

# ***GRID Enabling the Analysis of Multiprocess Random Effect Response Data in R***

- Daniel Grose -- University of Lancaster
  - Rob Crouchley -- University of Lancaster
  - Ties van Ark -- University of Lancaster
  - John Kewley -- CCLRC Daresbury
  - Adam Braimah -- CCLRC Daresbury
  - Mark Hayes -- University of Cambridge
- 
-

# *Introduction*

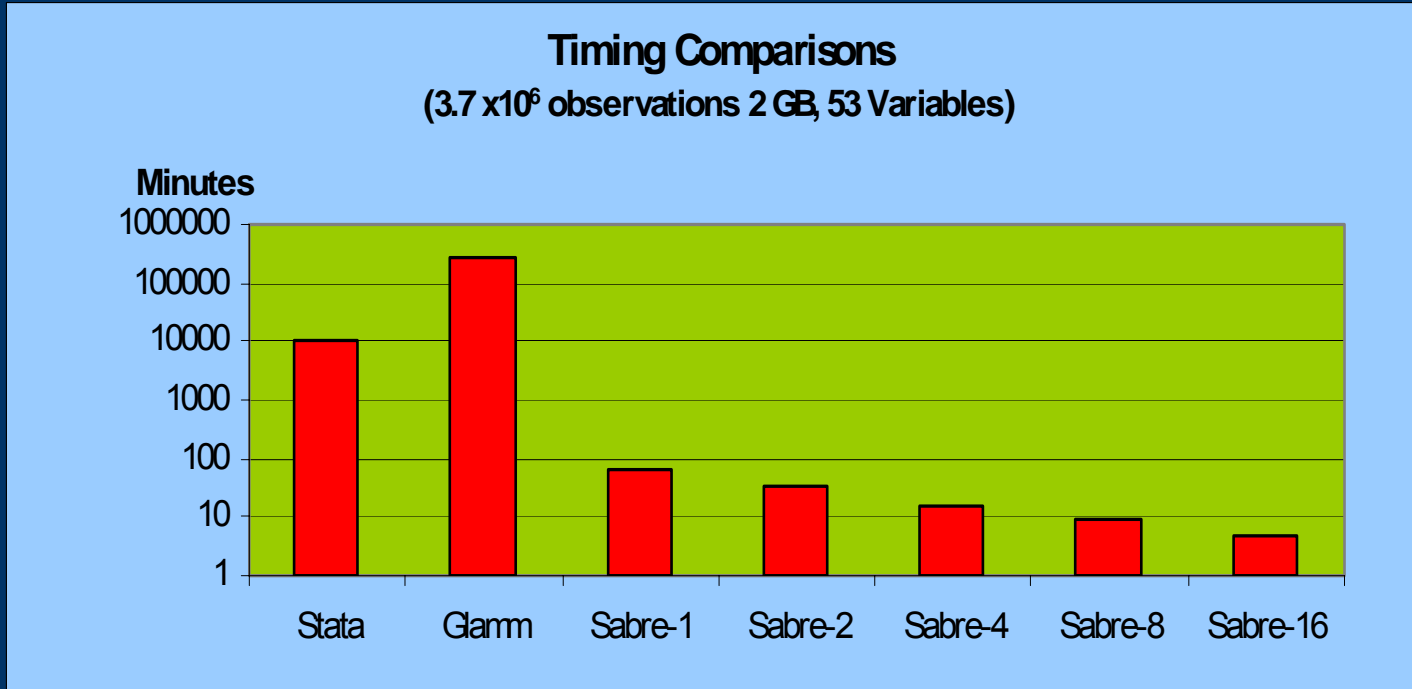
- Overview of paper. Three sections
    - SABRE : Software for the Analysis of Binary Recurrent Events.
    - GROWL : Grid Resources on a Workstation Library.
    - Integration of SABRE functionality into Statistics Software (R and Stata) using GROWL.
- 
-

# *SABRE – Software for the Analysis of Binary Recurrent Events*

- What is it ?
    - Programme for analysing binary, ordinal, count and recurrent events data. Employs fast numerical algorithms.
  - Some typical application areas.
    - Infertility in humans, animal husbandry.
    - Voting, trade union membership, economic activity and migration.
    - Absenteeism studies.
- 
-

# SABRE

- Why use it ?



# *Perceived Problems Associated with Employing SABRE*

- With the Software
    - Need to learn a new user interface
    - Have to change format of existing data
    - Difficult to integrate into existing work flows
  - To benefit from Sabre typically means using parallel version running on HPC, so ...
    - require knowledge of GRID middleware
    - often use unfamiliar operating systems and software
    - need to manage multiple accounts on different systems
- 
-

# Solution Part 1 - Integration

The screenshot displays the Intercooled Stata 9.0 interface. The main window shows a Data Editor with a table of data. The table has columns for 'cses' and 'mathach', and rows numbered 1 through 20. The 'cses' column contains values ranging from -1.1 to 0.35, and the 'mathach' column contains values ranging from 4.145 to 21.521. A 'Sabre Analysis' dialog box is open, showing the 'Variables' tab. The 'Variables' list includes 'school', 'student', 'minority', 'gender', 'ses', 'meansas', 'cses', 'mathach', 'size', 'meansasbycses'. The 'Case' is set to 'student', the 'Y-Variate' is 'mathach', and the 'R-Variate' is empty. The 'Constants' are set to 'cons', and the 'Number of Variables' are set to 2 for the first, 2 for the second, and 3 for the third. The 'Initial Values' field is empty. The 'Lag' checkbox is checked, and the 'Markov' checkbox is unchecked. The 'Save' and 'Done' buttons are visible at the bottom of the dialog box.

Intercooled Stata 9.0

File Edit Prefs Data Graphics Statistics User Window Help

Review Results

db sabre\_analysis  
db sabre\_analysis  
use "/home/grosed/  
edit

979-696-4601 (fax)

Data Editor

Preserve Restore Sort << >> Hide Delete...

cses[1] = -1.1

	cses	mathach
1	-1.1	5.876
2	-.16	19.708
3	-.1	20.349
4	-.24	8.781
5	.27	17.898
6	.45	4.583
7	-.19	-2.832
8	-.5700001	.523
9	-.46	1.527
10	-.03	21.521
11	-1.02	9.475
12	-.23	16.057
13	-.04	21.178
14	-.5599999	20.178
15	.76	20.349
16	-.25	20.508
17	.13	19.338
18	-1.1	4.145
19	.47	2.927
20	.35	16.405

Sabre Analysis

Variables Model Numerics Analysis and Results

Variables  
school student minority gender ses meansas cses mathach size meansasbycses

Case student Y-Variate mathach R-Variate Offset

Constants First cons Second Third

Number of Variables First 2 Second 2 Third 3

Initial Values

Lag  Markov

Save Done

/home/grosed

# Solution Part 1 - Integration

The screenshot displays the Intercooled Stata 9.0 interface. The main window shows a command window with the following text:

```
db sabre_analysis
db sabre_analysis
use "/home/grosed/
edit
```

The Data Editor window shows a table with the following data:

	cses	mathach
1	-1.1	5.876
2	-.16	19.708
3	-.1	20.349
4	-.24	8.781
5	.27	17.898
6	.45	4.583
7	-.19	-2.832
8	-.5700001	.523
9	-.46	1.527
10	-.03	21.521
11	-1.02	9.475
12	-.23	16.057
13	-.04	21.178
14	-.5599999	20.178
15	.76	20.349
16	-.25	20.508
17	.13	19.338
18	-1.1	4.145
19	.47	2.927
20	.35	16.405

The Sabre Analysis dialog box is open, showing the following settings:

- Model:  Univariate  Bivariate  Trivariate
- Family:
  - First:  Binomial  Poisson  Gaussian
  - Second:  Binomial  Poisson  Gaussian
  - Third:  Binomial  Poisson  Gaussian
- Link Function:
  - First:  Logit  Probit  Complementary
  - Second:  Logit  Probit  Complementary
  - Third:  Logit  Probit  Complementary
- Correlated  Ordered

The dialog box has "Save" and "Done" buttons at the bottom right.

## *Solution Part 2*

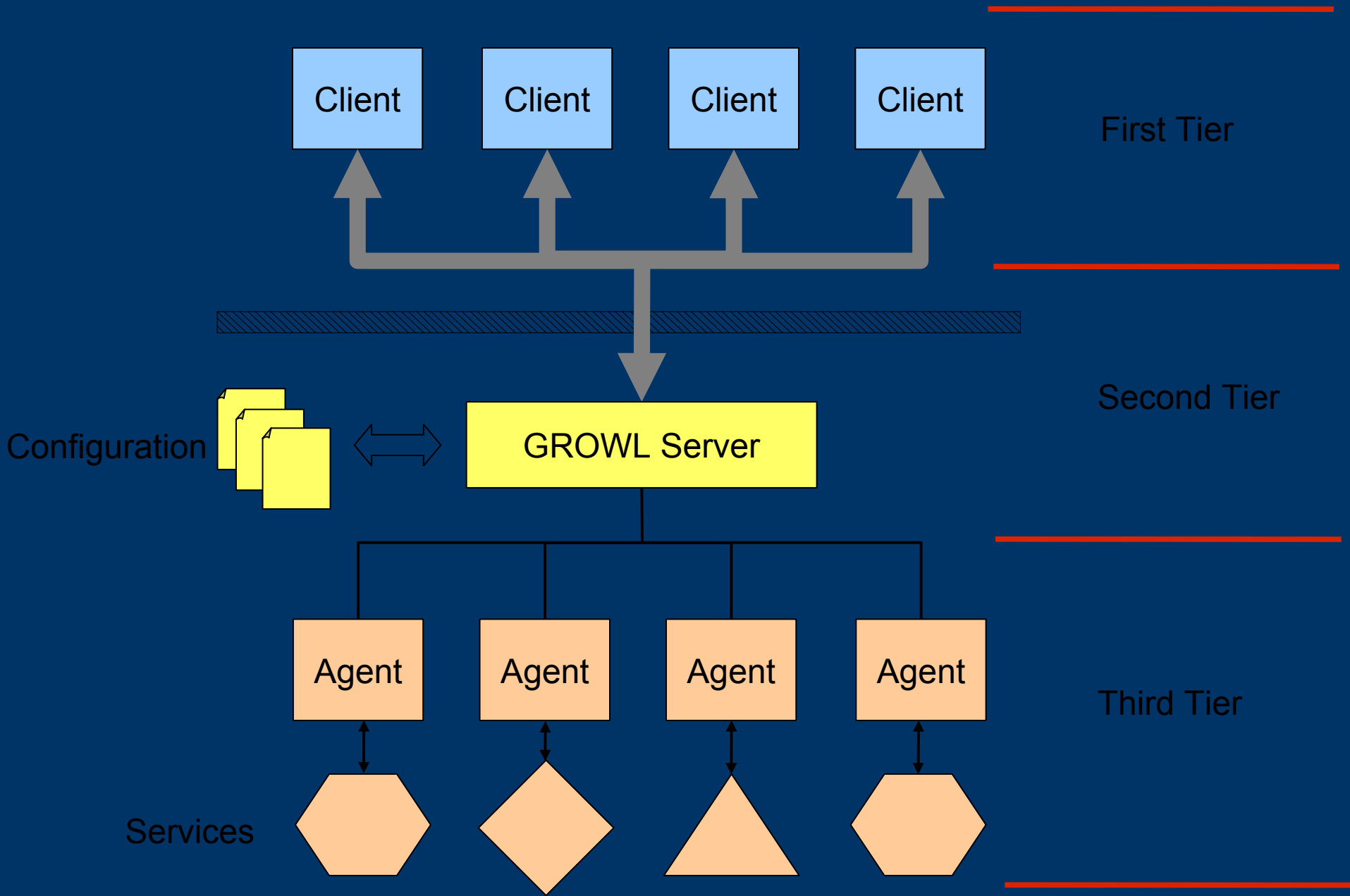
- Separation of Client and Server Logic
  - Why ?
    - Implementation of Service Logic may change to allow for improved algorithms, models or scheduling policies and so on
    - Many service vendors can provide the same service interfaces but with different local implementations, for example, by deployment of different middleware components, Globus vs Condor perhaps.
    - However, client interface should not change !!
- 
-

# *Solution Part 2 - Consequences*

- Client interface does not specify how service resources are employed
    - quality of service entirely depends on local service implementation and policies (scheduling and so on)
  - Modelling is Asynchronous in the User Environment
    - Client may be analysing several models simultaneously
    - Client interface must provide means of managing multiple models
- 
-

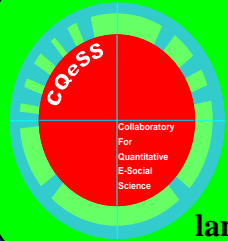
## *Solution Part 2 - How*

- Host Sabre as Secure Web Service
    - Service needs to be secure
    - Service needs to be persistent
    - Many services provided via a single host on a single port
    - Multiple clients
  - Difficult to do !!
    - Above features easy to host by employing generic GROWL server – allows the developer to concentrate just the service logic (algorithms, scheduling etc)
- 
-



# Summary

- Many useful and important software tools available, but can be difficult to use
    - Integrate functionality into existing client environments
  - Implementation of Service may change or vary according to host system
    - Implement all service logic in secure, persistent web service
  - Secure persistent web services difficult to develop
    - Host service using GROWL server
- 
-



e-science.  
lancs.ac.uk/cqess/

# SABRE

Lancaster's Statistical Software  
for e-Social Scientists

Software for the  
Analysis of  
Binary  
Recurrent  
Events  
[www.sabre.lancs.ac.uk](http://www.sabre.lancs.ac.uk)



Grid  
Resources  
On  
Workstation  
Library  
[www.growl.org.uk](http://www.growl.org.uk)

## SABRE

SABRE is a program specifically designed for the analysis of binary, ordinal, count recurrent events as are common in many surveys. SABRE's dedicated soft-ware ensures fast response times.



## SABRE + R

Adding SABRE as a plug-in to R allows Sabre commands to be processed from the R user interface. Configuration of models and preparation of data is then undertaken using the extensive functionality of R



## SABRE+ R+GROWL

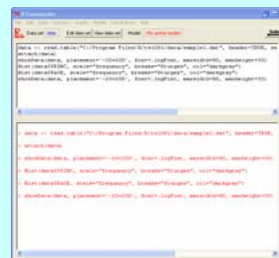
Using GROWL Components, SABRE commands invoked in R are executed in parallel on the GRID, making SABRE an excellent e-Social Science tool.

### Application area's

- Studies of infertility in humans.
- Animal husbandry.
- Absenteeism studies.
- Clustered sampling schemes.
- Studies of voting behavior, trade union membership, economic activity and migration.
- Demographic surveys.

### R Commander

The familiar R interface is being maintained by using SABRE as a plug in



### Grid Resources on Work Stations

GROWL employs a client/server architecture that hides the complexity of GRID middleware from the user. Client access to GROWL employs a secure (PKI/SSL) connection to a single port on the host system and clients are authenticated using the distinguished name extracted from their certificate. The use of a persistent server to access grid resources allows all of the service logic to be hosted by the server, making the client application, library or plugin extremely lightweight.

### Future developments

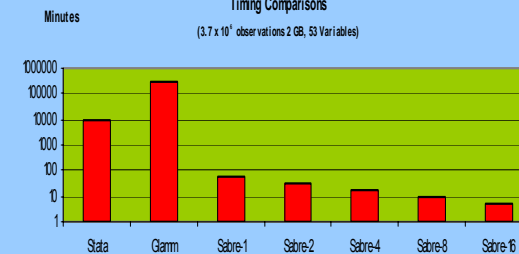
- Course material for the use of Sabre is currently being developed.
- It is planned to launch a Sabre/GROWL service on the North West Grid within the coming year. This will provide a utility based grid resource.
- Research into labour markets using Sabre/Growl.
- SABRE will become available as a plug in for STATA

### Acknowledgements:

Sabre was originally developed by Lancaster University's Centre for Applied Statistics, further development and use cases have been funded by the EPSRC, and ESRC as part of the NCeSS CQeSS node



Timing Comparisons  
( $3.7 \times 10^4$  observations 2 GB, 53 Variables)



# ***SABRE – Availability and Support***

- Web Site <http://sabre.lancs.ac.uk>
    - Full Command Documentation
    - Tutorials
    - Example Data
    - Publications
  - Downloads
    - “SabreR” binary R packages including documentation  
(end 06/2006)
    - “SabreStata” Stata plugin including documentation  
(end 07/2006)
    - Sabre source code
- 
-

