1

Statistical Testing of Medical Images

Emma O'Connor

&

Nick Fieller

Department of Probability & Statistics
University of Sheffield, UK

Sponsored by AstraZeneca, UK
18th January 2006

2

Acknowledgements

John Waterton

Andrew Holmes

& members of Imaging & Oncology Depts

AstraZeneca UK

3

Outline

- Introduction
- Available Data
- Exploratory analysis
  - histograms & kernel densities
- Main analysis
  - functional principal components & randomization tests
- Introduction to a second dataset
  - Comparison of analytic techniques
- What next?

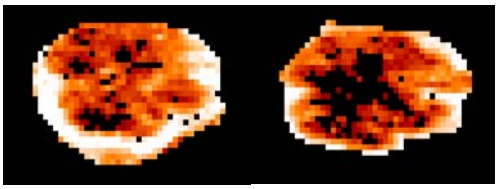
4

Introduction

- **General statistical objective**:
  - Explore avenues for statistical analysis of experiments when responses are images
- **Generic Problem:**
  - Measured response is an **'image'**
    - i.e values of 1 (or more) parameter(s) in pixels/voxels (+ spatial coordinates)
  - **N.B.** Not analyzing images *per se*

5

Pre Treatment Post Treatment



E14M34: 50mg dose

Maps of K^{trans} before and after treatment

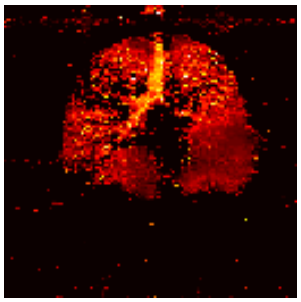
Lighter colour indicates higher permeability :–

note visible decrease in permeability post treatment

(Similar pairs of images for each of the 44 subjects)

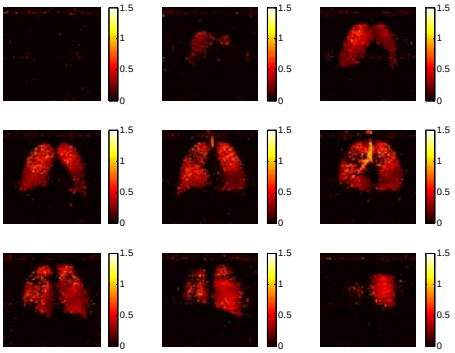
6

MRI using hyperpolarised helium-3 gas inhaled in the lungs – slice 6



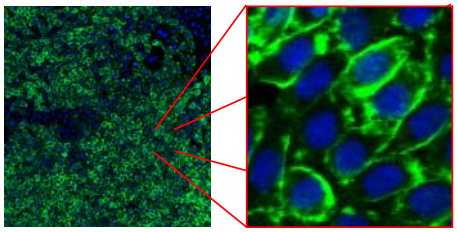
MRI using hyperpolarised helium-3 gas inhaled in the lungs - 9 slices from one scan

7



High throughput imaging of cell-based assays

8



9

- **Key Advantages of Magnetic Resonance Imaging:–**
  - Images are non-invasive
  - Subject can be monitored through time
    - Other measures of tumours require sacrifice of subject
  - Clear statistical advantages
    - Lessens ***inter-subject*** variability
    - Treatment comparison is ***intra-subject***

Available Data

10

- Three experiments formed a dose-response investigation
  - Four doses
    - 12.5mg, 25mg, 50mg, 100mg + control
- DCE-MRI of tumours before & 24 hrs after treatment
- Data provided:- **voxel values of K^{trans}**

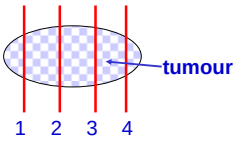
K^{trans} - reflects vascular permeability, flow, volume

11

- K^{trans} is the volume transfer constant between blood plasma and the extravascular extracellular space in the compartmental model used to quantify the changes in the DCE-MRI time series.
- K^{trans} has several physiologic interpretations, as it depends on vascular permeability, flow & vascular volume.

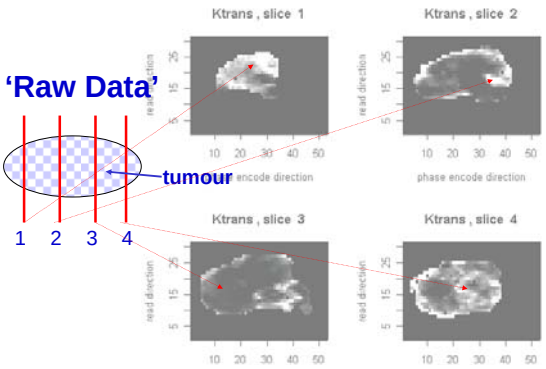
12

'Raw Data'



MRI scan focuses separately on four near-adjacent 'slices' measuring parameters in each voxel

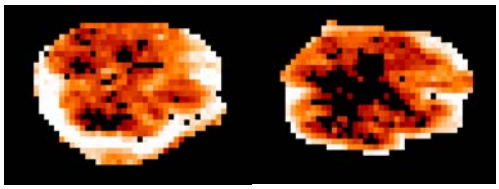
13



K^{trans} Maps of all four slices of *E12M48pre*

14

Pre Treatment Post Treatment



E14M34: 50mg dose

Maps of K^{trans} before and after treatment

Lighter colour indicates higher permeability :–

note visible decrease in permeability post treatment

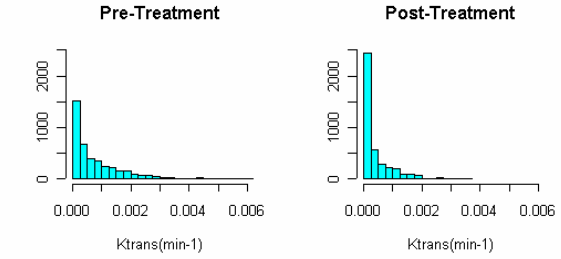
15

- **Technical Notes:**
  - Lack of registration between images
    - images 'before' & 'after' treatment are not aligned or 'registered' so cannot look at changes at voxel level
      - tumours are non-rigid
  - Lack of registration between slices
    - 2D slices are not perfectly registered so at best only partial information on proximity of voxels in different slices

16

Exploratory Analyses

- Images provided as four 2D slices
  - Pooled all values of K^{trans} from all four slices
    - One set of values pre- and post treatment
  - Many zero values of K^{trans}
    - (internal to image – ROI excludes 'external zeros')
  - Histograms of voxel values have very high positive skewness
  - ⇒ work with log(K^{trans})
    - or **log(ε+ K^{trans})** to avoid log(0)



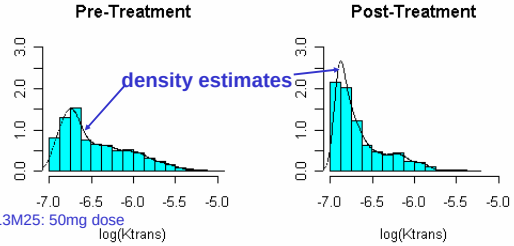
E13M25: 50mg dose

Histograms of values of K^{trans}

Note overall decrease in values post-treatment – typical in treated subjects

18

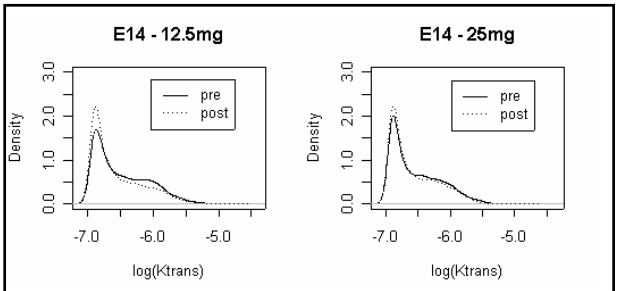
- *kernel density estimates* better than histograms
  - gives a 'smooth' version of the histogram



E13M25: 50mg dose

19

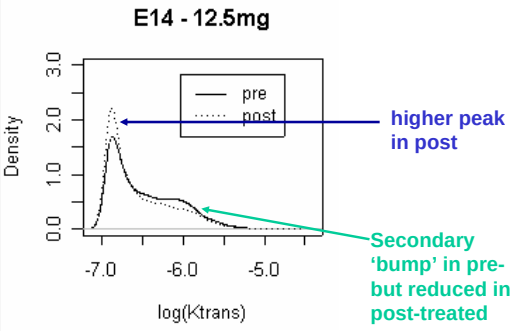
- **Notes:**
  - ♦ measured response: **image**
    ⇓
    sample distribution
    ⇓
    histogram
    ⇓
    Working unit:– kernel density estimate
  - ♦ **Gains:–** reduced random noise & comparability between samples
    - use common smoothing parameter
  - ♦ **Losses:–** actual values



Average densities for 12.5mg and 25mg dose groups

Note visible overall decrease in values post-treatment

21



These features generally typical in treated groups

22

Initial Analyses

- Simple sample summary statistics only capture changes in **location** and **scale**
- Exploratory analyses (kernel densities) showed more complex changes
- Kolmogorov-Smirnoff tests reveal small differences in distribution but say nothing about type of difference
  - not possible to see whether difference is clinically interesting or which sample is 'preferable'

23

Main Analyses

- Starting point:
  - ♦ measured response is a probability density
    - image ⇒ distribution ⇒ histogram ⇒ **density**
  - ♦ probability densities are **functions**
  - ♦ Key references:

Jim Ramsay & Bernard Silverman:

Functional Data Analysis

Applied Functional Data Analysis

(Springer, 1997 & 2002)

24

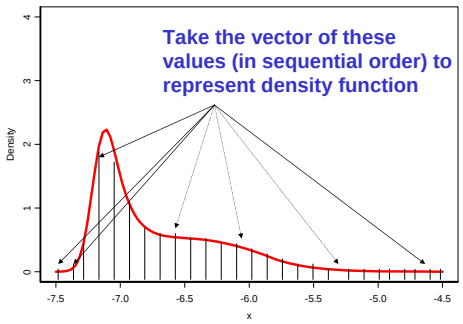
- **Key ideas:**
  - ♦ investigate components of **variability** between the functions
    - use a type of principal component analysis
  - ♦ relate these components to differences such as pre- / post- treatment or control / treated subjects or different dose levels

25

- **First step:**
  - need to summarize / parameterize the functions in a small number of values or parameters
    - note that kernel density estimates are *non-parametric* estimates
  - R & S recommend using a basis of B-splines to estimate log(density)
  - We use a simplistic approach and 'discretize' the density by taking its value at 100 equally spaced points

26

Discretization:



27

- measured response is a **vector**
  - image ⇒ distribution ⇒ histogram ⇒ density ⇒ **vector of 100 values**
  - **Losses:–** details of function by discretization
    - sensitivity analysis using grid of 50 or 200 values showed little difference in results
  - **Gains :–** all standard multivariate techniques such as PCA available
  - **But:–** need to interpret results of PCA in terms of density functions

28

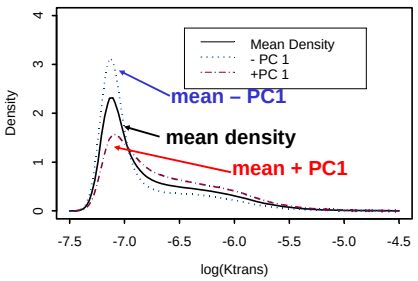
- ***Functional principal component analysis***
  - PCA on vector of 100 discrete values
  - Obtain first few (2 or 3) PCs
    - these are also vectors of 100 values
  - To interpret a PC we need to examine 'typical' densities that are varied from the overall mean in the direction of that PC
    - can do this by constructing a density function as overall mean density +/– PC
    - i.e. add / subtract PC vector from mean vector and then plot values as a density function

29

- Technical Notes:
  - The PC vector added/subtracted from the mean density needs to be a multiple within the limits of the data
    - Avoids negative densities
  - Need to rescale these 'new' densities to have unit area
    - Practically though, there is little difference in area

30

PC 1 (77.6%)



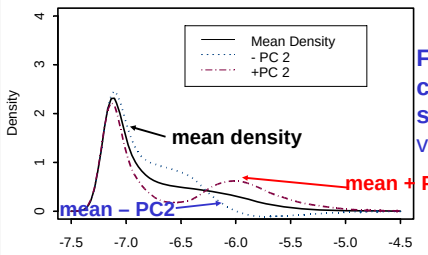
FPCA on complete data set: 77.6% of variation in PC1

PC1 reflects change in height of primary mode

low scores on PC1 ⇔ high peak, high score ⇔ low peak

PC 2 (8.5%)

31



FPCA on complete data set: 8.5% of variation in PC2

PC2 reflects presence of secondary 'bump'

low scores on PC2 ⇔ no bump, high score ⇔ bump

32

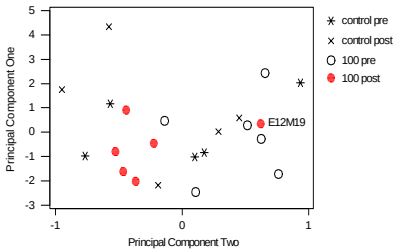
- FPCA (continued)
  - Signs of PCs are arbitrary:–
  - we have chosen signs so low values are **'desirable'** in current context
  - low PC1 ⇔ high peak of low values
  - low PC2 ⇔ no second 'bump' of high values
  - low scores on first 2 PCs ⇔ generally lower permeability & especially fewer fairly high values

33

- FPCA (continued)
  - Can calculate scores for each probability density on the PCs
  - image ⇒ distribution ⇒ histogram ⇒ density function ⇒ vector of 100 values **⇒ scores on first two PCs**
  - **Gains:–** can plot samples in a scatter plot
    - points towards bottom left corner are 'good'
      - (since they have low scores on PCs)
    - Can compare 'pre-' and 'post-' treated

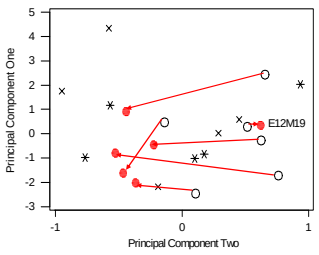
Discarding low order PCs reduces random noise

34



Score plot on first 2 PCs for Expt 12

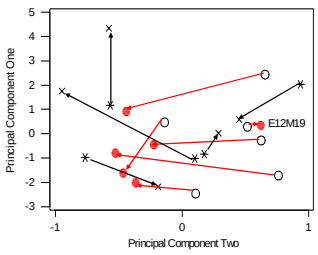
35



Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

Score plot on first 2 PCs for Expt 12

36

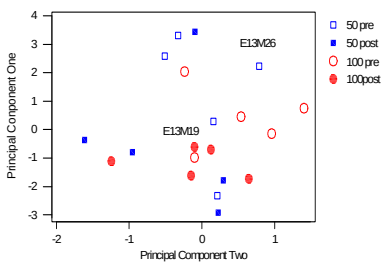


Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

No consistency in direction of change in controls

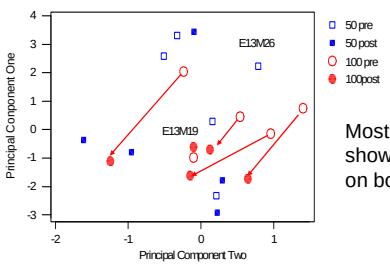
Score plot on first 2 PCs for Expt 12

37



Score plot on first 2 PCs for Expt 13

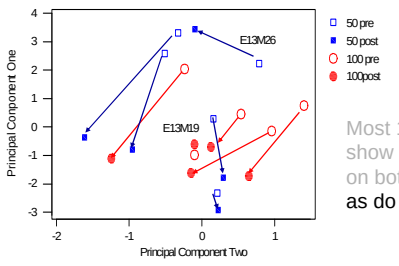
38



Most 100mg subjects show decrease in scores on both PC1 & PC2

Score plot on first 2 PCs for Expt 13

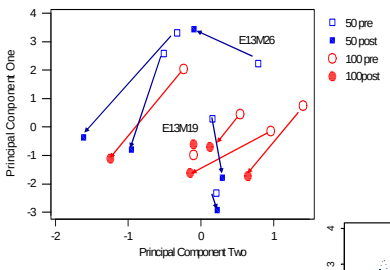
39



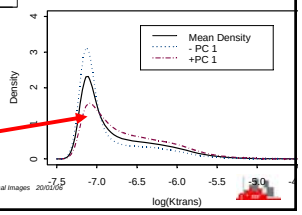
Most 100mg subjects show decrease in scores on both PC1 & PC2
as do most 50mg subjects

Score plot on first 2 PCs for Expt 13

40

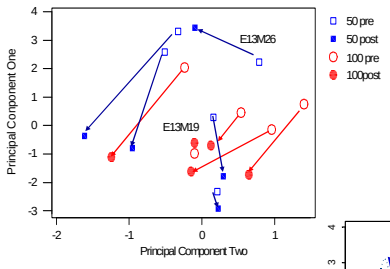


PC 1 (77.6%)



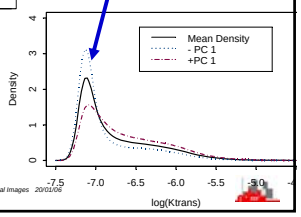
This shift on PC1 corresponds to change from this density pre treatment

41



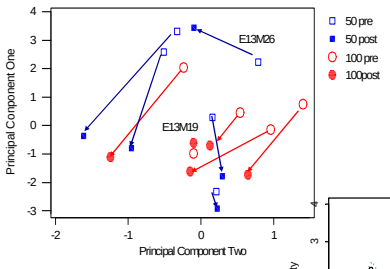
To this density post treatment

PC 1 (77.6%)

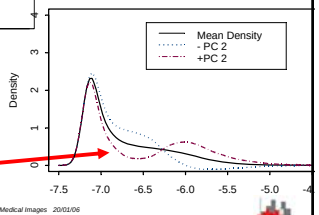


This shift on PC1 corresponds to change from this density pre treatment

42

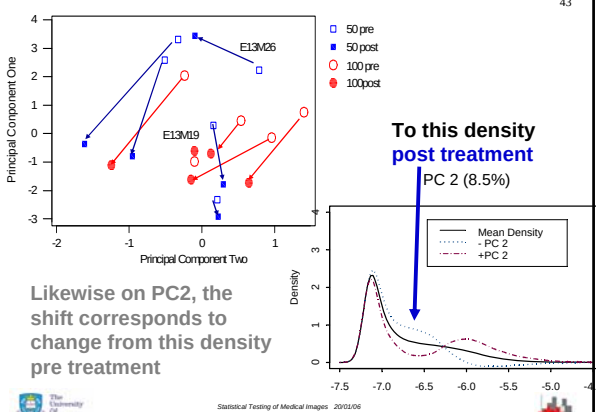


PC 2 (8.5%)



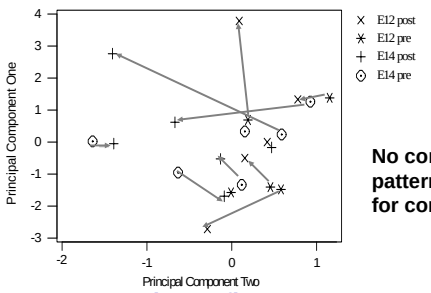
Likewise on PC2, the shift corresponds to change from this density pre treatment

43



Likewise on PC2, the shift corresponds to change from this density pre treatment

44

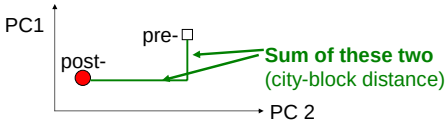


No consistent pattern of change for control subjects

Score plot on first 2 PCs control subjects

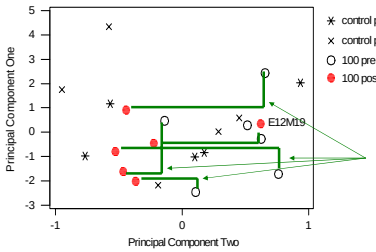
45

- Formal statistical testing:–
  - Calculate test statistic
  - Assess significance by a randomization test based on random relabelling of pre- & post-treated values
- Test statistic based on sum of differences in scores on PC1 & PC2



Sum of these two (city-block distance)

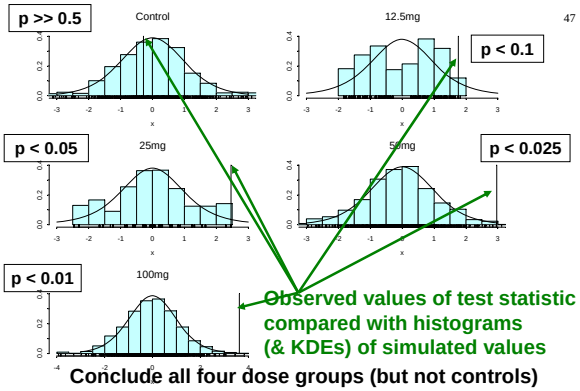
46



test statistic is sum of all these distances

Compare observed value with values from randomly interchanging pre- & post- labels
(near equivalent of paired t-test)

47



Observed values of test statistic compared with histograms (& KDEs) of simulated values

Conclude all four dose groups (but not controls) show clear evidence of effect of treatment

48

Interim Conclusions

- Overall picture is that treated subjects show a reduction in permeability both in overall level and in distribution of levels
  - higher proportion of low values
  - lower proportion of fairly high values
- Control subjects show no consistent change over 24 hours
- Additional insight into nature of drug effect

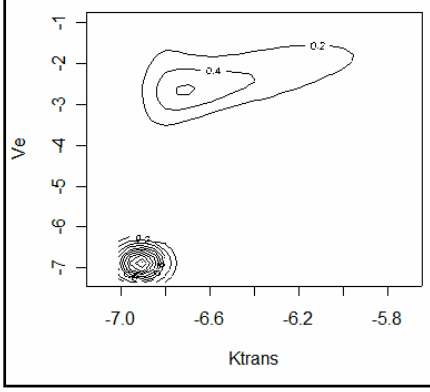
Next step: Multi-dimensional Analysis

49

- Theory extends to higher dimensions
  - Analyse K^{trans} and V_e simultaneously
- Method in brief...
  - Perform 2D kernel density estimation and discretize distributions over a 2D grid
  - Perform FPCA and interpret results similar to the 1D case, using CONTOUR plots

50

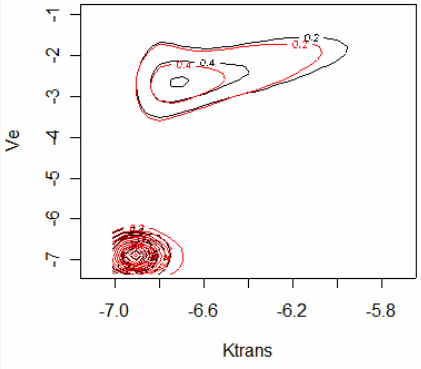
E12 - mean



Contour plot of the average density

51

E12: mean - PC1



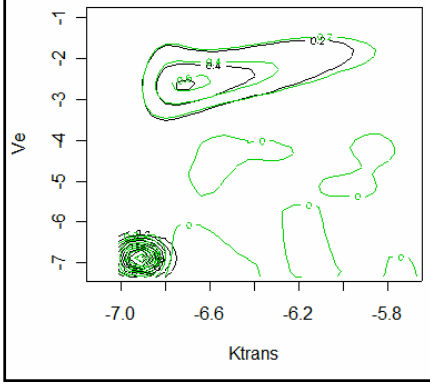
Contour plot of the average density

FPCA...

Average density minus PC1

52

E12: mean + PC1



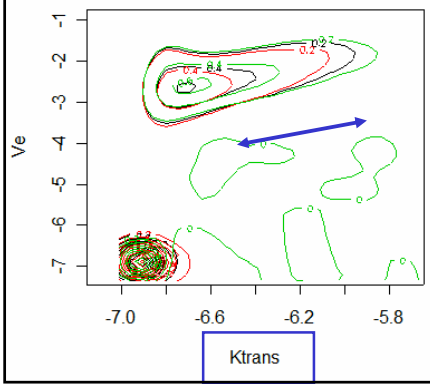
Contour plot of the average density

FPCA...

Average density plus PC1

53

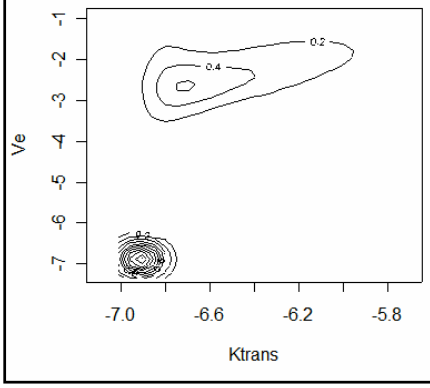
E12: mean +/- PC1



PC 1 shows change in K^{trans}

54

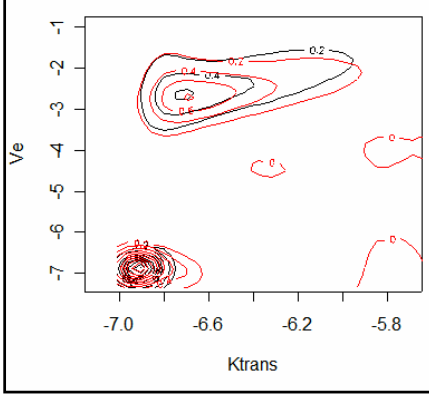
E12 - mean



Contour plot of the average density

55

E12: mean - PC2



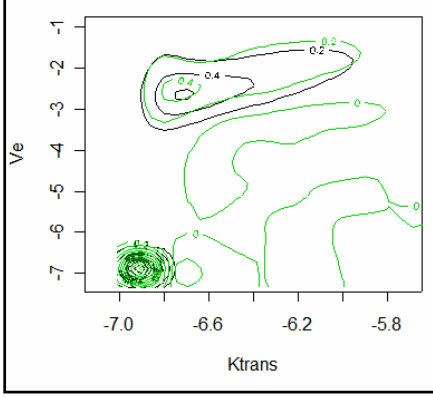
Contour plot of the average density

FPCA…

Average density minus PC2

56

E12: mean + PC2



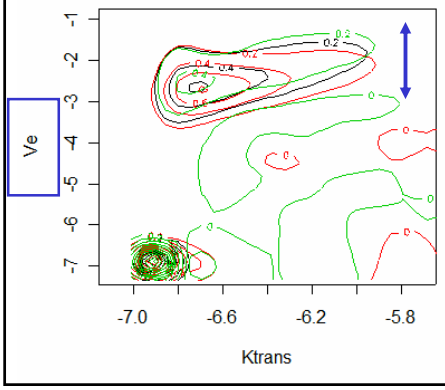
Contour plot of the average density

FPCA…

Average density plus PC2

57

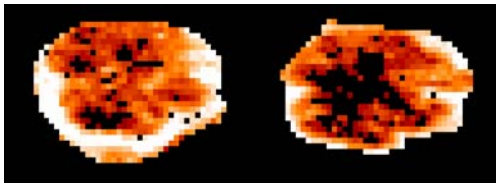
E12: mean +/- PC2



PC 2 shows change in V_e

58

Pre Treatment Post Treatment



E14M34: 50mg dose

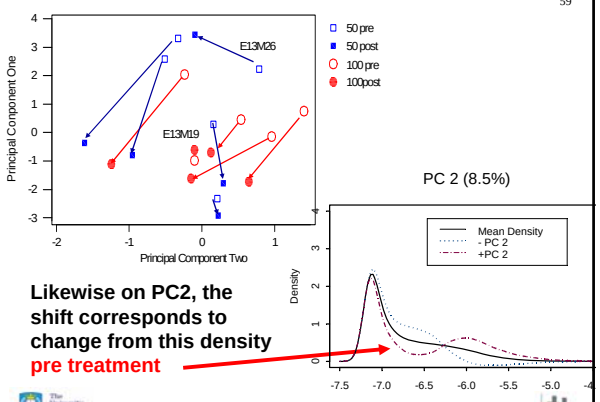
Maps of K^{trans} before and after treatment

Lighter colour indicates higher permeability :–

note visible decrease in permeability post treatment

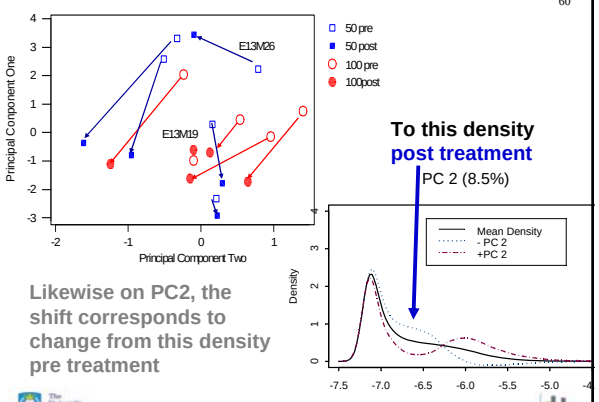
(Similar pairs of images for each of the 44 subjects)

59



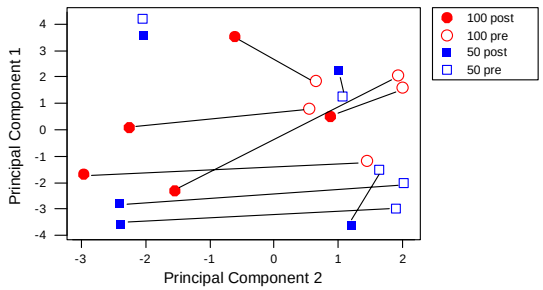
Likewise on PC2, the shift corresponds to change from this density pre treatment

60



Likewise on PC2, the shift corresponds to change from this density pre treatment

Experiment 13: Bivariate density: PC1 vs PC2

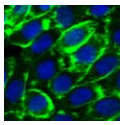


Most subjects show decrease in scores on both PC1 & PC2 – a decrease in K^{trans} and Ve post-treatment

- **Was this a one-off data set?**
  - Would technique work on other problems?
  - Can we say 'how well does the technique work'?
    (e.g. compare with some yardstick)
  - Can we extend the multi-parameter approach to other datasets?

Dataset 2: HTS

- Replicated dose-response data set
  - Measurements taken from each cell in every well
  - Two parameters of interest
    - **Ipeak** and **Ixpr**



- Distribution histograms of both variables show:
  - Clear change in location ('≡ mean') with dose
  - Systematic changes in skewness and spread

- **Two analyses:**
  - Non-parametric using *functional principal components*
    - FPCA takes groups of distributions and characterizes the modes of variation between those in that particular group
  - Parametric analysis using log-hyperbolic family
    Used previously in geology
    - particle sizes of sediments
  - statistical finance
    - values of derivatives and futures etc
  - Parametric analysis takes each distribution separately to estimate the hyperbolic parameters and we can assess how well the model fits each sample
  - Then observe the relationship with dose

Parametric Analysis

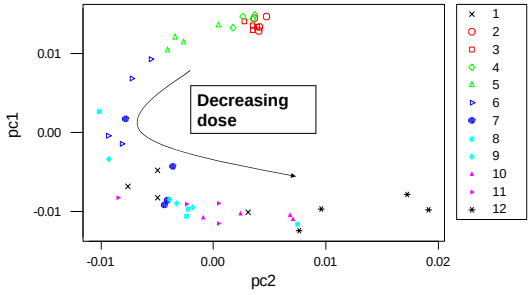
- Log hyperbolic distribution:
  - 4 parameters, various different parameterizations
  - very flexible unimodal family allowing both positive and negative skewness
  - we use a parameterization with
    - location parameter **μ**, scale parameter **δ**, **ζ** measures 'peakedness' and for fixed **ζ** the parameter **π** expresses the asymmetry of the distribution.
  - This distribution has the probability density function;

$$f(x) = \frac{1}{2\sqrt{(1+\pi^2)}K_1(\zeta)} \exp\left\{-\zeta\left[\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2} - \pi\frac{x-\mu}{\delta}\right]\right\}$$

    - where K1( ) is the modified Bessel function of the third kind with order 1

Results: FPCA

Ipeak: PC1 vs PC2 by Dose

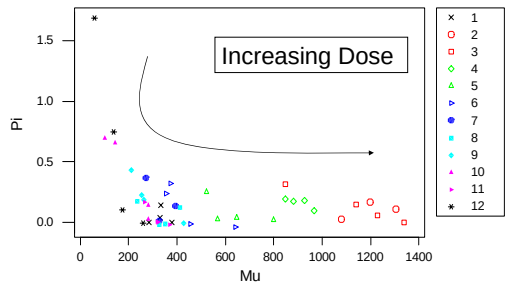


Clear dose-response relationship on PC1 with extra information on PC2

Results: Parametric Analysis

7

lpeak: Hyperbolic Estimates: pi vs Mu



Clear that this detects more than just a location change with Mu: also systematic change in skewness.

68

- Results consistent (happily!)
  - Both the non-parametric FPCA and the parametric log-hyperbolic pick up consistent changes in mean & in skewness with changing dose (i.e. dose-response)
  - This example provides good validation of the FPCA method to give confidence when it is the only one available

What Next?

69

- Verification of this method on a variety of datasets - ongoing
- Improving the 'discretization' step
  - Repeat the analysis using SPLINES / WAVELETS instead of 'discretizing' the kernel density estimate
    - Possibly more sensitive

http://www.pas-postgrads.group.shef.ac.uk/oconnor/