1

Using Medical Images as Primary Outcome Measures

Nick Fieller

Department of Probability & Statistics
University of Sheffield, UK

Newcastle University, 2nd February 2007

2

- **Joint work with Emma O'Connor**

- **Acknowledgements**

John Waterton
Andrew Holmes
& members of the Discovery, Oncology & Imaging Depts.
AstraZeneca

3

Outline

- Introduction
- Exploratory analysis
  - ♦ histograms & kernel densities
- Main analysis
  - ♦ functional principal components & randomization tests
- Extensions to bivariate images
- Parametric models
- Loose ends & refinements

4

Introduction

- Medical Images
  - NMR, X-Ray, Microscopy, …
  - ♦ key research tool in drug development
  - quickly reveal changes & nature of effects to scientists but not widely used as outcome measures in clinical trials
  - ♦ Some simple [simplistic?] exceptions
  - x-rays of broken legs
  - brain scans
  - … … …

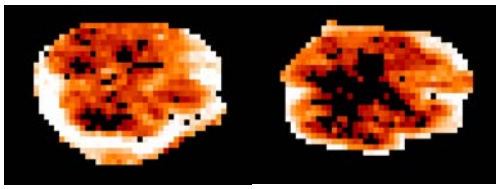
5

Examples

- NMR images of tumours
  - ♦ 3D images of complete tumour can reveal blood flow
    - changes 'obvious' 24 hours after treatment
- NMR images of lungs
  - [hyperpolarised helium-3]
  - ♦ reveals capacity & oxygen take-up
    - differences in CPD, smokers & normals
- High-throughput imaging of cell-based assays
  - *in vivo* dose-response assessment

6

Pre Treatment Post Treatment



E14M34: 50mg dose

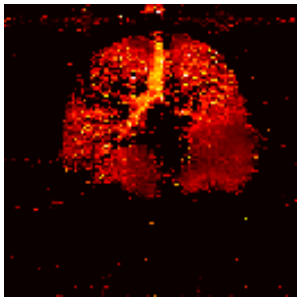
Maps of K^{trans} before and after treatment

Lighter colour indicates higher permeability :–

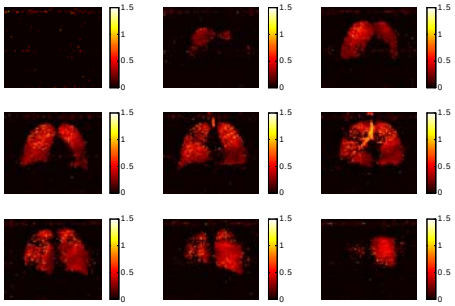
note visible decrease in permeability post treatment

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

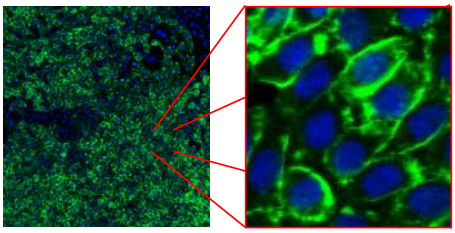
MRI using hyperpolarised helium-3 gas inhaled in the lungs – slice 6



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



High throughput imaging of cell-based assays



- These images are quick, cheap, non-invasive, clear & so are used in

  drug development

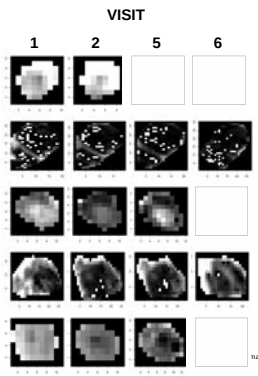
  - does it work?
  - how does it work?
  - can it be made better?
- But for formal statistical assessment some other measure is used
  - tumour weight, survival time, FEPV, … …

WHY?

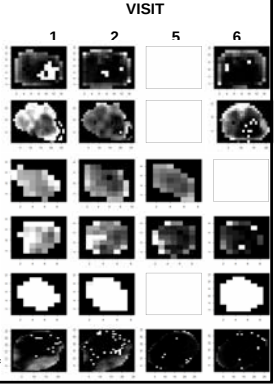
- need statistical techniques for describing & assessing ***variability*** between images
  - before & after treatment
  - ***between*** treatment groups
  - ***within*** treatment groups

(statistics is about describing & assessing variability in the context of uncertainty)

Individual **IAUC60** Maps – Dose 20



Individual **IAUC60** Maps – Dose 30



13

Example on Tumours

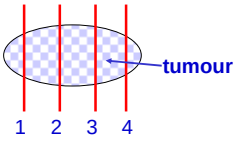
- 3 experiments:-
  change in vascular permeability in tumours after drug designed to inhibit angiogenesis
- Four doses
  - ♦ 12.5mg, 25mg, 50mg, 100mg + control
- 3D images of tumours by MRI
  before & 24 hrs **after** treatment
- Data provided:-

voxel values of K^{trans}

(& V_e & other parameters, see later)

14

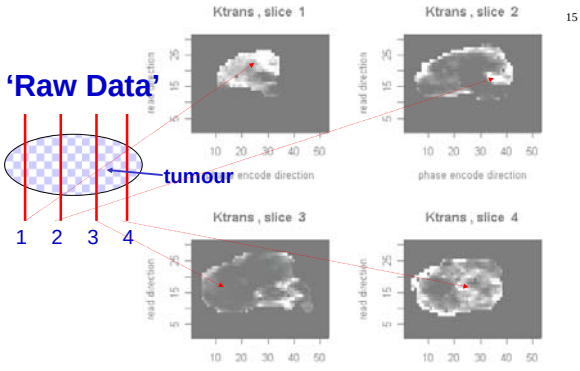
'Raw Data'



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

15

'Raw Data'

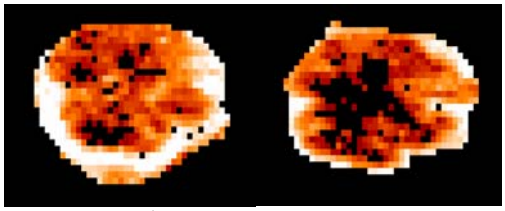


K^{trans} Maps of all four slices of *E12M48pre*

16

- **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 unlike images of the brain
  - ♦ Lack of registration between slices
    - Images provided as four 2D slices that are not perfectly registered so at best only partial information on proximity of voxels in different slices

17



Maps of K^{trans} before and after treatment

E14M34: 50mg dose

Images of tumour before and after treatment

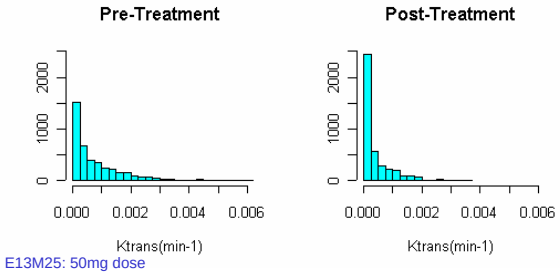
Lighter colour indicates higher permeability :–

note visible decrease in permeability post treatment

18

Exploratory analysis

- 3D images provided as four 2D slices
  - ♦ 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(1+ K^{trans}) or **log(ε+ K^{trans})** to avoid log(0)



E13M25: 50mg dose

Histograms of pixel values in slices 2 & 3 pooled

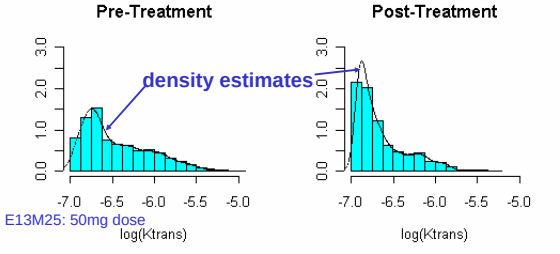
(note overall decrease in values post-treatment)

- **Notes:**
  - Measured response:– Image
  - Working unit:– sample distribution of K^{trans}
  - Losses
    - spatial information
      - (but see later loose end)
  - Gains
    - more familiar statistical object
    - allows calculation of 'mean responses' & assess variability relative to mean
      - (by pooling sample distributions)

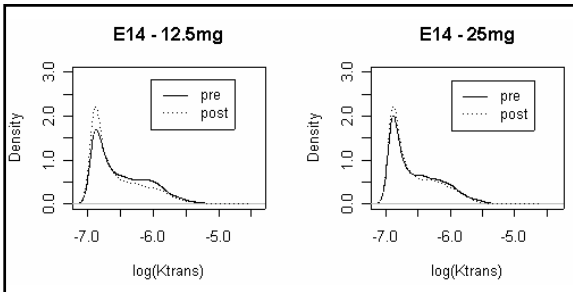
- **Notes(ct$^{d.}$)**
  - Ignoring spatial information masks ***dependence*** of voxel values
    - K^{trans}–values are not independent
    - Beware of statistical techniques which assume independence
    - 'sample size' is ***NOT*** #voxels
    - true sample size is #images, i.e. 88 in total from 44 individuals ~ 6 in each dose group

- **Notes(ct$^{d.}$)**
  - Working with distribution looses information on ***size*** of tumour
    - size change in 24 hours implausible
      - but could be an issue over weeks/months
  - Beware of effects of zero values of K^{trans}
    - (trace zeroes ←→ structural zeroes)
    - reliant on accuracy of segmentation techniques to determine ROI

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



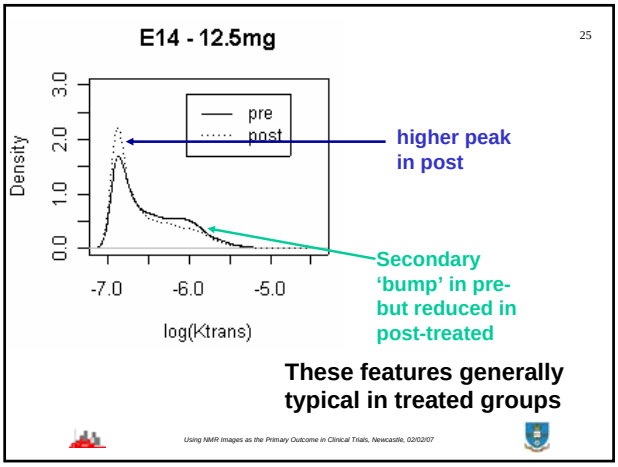
E13M25: 50mg dose



Average densities for 12.5mg and 25mg dose groups

Note visible overall decrease in values post-treatment

25

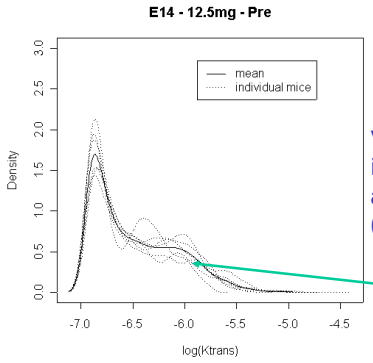


higher peak in post

Secondary 'bump' in pre- but reduced in post-treated

These features generally typical in treated groups

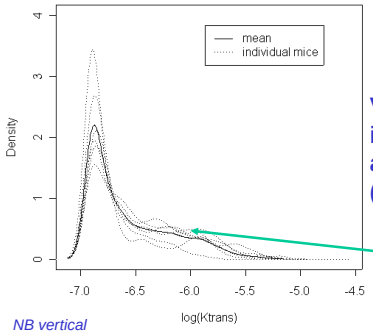
26



Variability of individuals about mean (pre-treatment)

Secondary bump

27



Variability of individuals about mean (post-treatment)

Secondary bump reduced / absent

NB vertical scale changed

28

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

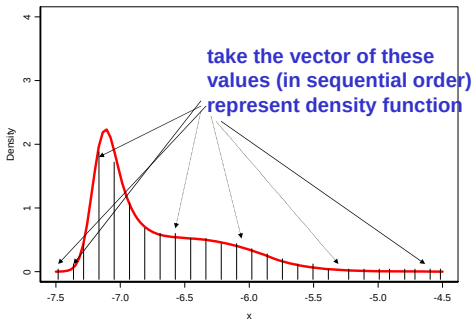
Applied Functional Data Analysis

(Springer, 1997 & 2002)

29

- **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

30



take the vector of these values (in sequential order) to represent density function

31

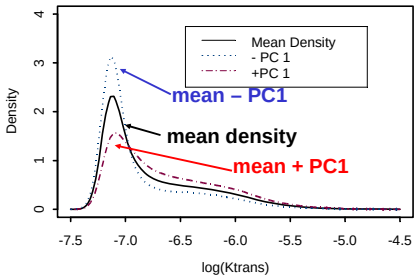
- measured response is a **vector**
  - ♦ image ⇒ distribution ⇒ histogram ⇒ density **⇒ vector of 100 values**
  - ♦ **Losses:–** details of function by discretization
    - sensitivity analysis using grid of 50 or 200 values shewed little difference in results
  - ♦ **Gains :–** all standard multivariate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

32

- 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

PC 1 (77.6%)

33



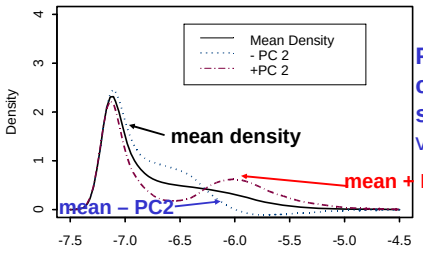
FPCA on complete data set: 77.6% of variation in PC1

PC1 reflects change in height of primary mode

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

PC 2 (8.5%)

34



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

35

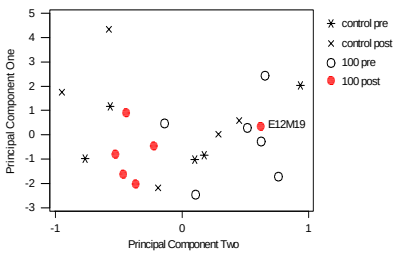
- FPCA (continued)
  - ♦ Signs of PCs are arbitrary:–
  - ♦ we have chosen signs so low values are **'desirable'** in current context
  - ♦ low PC1 ⇔ high peak of low values
  - ♦ low PC2 ⇔ no second 'bump' of high values
  - ♦ low scores on first 2 PCs ⇔ generally lower permeability & especially fewer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

36

- 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

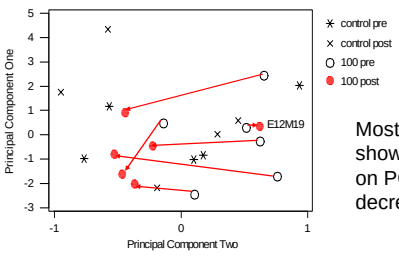
      Discarding low order PCs reduces random noise

37



Score plot on first 2 PCs for Expt 12

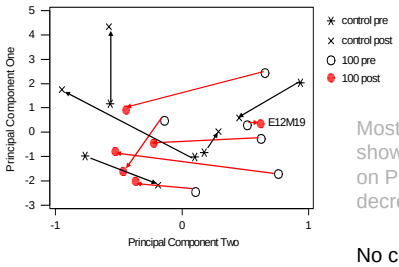
38



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

Score plot on first 2 PCs for Expt 12

39

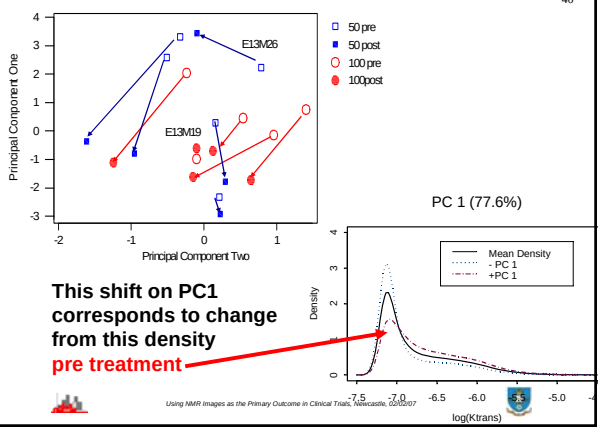


Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

No consistency in direction of change in controls

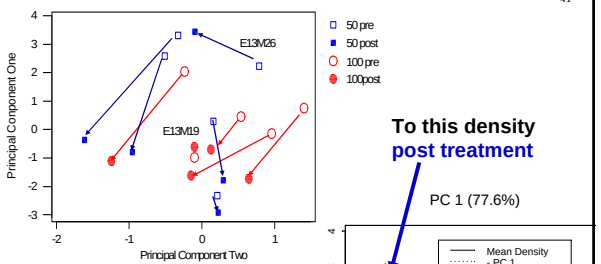
Score plot on first 2 PCs for Expt 12

40



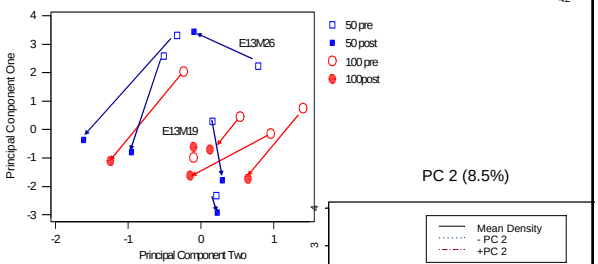
This shift on PC1 corresponds to change from this density pre treatment

41



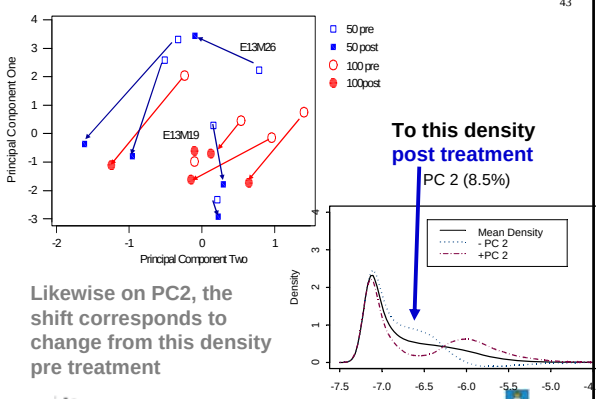
This shift on PC1 corresponds to change from this density pre treatment

42



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

43



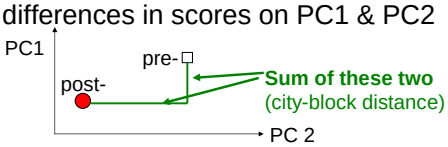
Likewise on PC2, the shift corresponds to change from this density pre treatment

44

- **Interim conclusions**
  - overall picture is that treated subjects shew 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 shew no consistent change over 24 hours

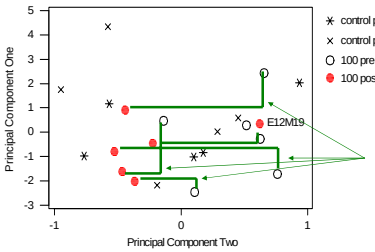
45

- 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)

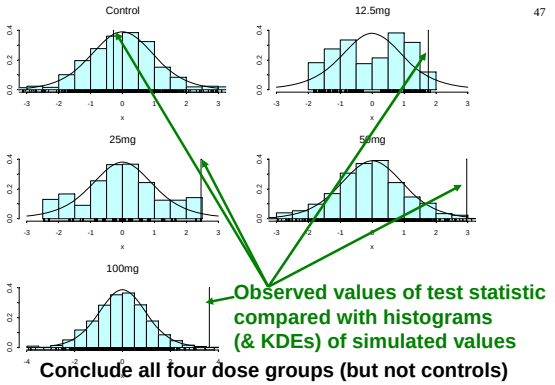
46



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) shew clear evidence of effect of treatment

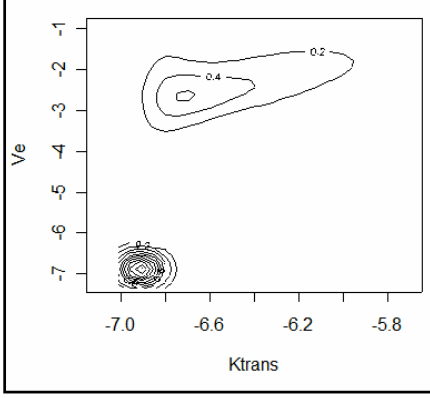
8

Next step: Multi-dimensional Analysis

- 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 plot

49

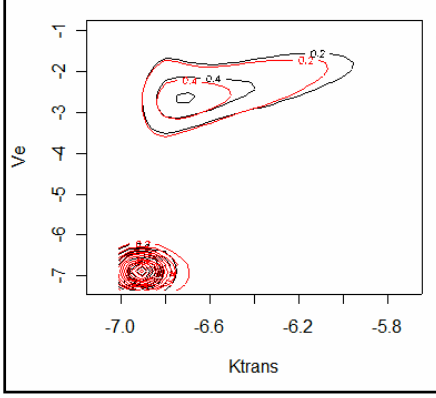
E12 - mean



Contour plot of the average density

50

E12: mean - PC1



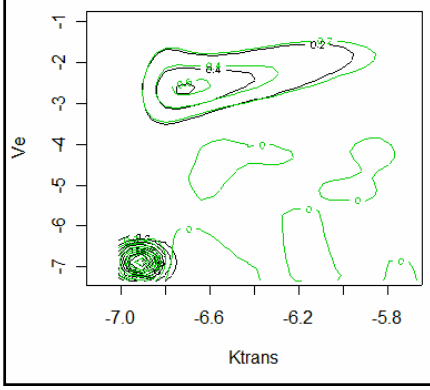
Contour plot of the average density

FPCA…

Average density minus PC1

51

E12: mean + PC1



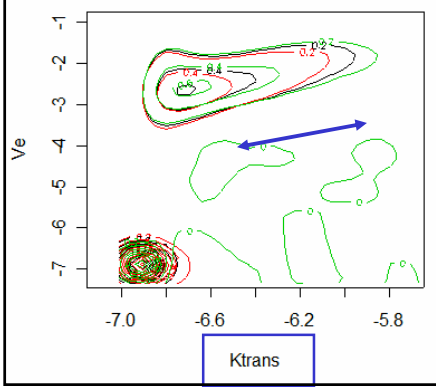
Contour plot of the average density

FPCA…

Average density plus PC1

52

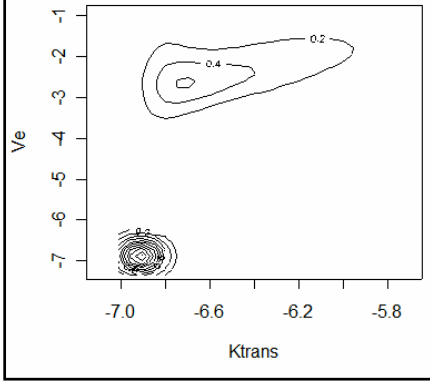
E12: mean +/- PC1



PC 1 shows change in K^{trans}

53

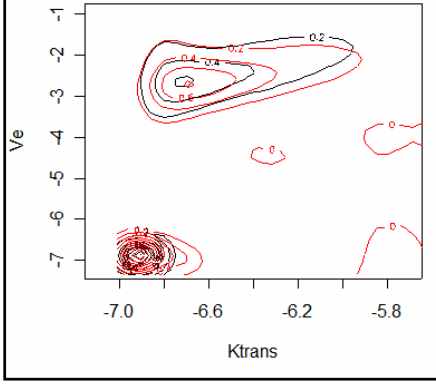
E12 - mean



Contour plot of the average density

54

E12: mean - PC2



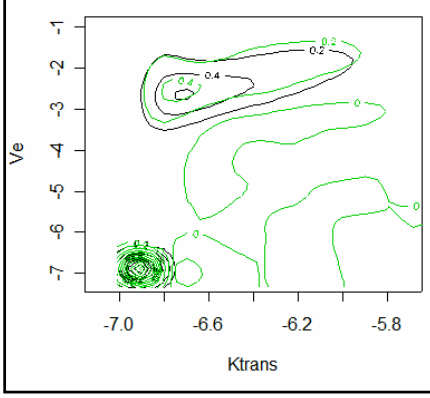
Contour plot of the average density

FPCA…

Average density minus PC2

55

E12: mean + PC2



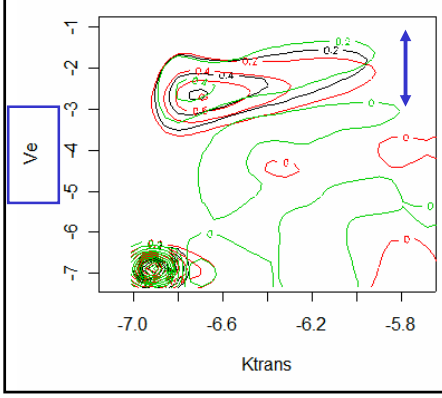
Contour plot of the average density

FPCA…

Average density plus PC2

56

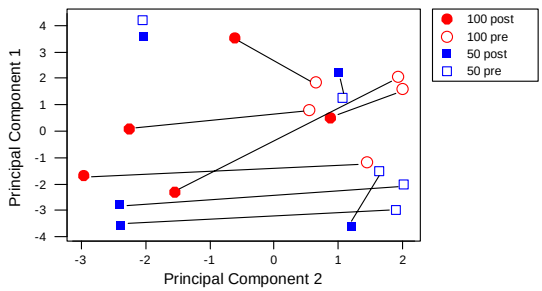
E12: mean +/- PC2



PC 2 shows change in V_e

Note change in V_e only occurs in voxels with high K^{trans} values

Experiment 13: Bivariate density: PC1 vs PC2

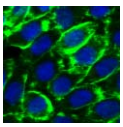


Most subjects show decrease in scores on both PC1 & PC2 – a decrease in K^{trans} and Ve post-treatment

58

Example of dose-response study: HTS with cell-based assay

- Replicated data
  - 12 dose levels (including controls)
  - Two parameters of interest
    - **Ipeak** and **Ixpr**

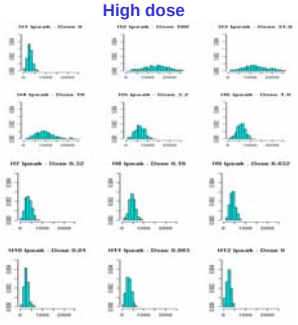


- Histograms of both variables show:
  - Clear change in location with dose
  - Systematic changes in skewness and spread

9

Distribution Histograms: Ipeak: Well H

High dose



Decreasing mean with decreasing dose

Decrease in variance with decreasing dose

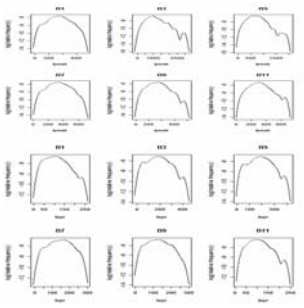
Not much visible change in skewness with decreasing dose

low dose

60

- **Two** analyses presented:
  - Non-parametric using FPCA
  - Parametric using log-hyperbolic distributions
    - Four parameters, allows skewness and 'heavy tails'
- 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)

Log Histograms: Ipeak: Well B

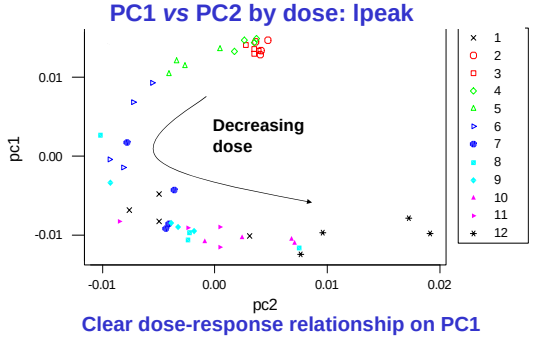


Very distinctive hyperbolae: some suggestion of secondary peak but this is exaggerated by vertical log scale

Suggests fitting log hyperbolic distributions and looking at parameter estimates: we plot
- μ (location) and
- π (skewness)

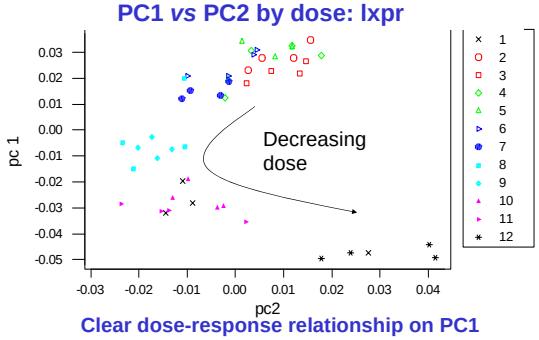
N.B. differing horizontal axes

non-parametric FPCA analysis
PC1 *vs* PC2 by dose: Ipeak

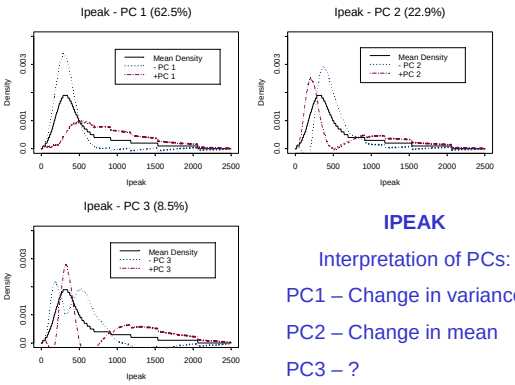


Clear dose-response relationship on PC1
with extra information on PC2 (interpretation of PCs later)

non-parametric FPCA analysis
PC1 *vs* PC2 by dose: lxpr



Clear dose-response relationship on PC1
with extra information on PC2 (interpretation of PCs later)



IPEAK

Interpretation of PCs:

PC1 – Change in variance

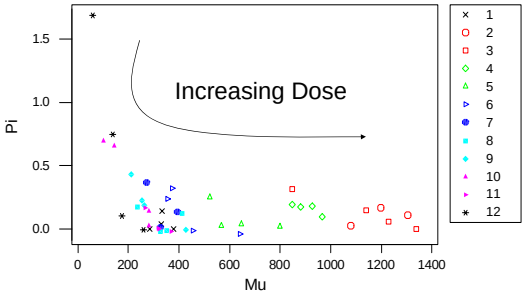
PC2 – Change in mean

PC3 – ?

FPCA Interpretation

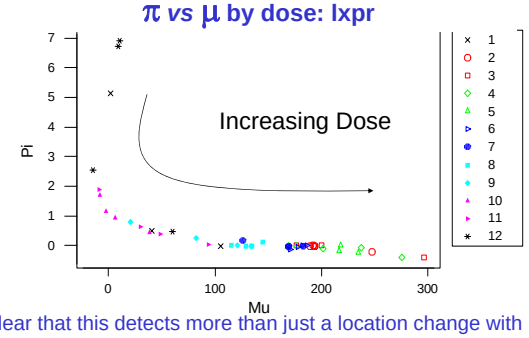
- PC1 separates the middle range doses
- PC2 separates the low doses
- PC3 separates the high doses

parametric analysis with log-hyperbolic
π *vs* μ by dose: Ipeak



parametric analysis with log-hyperbolic

π *vs* μ by dose: lxpr



Clear that this detects more than just a location change with μ: also systematic change in skewness (π)
Yet to know whether this is biologically of interest

Some differences between analyses:

68

- Parametric analysis takes **each** distribution separately to estimate the hyperbolic parameters & we can assess how well the model fits each sample
  We then observe the relationship with dose
- Non-parameteric FPCA takes **groups** of distributions & characterizes the modes of variation between those in that particular group

Loose ends & refinements

69

- Density estimation
  - needed? why not use histograms?
    - yes unless very large number of observations
      - e.g. brain scans, 20,000 pixels
    - 'smoothing' of histogram important
      - not sensitive to particular choices
- Alternatives to kernel densities?
  - doesn't matter
    - wavelets, splines give similar analyses
      - might be easier to extend into > 2 dimensions?

70

- Issues on inference
  - especially with FPCA
    - analysis may be dependent on ***groups*** of observations
      - e.g. should PCA be including/excluding controls??
  - especially with 'testing'
    - prefer just to describe but need a more formal test to show efficacy in a formal clinical trial

71

72