An Application of Reification for a Rainfall-Runoff Computer Model

Leanna House

(acknowledge Michael Goldstein, Allen Seheult, and Ian Vernon)

December 16, 2009

For the next 30 minutes...

- Background Computer models/Simulators
- What the heck is *Reification*?
- Motivation for this work - justify reification
- Application: Rainfall-Runoff Model
- Discussion

Why might we use simulated data?

To assess the behavior of complex physical systems,

we often start by collecting data.

Simulator based inferences include...

- When we use simulated data to assess characteristics of a complex system we have/model the following:

  z : Real world observation(s)

  y : Characteristic(s) of the complex system

  f : Deterministic simulator or computer model

  X : The input variables

- For a Design of Experiments (DOE), we have

$$
\begin{aligned}
X &= \{x_1, ..., x_n\} \\
\{f\}_{[n]} &= \{f(x_1), ..., f(x_n)\}
\end{aligned}
$$

- Example: Assess water runoff from Swiss catchment

  z_t : Measured run-off at time t

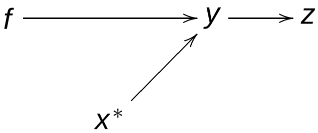
  y_t : Actual run-off at time t

  $f_t(x_i)$: Simulated run-off at time t given $X = x_i$

Virginia Tech. VT

Hierarchical model

- Conditional independence is often desired
- Given a tuned input x^*,



- And, we might have the following hierarchical model

$$
\begin{aligned}
z &= h(y) + e_z, \quad e_z \sim \text{No}(0, \sigma_z^2) \\
y &= f(x^*) + e_y, \quad e_y \sim \text{No}(0, \sigma_y^2) \\
x^* &\sim \text{Un}(a, b) \\
f(x) &= g(x)\beta + e(x), \quad e(x) \sim \text{GP}(0, \Sigma_f)
\end{aligned}
$$

One Source of Model Uncertainty

The model f itself is uncertain!

- $f(x)$ is inevitably *wrong*
  $\rightarrow$ depends on theories, assumptions, comp. resources...
- In turn, the orthogonal assumption:

  $$\mathrm{Cov}[\epsilon_y, \{f, x^*\}] \neq 0$$

  is rarely, if ever, appropriate.
- Existence of structural error in f; f does not mimic y perfectly
- What do we do?
  a) Ignore the data weakness?
  b) Re-code f and repeat a simulation experiment?

Reification (Goldstein and Rougier, 2008)

- Answer: Take advantage of being a Bayesian and elicit good, prior, expert judgments about the improved model, f^*
- The orthogonal assumption may hold for a hypothetical, ideal model $f^*(\tilde{x}^{**})$ at it's best input ($\tilde{x}$ may represent the same variables as x or the union of x and other variables)

$$y = f^*(\tilde{x}^{**}) + \epsilon_y^*, \quad \text{where } \epsilon_y^* \perp \{f, f^*, x^{**}\}.$$

- So, f^* represents our *reified* model (where, f is our *realized* model)
- M-W Dictionary: `To reify is to regard (something abstract) as a material or concrete thing`
- How do we come up with a good f^*?

Two Types of Reification: General

- A model reified generally is simply a realized model with more uncertainty.

$$f^*(x) \quad = \quad g(x)\beta^* + e^*(x) \tag{1}$$

where $\mathrm{E}[\beta^*] = \mathrm{E}[\beta]$ and $\mathrm{E}[e(x)] = \mathrm{E}[e^*(x)]$, but

$$\mathrm{Var}[\beta^*] = \Sigma^*_\beta > \Sigma_\beta, \qquad \mathrm{Var}[e^*(x)] = \Sigma^*_{e(x)} > \Sigma_{e(x)}$$

- Note, emulator (1) can be written as

$$f^*(x) \quad = \quad g(x)(\beta + \beta^+) + e(x) + (e^*(x) - e(x))$$

where $\mathrm{E}[\beta^+] = 0$ and $\mathrm{Var}[\beta^+] = \Sigma^+_\beta$. Judgments about the change in β and $e(x)$ might be easier for an expert than direct judgments about β^* and $e^*(x)$.

Two Types of Reification: Structural

- Structural reification adds one or more hypothetical input variables ν to a current, realized simulator/emulator.
- In some cases, there exists a value ν_0, such that the reified outcome $f^*(x, \nu)$ equals the realized outcome $f(x)$ when $\nu = \nu_0$.

$$\begin{aligned} f^*(x,\nu) &= g^*(x,\nu)\beta^* + e^*(x,\nu) \\ &= g(x)\beta + e(x) + g^+(x,\nu)\beta^+ + e^+(x,\nu) \\ &= f(x) + g^+(x,\nu)\beta^+ + e^+(x,\nu), \end{aligned} \tag{2}$$

where, $f(x)$ is learned from the data; $g(x, \nu_0) = 0$; and $e(x, \nu_0) = 0$.

Virginia Tech.

Combine reification methods

- Combine:
  1. Reify f structurally and emulate by $f'(x)$
  2. Reify $f'(x)$ generally and emulate by $f^*(x)$
- We have the following hierarchical model,

$$\begin{aligned} f^*(x)|f'(x), x, \nu &= f'(x) + g^{+\prime}(x,\nu)\beta^{+*} + e^{+*}(x,\nu) \\ f'(x)|f(x), x, \nu &= f(x) + g^{+}(x,\nu)\beta^{+\prime} + e^{+\prime}(x,\nu) \\ f(x)|x, \nu &= g(x)\beta + e(x) \quad \text{where,} \end{aligned}$$

- If $\mathrm{E}[\beta^{+*}] = \mathrm{E}[e^{+*}(x,\nu)] = \mathrm{E}[\beta^{+\prime}] = \mathrm{E}[e^{+\prime}(x,\nu)] = 0$,

$$\begin{aligned} \mathrm{E}[f^*(x)|f'(x), x, \nu] &= f'(x) \\ \mathrm{E}[f'(x)|f(x), x, \nu] &= f(x) \\ \mathrm{E}[f(x)|x] &= g(x)\mathrm{E}[\beta] + \mathrm{E}[e(x)]. \end{aligned}$$

Criticisms of Reification/Motivation of Work

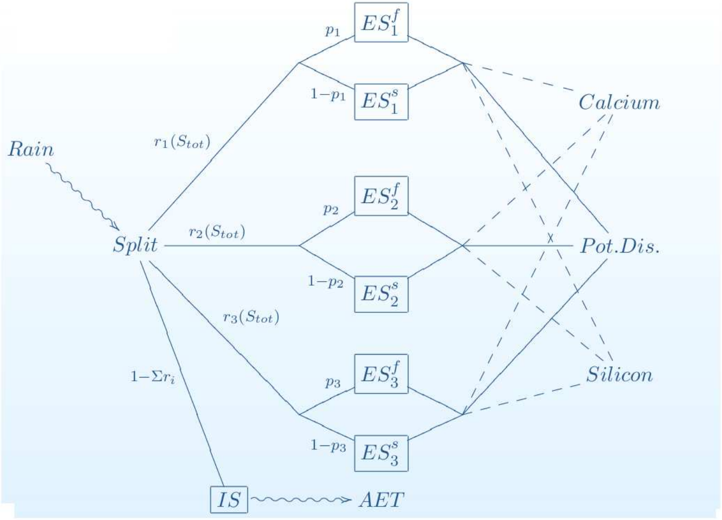
- Quantitative comparisons between a realized and reified model do not exist because reified models are hypothetical
- Only expert critiques can the validate the subjective judgments used to specify a reified model
- So, I asked,

  How do we elicit a good f^* from experts that other experts and non-experts might accept?

- One idea:
  1. Create a reduced vs. complete scenario
  2. Reify the reduced model *honestly*
  3. Compare the reified and complete models
  4. If the reification is successful, repeat for the complete model
- Example...

Graph of the Rainfall-Runoff Model

Three compartment mixing model with parallel transfer
(Iorgulescu, I., Beven, K.J., and Musy, A., 2005)

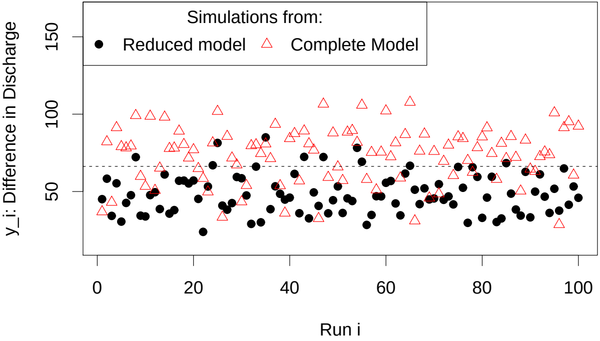


Details of the Complete Rainfall-Runoff Model

- Research Goal: Assess the true hydrological process for how rainfall becomes runoff, based on estimates of total discharge and tracer concentrations from real-world measurements and computer simulations,.
- Rainfall flows into one of 3 compartments (DP, GW, AS)
- Inputs: 6 parameters per compartment $p^{(k)},\ c_f^{(k)},\ c_s^{(k)},\ a^{(k)},\ b^{(k)},\ k^{(k)}, k \in [PD, GW, AS]$ and one constraint totaling to 17 inputs.
- Forcing functions: measured rainfall $rain(t)$ and assumed evaporation rates $AET(t)$, $t \in [1, ..., 839]$ hours
- Outputs include 3 time series: water discharged (PD), and the concentration of two tracer chemicals (Ca and Si) at each hour.
- For now, consider only PD(620)

Reduce the RR Model

- Exclude compartment AS; 11, not 17, inputs
- We create a 17 dimensional Latin hypercube (LHC) design, and remove the appropriate columns to create an 11-dimensional LHC



- Thus, results from the complete design, could represent what we might have had from the reduced model *if* it were complete.

Reify (honestly) the RR Model

- Judgment:
  Conditional on equal rainfalls (water input) and input parameters,
  1. More compartments $\rightarrow$ more discharge
  2. Compartment DP behaves similarly to AS
- So, a reified model (which includes AS) might predict 50% more discharge than the realized model or double the output from DP.
- We include these judgments by reifying the reduced RR model

General reification of RR Model

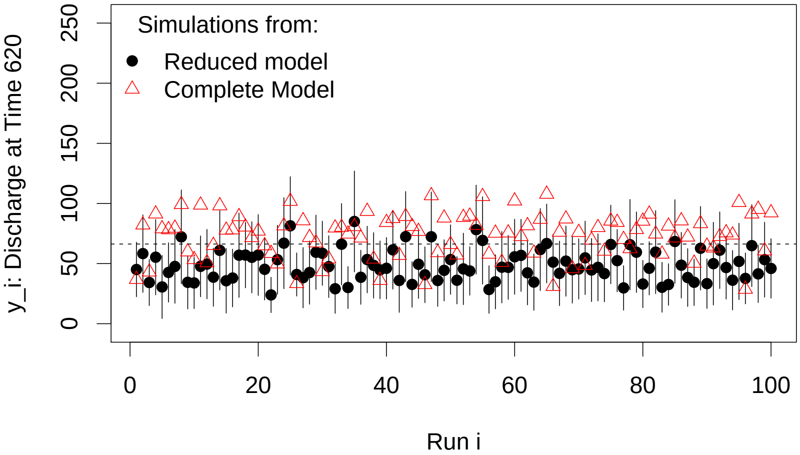
$$f^*(x) = g(x)(\beta + \beta^{+*}) + h(x)(A + A^{+*}) + e$$

where, $\mathrm{E}[\beta^{+*}] = \mathrm{E}[A^{+*}] = 0$ and the variances $\mathrm{Var}[\beta^{+*}]$ and $\mathrm{Var}[A^{+*}]$ are greater than zero. We estimate the variances of β^{+*} and A^{+*} based on $\mathrm{Var}[\beta]$, $\mathrm{Var}[A]$, and pre-specified constants $c_i \in \Re^+$ for $i \in [1, ..., 4]$:

$$\begin{aligned} \mathrm{Var}[\beta^{+*}] &= c_1 \mathrm{Var}[\beta] + c_2 \\ \mathrm{Var}[A^{+*}] &= c_3 \mathrm{Var}[A] + c_4. \end{aligned}$$

The size of the constants c_i reflect our judgment that the reified outcomes could equal approximately 1.5 times the realized outcomes.

Graph of general reification of RR Model



56% of the intervals contain the complete model outcomes.

Structure reification of RR Model

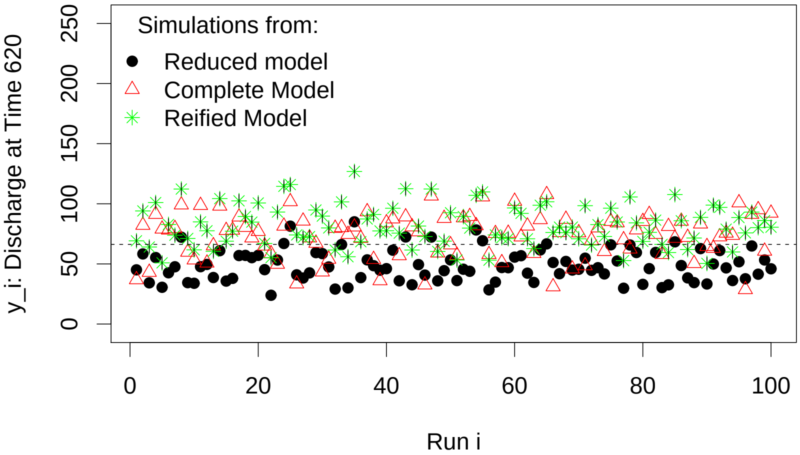
- Two options:
  1. Add six new parameters to emulator: $p^{(\text{new})}$, $c_f^{(\text{new})}$, $c_s^{(\text{new})}$, $k^{(\text{new})}$, $a^{(\text{new})}$, $b^{(\text{new})}$
  2. Exaggerate the effect of a compartment (DP) that is both currently in the realized model and expected to behave similarly to the new compartment, and add a k-parameter so that $k^{(\text{DP})}+k^{(\text{GW})}+k^{(\text{new})} = 1$.
- We chose 2. $f'(x,\nu) = f(x) + g^{+\prime}(x,\nu)\beta^{+\prime} + e^{+\prime}(x,\nu)$

$$
\begin{aligned}
g^{+\prime}(x,\nu) &= [\nu_1 g(x),\ \nu_1\nu_2(1-k^{(\text{DP})})] \\
e^{+\prime}(x,\nu) &= \nu_1(\nu_3 e(x) + \delta^+) \\
\delta^+ &\sim \pi(0,\nu_4)
\end{aligned}
$$

$$
\beta_j^{+\prime} = \begin{cases} 0 & \text{if } x_j \in (p^{(\text{GW})}, c_f^{(\text{GW})}, c_f^{(\text{GW})}, a^{(\text{GW})}, b^{(\text{GW})}) \\ \beta_j + s_j & \text{if } j = 0 \text{ or } x_j \in (p^{(\text{DP})}, c_f^{(\text{DP})}, c_f^{(\text{DP})}, a^{(\text{DP})}, b^{(\text{DP})}, k^{(\text{DP})}) \\ \beta_{k^{(\text{DP})}} + s_j & \text{if } j = 12, \end{cases}
$$

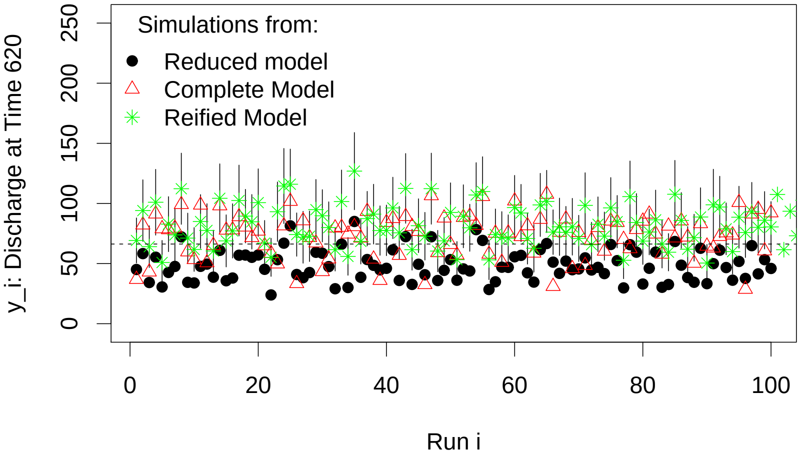
where $\mathrm{E}[s_j] = 0$ and $\mathrm{Var}[s_j] = \nu_5$.

Graph of structure reification of RR Model



21% of the intervals contain the three-compartment model outcomes.

Graph of structure and general reification of RR Model



We cover 67% of the complete model outcomes!

Summary/Discussion

- Computer Models and Reification
- I propose a way to, possibly, build confidence in reification
- Possibly justification in it's own right
- Possibly a check that should be used whenever opting to reify a model
- If the latter, the model must be accessible
- Didn't show how to use the reified model for inference
- Future: Extend to the multivariate outputs

Thank you!