Interfacing physical experiments and computer models

Preliminary remarks

Tony O'Hagan

Linking simulators to reality

- Before we can think about designing observational studies we have to think about how the simulation model relates to reality
- It is absolutely vital to recognise that all models are wrong
- It is also important to distinguish between two objectives
  - Calibration – learning about parameters in the simulator
  - Prediction – and in this case there's a distinction between
    - Interpolation = prediction within the domain of observations
    - Extrapolation = prediction outside that domain

Traditional formulation

- The problem
  - We have a simulator that predicts a real world phenomenon
  - We have some observations of the real world
- Formally, the simulator takes two kinds of inputs
  - Calibration parameters θ
  - Control inputs x
  - Simulator is written $y = f(x, \theta)$
  - Observation z_i is obtained at control input values x_i
    - Experimental design involves choosing those x_i points in x space
- It is usual to write

$$z_i = f(x_i, \theta) + \varepsilon_i$$

  - where ε_i are independent observation errors
- But this is **wrong** because the model is wrong

Model discrepancy

- It is necessary to acknowledge model discrepancy
  - There is a difference between the model with best/true parameter values and reality

$$y = f(x, \theta) + \delta(x) + \varepsilon$$

  - where $\delta(x)$ accounts for this discrepancy
- Kennedy and O'Hagan (JRSSB, 2001) introduced this model discrepancy
  - Modelled it as a zero-mean Gaussian process
    - Conditional on some hyperparameters
  - They claimed it acknowledges additional uncertainty
  - And mitigates against over-fitting of θ
  - Subsequent experience suggests that caution is needed

The bottom line

- If you don't include some representation of model discrepancy, you can expect to get nonsense
  - That applies to experimental design, too
- Even if you do include model discrepancy, it's essential to **think** about it and model it carefully
  - Especially for calibration
  - Use knowledge about aspects of reality that are not adequately represented in the model
- Even if you do think about it and model it carefully,
  - You will not be able to learn the true physical values of calibration parameters, or to make accurate extrapolations
    - Not even with an unlimited number of physical observations
  - But interpolation predictions may be quite accurate
    - And increasingly so with more observations

And finally

- I am not aware of any work on design of physical experiments that includes any formal representation of model discrepancy
  - Whether for purposes of prediction or calibration
- But I hope to be corrected!