Model Evaluation

Some Methodologies and Opportunities

- IMPRINT, ACT-R
- Modeling with and without 'first principles'
- Multivariate visualization
- Multivariate optimization

Bengt Fornberg

University of Colorado, Boulder
Department of Applied Mathematics

Presentation at "Training MURI" kickoff meeting, Army Research Institute, Washington DC, November 4, 2005.

IMPRINT, ACT-R

Some underlying principles:

- Decompose very complex human responses into conceptual 'modules'
- Cross-connect 'modules' to create powerful simulation tool with intuitive user interface
- Modeling deepens understanding of genuine mental mechanisms
- Provides ability to predict, optimize, explore wide-ranging 'what-if' scenarios, etc.

IMPRINT and ACT-R are extremely capable systems, ideally suited for the task, widely used, and fully able to both model and predict many human responses.

Why is there a need for additional modeling considerations?

Don't fix what works well. Don't go off in different directions within the team.

However, important to

- Contrast/compare very different modeling approaches
  - Radial basis functions
  - Multivariate visualization
- Assess computational feasibility, reliability, and complexity with regard to
  - Running of models
  - Visualization
  - Performing global optimization of functions of many variables

Two main modeling methodologies

By First Principles

Split complete task into 'modules', where each one is relatively understandable - then let these interact in some computational network.

Extremely desirable when works

- Can greatly help understanding, and can lend support to assumptions that have been made (or behavior of network can suggest improvements in 'modules').

Potential problems:

- Taxonomy might at times be uncertain,
- Some processes just might not be possible to decompose in 'modules' with explicit governing equations

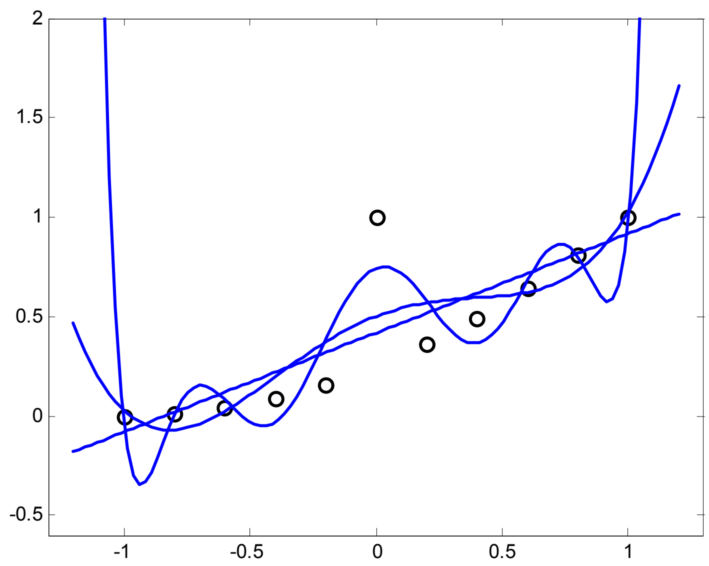
By 'Brute force' data fitting

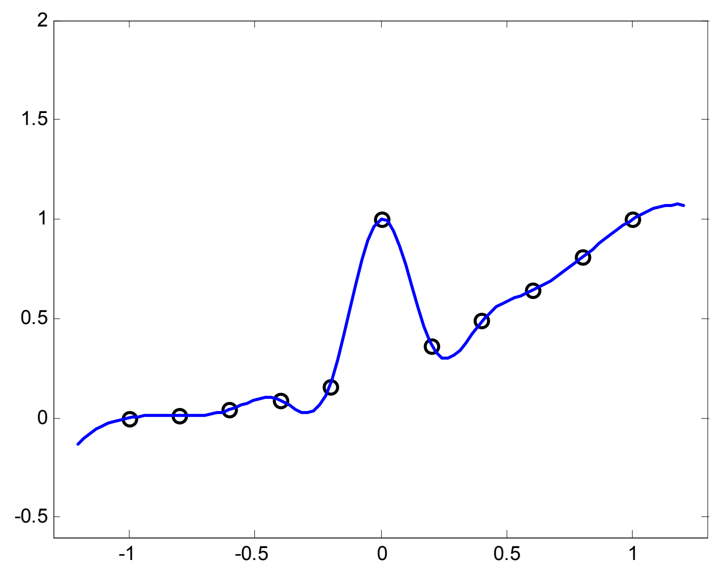
Recently discovered opportunities (examples of 'Data Mining').

- Neural networks — Successful in many cases which are very hard to cast in mathematical language
- Radial basis functions (RBF)

Capable of 'catching' entirely unanticipated relations, deeply hidden in massive data sets, with little danger that any preconceived assumptions will provide an unwanted bias in a model.

Simplest imaginable modeling situation? Data in 1-D





Polynomial regression

- Low degree contains no details
- High degree bad at ends

Splines

- Catches structures, whether anticipated or not
- In past, no generalization to multiple dimensions
  Now there is: Radial Basis Functions.

Extrapolation particularly difficult - Data are just not available...

Example: Evaluate different training methodologies while a course is given:

- Which gives best retention 5 years later?
- How do different aspects of retention depend on various training factors?

Small 1-D example: Data from present MURI project

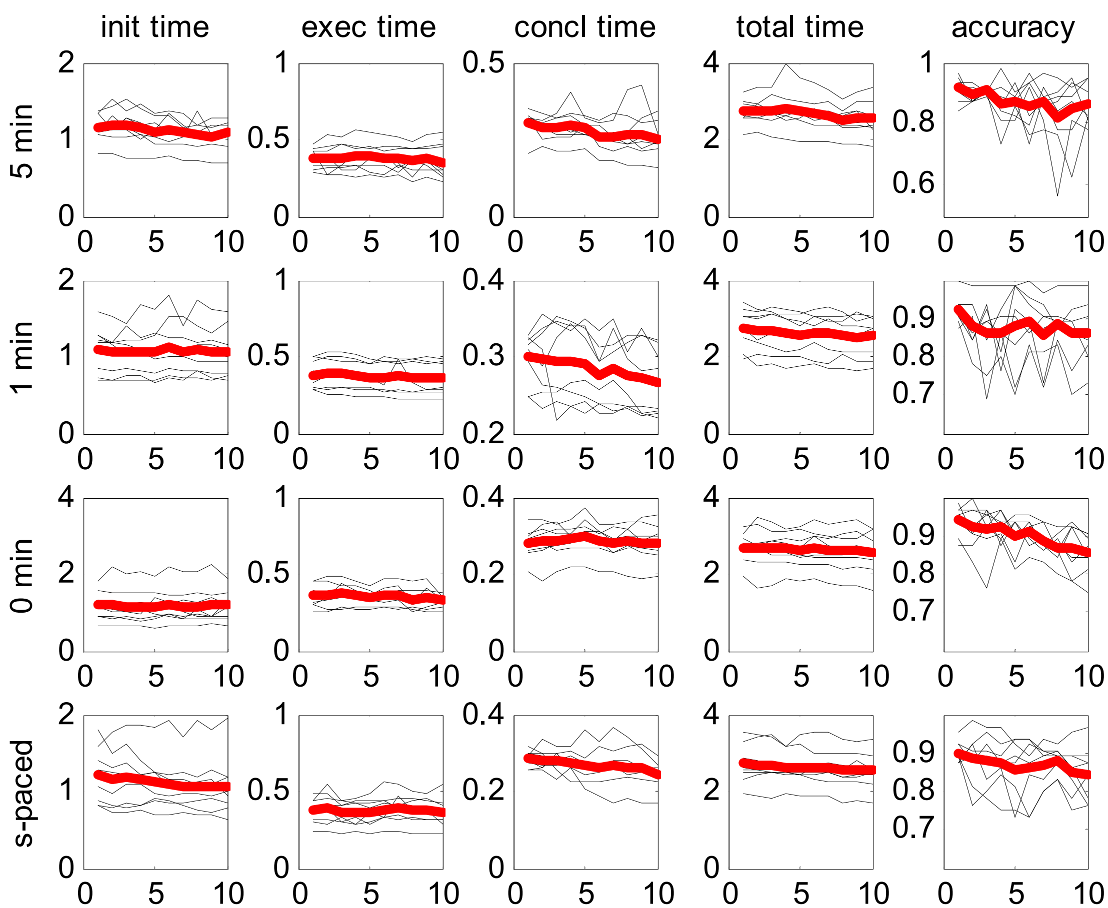
(Erica Wohldmann, James Kole)

10-part training sessions, different breaks (5 min, 1 min, 0 min, self-paced), 8 participants in each

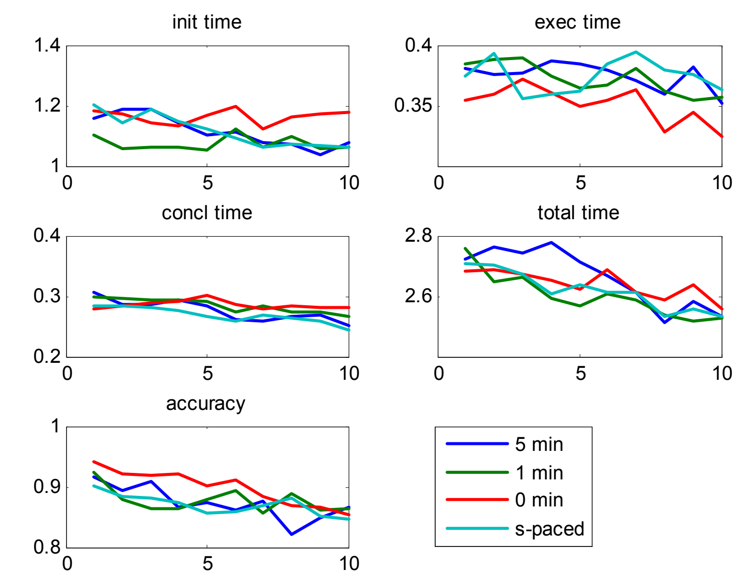
Measured: Initiation time, execution time, total time, accuracy

Questions: What is the long term retention? Are the data there? What data should be added?

Raw data: (black lines - averages in red)



Different averages:



Example of performed modeling

- Assume some 'true' time evolution
- Add lots of noise
- Try to model based just on noisy data
- Compare against 'true' solution

What about modeling data in, say, 6-D?

- RBF methodology capable of doing automatically what splines do in 1-D.
    - Permits multidimensional visualization
    - Can be used for interpolation or regression (to filter out noise)
    - Comparatively robust also for extrapolation

Schematic illustration of RBF use

Data measurements:

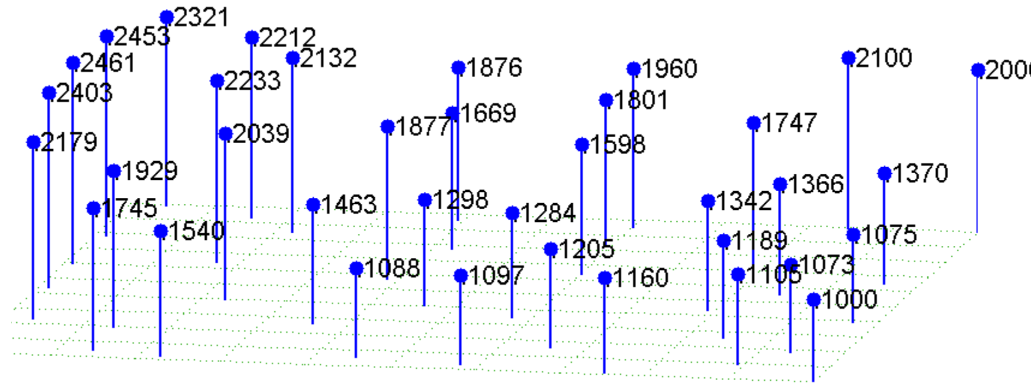
Different measured input quantities						Different variables to model			
1	2	3	4	...	*N*	1	2	...	*M*
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...

- An RBF model can provide answers to questions such as how different outcomes depend on different variables separately.

- Fully effective also if actual relation forms some low-dimensional manifold in a higher-dimensional data space

Brief history of RBF

1970 RBF proposed for drawing contour lines on maps from scattered elevation measurements (R.L. Hardy) ⇒



1985 Methodology proven to be robust for any number of points in any number of dimensions

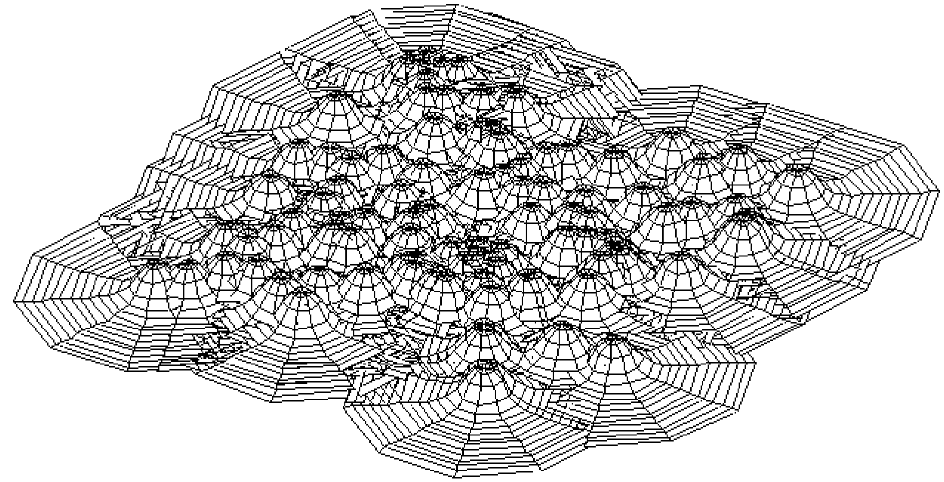
Intense work on RBF starts at several centers.

1990 RBF first recognized as a powerful tool to solve many PDEs

1993 First industrial use of RBF for modeling / data mining (Exxon and Shell)

2000+ Literature exceeds 2000 research papers
4 books published
Fast computational algorithms emerging
Vast number of applications

Mathematical concept of RBF



In simplest form:

Approximate scattered data by a linear combination of translates of a single 'radial function', one placed at each data location.

$$s(\underline{x}) = \sum_{k=1}^{N} \lambda_k \, \phi(\| \underline{x} - \underline{x}_k \|)$$

Coefficients λ_k found by solving a $N \times N$ (guaranteed nonsingular) linear system.

<u>Illustration:</u> Gaussians: $\phi(r) = e^{-(\varepsilon r)^2}$ (ε large)

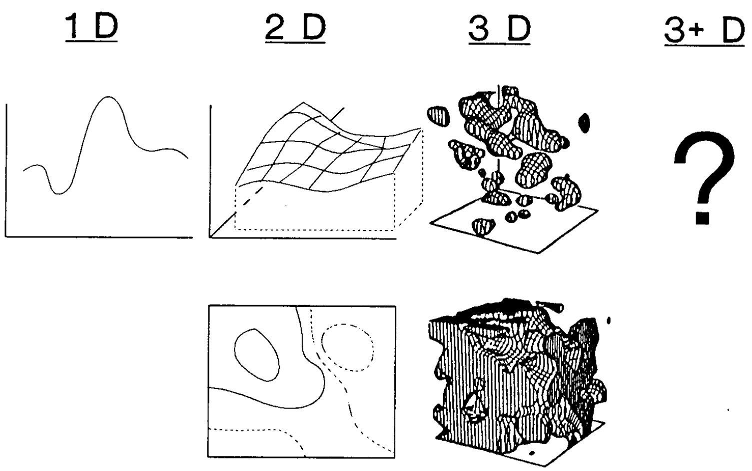
<u>Usually better:</u> Multiquadrics: $\phi(r) = \sqrt{1 + (\varepsilon r)^2}$ (ε small)

Can one visualize functions of many variables?

Sheet of paper (or computer screen) is 2-D

World we live in is 3-D

No reason whatsoever that some process might not depend on, say, 5 or 10 variables

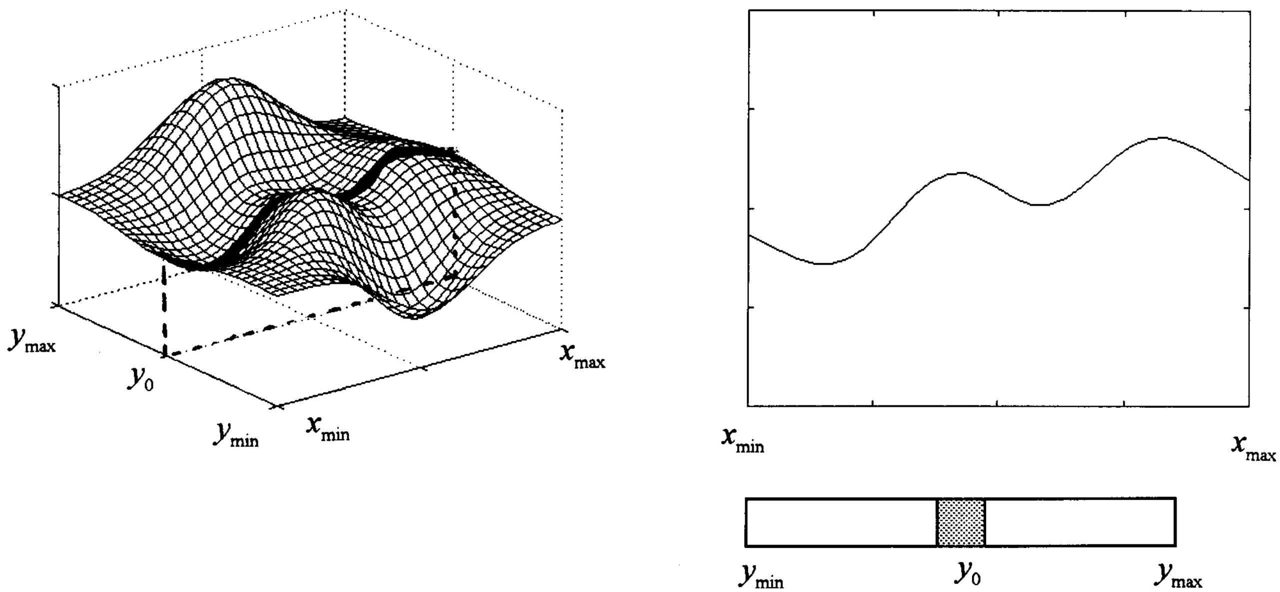


Is there any way to visualize multi-variate functions?

For example: Is it possible to convey to the human mind - via a 2-D computer screen - what the 'character' might be of some 5-D relationship?

Multivariate visualization

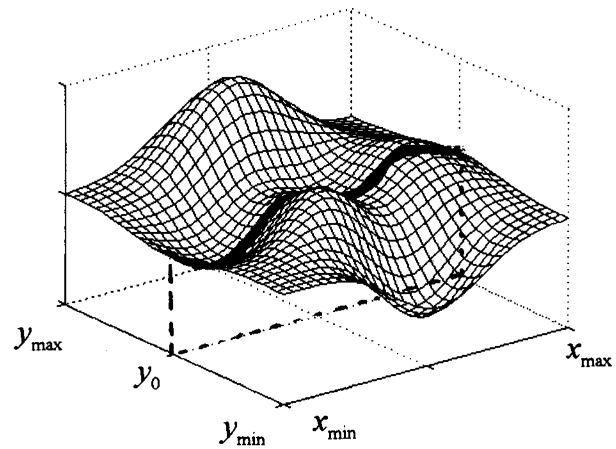
Illustration of how a 2-variate function can be visualized in 1-D:



Similarly, functions in *n-D* can be represented by RBFs and shown as 2-D surfaces together with *n* - 2 'sliders' (for traveling through the data volume).

Multivariate optimization

Already for a function of two variables, searching for GLOBAL optimum is NOT trivial.



Say, function of 10 variables:
Equally accurate representation in all directions would require 30^{10} points...

Computational complexity potentially a major issue

Novel opportunities: (both 'borrowed' from analogies in nature)

- Genetic algorithms — Evolution of species infinitely more complex (and successful) than anything undertaken mathematically - try to mimick it.

- Simulated annealing — Slow cooling can give perfect crystals (global energy minima)

    - Traveling salesman problem
    - Integrated circuit layouts
    - Protein folding

  Named in 2000 as 'One of the top 10 algorithms of the 20th century'

  (Computing in Science and Engineering, American Institute of Physics, and IEEE Computer Society)

Conclusions

The methodologies in IMPRINT and ACT-R for modeling and visualization will likely prove robust, practical to use, computationally economical, and fully sufficient.

However, it is important to contrast with other approaches:

- Certain 'modules' might be too complex for a 'first principle' approach

- 'Data mining' may reveal relations which were unforeseen - and possibly not interpretable in all taxonomies even after having been observed.

- If a well working model becomes too slow to compute, one may want to model the model - which may require a quite different approach.

- It may be easier (and more convincing) to do some modeling outside of the main tool set.

- Visualization and optimization of multivariate relationships can contain significant computational challenges.