Methodology Panel

Development, Assessment and Utilization of Complex Computer Models

Kickoff Workshop

Research Triangle Park, September 11, 2006

Panelists

Michael Goldstein (U. Durham)

Tony O'Hagan (U. Sheffield)

Henry Wynn (London School of Economics)

Susie Bayarri (U. Valencia)

Calibration/Validation/Prediction

often for statisticians $\rightsquigarrow$ a CompMod can be treated as a black-box:

- feed it with inputs $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{u})$ (large dimension):

  $\boldsymbol{x} \rightsquigarrow$ controllable

  $\boldsymbol{u} \rightsquigarrow$ unknown parameters (tuning, calibration)

- produces output $y^M(\boldsymbol{x}, \boldsymbol{u})$ (expensive)

- CompMod $y^M(\boldsymbol{x}, \boldsymbol{u}) \rightsquigarrow$ surrogate of 'reality' $y^R(\boldsymbol{x})$ (reality 'knows' u)
- Model is run at inputs $(\boldsymbol{x}, \boldsymbol{u}) \in D^M \rightsquigarrow \boldsymbol{y}^M$
  Field data is collected at inputs $\boldsymbol{x} \in D^F \rightsquigarrow \boldsymbol{y}^F$
- Calibration $\rightsquigarrow$ learn about $\boldsymbol{u}$
  Prediction $\rightsquigarrow$ learn about $y^R(\boldsymbol{x})$
- Most important: Take into account *all* uncertainties:
  - Give $\hat{\boldsymbol{u}}$ and uncertainty (confidence) bands
  - Give $\hat{y}^R(\boldsymbol{x})$ and uncertainty (confidence) bands (Laura Swiler's talk)

- Validation $\rightsquigarrow$ is y^M *useful* for the *intended* use ??

  . (see also Wendy Parker and Laura Laura Swiler's talks)

  Note: Usual question "is the model correct?" not good

- Validation $\approx$ learn about $bias$ $=$ reality – model

- useful framework to achieve these goals:

$$
\begin{aligned}
y^R(\boldsymbol{x}) &= y^M(\boldsymbol{x}, \boldsymbol{u}) + b_u(\boldsymbol{x}) \\
y^F(\boldsymbol{x}) &= y^R(\boldsymbol{x}) + \epsilon
\end{aligned}
$$

$+$ Bayes (incorporates uncertainties)

- function b_u unknown $\rightsquigarrow$ a prior on it: $\pi(b_u \mid \boldsymbol{\theta}^b)$
- often y^M very slow $\rightsquigarrow$ unknown function (known only at some few inputs) $\rightsquigarrow$ a prior on it: $\pi(y^M \mid \boldsymbol{\theta}^M)$ (when 'fitted' $\rightsquigarrow$ emulator, surrogator, fast simulator, meta-model, ...)
- from joint posterior one gets:
  - posterior of $y^M \rightsquigarrow$ 'emulator'
  - posterior of $\boldsymbol{u} \rightsquigarrow$ calibration
  - posterior of $y^R \rightsquigarrow$ prediction
  - posterior of $b_u \rightsquigarrow$ validation
- basic framework seems simple; but lots of issues.

- Design $\rightsquigarrow$ choose D^M, D^F. Should 'extreme' values of $\boldsymbol{u}$ be used? How do we interpret run failures? can we identify 'unfeasible' regions of $(\boldsymbol{x}, \boldsymbol{u})$ values?
- Surrogator/fast simulator (non parametric prior on y^M) usually $\rightsquigarrow$ 'fitted' with little data
  - GASP (Gaussian separable processes) most popular
  - Other Processes: Lee, Morris, Wilkinson, Wolpert, ...
  - Other (statistical) possibilities? (looking inside the 'black box'?) (D. Wilkinson talk)
  - Simpler/rougher models? (D. Wilkinson talk)
- What to do with HUGE input spaces? (GASP does not scale to large dimensions) more research needed.

- High correlation among parameters; Confounding between $\boldsymbol{u}$ and b_u (and others)
  - How to report individual inferences?
  - How to interpret them?
  - Note: Prediction of reality is not affected
- Problem aggravated when $\boldsymbol{x}$ is also uncertain $\leadsto$ more parameters and more confounding
- Lots of hyperparameters ($\boldsymbol{\theta}^M, \boldsymbol{\theta}^b$, variances, means, ...)
  - Which priors to use for *automatic* use?
  - Because of confounding, many will have to be proper: which ones should not be automatic? (incorporate external information)

- Functional outputs: a couple possibilities
  - Add (discretized) ‘time’ as another input (and do something clever with GASP)
  - Expand function (e.g. wavelet basis) and apply methodology independently to each coefficient $\leadsto$ correlated errors?
  - Tempor-spacial models?

- Multivariate outputs
  - Multivariate priors

    Some GASP generalizations; more work needed
  - Hierarchical models:

    $y_1^M, \ldots, y_k^M$ related

    $b_{u1}, \ldots, b_{uk}$ related

    * What kind or relation? how strong?
    * priors? which ones need external information?

    (see also Gang Han's talk)

- Prediction for untried (o altered) scenarios (L. Swiler)
  - scheme works well in interpolation
  - how to extrapolate bias?
  - additive vs. multiplicative bias? others?
  - hierarchical models?
- MCMC. Poor identifiability results in serious numerical problems. Lots of work is needed here. (D. Wilkinson talk)
- Deliverable Software $\Rightarrow$ automatic enough $\Rightarrow$ needs
  - most of the priors $\rightsquigarrow$ 'automatic'
  - MCMC also 'automatic'

(L. Swiler's talk)

- Approximations?
  With tons and tons of (highly correlated/confounded) parameters, and little external prior information about most of them $\rightsquigarrow$ full Bayes might not be feasible $\rightsquigarrow$ approximations?
  - Which parameters are 'safest' to be MLE-ed? impact?
  - 'Modular' approaches?: learn about prior on y^M based only on $\boldsymbol{y}^M$; learn about $\boldsymbol{u}$, variances and everything else based on all data $\boldsymbol{y}^M, \boldsymbol{y}^F$
- DAU of stochastic computer models
  - 'emulators' for SCM
  - build stochasticity into CompMod (D. Wilkinson, M.West)

in summary ...

- huge, difficult problems
- way too little data (model runs, and/or field data) for the complexity of the models

... great challenges for working groups!

but there is only so much that we can do ;)