Using R for the design and analysis of computer experiments with the Nimrod toolkit

Neil Diamond[1,*], **David Abramson**[2], **Tom Peachey**[2]

1. Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia
2. Caulfield School of Information Technology, Monash University, Melbourne, Australia

* Contact author: neil.diamond@buseco.monash.edu.au

Keywords: Computer Experiments, Parameter Sweeps, Workflow Engines

The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering (see, for example, Santner et. al., 2003). There is much more to this than merely choosing the design and analysing the resulting data. Computer Scientists at Monash University's eScience and Grid Engineering Laboratory have developed the Nimrod suite of tools (Monash eScience and Grid Engineering Laboratory, 2009) that automates the formulation, running, and collation of the individual experiments and includes a distributed scheduling component that can manage the scheduling of individual components in a local area network. Nimrod contains tools to perform a complete parameter sweep across all possible combinations (Nimrod/G), search using non-linear optimization algorithms (Nimrod/O), or use fractional factorial design techniques (Nimrod/E).

There are a number of workflow engines which provide scientists with an environment with which they can manage data, the workflows of the various analytical steps in their investigation, and summaries of findings. Although existing workflow systems can specify arbitrary parallel programs, they are typically not effective with large and variable parallelism. Similarly, Nimrod was not designed to execute arbitrary workflows. Thus, it is difficult to run sweeps over workflows, and workflows containing sweeps. To overcome these problems, a new tool (Nimrod/K) is being developed, based on the Kepler workflow engine (Kepler Core, 2009). It leverages a number of the techniques developed in the earlier Nimrod tools for distributing tasks to the Grid.

Kepler allows the user to specify R expressions and access R objects as part of the scientific workflow. This talk will describe how existing R packages have been used and extended both to help in the design of the computer experiment, and in the analysis and display of the results.

References

Santner T.J., Williams, B.J., and Notz, William.I. (2003). *The Design and Analysis of Computer Experiments.* Springer: New York.

Monash eScience and Grid Engineering Laboratory (2009). The Nimrod Toolkit., `http://messagelab.monash.edu.au/Nimrod/`.

Kepler Core (2009). Kepler Project, `http://www.kepler-project.org/`.