Plenary discussion of: DISCREPANCY BETWEEN CODE AND REALITY

Led by: Peter Craig

Modelling Inadequacy:

Computer Code	$y = y(q)$
Nature	y + measurement error

What connects them? Three 'spanning' views

1. One extreme view is that $y\ (q)$ perfectly describes nature.
2. $y(q)$ may be informative about y ….. and q may be 'motivated by nature'.
3. $y(q)$ may be informative without q being motivated by nature.

Forget view 1!

(a) Assume the existence of q_0 which is the 'best value' of q. 'Best' might be given by:

- (i) true $\leftarrow$ abandon!
- *(ii)* Best fit to 'real' parameters
- (iii) $y(q_0)$ best fit to y
- (iv) $y(q_0)$ is a sufficient statistic to represent the simulator
- (v) ……?

There followed a discussion trying to operationalise the definition of q.

- (vi) A value of q which gives $y(q)$ which has good predictive properties –
  This led to:

  Why are we modelling?

  The definition of q_0 may depend on the purpose of modelling

  $y = r.y(q) + h$ with h being the discrepancy and independent of q_0

(b) Write $y(q) = y + e(q)$ [y and e independent???]
If simulator does not match nature well, $e(q)$ is very (auto)-correlated. In effect:

$$e(q) = h + e(q)$$

N.B. Often know some – the history of y.

The q's are motivated by the model builder's understanding of the process of nature.

The Discrepancy

(a) What models are appropriate for h? It will be multi-dimensional.

(b) h is not necessarily – usually not – stationary. How do we handle the idea that $y(q^0)$ is some fit of y?

Discussion by Tony O'Hagen

$$z_i = r.h(x_i, q) + d(x_i) + e_i$$

where

z_i	–	reality
r	–	code scaler
h	–	code
x_i	–	control inputs
d	–	model inadequacy
e_I	–	observation error/random variation

Q. What do we mean by true values of calibration parameters?

A. No more or less than we mean by 'true' values of parameters in a (non-linear) regression model.

The person who wrote the code had some meaning in mind for the parameters. Modellers can agree on the conceptual definition of a quantity yet use it 'differently' in different models.

If you allow q to be loose, variability will be picked up in $h(x, q)$ but q will become less interpretable.

Smoothness

In general we should expect reality to be 'rougher' than the code: different roughness parameters, even different kinds of co-variance kernels.

Mike Goldstein

Experts often have a meaning for 'acceptable' discrepancy which relates to expected correlation structure.

General Discussion

- Discrepancy may have a physical meaning to modeller about a process not captured in the model.
- Some times direction of change is more important than the absolute value of y.
- We are getting back to methodology of science. In practice, scientists act *as if* a model is true when they also *accept* it to be false.
- Idea of a good model is related to prediction and to the decision to be based on the prediction.
- We work in a variety of contexts. Some science is about explanation but basic science eventually feeds predictive models.
- There are many varieties of predictions. There is also post-diction (e.g. what the climate was 10 years ago?)

- Observables may be as much mental constructions as parameters.
- The good thing about this workshop is the bringing together of statisticians and modellers.
- A computer code needs to be able to predict an experimental result within ‘expected error’.
- No, usually the computer code plus an expert forecast. The real validation of a computer code is whether it helps the expert forecast better.

Emulators

Led by Peter Challenor

- An emulator is a (statistical) faster model of a complex slow code.
- How do we design the experiment to provide the data to fit the complex models?
- Emulators in large scale models have a problem if the impact data/parameters are too numerous.
- Meta-models another name for emulators.
- 10,000 variables – need to screen out/discover important variables.
- Use meta-models to check for ‘sensible’ behaviour, i.e. understand the system.
- Is it worth getting a register together of experience with emulators?
- An emulator can be very different from a statistical regression model. May be derived analytically as a specific case of/approximation to a more complex model.
- Is simply treating a model as a ‘black box’ useful in this case? Someone built the model and knows why he put the components together.
- Three questions:
  - What do we want out of the behaviours of an emulator on the subspace of inputs considered?
  - What form should the emulator take?
  - How should the emulators be determined?
- Building a good emulator is analysis specific.
- Designing the experiment to provide data to fit emulator may be different to real world experimental design because of absence of random error? Ans. Latin hypercube is appropriate but so are other methods. Nonetheless, the point is right: the choice of the design should be matched to the purpose of the emulation.

Sensitivity Analysis – Pros and Cons

I A taxonomy of techniques (Jon Helton, Tony O'Hagan)

A Local Sensitivity Methods

1. Differential Analysis: Calculate derivatives with respect to the uncertain inputs and look at gradients. "Adjoint Methods" are used for efficiently calculating derivatives. Differential Analysis is sometimes called Perturbation Analysis (Ho, Harvard), One difficulty is numerical difficulties in calculating derivatives. Ref: Can Cacuc, Ed Ablo, Also see the program GRESS .

B. Global Sensitivity Methods

1. Monte Carlo: Sample from the distribution of the uncertain inputs – induces distribution on the output.
2. Response Surface Methodology: Use a response surface design (Box-Behnken, Central Composite, etc) to fit a response surface to the response as a function of the inputs. Squares of standardised regression coefficients represent the variance "explained" by the corresponding input variables and are used as a measure of importance.

   (a) Main effect and interaction plots (based on averages over all other variables but one, two etc) can be used as a measure of importance. These sorts of measures can be applied to other types of statistical models, such as kriging models.

   (b) (From David Draper's talk). Projection pursuit models rather than polynomial models can be used to explore "importance".

3. Fourier Amplitude Sensitivity Test (FAST): Fit Fourier series: amplitudes give information about variable importance. Sobel indices (integrating over uncertain inputs to get variances) can be helpful, but are computationally demanding. Ref: Saltelli.

II Additional Comments

A. Expert opinion as to how sensitive code is to an input – another measure of sensitivity to consider.

B. Three reasons for using sensitivity analysis.

1. Determining if the computations made by the model are sensible (give sensible answers)
2. Determining how the model "works".
3. Determining if disagreement between experts matters. (Do they disagree about a sensitive or insensitive input?).

C. A "second order" sensitivity analysis: how sensitive are the conclusions made from the model (as opposed to the output) to the inputs?