Detecting drug-related signatures in the electroencephalogram using Extended Semi-Linear Canonical Correlation Analysis (ESLCCA)

Phil Brain & Magnus Ivarsson
Pfizer Global R & D, Sandwich, UK

Genstat 2010 (14/7/2010)

Overview

- What are we trying to achieve?
- Electroencephalogram (EEG) experiments
- What the data looks like
- ESLCCA
- Application to EEG data
- Conclusions

What are we trying to achieve?

- Develop a **non-invasive translatable** method of identifying and profiling the potential effect of the drug on the brain.

Current Methods

Drug → **Blood** → **Brain**

Blood: Measurable

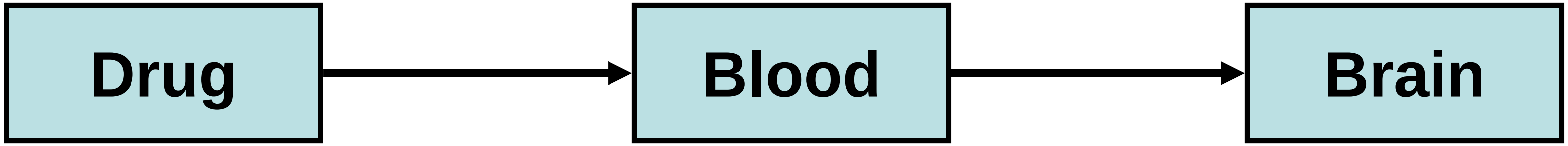
Brain: Not directly measurable

Subjective

Definition of the electroencephalogram (EEG)

- Measure of electrical activity produced by the brain as recorded from electrodes on the scalp over time
- Looking at **differences** in voltages between baseline-electrode and other electrodes

Aim



Drug → **Blood** → **Brain**

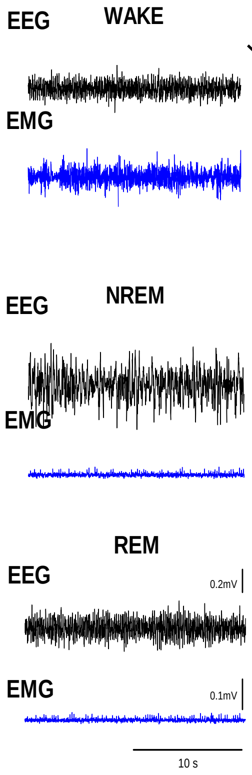
Measurable

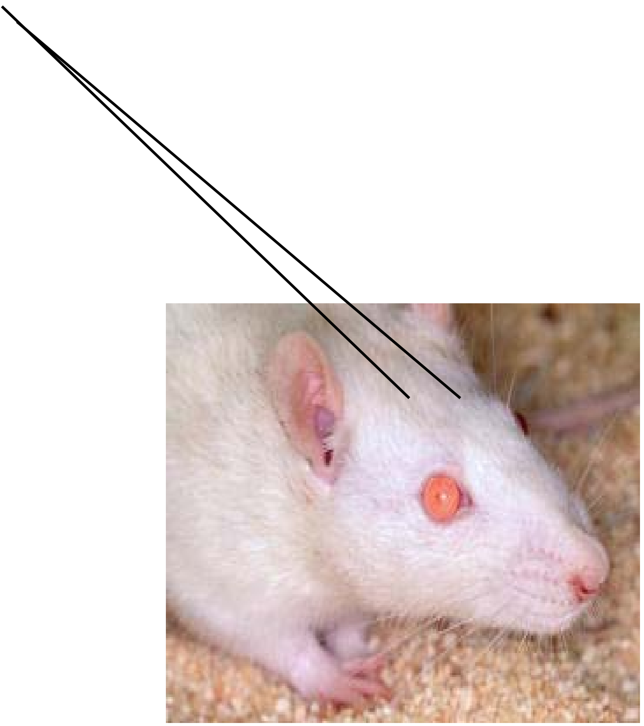
Measurable

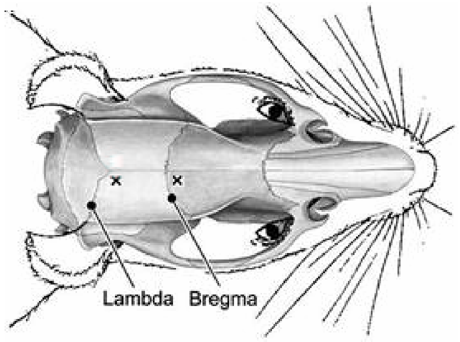
Objective

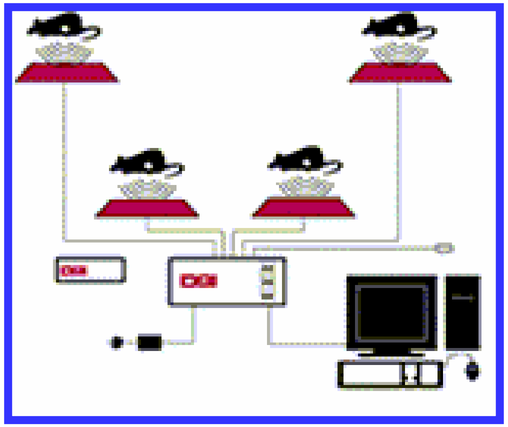
Electroencephalogram (EEG) experiments – in rat

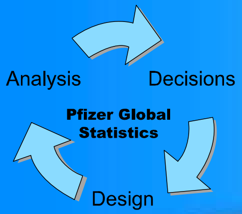
- Rat EEG data is sampled with telemetry
- Maximum of 3 electrodes in parallel
- Objective automatic analysis of three vigilance stages – WAKE, NREM and REM











Analysis

Decisions

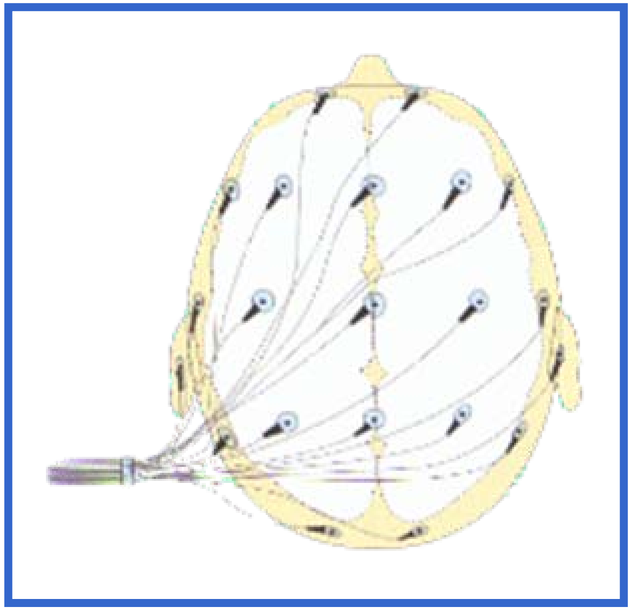
Pfizer Global Statistics

Design

EEG experiments – in human

- 1-256 recording electrodes (21 is standard)
- Electrodes as placed in standard positions on the scalp (10-20 system)
- Digital EEG makes it possible for further analysis

21 channels



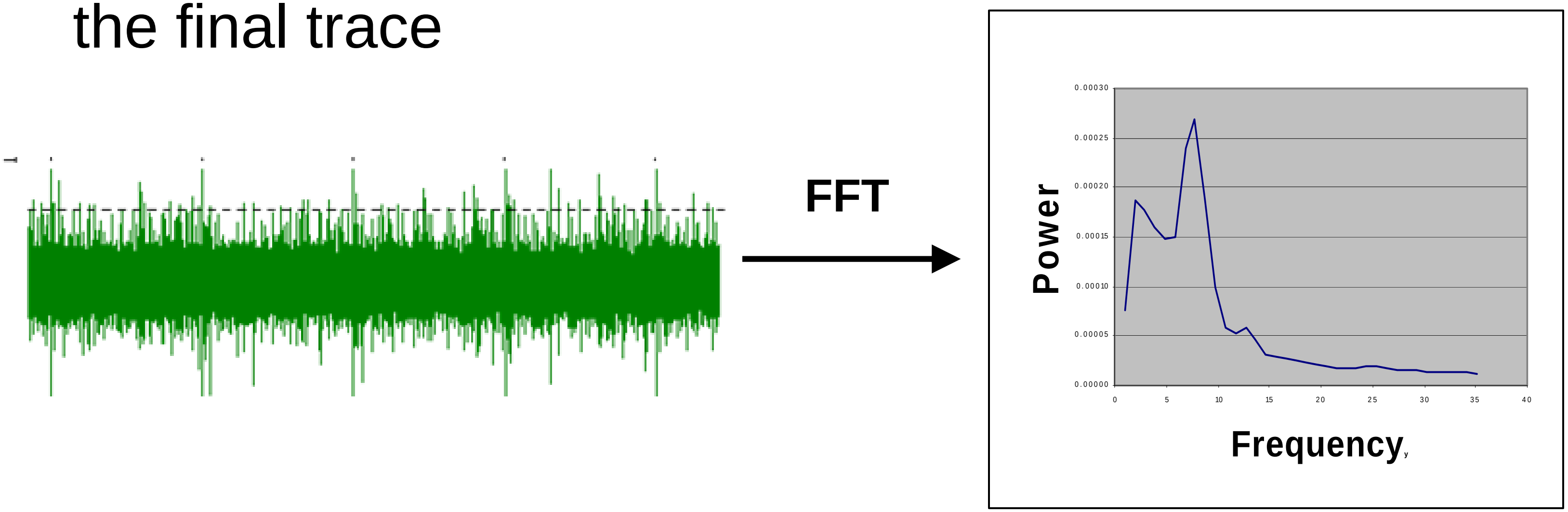
256 channels

What the data looks like

- Experiments involve 8 rats
- 4-period cross-over for each rat
- In each period EEG measurements made over 12 hours
- 4 doses of drug
    - **Control**
    - **Low Dose**
    - **Medium Dose**
    - **High Dose**
- To date – data from some 11 drug x light/dark combinations; many more in the pipeline

What the data looks like

- EEG data is recorded continuously over time (2 electrodes per rat); difference in the two voltages gives the final trace



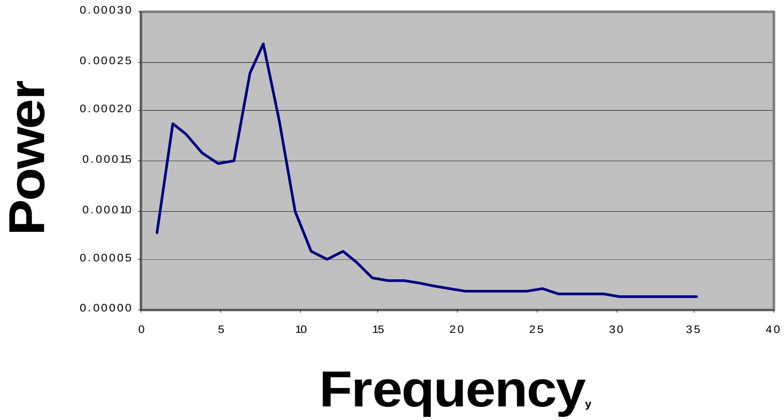
- Use Fast Fourier Transform to find the “amount” of each frequency for each 12 sec time “slice”

Typical data from an experiment

			Frequency						
Rat	Treatment	Time (sec)	0.98	1.96	2.93	3.91	4.89	5.87	
39	Control	0	5.17255E-05	0.000106758	7.28047E-05	5.56779E-05	5.96103E-05	7.08187E-05	-
39	Control	300	4.85099E-05	0.000102449	7.81046E-05	5.99523E-05	4.93621E-05	4.60752E-05	-
39	Control	600	5.25064E-05	0.00011139	8.70883E-05	6.94745E-05	5.57209E-05	6.41247E-05	-
39	Control	900	4.5226E-05	0.000100314	7.73427E-05	5.83909E-05	5.091E-05	4.80755E-05	-
39	Control	1200	3.03898E-05	7.8413E-05	7.60568E-05	6.69359E-05	5.44237E-05	4.62054E-05	-
:	:	:	:	:	:	:	:	:	

- The data is in the form of the Power at every frequency, at each time, so can be conveniently viewed as a matrix.
- Each row of the matrix is a Power Spectra
- An individual matrix is obtained for each rat in the study
- Typically 144 Power Spectra/Dose/Rat over 12hours (5 min. bins)

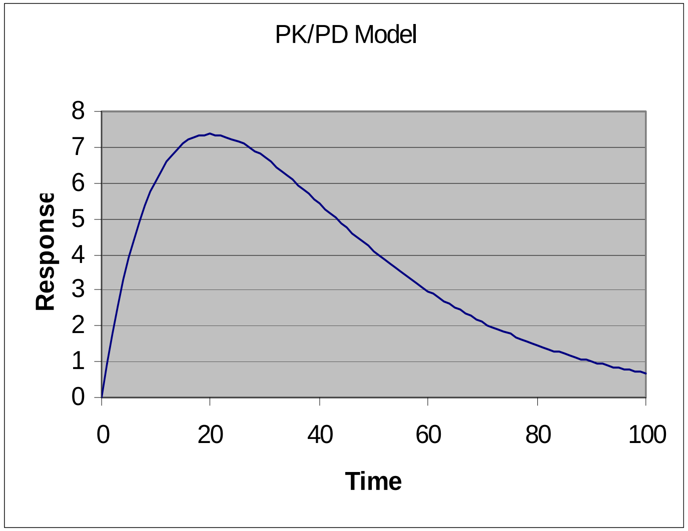
An example plot of a Power Spectra



Modelling the Spectra – Assumptions - the PK Model

- **Assumption 1** - the drug level in the brain will follow a Pharmacokinetics (PK) model
- **Assumption 2** - the **brain response** is proportional to dose level (PD)

- **Then** expected behaviour over time (somewhere in the data!) is:-



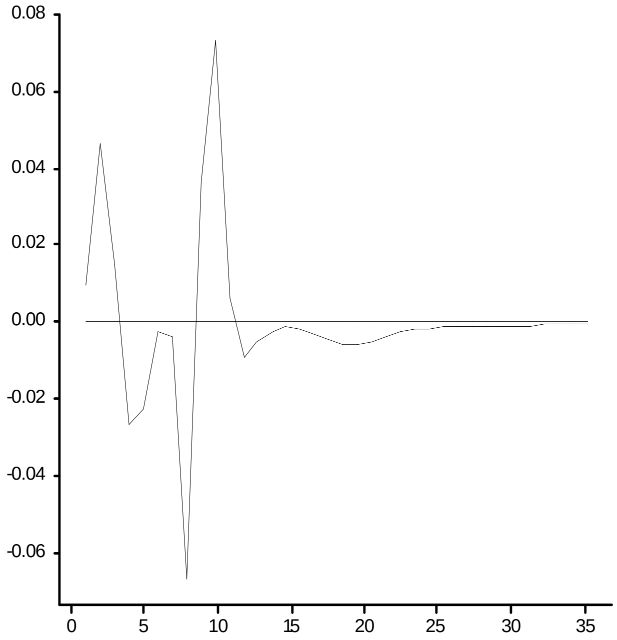
- A **suitable PK Equation** for the level of drug in the brain is the two compartmental model

$$Y = A*(\exp(-p*t) - \exp(-q*t))$$

- Or a **special case** ($p \cong q=k$) $Y = A*t*\exp(-k*t)$

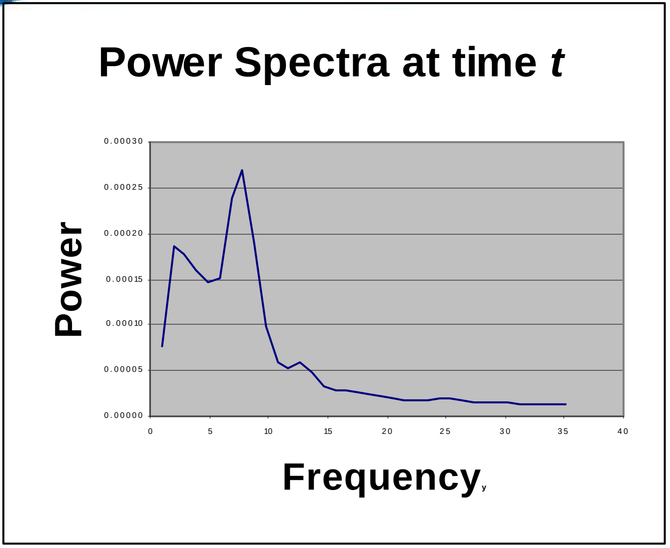
Modelling the Spectra - Assumptions

- **Assumption 3** – For a given **compound** there is an underlying linear combination of the Power at different frequencies that measures the response to **dose changing over time** (the **Signature,** or the **Weights**).
- Our aim - to find the underlying **Signature**
- The weighted mean power corresponds to the dose of the compound in the brain.



A possible “**Signature**” versus frequency for a single **rat** & **Compound**

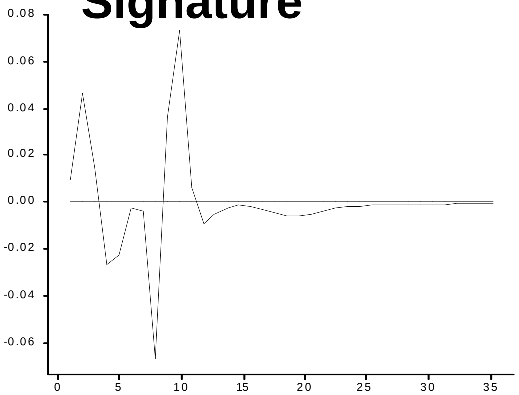
Combining the assumptions in practice (in pictures)

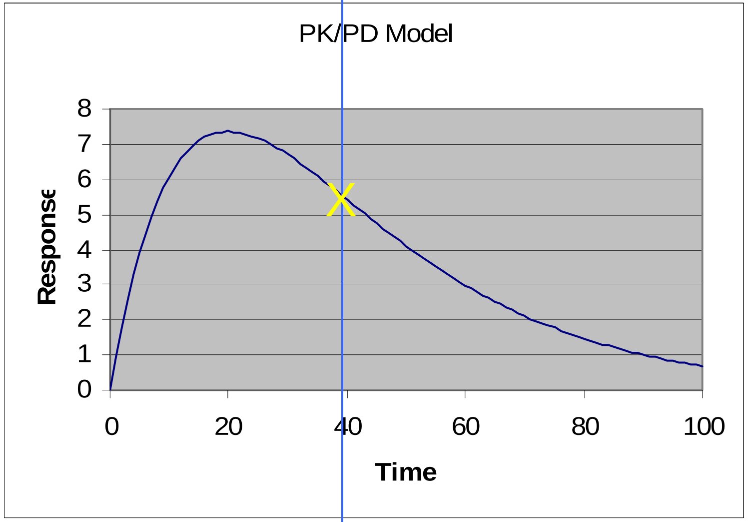


Predicted Response at time *t*

Sum of X

Signature





Modelling the Spectra - Extended Semi-Linear Canonical Correlation Analysis (ESLCCA)

- **Canonical Correlation Analysis** – given two matrices ***Y*** & ***X*,** find vectors ***a*** and ***b*** such that Corr(***Ya***, ***Xb***) is a maximum.

- For our case ***Y*** is the **Power Spectrum Matrix** & ***X*** is the PK/PD model versus time
- ***a*** is the Drug Signature

- ESLCCA estimates the non-linear parameters of the PK/PD model using an iterative process to **maximise** the Canonical Correlation.

- ESLCCA is **aimed** at finding a PK/PD type model.

ESLCCA In Pictures (1 - LHS)

Aim – Maximise correlation between (Y x a) & (X x b)

1) Calculate Y x a (LHS)

Matrix of Powers (Time/Dose x Frequency)

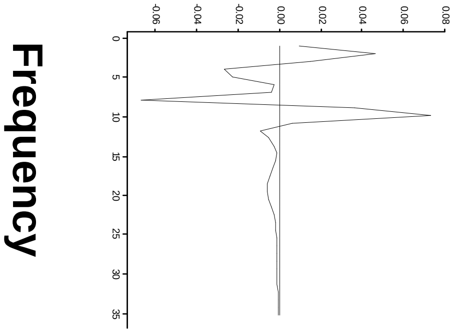
Dose	Time	Frequency			
		1	2	...	35
V	0	×	×	...	×
	5	×	×	...	×
	.	.	.	...	.
	720	×	×	...	×
L	0	×	×	...	×
	5	×	×	...	×
	.	.	.	...	.
	720	×	×	...	×
M	0	×	×	...	×
	5	×	×	...	×
	.	.	.	...	.
	720	×	×	...	×
H	0	×	×	...	×
	5	×	×	...	×
	.	.	.	...	.
	720	×	×	...	×

Y - Matrix

Vector of Weights (Signature)

Frequency	Signature
1	X
2	X
...	
35	X

a - Vector



Y x a = Y x a

144 x 35 x 35 x 1 = 144 x 1

ESLCCA In Pictures (2 - RHS)

Aim – Maximise correlation between (Y x a) & (X x b)

2) Calculate X x b (RHS)

		Intercepts				Time Response		
Dose	**Time**	**V**	**L**	**M**	**H**	**L**	**M**	**H**
V	**0**	1	0	0	0	0	0	0
	5	1	0	0	0	0	0	0
	.	.	.	.	.	.	.	.
	720	1	0	0	0	0	0	0
L	**0**	0	1	0	0	$f_{Low}(0)$	0	0
	5	0	1	0	0	$f_{Low}(5)$	0	0
	.	.	.	.	.	.	.	.
	720	0	1	0	0	$f_{Low}(720)$	0	0
M	**0**	0	0	1	0	0	$f_{Medium}(0)$	0
	5	0	0	1	0	0	$f_{Medium}(5)$	0
	.	.	.	.	.	.	.	.
	720	0	0	1	0	0	$f_{Medium}(720)$	0
H	**0**	0	0	0	1	0	0	$f_{High}(0)$
	5	0	0	0	1	0	0	$f_{High}(5)$
	.	.	.	.	.	.	.	.
	720	0	0	0	1	0	0	$f_{High}(720)$

X - Matrix

Coefficient	**Dose**	**b**
Intercept	**Vehicle**	x
	Low	x
	Medium	x
	High	x
Slope	**Low**	x
	Medium	x
	High	x

b - Vector

X x b = X x b

144 x 35 x 35 x 1 = 144 x 1

($f_{Low}(t)$, etc = t * exp(-k*t), where t = Time & k is to be estimated)

ESLCCA In Pictures (Correlating (Y x a) & (X x b))

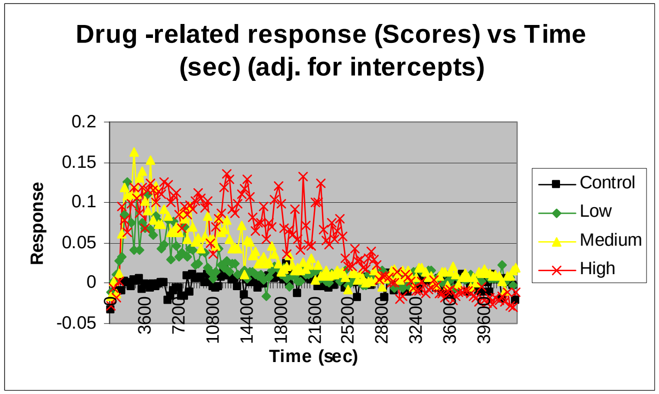
Aim – Maximise correlation between (Y x a) & (X x b)

3) Maximise Correlation between Y x a & X x b

- For known ***k*** CCA estimates ***a*** & ***b*** to maximise the Correlation
- Iterate on ***k*** to maximise the CCA Correlation

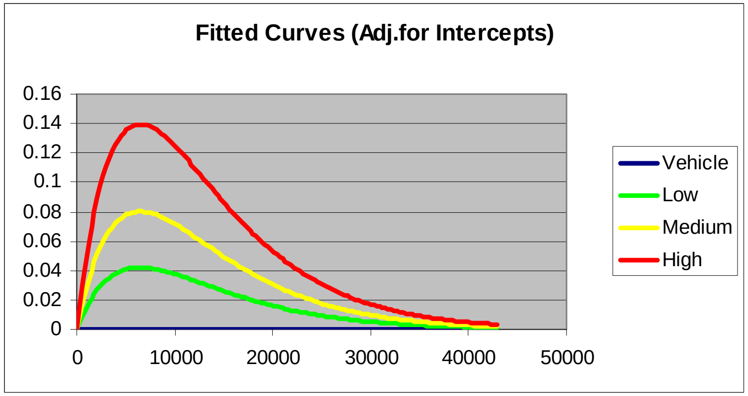
Observed PK/PD curve

(Y x a)



Predicted PK/PD Curves

(X x b)



ESLCCA in Genstat

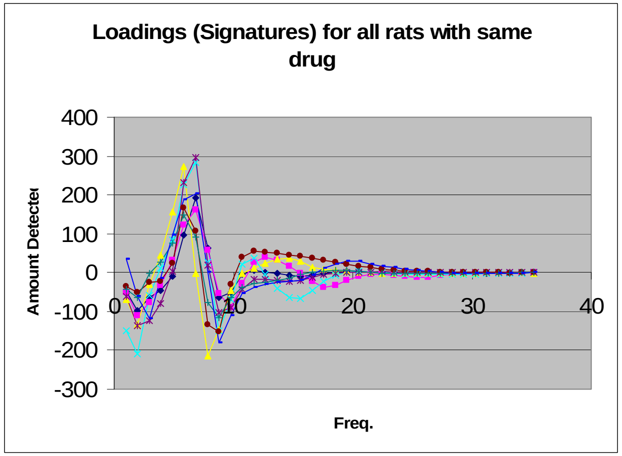
“Maximising the Partial Canonical Correlation Coefficient"

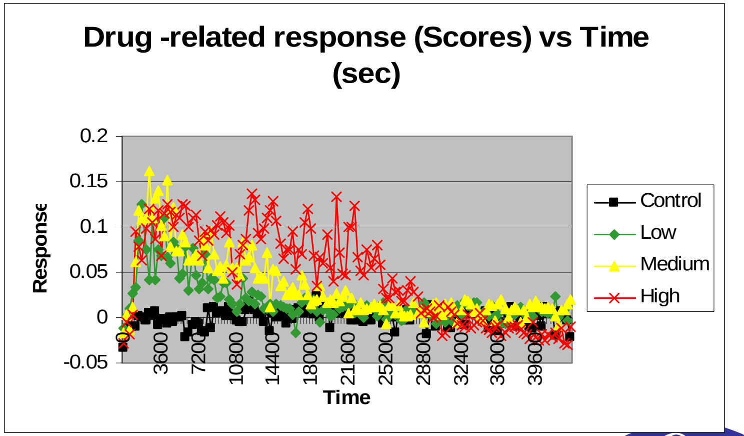
```
expression [value=(F$[*;1...3]=(exp(-t/exp(lTmax1[1]))-exp(-t/exp(lTmax2[1]))) \
    *(Treat.eq.STREAT['levels']$[2...nlev]))] F_function[1]
expression
    [value=(CorrMat=SSYY*+(t(RY)*+R*+F)*+ginv(t(R*+F)*+R*+F)*+(t(R*+F)*+
    RY)*+SSYY)] E[1]
expression [value=(R2[i]=eval(CorrMat))] E[2]
expression [value=(LogCorr[i]=log(1-R2[i]$[1]))] E[3]

model [function=LogCorr[i]]
rcycle lTmax1[1],lTmax2[1];init=8.5,9;step=2(*);lower=2(Lower);upper=2(Upper)
fitnonlin [pr=mon,m,s,e;calc=F_function[1],E[1,2,3]]
```

Example - Drug Study (Loadings & Scores)

- **Signatures** (vector ***a***) for 8 individual rats for a chosen compound
- This gives the contribution of the **power** at each frequency to the PK curve over time
- **Scores** for a single rat; evidence of PK/PD model behaviour

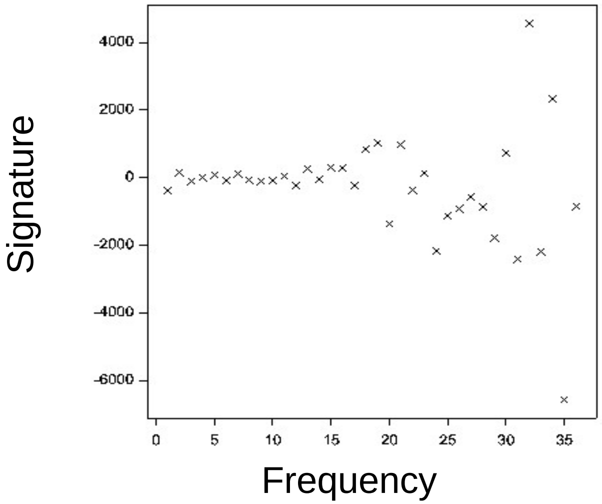

Loadings (Signatures) for all rats with same drug


Drug -related response (Scores) vs Time (sec)

Behind the scenes…

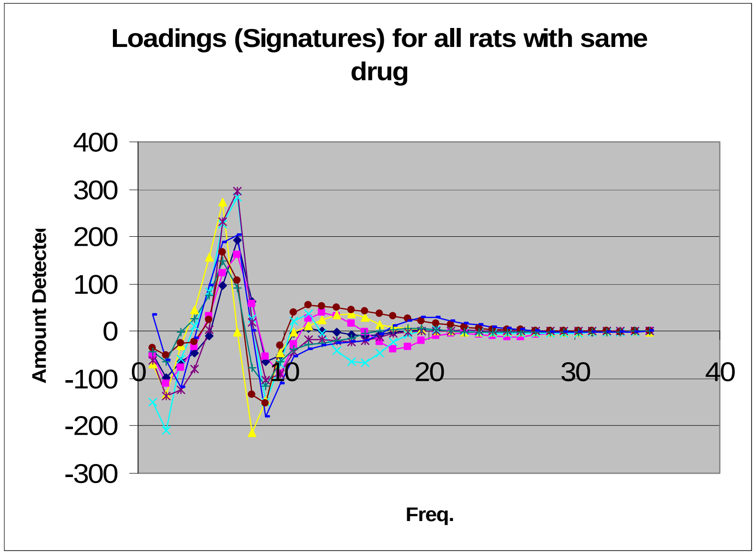
- Need to smooth the Y-matrix pre-analysis
  - ESLCCA detects the time profile well
  - But the signature is bad

Drug signature from unsmoothed data



Behind the scenes (contd)…

- Need to smooth the Y-matrix pre-analysis, otherwise
  - ESLCCA detects the time profile well
  - But the signature is bad
- Smooth by taking SVD of Time x Frequency matrix of powers before ESLCCA & predict matrix using less roots (typically 4 or 5).
- Signatures then consistent between rats.

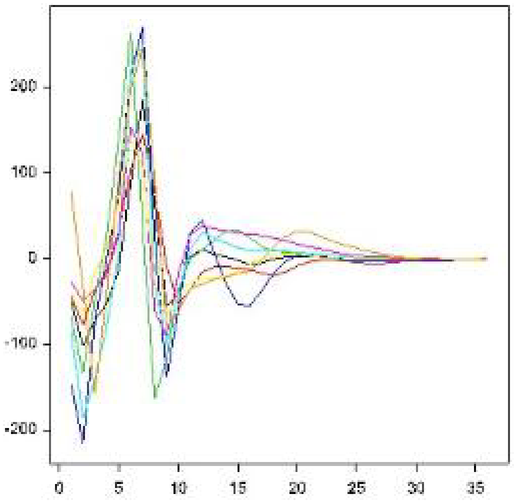


Loadings (Signatures) for all rats with same drug

Examples – A known active drug

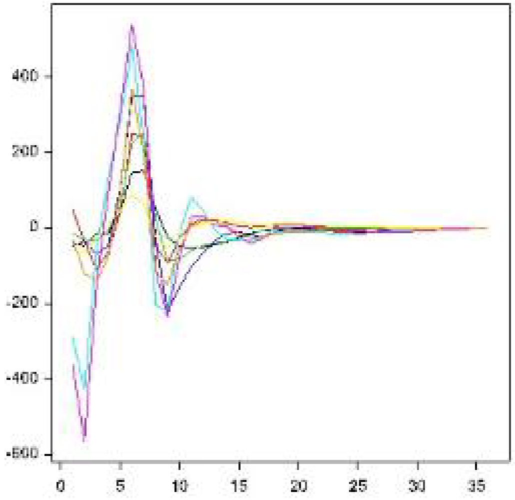
Signature of an **active** drug in:-

Dark Phase



Frequency

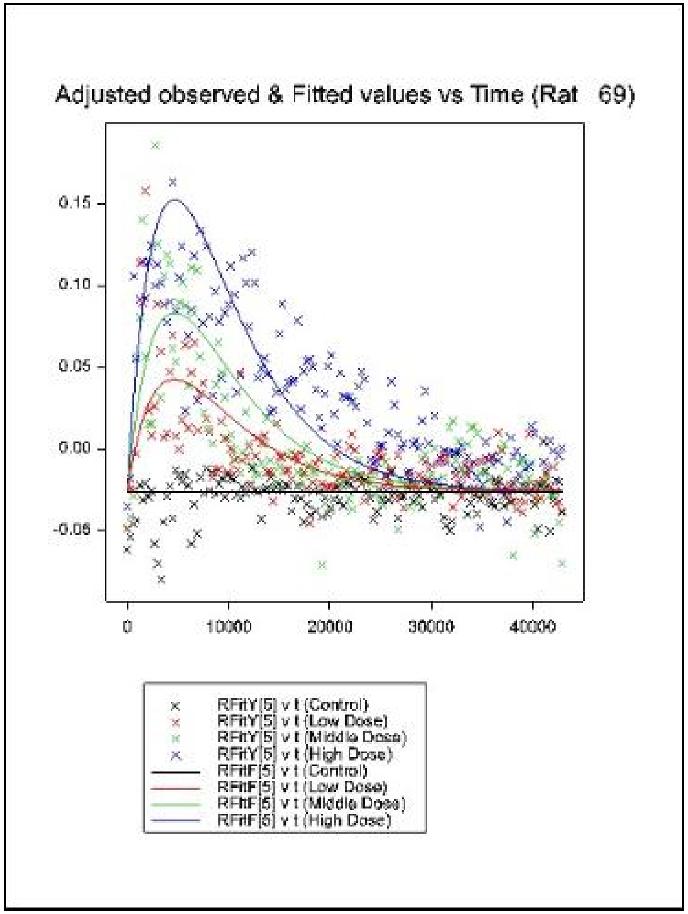
Light Phase

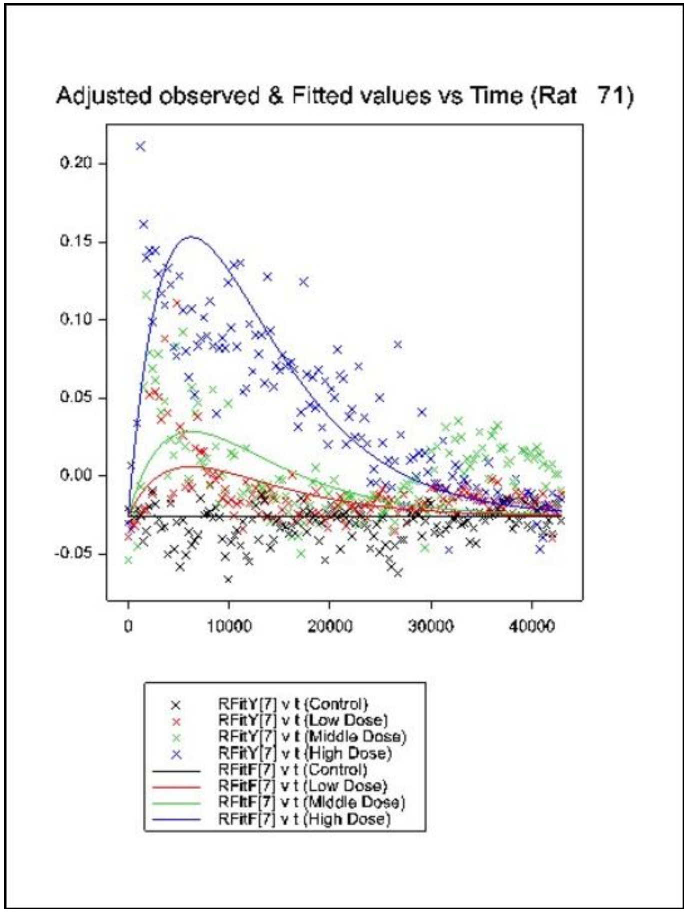


Frequency

Examples – A known active drug

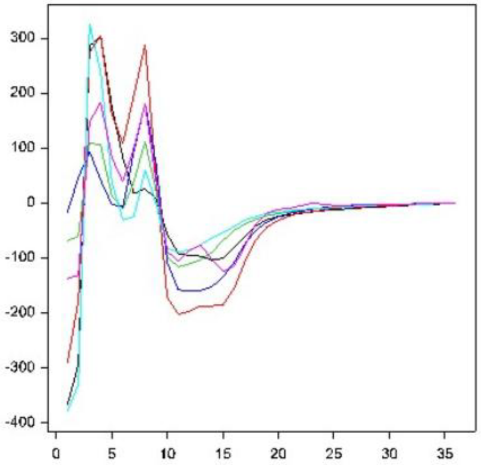
“Observed” & Fitted curves



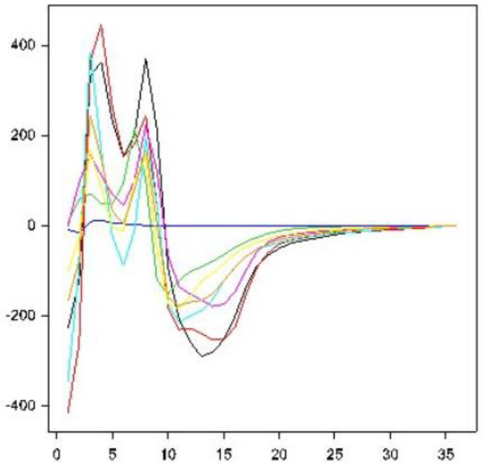


Examples – Three Known inactive drugs

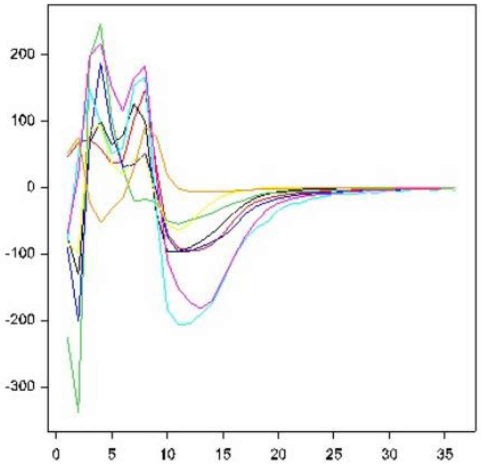
Compound A



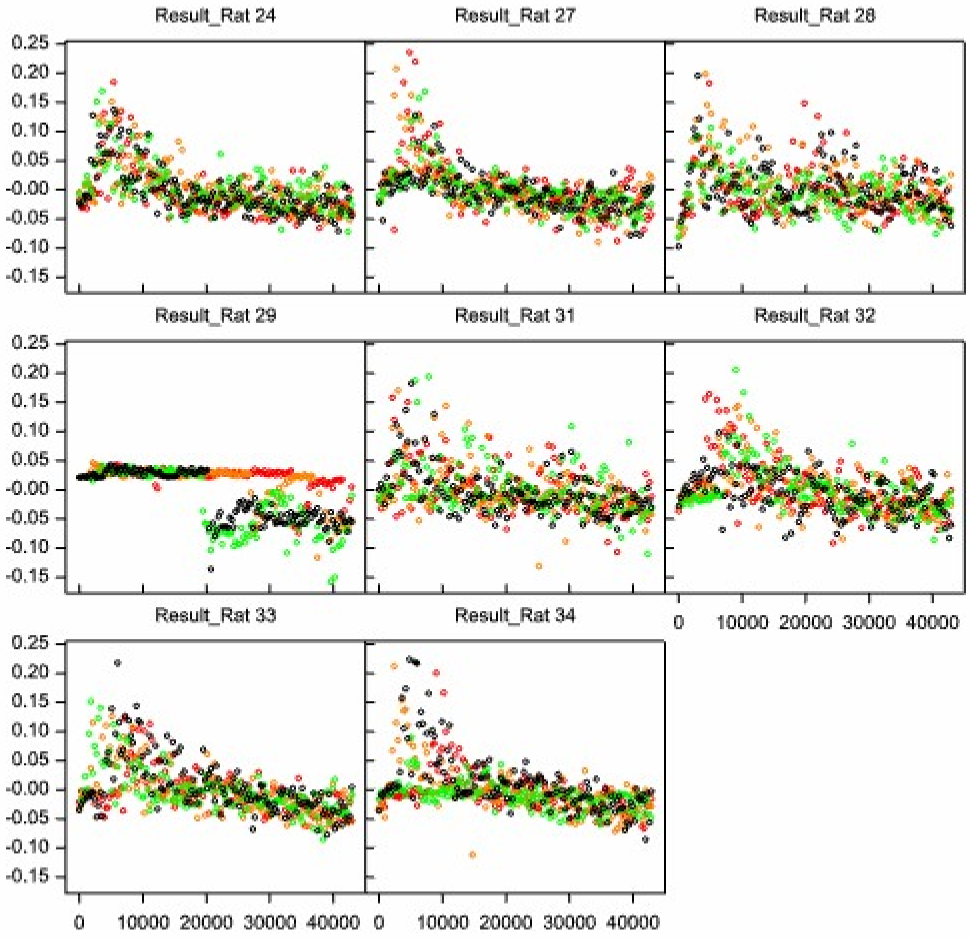
Compound B



Compound C



Examples – Three Known inactive drugs (Y-variate for Drug B)



- •ESLCCA detects an effect which isn’t different between Vehicle, Low, Med & High
- •Handling effect causing release of stress hormone?

Next stages

- Looking for a translatable biomarker (i.e. a method that can be used to detect drug penetration into the brain in rat & human)
- This works well for rats…
- …but what about human?
- Simulate study in human by choosing a set of “clinical” time points (about 7)
- Use approach on cut-down set – does it work? (Yes)
- Get real clinical data & try out with human data (in progress).

Discussion

- Very easy to extend to optimising the Partial Correlation Coefficient (so, for example, we can adjust for covariates).
- Once we’ve got the “new” y-variate (eg Y x a), can we use this to formally decide on model complexity? (eg Critical Exponential vs Double Exponential?)
- In this case we know that log(Power) is approx. Normally distributed – can we correlate
  - Ln(Y x a) & ln(X x b) rather than (Y x a) & (X x b)?
- Interested in hearing of other possible applications

Discussion

- Availability of algorithm:-
  - Genstat Procedure?
  - Postdoc @ Pfizer has coded it in R.

- Paper on approach submitted to Statistics in Biopharmaceutical Research “Analysing electroencephalogram (EEG) data using Extended Semi-Linear Canonical Correlation Analysis”