A Study of Grid Applications: Scheduling Perspective

Aleksandar Lazarević, Dr. Lionel Sacks

ACSE Group

Dept. of Electronic and Electrical Engineering
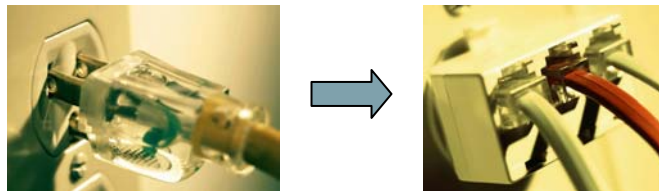
University College London

Outline

- Introduction to Grid Computing
- Study of Grid Applications
- Implications for Grid Scheduling

Grid Computing

- Utility platform

- Computational Service

- Range of Hardware

- Range of Uses/Applications

Mission Statement

- Study Grid Usage/Applications
  - Explore job & resource utilisation statistics
  - Patterns of user behaviour and workflows
  - Correlation with historical data
- Explore Resource Management implications
  - On higher level: planning, provisioning and SLA
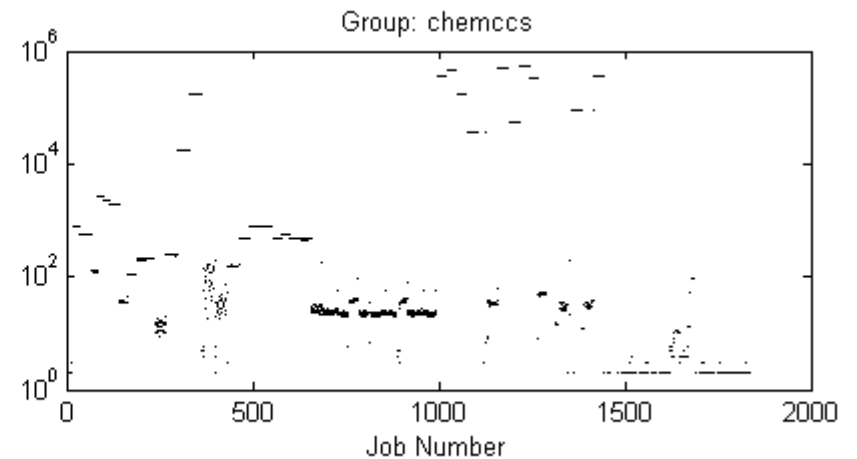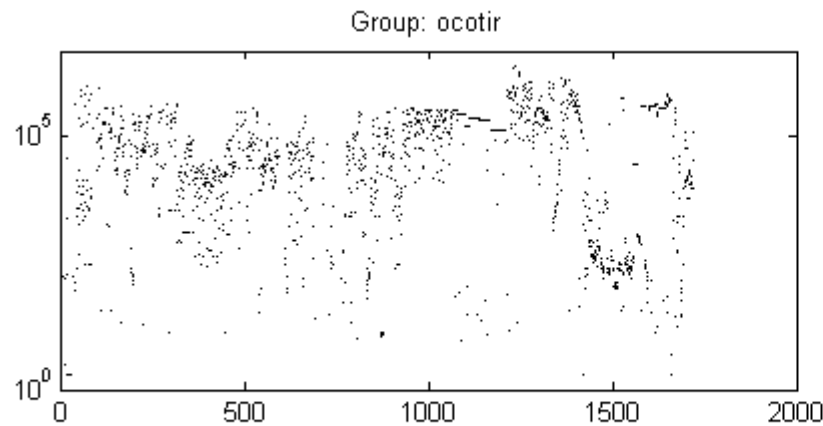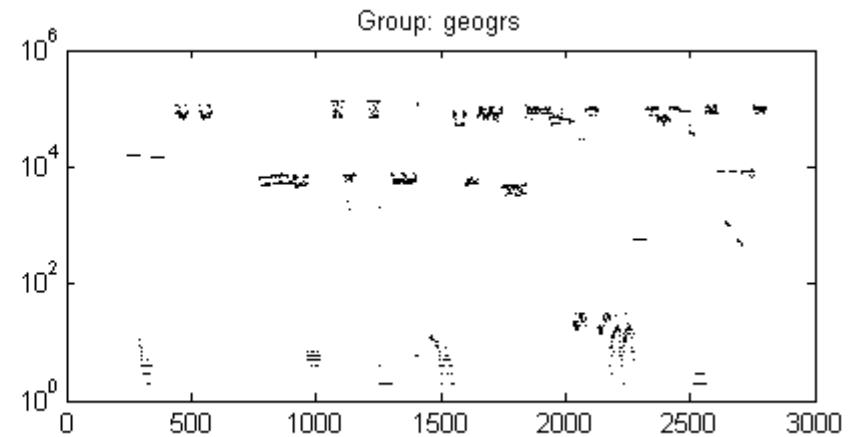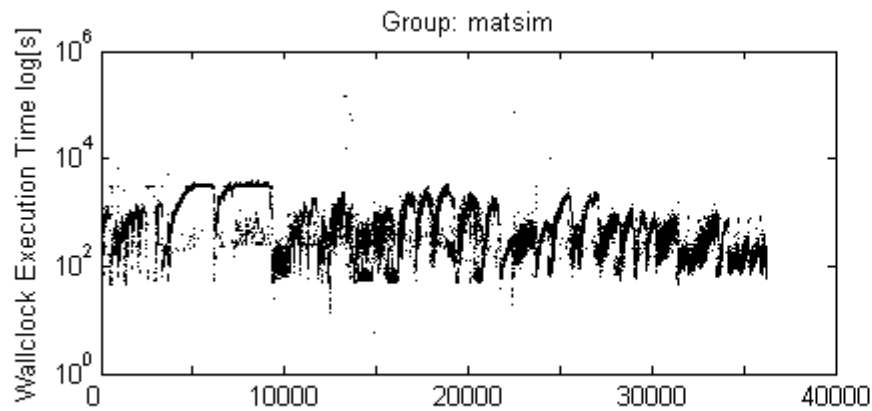 - On lower level: admission control and **<u>scheduling</u>**

UCL Grid Cluster

- 6 months of job accounting data from UCL's Central Computing Cluster

- 25 eScience projects, ~ 50,000 jobs

- Collect meta-data on job submitter, submit time and node, scheduling delay, real CPU time, **wallclock execution time** etc.
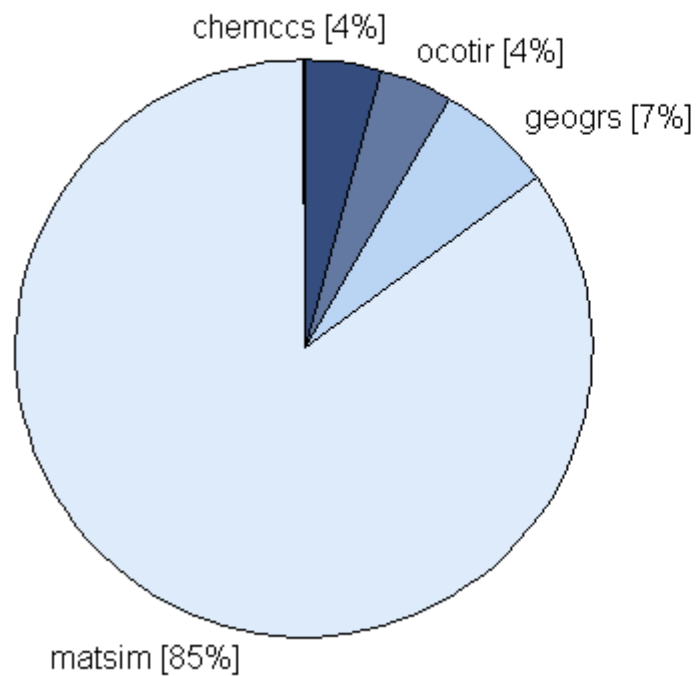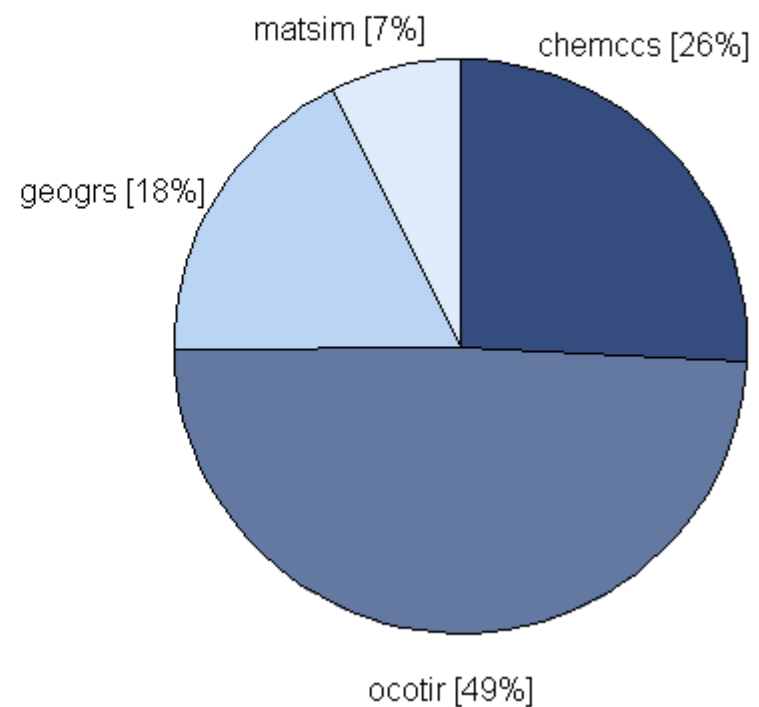
Overall Job Execution Times
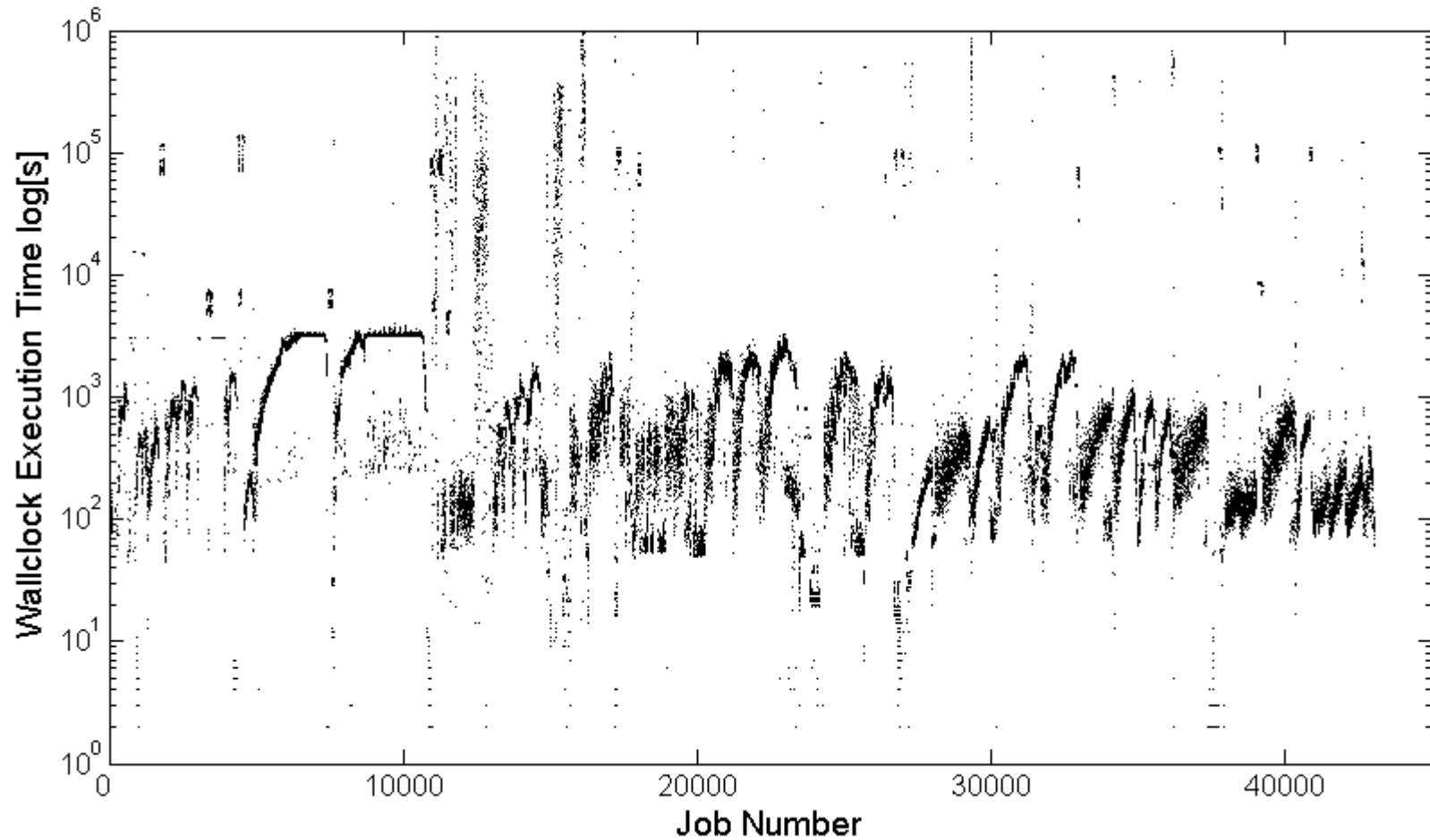
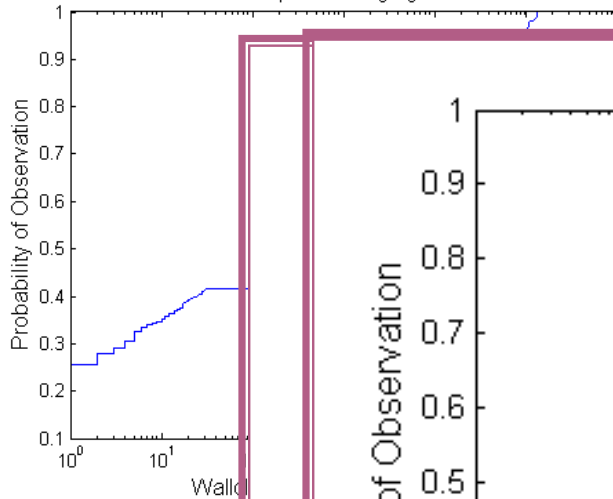# Group Execution Times

Group Job Distribution

Cumulative Execution Time Distributions
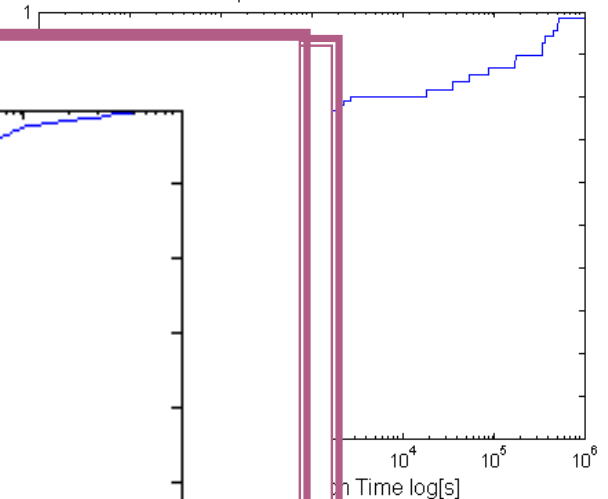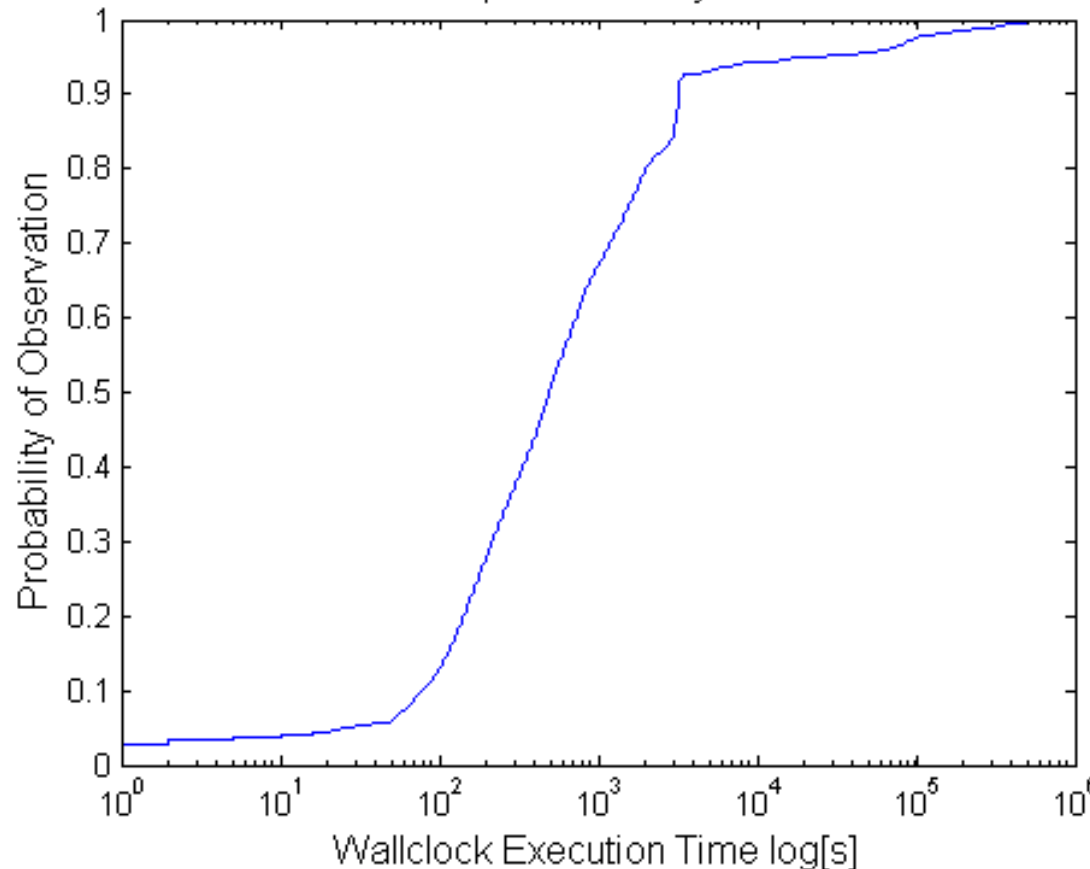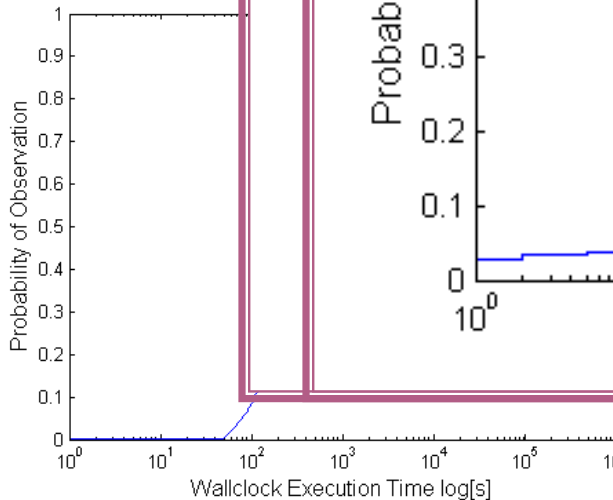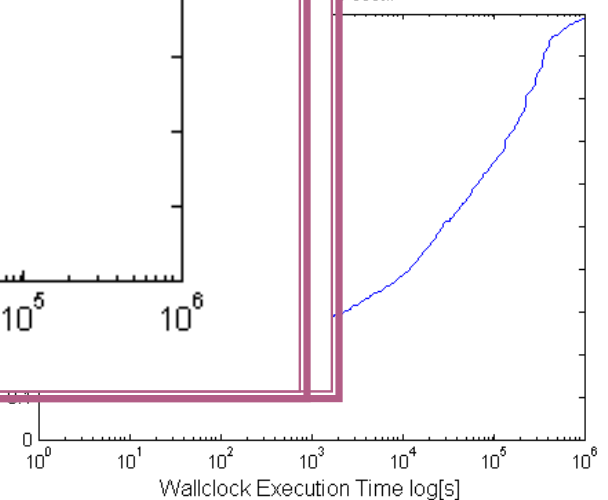
Execution Time Autocorrelation

Wallclock Execution Times Autocorrelation Function

Lag	Corr.	S.E.	Q
1	+.808	.0048	280E2
2	+.734	.0048	511E2
3	+.683	.0048	711E2
4	+.646	.0048	890E2
5	+.625	.0048	106E
6	+.648	.0048	124E3
7	+.620	.0048	140E3
8	+.605	.0048	156E3
9	+.557	.0048	169E3
10	+.544	.0048	182E3
11	+.506	.0048	193E3
12	+.505	.0048	204E3
13	+.483	.0048	214E3
14	+.461	.0048	223E3
15	+.435	.0048	231E3
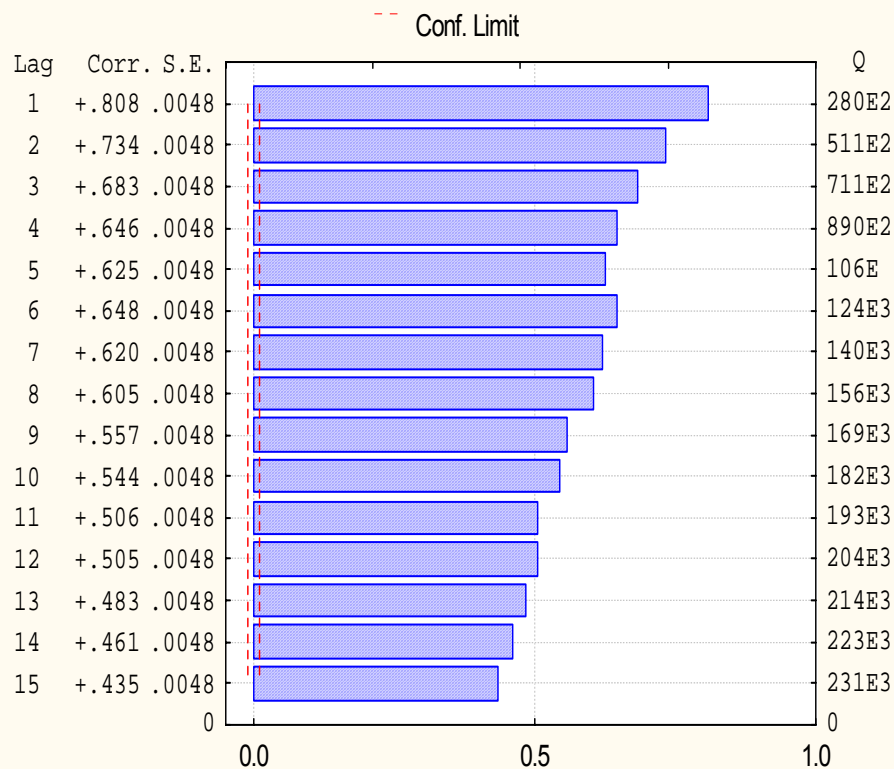
Wallclock Exe Times Partial Autocorrelation Function

Lag	Corr.	S.E.
1	+.808	.0048
2	+.236	.0048
3	+.119	.0048
4	+.088	.0048
5	+.094	.0048
6	+.196	.0048
7	+.004	.0048
8	+.039	.0048
9	-.069	.0048
10	+.050	.0048
11	-.046	.0048
12	+.048	.0048
13	-.020	.0048
14	-.030	.0048
15	-.009	.0048
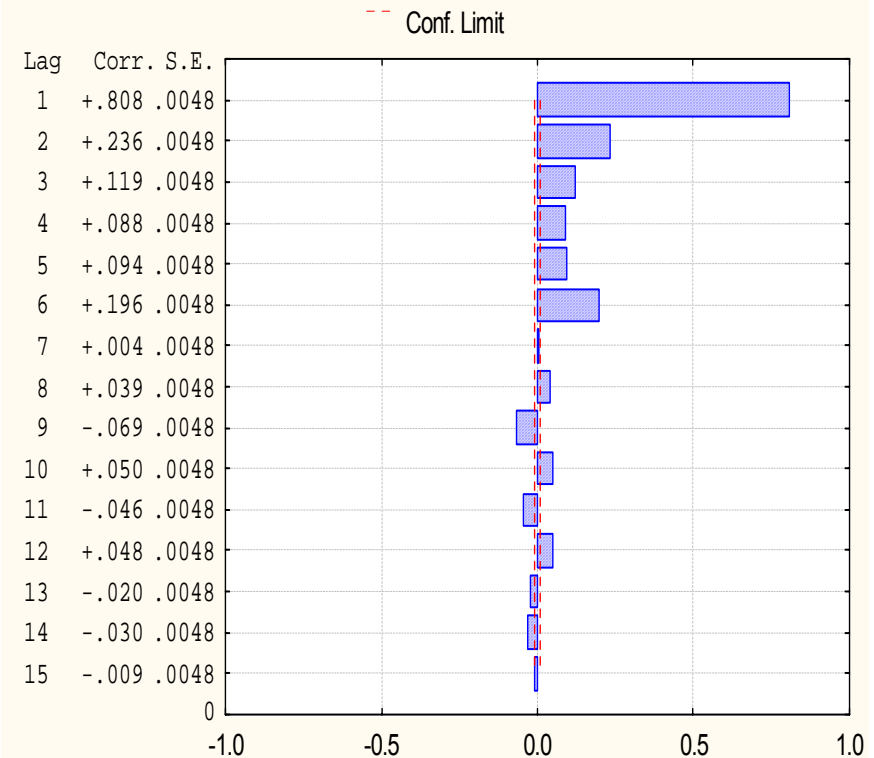
Scheduling Implications

- Could we anticipate job duration & resource requirements trends?

- Could we predict job duration of specific jobs in the queue?

- Could we offer an intuitive "deadline" scheduler with low administration cost?

- Could users live with probabilistic guarantees?

Conclusions

- Observed emergence of workflow patterns and their correlation with job meta-data

- Heterogeneous application set, size of jobs related to the resource size and expected performance

- High levels of autocorrelation could make statistical modelling feasible

Q & A

- a.lazarevic@ee.ucl.ac.uk

- www.ee.ucl.ac.uk/~alazarev/

- Self-Organising Grid Resource Management (SO-GRM) Project:
  www.ee.ucl.ac.uk/acse/so-grm/index.htm