# 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
2. Caulfield School of Information Technology

MONASH University

# Computer Experiments

- The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering.

# Computer Experiments

- The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering.
- At least two books on the topic:

# Computer Experiments

- The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering.
- At least two books on the topic:
    - The Design and Analysis of Computer Experiments. T. J. Santner, B. J. Williamns, W.I Notz. (2003), Springer: New York.

# Computer Experiments

- The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering.
- At least two books on the topic:
    - The Design and Analysis of Computer Experiments. T. J. Santner, B. J. Williamns, W.I Notz. (2003), Springer: New York.
    - Design and Modeling for Computer Experiments. K-T. Fang, R. Li, A. Sudjianto. (2006), Chapman & Hall/CRC: London.

# Computer Experiments

- The design and analysis of computer experiments to explore the behavior of complex systems is becoming increasingly important in science and engineering.
- At least two books on the topic:
  - The Design and Analysis of Computer Experiments. T. J. Santner, B. J. Williamns, W.I Notz. (2003), Springer: New York.
  - Design and Modeling for Computer Experiments. K-T. Fang, R. Li, A. Sudjianto. (2006), Chapman & Hall/CRC: London.
- Some R packages-more on that later.

# Nimrod

- Developed by Computer Scientists at Monash University's eScience and Grid Engineering Laboratory.

# Nimrod

- Developed by Computer Scientists at Monash University's eScience and Grid Engineering Laboratory.
- Automates the formulation, running, and collation of the individual experiments.

# Nimrod

- Developed by Computer Scientists at Monash University's eScience and Grid Engineering Laboratory.
- Automates the formulation, running, and collation of the individual experiments.
- Includes a distributed scheduling component that can manage the scheduling of individual jobs.

# Nimrod Set of Tools

Nimrod contains tools to

- perform a complete parameter sweep across all possible combinations (Nimrod/G),

# Nimrod Set of Tools

Nimrod contains tools to

- perform a complete parameter sweep across all possible combinations (Nimrod/G),
- search using non-linear optimization algorithms (Nimrod/O),

# Nimrod Set of Tools

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).

# Nimrod Set of Tools

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).

# Nimrod Set of Tools

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).

These can be run stand-alone or accessed via the Nimrod portal

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics
- Ecology

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics
- Ecology
- Quantum Chemistry

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics
- Ecology
- Quantum Chemistry
- CAD Digital Simulation

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics
- Ecology
- Quantum Chemistry
- CAD Digital Simulation
- Antenna Design

# Nimrod Applications

Nimrod has been used in an extensive range of applications

- Air Pollution Studies
- Laser Physics
- Ecology
- Quantum Chemistry
- CAD Digital Simulation
- Antenna Design
- Cardiac Modelling

# Workflow Engines

- 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.

# Workflow Engines

- 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.

# Workflow Engines

- 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.

# Workflow Engines

- 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.

# Nimrod/K

- To overcome these problems, a new tool (Nimrod/K) is being developed, based on the Kepler workflow engine (Kepler Core, 2009).

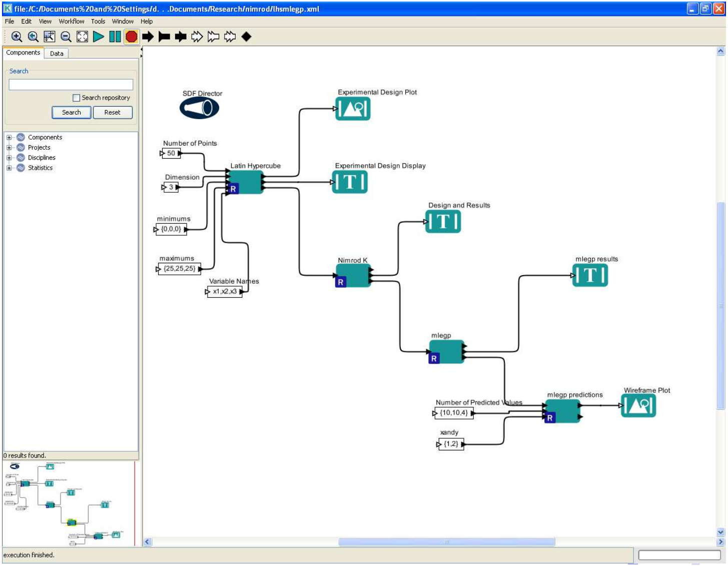
# Nimrod/K

- 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.

# Nimrod/K

- 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.

# Example Workflow



# Statistical Approach to Computer Experiments

- Unlike physical experiments, repeated experiments give the same results.

# Statistical Approach to Computer Experiments

- Unlike physical experiments, repeated experiments give the same results.
- Model the output as the realisation of a stochastic process with a correlation structure that depends on a distance to other points in the experiment.

# Statistical Approach to Computer Experiments

- Unlike physical experiments, repeated experiments give the same results.
- Model the output as the realisation of a stochastic process with a correlation structure that depends on a distance to other points in the experiment.
- Allows estimates of untried experiments.

# Statistical Approach to Computer Experiments

- Unlike physical experiments, repeated experiments give the same results.
- Model the output as the realisation of a stochastic process with a correlation structure that depends on a distance to other points in the experiment.
- Allows estimates of untried experiments.
- Gives an estimate of the uncertainty.

# Computer Experiments-Designs

- Simplest method-Latin Hypercubes

# Computer Experiments-Designs

- Simplest method-Latin Hypercubes
- Other more sophisticated methods include Orthogonal Arrays and Scrambled Nets.

# Computer Experiments-Designs

- Simplest method-Latin Hypercubes
- Other more sophisticated methods include Orthogonal Arrays and Scrambled Nets.
- Various space filling designs.

# Computer Experiments-Model

$$\text{Response} = \text{Linear Model} + \text{Departure}$$

$$y(x) = \beta + z(x)$$

$$E(z(x) = 0$$

$$\text{Cov}(z(t), z(u)) = \sigma_z^2 \prod_{j=1}^{d} R_j(t_j, u_j)$$

$$R_j(t_j, u_j) = \exp\left[-\theta_j (t_j - u_j)^{p_j}\right]$$

# MLE of $\theta, p, \beta$, and $\sigma^2$

Reduces to numerically optimising

$$-\frac{1}{2}(n \ln \hat{\sigma}^2 + \ln \det R_D)$$

$$R_D = \text{Matrix of correlations for design points}$$

$$\begin{aligned}
\hat{\beta} &= (1^T R_D^{-1} 1^T)^{-1} 1^T R_D^{-1} y \\
\hat{\sigma}^2 &= \frac{1}{n}(y - 1\hat{\beta})^T R_D^{-1}(y - 1\hat{\beta})
\end{aligned}$$

# Best Linear Unbiased Predictor for an untried x

$$\hat{y}_x = \hat{\beta} + r^T(x) R_D^{-1} (y - 1\hat{\beta})$$

where

$$r(x) = [R(x_1, x), R(x_2, x), \ldots, R(x_n, x)]^T$$

Design point : $[x_1, x_2, \ldots, x_n]$ Untried Input : $x$

Interpolates the data points.

# Implementations in R

- BACCO

# Implementations in R

- BACCO
    - Emulator

# Implementations in R

- BACCO
    - Emulator
    - Approximator

# Implementations in R

- BACCO
  - Emulator
  - Approximator
  - Calibrator

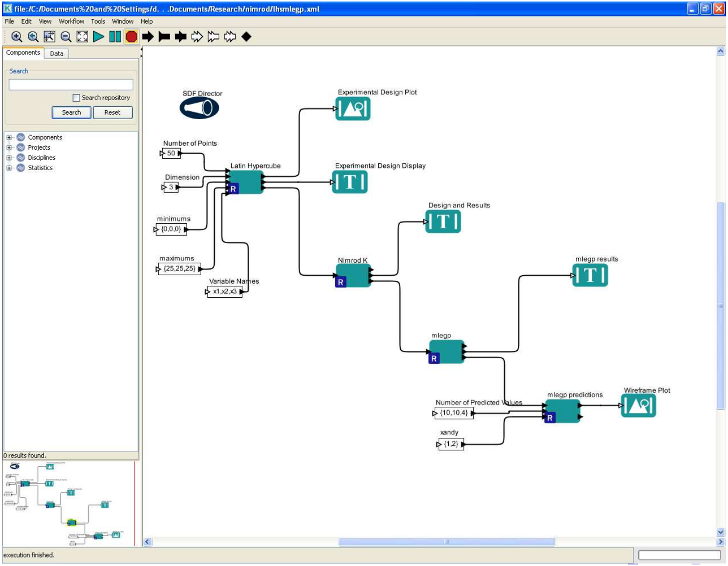
# Implementations in R

- BACCO
    - Emulator
    - Approximator
    - Calibrator
- mlegp: an R package for Gaussian process modeling and sensitivity analysis

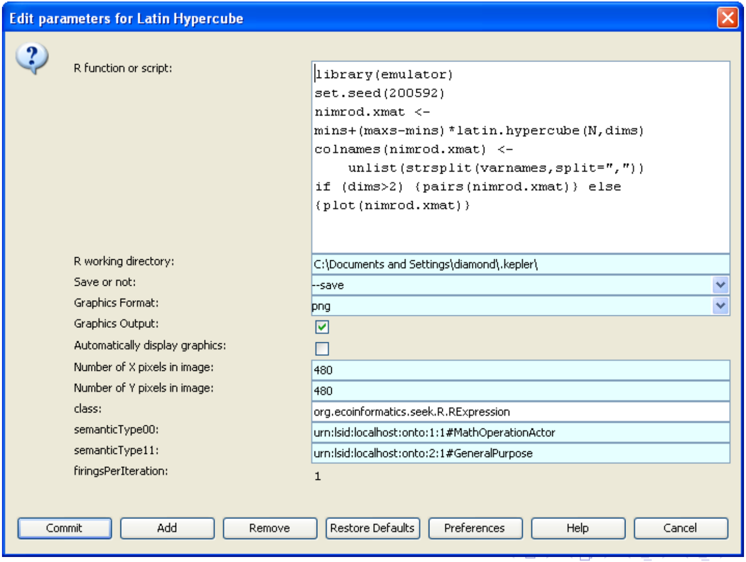
# Implementations in R

- BACCO
  - Emulator
  - Approximator
  - Calibrator
- mlegp: an R package for Gaussian process modeling and sensitivity analysis
- Certainly others . . .

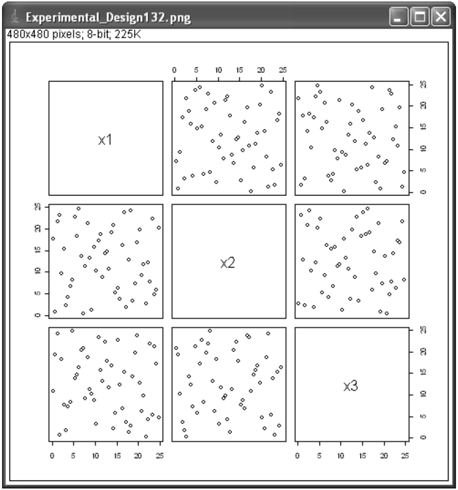
# Example Workflow



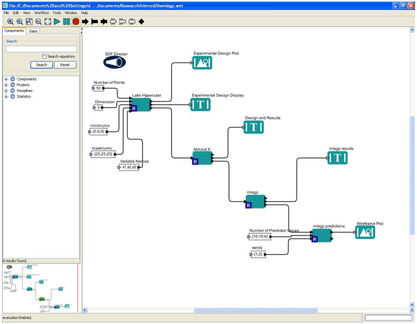
# Latin Hypercube Actor



# Latin Hypercube Design

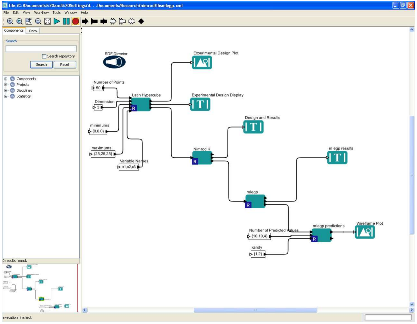


# Nimrod/K Actor



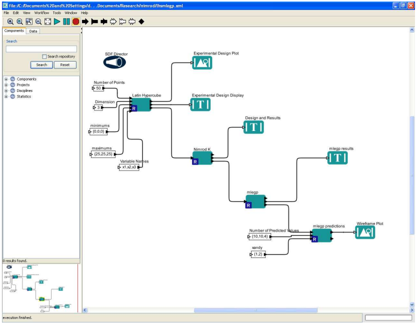

- Nimrod takes the experimental design and controls the running of the experiments and collation of results.

# Nimrod/K Actor



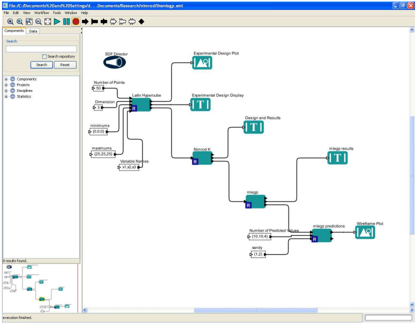

- Nimrod takes the experimental design and controls the running of the experiments and collation of results.
- Passes the results onto mlegp actor which fits the Gaussian model to the data.

# mlegp predictions Actor



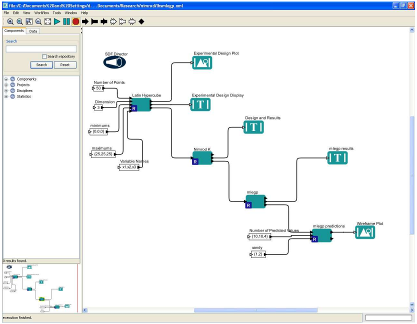

- Takes fitted model and predicts at a grid of untried inputs.

# mlegp predictions Actor



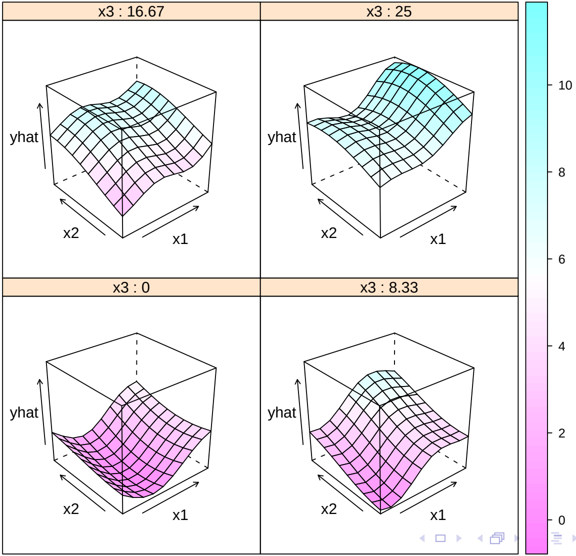
- Takes fitted model and predicts at a grid of untried inputs.
- Inputs are the granularity of the grid, and which are the primary and conditioning inputs.

# mlegp predictions Actor




- Takes fitted model and predicts at a grid of untried inputs.
- Inputs are the granularity of the grid, and which are the primary and conditioning inputs.
- Uses Lattice graphics to produce a visualisation of the surface.

# Visualisation



# Key Message

- Computer Experiments are very important.

# Key Message

- Computer Experiments are very important.
- Many tools in R both to design and analyse computer experiments.

# Key Message

- Computer Experiments are very important.
- Many tools in R both to design and analyse computer experiments.
- Nimrod tools are convenient in managing the execution of the computer experiments.

# Key Message

- Computer Experiments are very important.
- Many tools in R both to design and analyse computer experiments.
- Nimrod tools are convenient in managing the execution of the computer experiments.
- Using Nimrod/K takes advantage of the Kepler workflow engine.

# Key Message

- Computer Experiments are very important.
- Many tools in R both to design and analyse computer experiments.
- Nimrod tools are convenient in managing the execution of the computer experiments.
- Using Nimrod/K takes advantage of the Kepler workflow engine.
- Kepler and R are integrated, making it easy to use existing packages in R for computer experiments, and extends their usefulness.

# MeSsAGE Lab

Monash eScience and Grid Engineering Laboratory

`http://messagelab.monash.edu.au/`