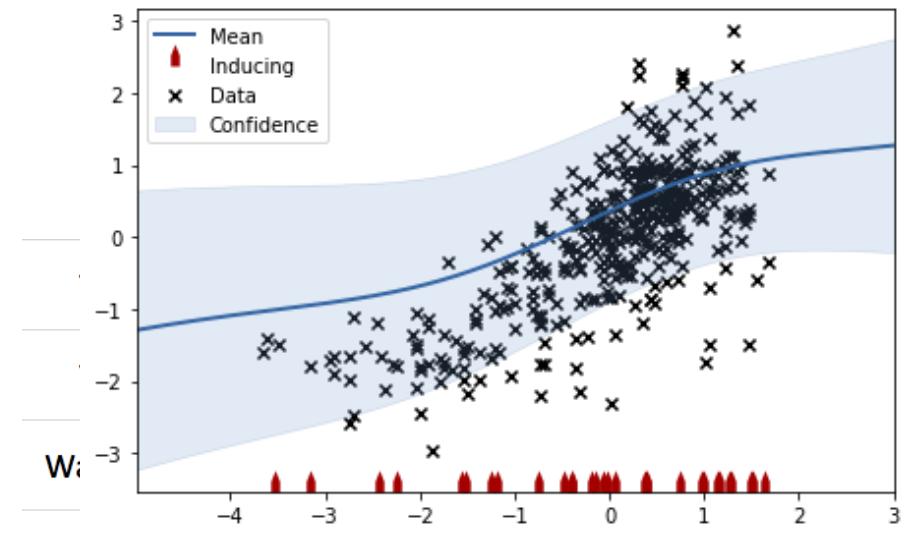
Brighter values are where the lengthscale is large - effectively linear.

Most of the brain shows non-linear changes – RBF variance is higher at tissue boundaries



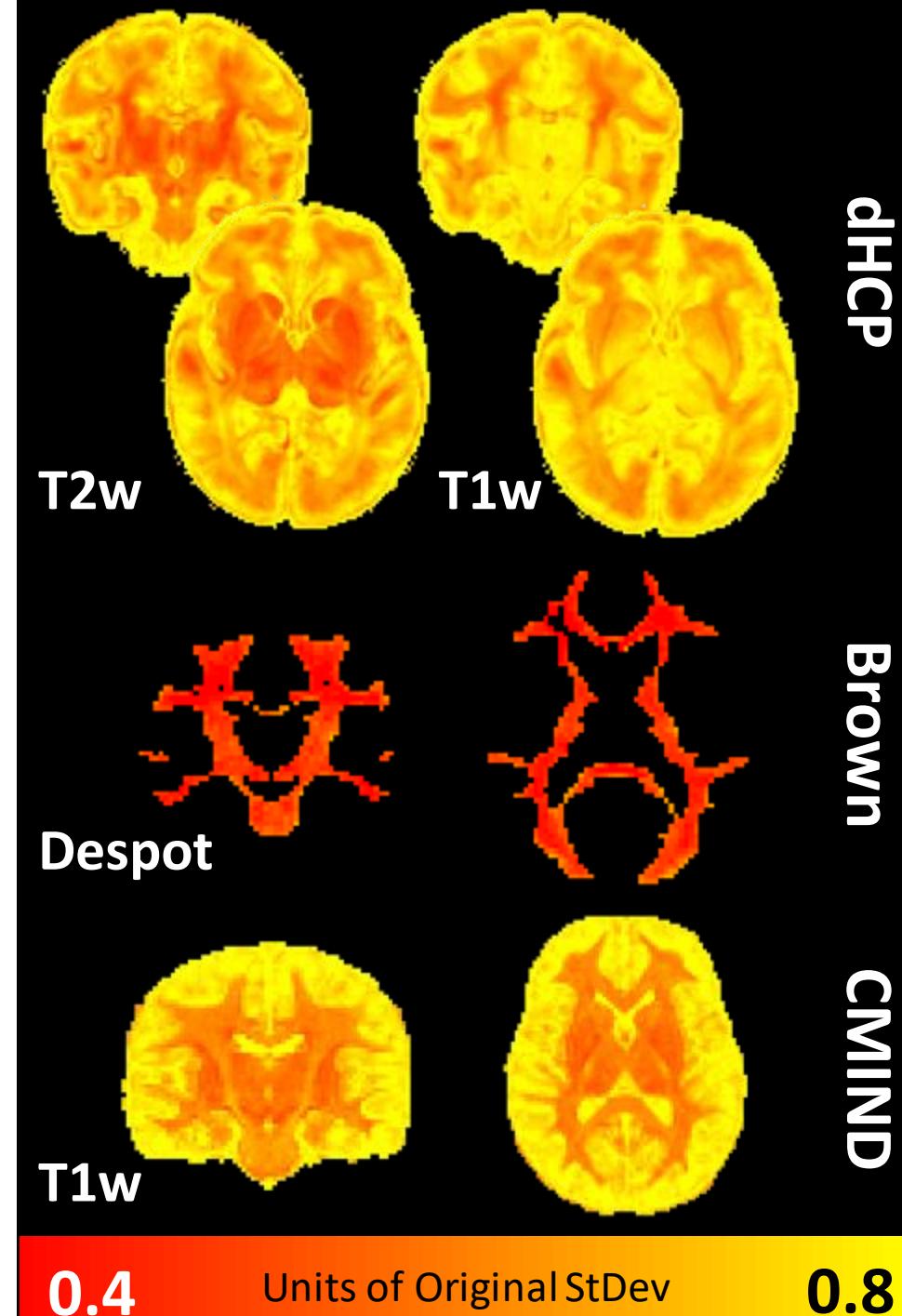
Does multi-output help

- The approach to multi-output regression taken here is the intrinsic coregionalization model (Alvarez et al 2012).
- This matrix (B) represents the information sharing between outputs
- If there off diagonal values are low, there's probably very little info sharing between outputs
- For here, intensity information is collaborative, registration info less so.



GPR well for diff age ranges

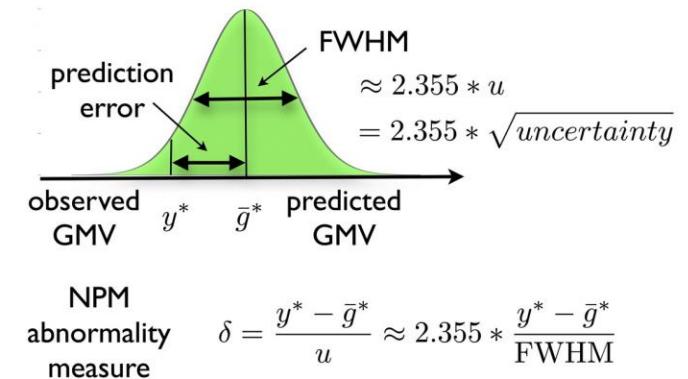
- Captures changing image intensity in neonatal images
- Predicted images look pretty similar to actual (residuals between predicted and actual are mostly normal)
- Uses all the data in the **raw weighted images** (though for other scanners / sites, this may need to have some scaling)
- Big increase in performance with quantitative (though relative to stdev!)



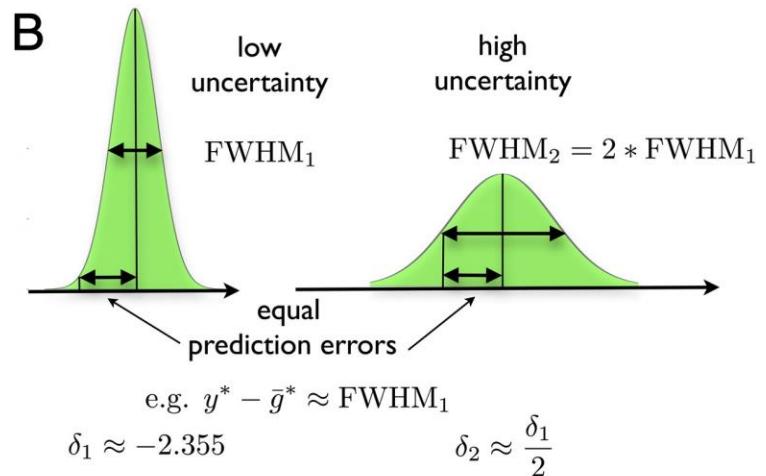
What is it good for?

- So here we look for unexpected increases in the signal of the actual image compared to the predicted one (prediction error) and this is scaled to a simple Z score.
- The Z score is different from voxel to voxel and timepoint to timepoint
- Importantly uses the full range of intensity values of an image, not just “grey” or “white” matter

A

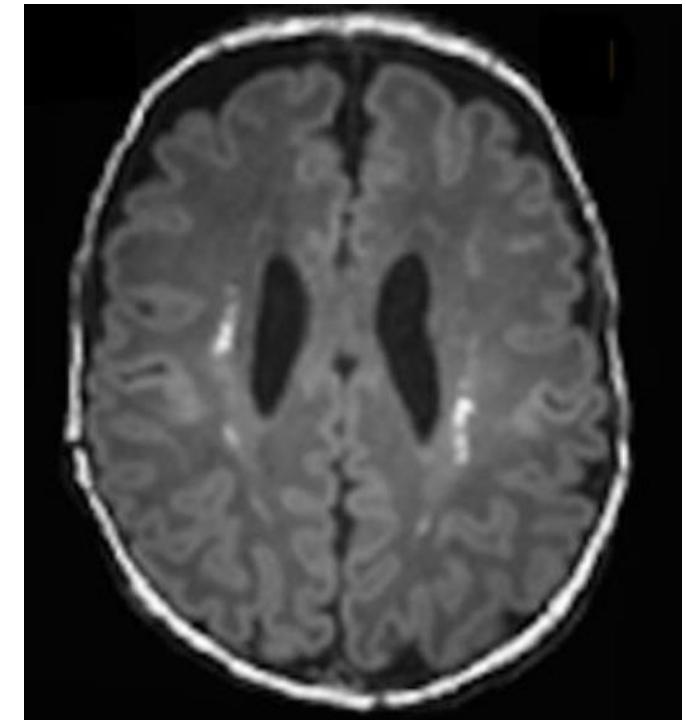


B



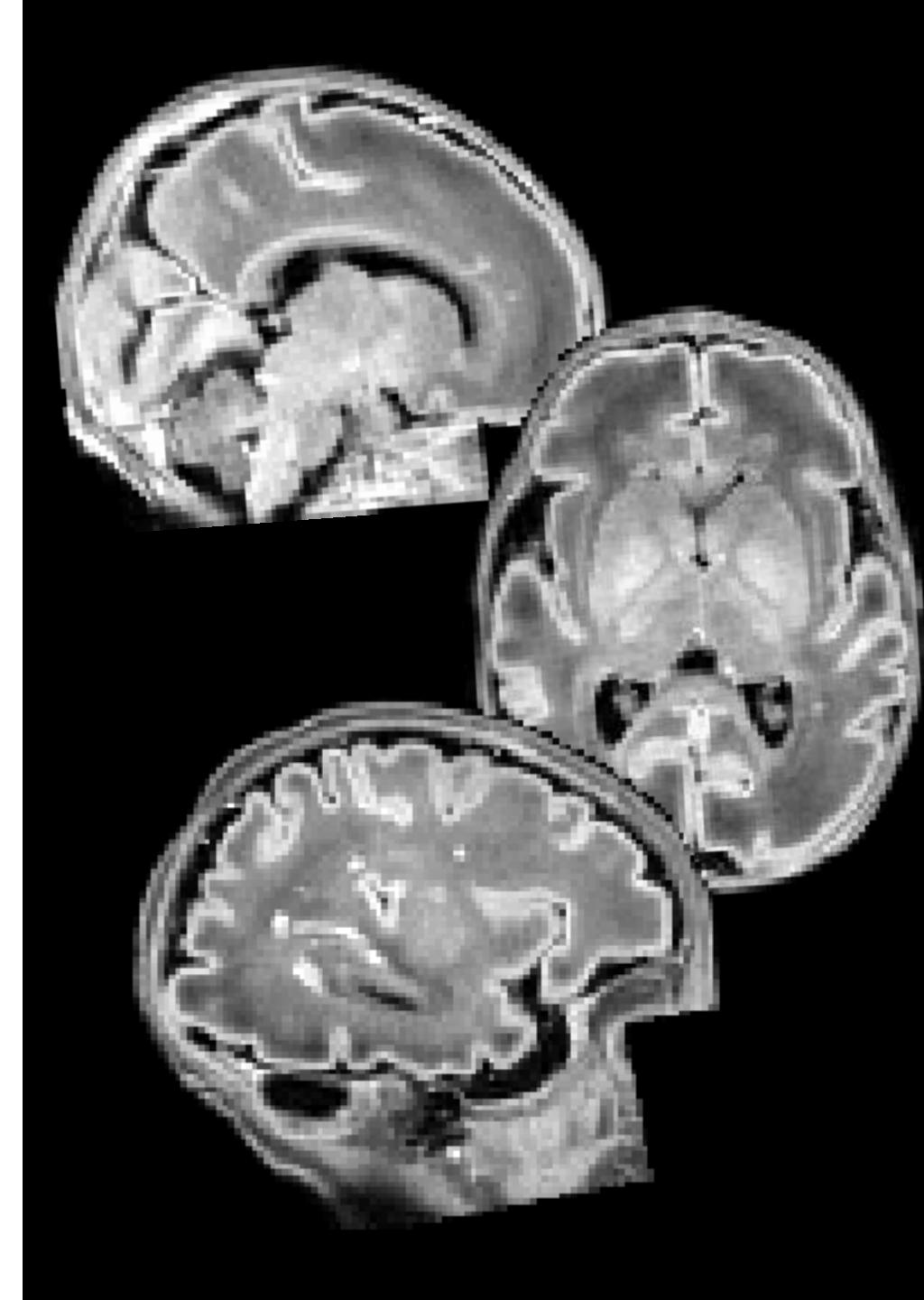
What is it good for?

- The aim of all of this is to move towards single subject inference – try something obvious
- As a test case look to detect abnormalities in the dHCP cohort – punctate white matter lesions
- Punctate white matter lesions are variable in site and number, small
- As in disorders like Multiple Sclerosis, identification and delineation is very difficult and time consuming

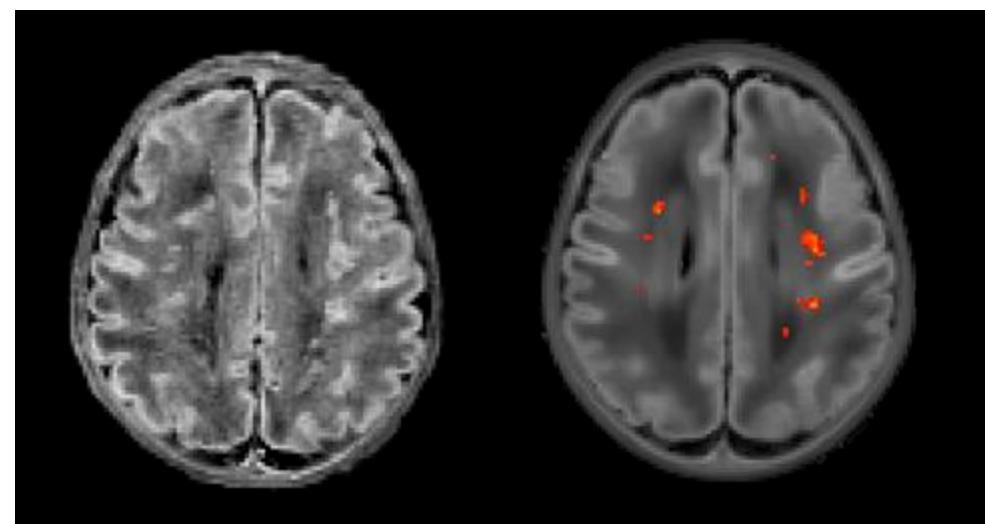
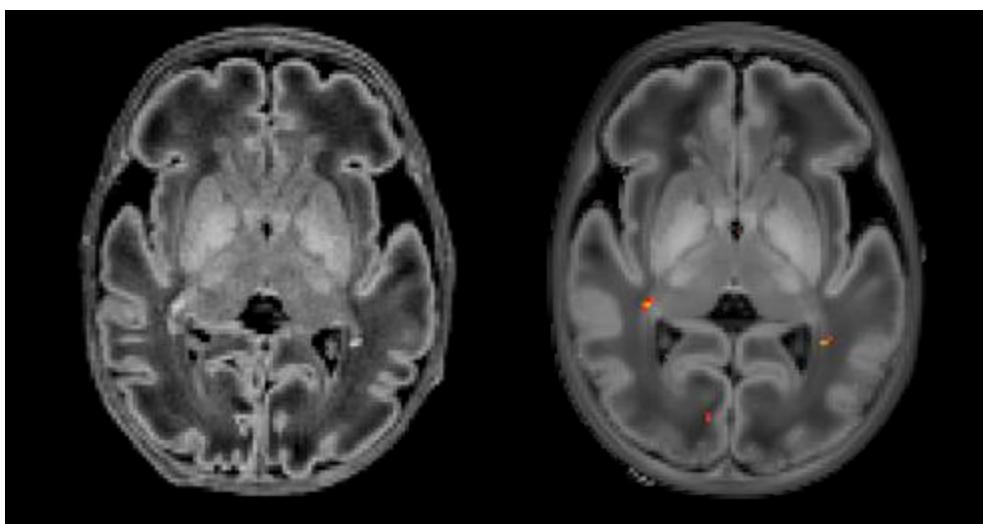
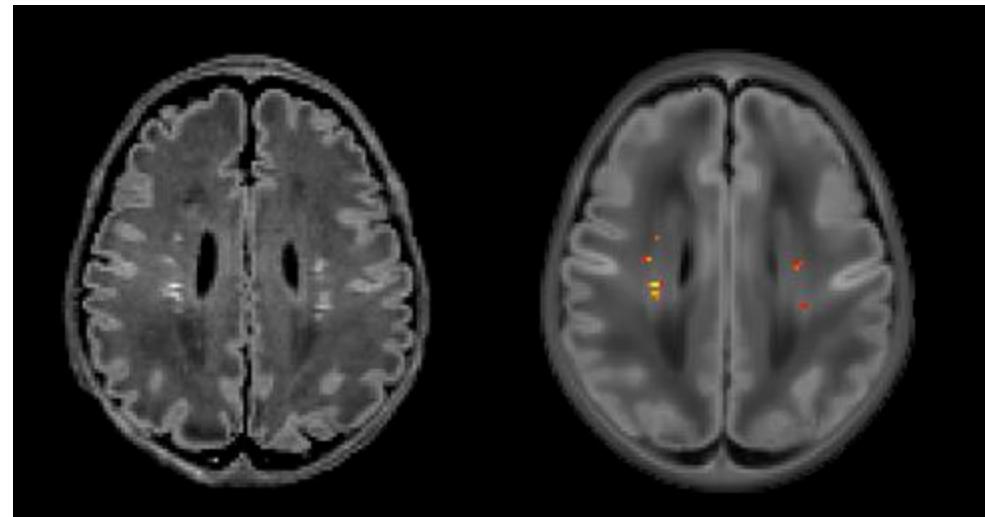
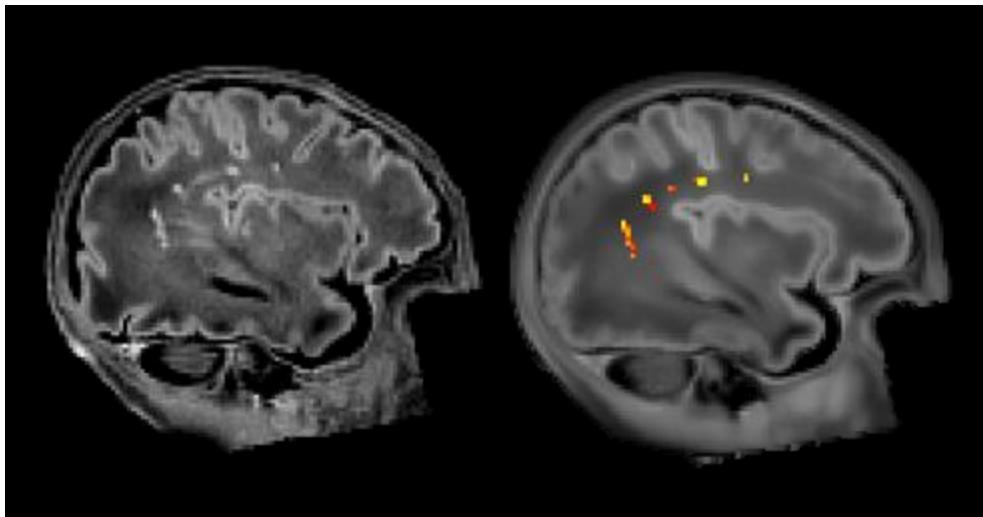


Clinically useful?

- Punctate white matter lesions
- Usually small T1 hyper-intensities
- Can be very bright or reasonably subtle
- Associated with neurodevelopmental outcome in preterm neonates (Tusor et al 2017, Guo et al 2017) and are more common in term infants with neonatal illness (Guo et al 2018)



What is it good for? Single Subject vs Group



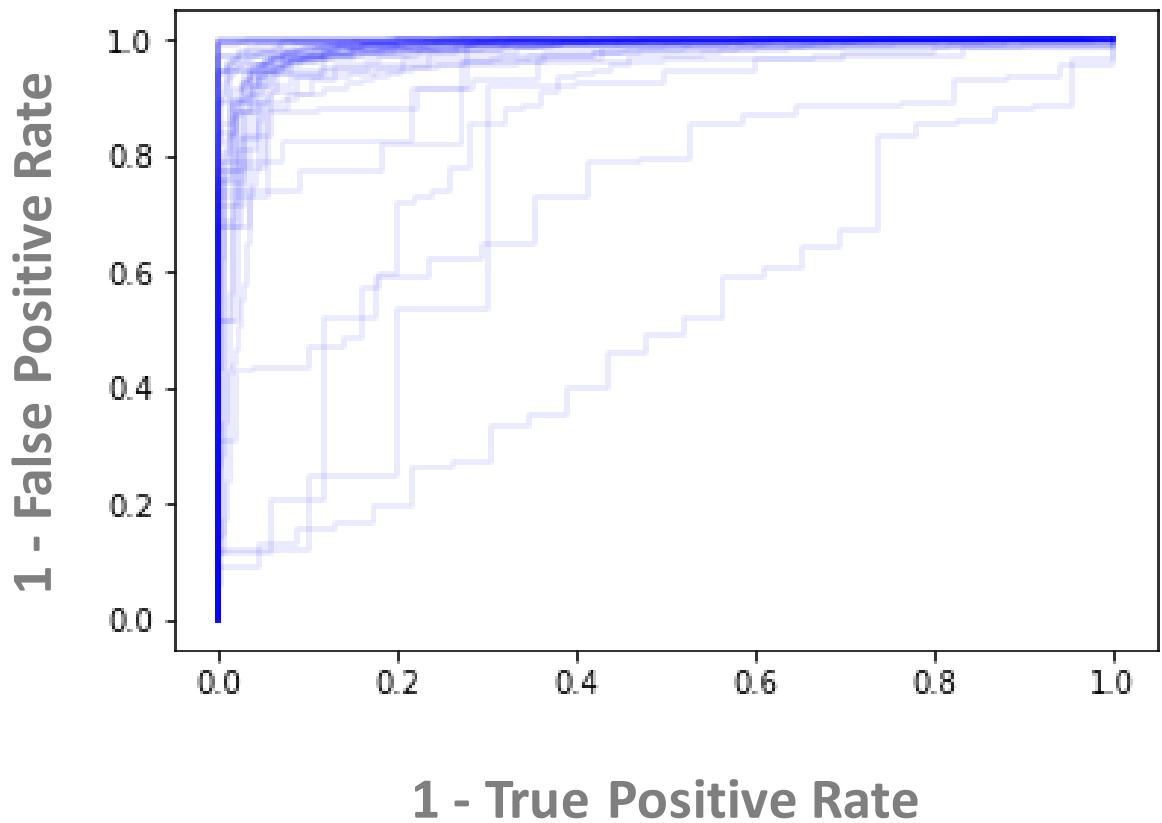
$Z > 3$ which has
very few false
positives

$Z > 5$ quite
specific to
pathology
even with
motion

No
smoothing!

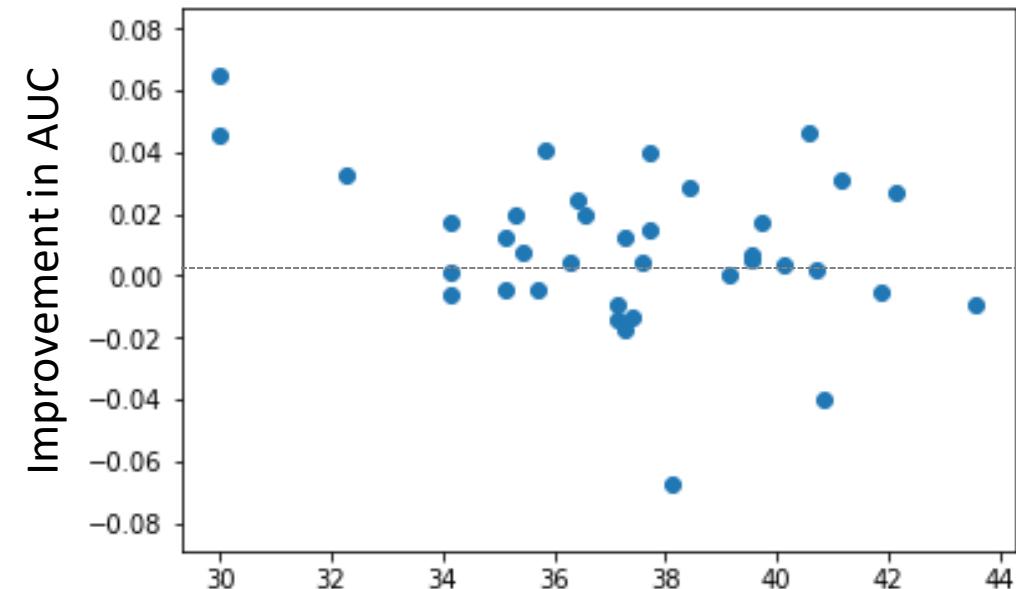
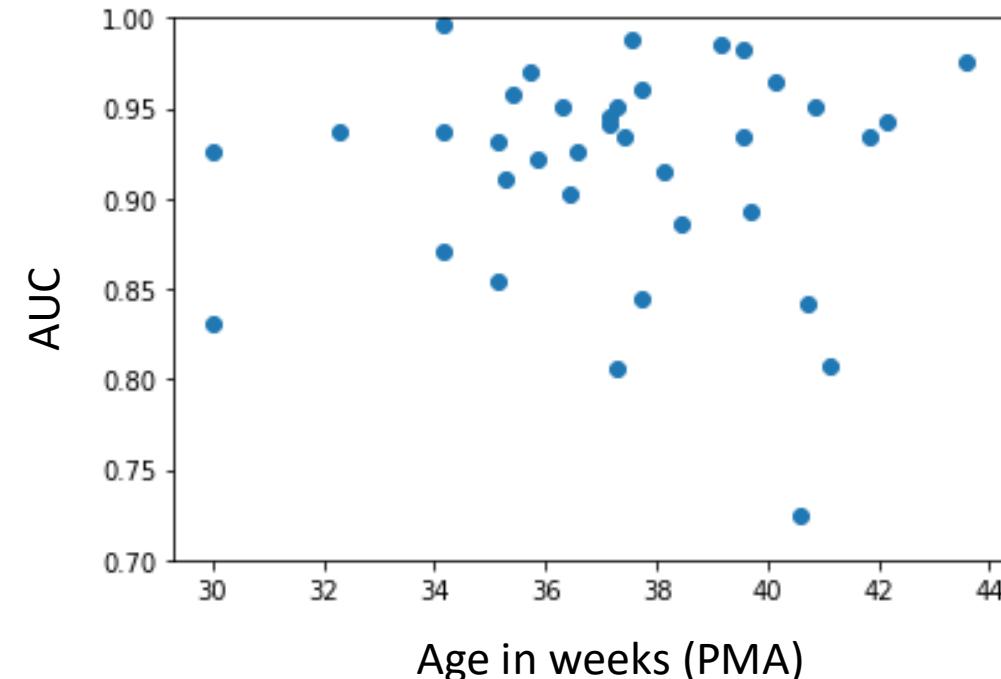
Clinically useful?

- Of the 446 individual neonates in the initial sample, 42 have visually identifiable small white matter hyperintensities.
- In the 42, the abnormal tissue was manually outlined
- AUC Z Score: 0.9
- AUC GPR Model: 0.925
- AUC GPR Enhance: 0.955



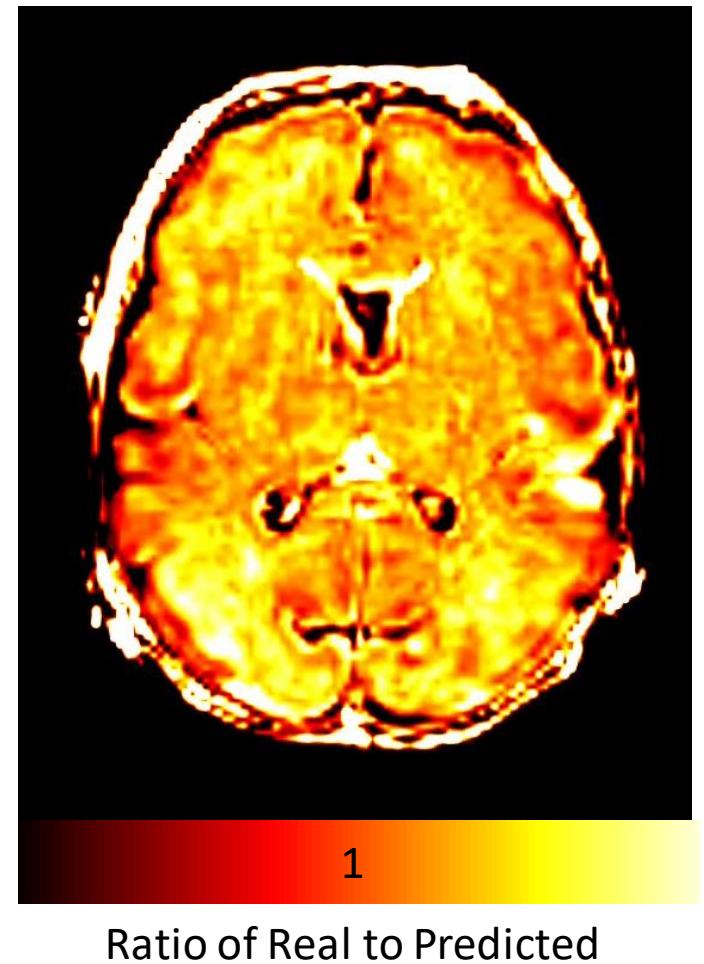
Practically useful?

- The accuracies probably don't matter that much (for a proof of principle)
- What matters is the **age dependence** of AUC disappears by using a GP model
- The AUC values reflect a whole brain mask and no image smoothing (important for punctates in particular)
- Small improvement in accuracy for most but improvement much larger at younger ages



This *might* work for varied clinical imaging

- Neonatal acquisition was controlled as much as possible in terms of acquisition. Values between subjects are ~comparable
- RF pulse may have varied to achieve FA from individual to individual (this scales image intensity)
- In typical clinical imaging, this is very hard to control - you can (successfully) scale images OR
- Alternative: Embrace ignorance and investigate not just a voxel but take into account neighbourhood information as well – can be complementary



More effort for increasingly little reward?

- Well, no. From modest (low) thresholds, you get quite good detection.
- Data driven way to address age effects
- Accuracy can / should improve with increased sample / age density / cross-validation of kernel function (e.g. linear, wiggly, etc)
- How to approach site effects in large cohort studies
 1. Include site in the model (or even FA / TR etc)
 2. Make different sites different outputs
 3. Use genetic batch techniques to normalise data prior to analysis
 4. Use new data with quantitative MRI & image synthesis to try and match to site

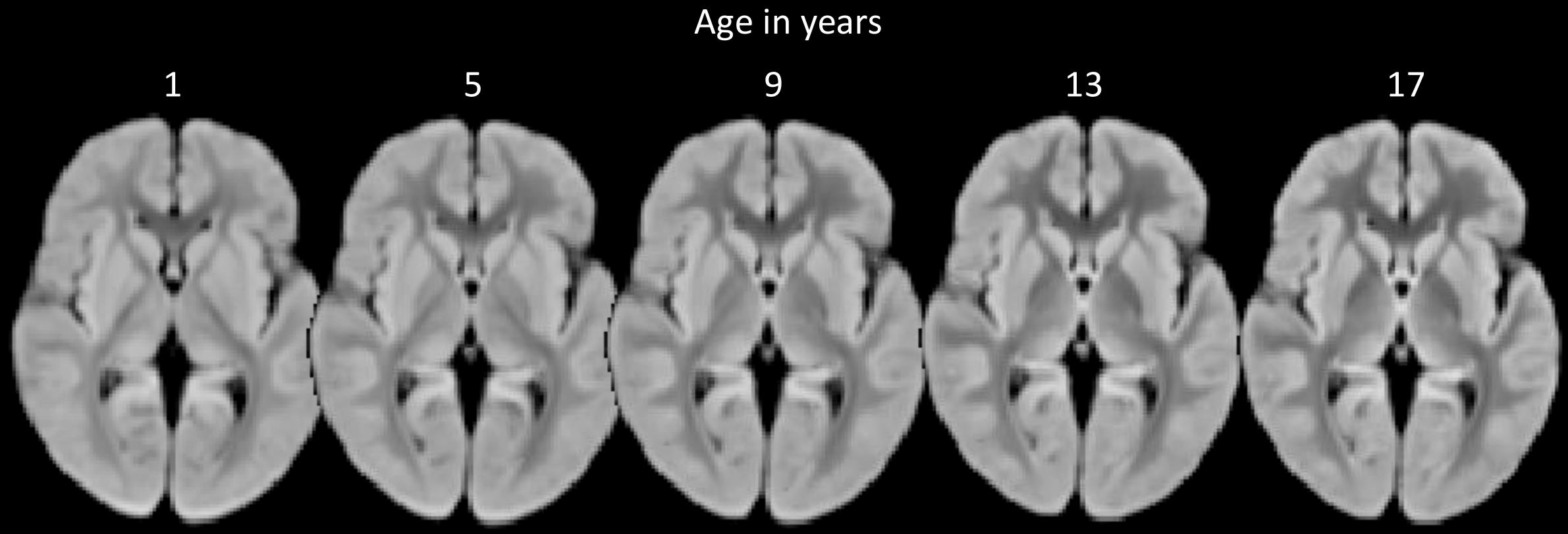
Improvements (some quite basic!)

- Image smoothing improves performance in some applications (Ziegler et al, 2014) but would be deleterious for the application here (small punctate lesions would disappear but less false positives)
- Increasing the rank of the Intrinsic Coregionalisation Model doesn't really change performance vs computation time
- Gaussian Processes quite sensitive to outliers but Student's T type processes may be more robust (though adds an effective df parameter...)

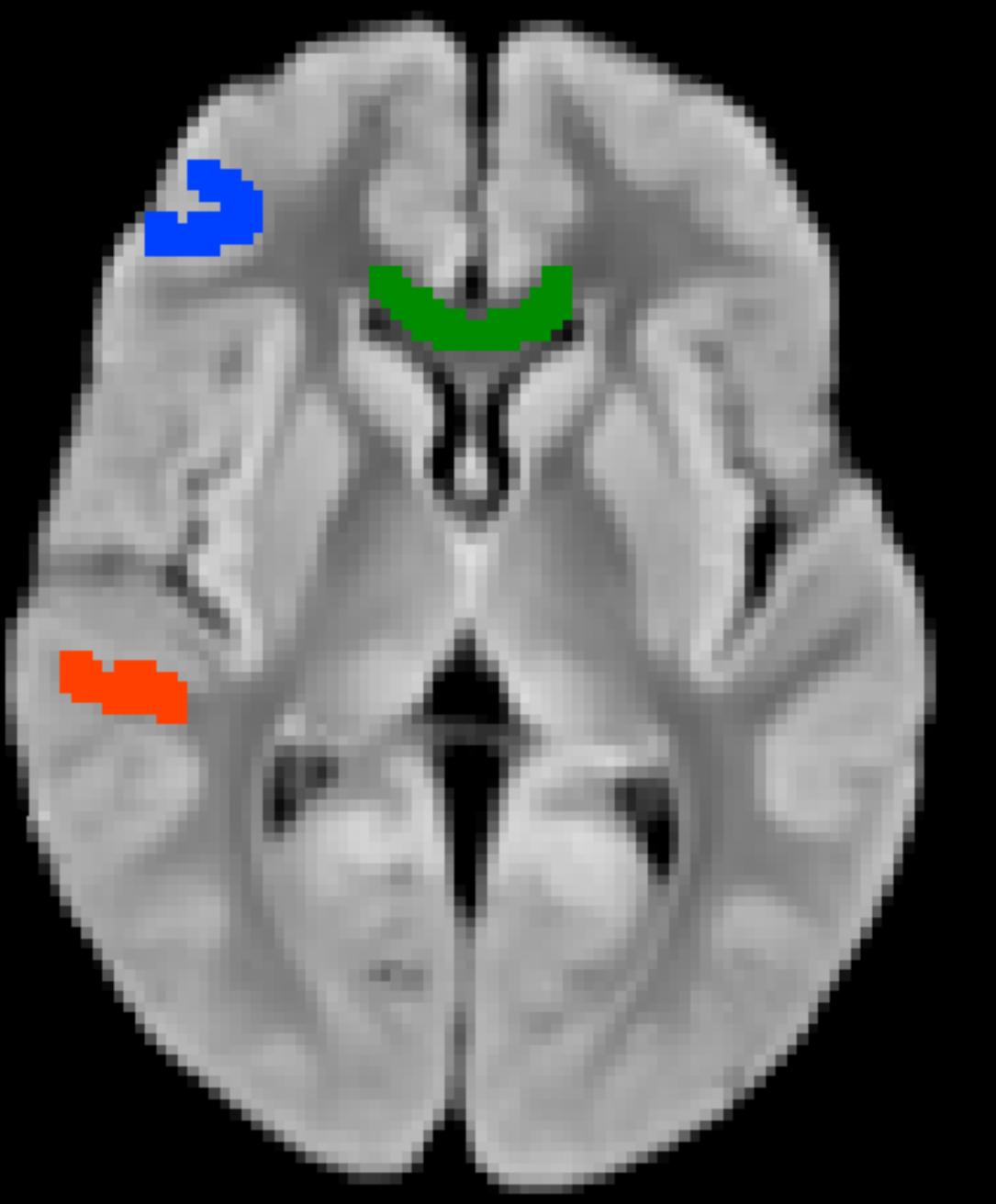
Can we use a similar approach in epilepsy?

- Great Ormond Street Hospital / Institute of Child Health
- 102 children aged 1-18 years scanned as part of their clinical care
- All have focal epilepsy **but** the cause and localisation varies
(malformations of cortical development, FCDs, TLE, DNET, unknown)
- Here the question is whether we can detect abnormal tissue in one patient compared only to other patients – no control children
- Otherwise, conceptually the same as VBM approaches that have been tried in normalise FLAIR comparisons (Focke et al, 2008, 2009)

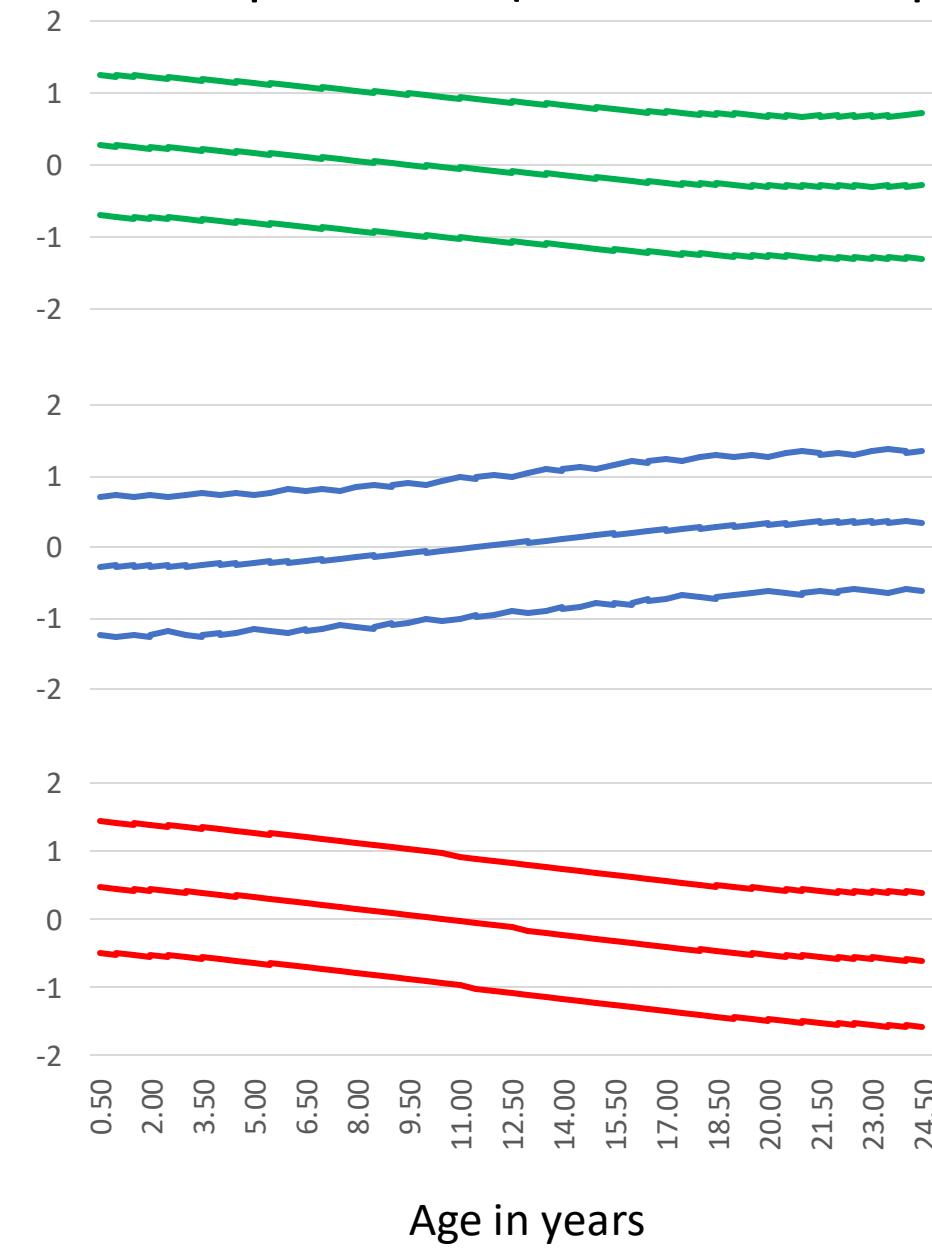
Brain image intensity over development (trained from patient data only)



Predictions of clinical FLAIR for age (synthetic)

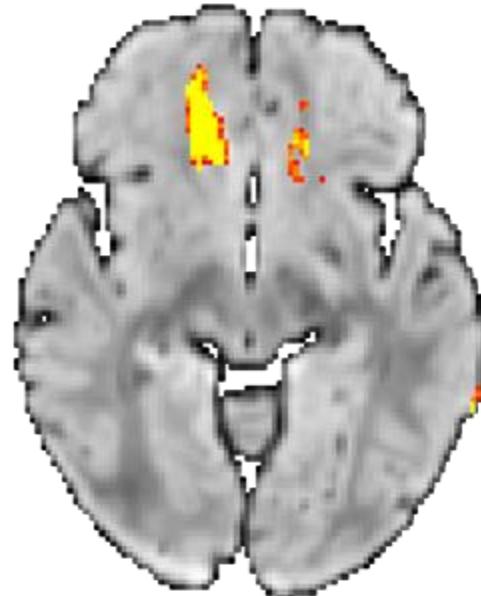


Growth Curves with (normalised) FLAIR: derived from hospital data (children with epilepsy)

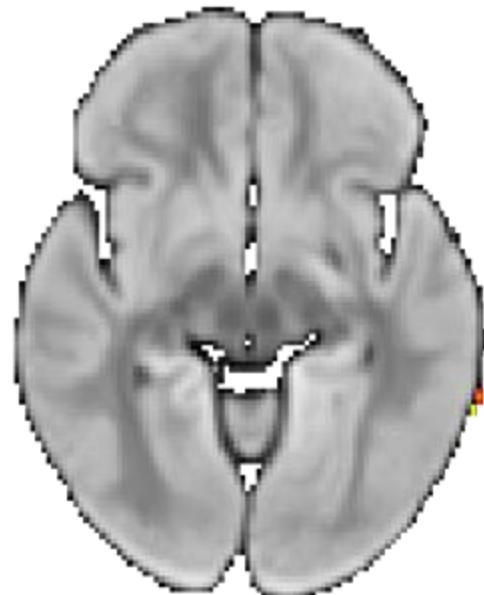


Example Cases – where do individual children fall against the FLAIR intensity growth curve?

Observed image

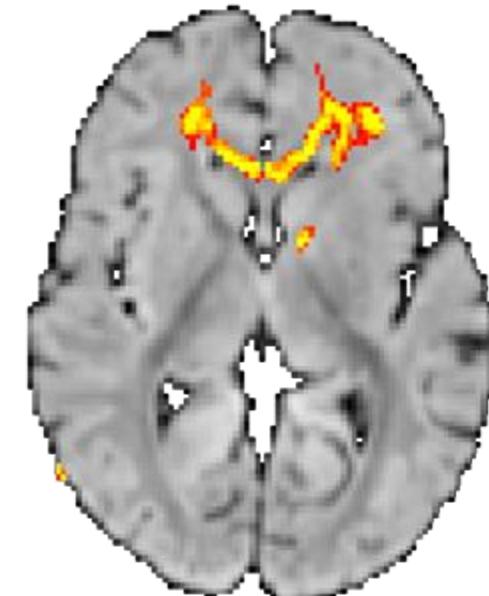


Predicted image

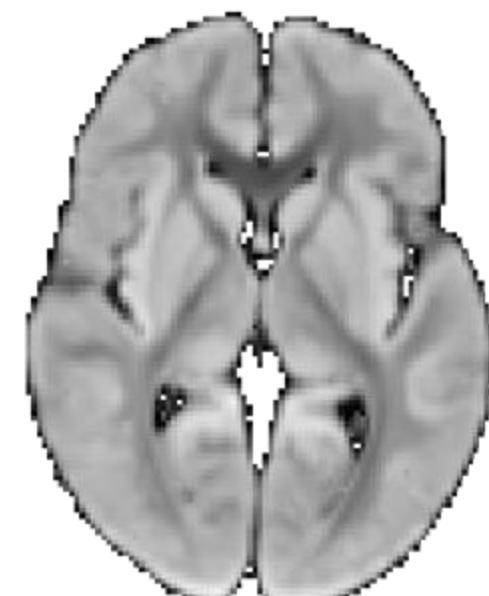


2 Year Old
(no report)

Observed image



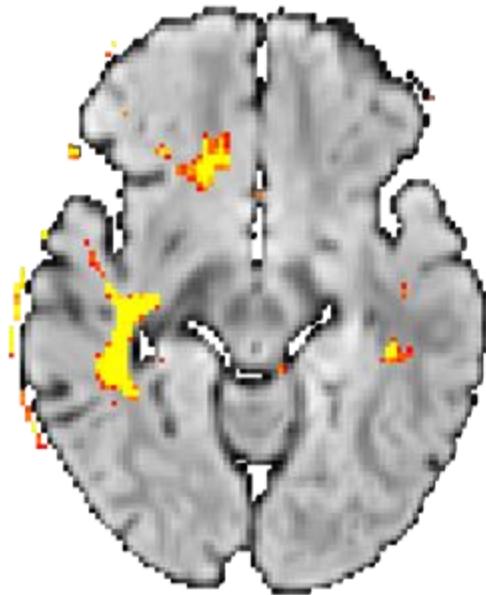
Predicted image



2 Year Old
Left frontal abnormal

Example Cases

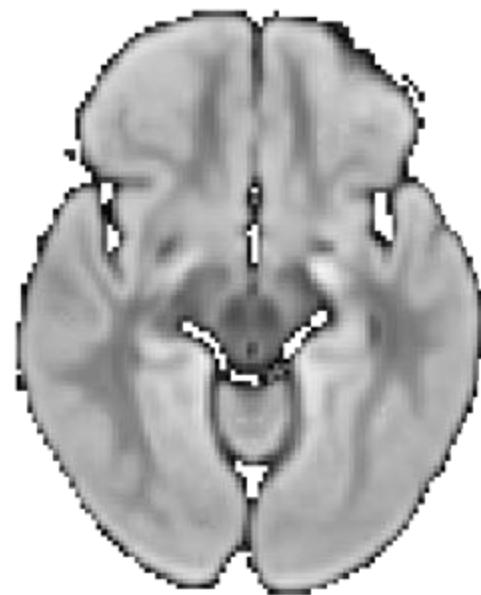
Observed image



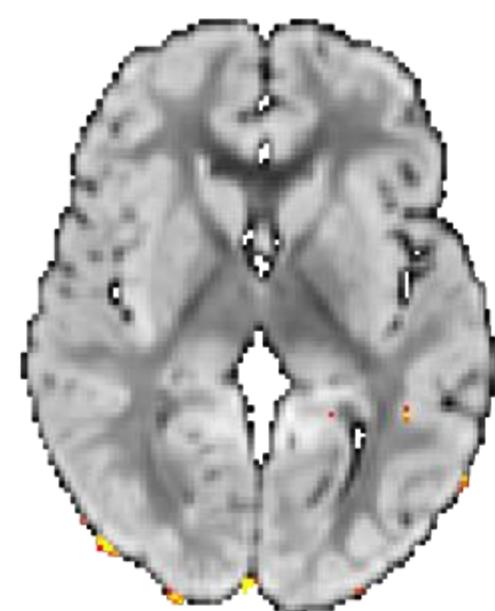
5 years old

Bilateral TLE & extensive lateralised white matter
damage

Predicted image



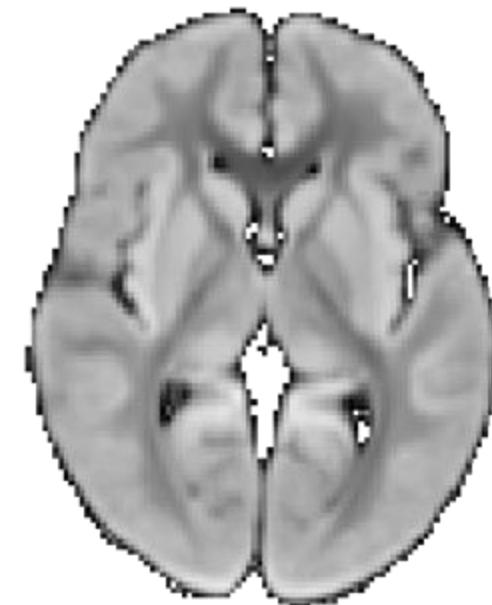
Observed image



4 years old

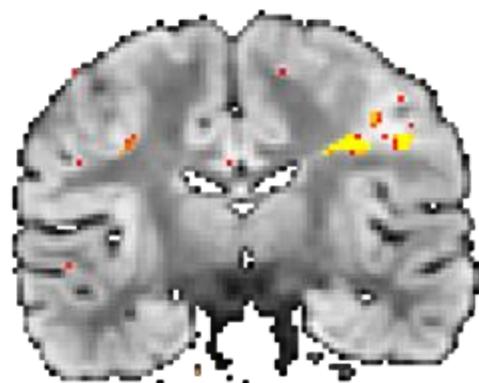
MRI negative

Predicted image



Example Cases

Observed image

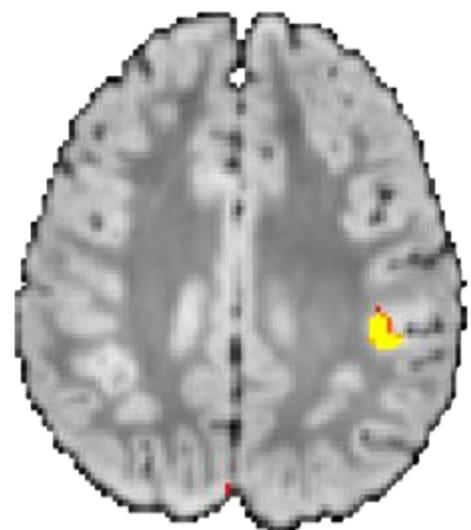


Predicted image



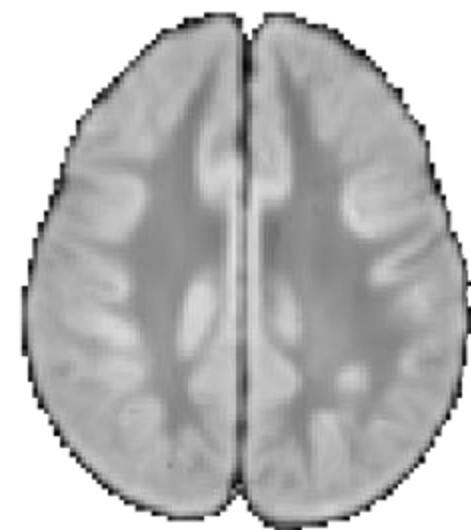
7 Year Old
Clear Transmantle Sign

Observed image



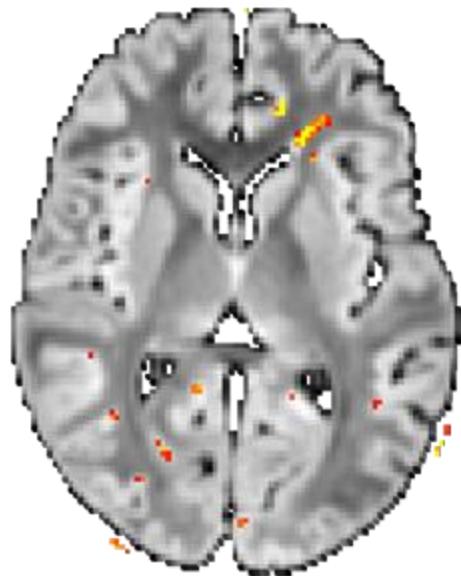
7 Year Old
Bilateral FCD (only see one)

Predicted image



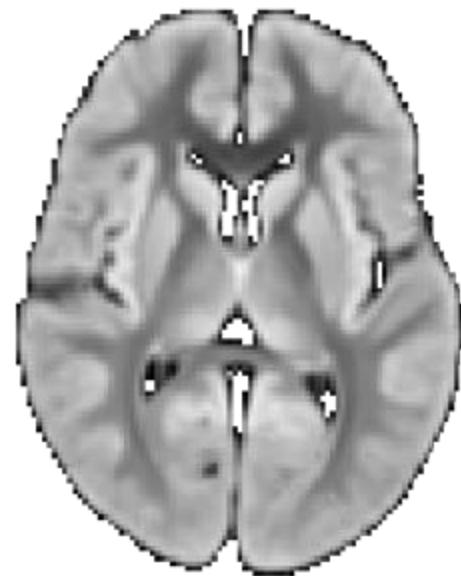
Example Cases

Observed image

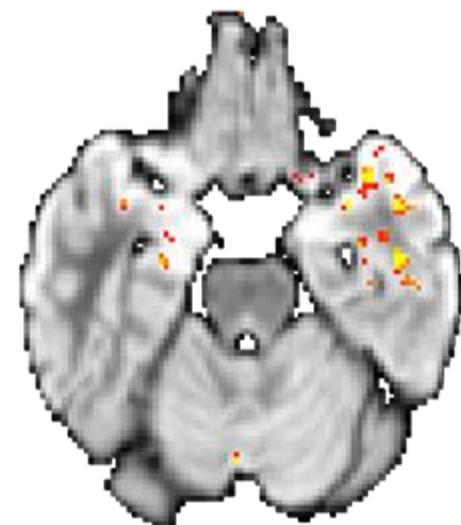


MRI Negative
17 years old

Predicted image

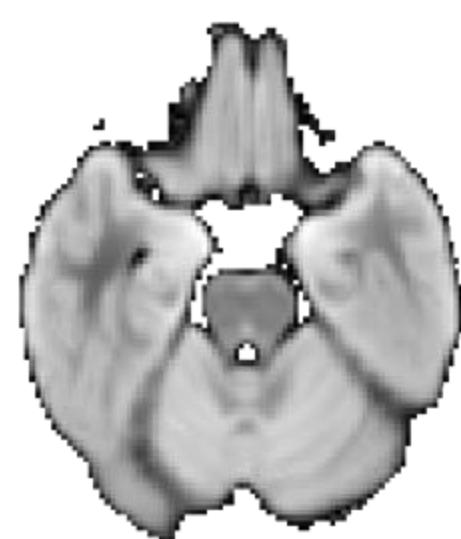


Observed image



Left temporal pole subcortical
17 years old

Predicted image

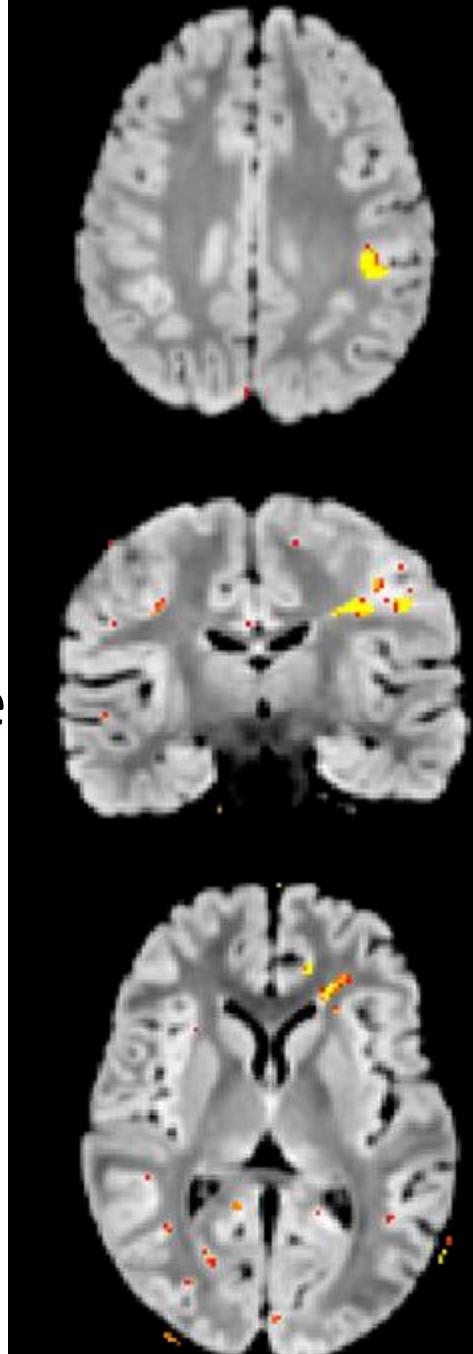


...Very preliminary conclusions

8 very selected cases.

Detecting a white matter transmantle sign is easy

1. MRI negative tends to stay MRI negative
2. Cortex is trickier – this system uses the entire MRI volume but surface based approaches are clearly preferable here
3. Doesn't use a healthy control reference group and only the standard epilepsy clinical protocol
4. Works (so far) with a very wide age range
5. Highlights features, doesn't classify





IoPPN
Gráinne
McAlonan
Declan Murphy
Steve Williams



Perinatal
Imaging
Serena Counsell
Mary Rutherford
Max Pietsch
Paul Aljabar
David Edwards



Developmental
Imaging &
Biosciences
Chris Clark
David Carmichael
Sara Lorio

Sackler
INSTITUTE FOR TRANSLATIONAL NEURODEVELOPMENT
translating basic science into treatment advances

