An Application of Reification to a Rainfall-Runoff Computer Model

Leanna House

March 12, 2009

Abstract

Michael Goldstein's and Jonathon Rougier's 2007 paper, "Reified Bayesian Modelling and Inference for Physical Models" is challenging. Many people who do not have a Bayes linear background and/or are not familiar with subjective Bayesian methods have a difficult time understanding how to use the method suggested in Goldstein and Rougier [2008]. Thus, one goal for this paper is to distribute a laymen's version of reification for people who would like a second source on the matter. Additionally, some criticize reification because subjective model choices are hard to justify. However, the choice to not reify a model that is almost certainly inadequate is just as subjective and hard to justify. Thus, the second goal of this paper is exemplify the significance of reification using a hydrological computer model.

Key words: Computer Models, Reification, model discrepancy, application

1 Notation

Consider a computer experiment with n runs of a model f which depends upon a p-vector x as input and simulates q features of a physical process. Let y represent the o-vector

$y = [y_1, ..., y_o]^T$ of physical features; $f(x_i)$, $i \in [1, ..., n]$, represent a q-vector of model outcomes for input $x = x_i$; X represent a $n \times p$ matrix of inputs used for a computer experiment; F represent the $n \times q$ matrix of model outcomes from the experiment defined by X; and S be a shorthand notation for the set of all inputs and outcomes $\{F; X\}$ from one computer experiment. All of the features in y may or not be measurable; e.g., some features in y may correspond to future values, such as, global average temperature in year 2010. Thus, let the vector z contain the m observations ($m \leq q$) of either the measurable features or transformations of the measurable features, $z = [z_1, ..., z_m]^T$. Goldstein and Rougier (2008) model z as

$$z = Hy_h + \epsilon_z, \tag{1}$$

where H is a transformation matrix (e.g. identity matrix), y_h contains only measurable or *historical* elements of y (we will later refer to the unmeasurable or *future* elements as y_f), and ϵ_z represents measurement error that is orthogonal to y_h.

Often computer models are calibrated and, depending upon the procedure, a value for x is selected that minimizes either $f(x) - z$ or $f(x) - y$. Since Goldstein and Rougier (2008) model z by (1), the selections are equivalent regardless of the method and are called *best input* or *tuned input*. The general notation for a best input is x^*, but may vary depending upon the model for which x^* is a tuned input. For example, for model $f(x)$, the best input is x^* and for model $f'(x)$ it's $x^{*\prime}$. However, Goldstein and Rougier (2008) will refer to a model by $f^*(x)$, but the best input keeps the original notation x^*. Thus, for this paper, we introduce x^{**} which is the best input for model $f^*(x)$.

2 Purpose of reification

A computer model is rarely developed without relying on untestable, (although theoretically justifiable) assumptions, incorporating at least one simplification of known scientific principles, and considering computational limitations. Thus, researchers are seldom 100% certain that a model f can mimic the process which may result in y perfectly. On the other hand, researchers do know with reasonable certainty ways in which a model can be improved, and thus are presented with a decision before making inferences. Given the outcomes from an imperfect model, researchers may choose to either analyze the results knowing various data weaknesses or invest additional resources to re-code the model and repeat a simulation experiment to formulate inferences given new, up-versioned results. The first choice is tempting, but extremely unsatisfactory, while, the second is often impractical to implement. Comprehensive model improvements are typically too expensive, if not impossible, to administer. So, what is a modeler to do? The answer is to reify.

Inaccurate computer models are unsatisfactory because the standard characterization of y (for best input $x = x^*$)

$$y = f(x^*) + \epsilon_y, \quad \mathrm{E}[\epsilon_y] = 0, \quad \epsilon_y \perp \{f, x^*\}$$

is inappropriate; the orthogonality condition for ϵ_y (i.e., $\mathrm{Cov}[\epsilon_y, \{f, x^*\}] = 0$) does not hold, so the uncertainty of model based inferences will be underestimated. However, the orthogonal assumption may hold for a hypothetical, ideal model f^* that depends upon $\tilde{x}$ ($\tilde{x}$ may represent the same variables as x or the union of x and other variables) and is solved at the best input $\tilde{x}^{**}$,

$$y = f^*(\tilde{x}^{**}) + \epsilon_y^*, \quad \text{where} \;\; \epsilon_y^* \perp \{f, f^*, x^{**}\}. \tag{2}$$

Thus, conditional on $f(x)$, if an expert can 1) conceive of the new model $f^*(\tilde{x})$, and 2) quantify the difference $f^*(\tilde{x}) - f(x)|\tilde{x}, x$ (e.g., characterize via a statistical model), then legitimate inferences based on $f(x)$ are possible. Goldstein and Rougier (2008) refer to $f^*(\tilde{x})$ as a *reified simulator*, and we will often refer to $f(x)$ as the *realized simulator*.

3 Use of Emulators

Reified simulators are, by definition, hypothetical. They are not physically implemented and how they differ from realized models is naturally uncertain. In effect, reified simulators are random, unknown quantities and require stochastic representations to be discussed. Goldstein and Rougier (2008) exemplify the differences between realized and reified simulators via the difference in their emulators. In the next section we use emulators to clarify two methods of reifying a computer model which can be emulated with the following statistical model:

$$f(x) = g(x)\beta + e(x) \tag{3}$$

where $f(x)$ is $1 \times q$; g is a predetermined function of x and transforms the $1 \times p$-vector x to an $1 \times r_g$ vector $g(x)$; β represents a $r_g \times q$ matrix of model coefficients which has expectation and variance $\mathrm{E}[\beta] = \mu_\beta$ and $\mathrm{Var}[\beta] = \Sigma_\beta$ respectively; and $e(x)$ is a $1 \times q$ vector of residuals which depend upon x. The residuals $e = \{e(x_1), ..., e(x_n)\}$ are typically assumed to be orthogonal to β and have expectation zero, variance $\Sigma_{e(x)}$ per x, and an isotropic stationary correlation structure such as

$$\mathrm{Corr}[e(x), e(x')] = \prod_{l=1}^{p} \exp\{-\theta(x_l - x'_l)^2\}.$$

4 Two types of reification

Goldstein and Rougier (2008), introduce two kinds of reification, *general* and *structural*. Both methods are explained within the following context.

> Suppose the computer model $f(x)$ failed to include a variable ω which may or may not interact with x. In which case, a new version of emulator (3) is needed and Goldstein and Rougier (2008) suggest
>
> $$f^*(x,\omega) = g(x)\beta^* + h(x,\omega)\theta^* + e^*(x,\omega) \tag{4}$$
>
> where $h(x,\omega)$ is a predetermined function of both x and ω and has dimension $1 \times r_h$; β^* has the same meaning and dimension as Equation (3), but may have a different expectation and variance; θ^* represents a $r_h \times q$ matrix of model coefficients; and $e^*(x,\omega)$ is a $1 \times q$ vector of residuals which depends upon x and ω.

If we presume $h(x,\omega)\theta^*$ models a source of variation in $f(x)$ that is uncorrelated with $g(x)\beta^*$, then a realized simulator contains information about the reified simulator through the emulator components, β and $e(x)$; β and $e(x)$ can partially inform us about the reified emulator components β^* and $e^*(x,\omega)$ respectively. The final uncertainty assessment for $f^*(x,\omega)$ depends upon the degree to which we can characterize ω in $h(x,\omega)\theta^*$ and the differences, β^*- β and $e^*(x,\omega)$ - $e(x)$. In fact, general and structural reification methods differ according to these characterizations (or lack thereof).

4.1 General

When little is understood about ω, general reification is appropriate. A generally reified model is simply a realized model with more uncertainty. Since we cannot learn from experts

about the specific impacts of ω on $f(x)$, the cross product $h(x,\omega)\theta^*$ in Equation (4) represents an additional, unknown, uncharacterizable, source of uncertainty. Thus, in the name of being conservative, we inflate the uncertainty of $f(x)$ through β and $e(x)$. This is already justified because, in our modeling assumptions, we stated that the reified model depends on x and thus the emulator of $f^*(x,\omega)$ should depend partially on β and $e(x)$.

A generally reified emulator will have the same form as a realized emulator, and the difference between the two rests in the statistical assessments for the emulator terms. In other words, just like Equation (3), an emulator for $f^*(x)$ is

$$f^*(x) \quad = \quad g(x)\beta^* + e^*(x) \tag{5}$$

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

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

because emulator (5) 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 $e^*(x)$.

4.2 Structural

The second is *structural* reification which actually does add one or more hypothetical input variables ω to a current, realized simulator. Thus, the corresponding reified emulator includes both the terms of a realized emulator and entirely new coefficients to account for the additional variables defined in Equation (4). If an expert is extremely confident about

the effects of ω on $f(x)$, then the expectation and variance for the emulator terms θ^* and $e^*(x, \omega)$ could be specified directly. Although, in the event that they are not, we can rely on the simulated output if we consider ω to have a special property in the reified model. Variables with the special property are denoted as ν in Goldstein and Rougier (2008).

For some latent variables which we denote as ν (rather than ω), let there exist a value ν_0, such that the reified outcome $f^*(x, \nu)$ equals the realized outcome $f(x)$ when $\nu = \nu_0$. In such cases, the reified simulator augments the original simulator as

$$\begin{aligned} f^*(x) &= g^*(x, \nu)\beta^* + e^*(x, \nu) \\ &= g(x)\beta + e(x) + g^+(x, \nu)\beta^+ + e^+(x, \nu) \\ &= f(x) + g^+(x, \nu)\beta^+ + e^+(x, \nu), \end{aligned} \tag{6}$$

where $f(x)$ is learned from the data, $g(x, \nu_0) = 0$ and $e(x, \nu_0) = 0$. For any other value of ν, $\nu_i \in \texttt{span}(\nu)$, a reified simulator given (X, ν_i) would change the realized outcomes F according to $g^+(x, \nu)\beta^+ + e^+(x, \nu)$.

4.3 Combined reification

The primary application in Goldstein and Rougier (2008) exemplifies both kinds of reification. First, a realized emulator is reified structurally and emulated by $f'(x)$. Subsequently, $f'(x)$ is reified generally and emulated by $f^*(x)$ in GR-Section 6.3.

4.4 Expectation and variance of $f(x)$, $f'(x,\nu)$, and $f^*(x,\nu)$

If we choose to use both reification methods, we have the following hierarchical model,

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

where

$$\begin{aligned} \mathrm{E}[f^*(x)|f'(x),x,\nu] &= f'(x) + g^{+\prime}(x,\nu)\mathrm{E}[\beta^{+*}] + \mathrm{E}[e^{+*}(x,\nu)] \\ \mathrm{E}[f'(x)|f(x),x,\nu] &= f(x) + g^{+\prime}(x,\nu)\mathrm{E}[\beta^{+\prime}] + \mathrm{E}[e^{+\prime}(x,\nu)] \\ \mathrm{E}[f(x)|x] &= g(x)\mathrm{E}[\beta] + \mathrm{E}[e(x)] \end{aligned} \tag{7}$$

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

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

Additionally, we have

$$\begin{aligned} \mathrm{Var}[f^*(x)|f'(x),x,\nu] &= g'(x,\nu)\mathrm{Var}[\beta^{+*}]g'(x,\nu) + \mathrm{Var}[e^{+*}(x,\nu)] \\ \mathrm{Var}[f'(x)|f(x),x,\nu] &= g^{+\prime}(x,\nu)\mathrm{Var}[\beta^{+\prime}]g^{+\prime}(x,\nu) + \mathrm{Var}[e^{+\prime}(x,\nu)] \\ \mathrm{Var}[f(x)|x] &= g(x)\mathrm{Var}[\beta]g(x) + \mathrm{Var}[e(x)]. \end{aligned} \tag{8}$$

Notice that the stages differ in expectation and variance according to the moments of β, $e(x)$, $\beta^{+\prime}$, $e^{+\prime}(x,\nu)$,β^{+*}, and $e^{+*}(x,\nu)$. Thus, rather than build a hierarchical model for the outcomes, we could build a hierarchical model for the model components that reflects the relationship among the outcomes. In fact, the definitions of Δ' and Δ^* in Equation (17) of Goldstein and Rougier (2008) can be rewritten as

$$\Delta' = \mathrm{E}[\mathrm{Var}[f'(x)|f(x),x,\nu]]$$
$$\Delta^* = \mathrm{E}[\mathrm{Var}[f^*(x)|f'(x),x,\nu]],$$

and all of the complex calculations explained in GR-Sections 6.2.1-6.2.4 pertain to the specifications of $g^{+\prime}(x,\nu)$, $\beta^{+\prime}$ and $e^{+\prime}(x,\nu)$ as defined in Equation (6).

In the hierarchical form, we might also see why Goldstein and Rougier (2008) emphasize the orthogonality conditions in Equations (6) and (7) and Figure 12. With the above conditional independence structure, learning the marginal expectation and variance for $f^*(x)|x,\nu$ follows directly from the laws of conditional expectation and variance. For example,

$$\begin{aligned}
\mathrm{E}[f^*(x)|f(x),x,\nu] &= f(x) + g^{+\prime}(x,\nu)\mathrm{E}[\beta^{+\prime}] + \mathrm{E}[e^{+\prime}(x,\nu)] \\
\mathrm{Var}[f^*(x)|f(x),x,\nu] &= \mathrm{E}\Big[\mathrm{Var}[f^*(x)|f(x),f'(x),x,\nu]\Big] + \mathrm{Var}\Big[\mathrm{E}[f^*(x)|f(x),f'(x),x,\nu]\Big] \\
&= \mathrm{E}[Equation(8)] + \mathrm{Var}[Equation(7)] \\
&= \mathrm{E}\Big[g'(x,\nu)\mathrm{Var}[\beta^{+*}]g'(x,\nu) + \mathrm{Var}[e^{+*}(x,\nu)]\Big] \\
&\quad + \mathrm{Var}\Big[f'(x) + g'(x,\nu)\mathrm{E}[\beta^{+*}] + \mathrm{E}[e^{+*}(x,\nu)]\Big] \\
&= g'(x,\nu)\mathrm{Var}[\beta^{+*}]g'(x,\nu) + \mathrm{Var}[e^{+*}(x,\nu)] + \mathrm{Var}\Big[f'(x)|f(x)\Big],
\end{aligned}$$

so,

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

5 Model based inference with reification

Although reification is a novel concept and a key point in Goldstein and Rougier (2008), it is not the primary goal of the paper. Rather, the goal of Goldstein and Rougier (2008) is to develop informed inferences about y based on $z = \tilde{z}$ (note: the observed data $\tilde{z}$ are realizations of the random variable z as stated in GR-Section 4.2). In particular, they use real-world data to update carefully formulated prior judgements about y which are representative of the current state of understanding for the true system.

In Bayes linear terms, this means that Goldstein and Rougier (2008) calculate the expectations and variance of y *adjusted* by z

$$\mathrm{E}_{z=\tilde{z}}[y] \quad \text{and} \quad \mathrm{Var}_{z=\tilde{z}}[y].$$

These calculations require the prior specifications listed in table 1.

$\mathrm{E}[z]$,	$\mathrm{Var}[z]$,	$\mathrm{E}[y]$,	$\mathrm{Var}[y]$,	$\mathrm{Cov}[y,z]$

Table 1: Prior specifications needed to learn about y.

Given that we specify model (1), the prior moments for z are as follows: $\mathrm{E}[z] = H\mathrm{E}[y]$ and $\mathrm{Var}[z] = H\mathrm{Var}[y]H^T + \mathrm{Var}[\epsilon_z]$, where $\mathrm{Var}[\epsilon_z]$ represents the variance of measurement error and is often straightforward to specify. Prior moments of y however, stem from the computer

model and expert judgement. Namely, information about the true system that is contained in S and expert judgments concerning the realized model, reified versions of the model, and the best input infiltrate to the model-based inference via the prior specifications of y.

To quantify $\mathrm{E}[y]$ and $\mathrm{Var}[y]$, we choose to model y by Equation (2) with definition (6)

$$y|f^*(x,\nu),x^{**},\nu^{**},\epsilon_y^* = g^*(x^{**},\nu^{**})\beta^* + e^*(x^{**},\nu^{**}) + \epsilon_y^* \tag{9}$$

so that

$$\begin{aligned} \mathrm{E}[y|f^*(x,\nu),x^{**},\nu^{**}] &= g^*(x^{**},\nu^{**})\beta^* + e^*(x^{**},\nu^{**}) && (10)\\ \mathrm{Var}[y|f^*(x,\nu),x^{**},\nu^{**}] &= \Sigma_{\epsilon_y^*} && (11) \end{aligned}$$

However, the prior quantities listed in Table 1, are marginal, not conditional, expectations and variances. Thus, we apply the laws of conditional expectation and variance to marginalize the prior quantities for y:

$$\begin{aligned} \mathrm{E}[y|x^{**},\nu^{**}] &= g^*(x^{**},\nu^{**})\mathrm{E}[\beta^*] + 0 && (12)\\ \mathrm{Var}\Big[y|x^{**},\nu^{**}\Big] &= \mathrm{E}\Big[\mathrm{Var}\Big[y|f^*(x,\nu),x^{**},\nu^{**}\Big]\Big] + \mathrm{Var}\Big[\mathrm{E}\Big[y|f^*(x,\nu),x^{**},\nu^{**}\Big]\Big] \\ &= \mathrm{E}\Big[\Sigma_{\epsilon_y^*}\Big] + \mathrm{Var}\Big[g^*(x^{**},\nu^{**})\beta^* + e^*(x^{**},\nu^{**})\Big] \\ &= \Sigma_{\epsilon_y^*} + g^*(x^{**},\nu^{**})\mathrm{Var}[\beta^*]g(x^{**},\nu^{**})^T + \mathrm{Var}\Big[e^*(x^{**},\nu^{**})\Big], && (13) \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}[y] &= \mathrm{E}\Big[g^*\Big(x^{**},\nu^{**}\Big)\mathrm{E}[\beta^*]\Big] && (14)\\ \mathrm{Var}\Big[y\Big] &= \mathrm{E}\Big[\mathrm{Var}\Big[y|x^{**},\nu^{**}\Big]\Big] + \mathrm{Var}\Big[\mathrm{E}\Big[y|x^{**},\nu^{**}\Big]\Big] \\ &= \mathrm{E}\Big[Equation(13)\Big] + \mathrm{Var}\Big[Equation(12)\Big]. && (15) \end{aligned}$$

The marginal quantities for y and z rely on the elicitation from experts of the expectation and variance of ϵ_z, $g^*(x^{**}, \nu^{**})$, β^*, and ϵ_y^*. In the next section, we explain how to make these specifications within the context of an example.

6 Application

Since most simulators are imperfect, Goldstein and Rougier's reification methods apply to almost any deterministic or stochastic computer model. However, one criticism of reification is that model validation is, at best, challenging; quantitative comparisons between a realized and reified model do not exist because reified models are, by definition, hypothetical. Only expert critiques can the validate the subject judgements used to specify a reified model.

For this section, we create an artificial scenario which allows us to validate reification with both data and expert judgment. We simplified a hydrological computer model to create a second, unsatisfactory simulator; we refer to each model respectively as "complete" and "reduced". Knowing that an improved version of the reduced model is the complete model, we can reify the reduced model and compare our results to the complete model.

6.1 Complete and Reduced Model Description

The complete model is a rainfall-runoff (RR) model developed by Iorgulescu et al. [2005] that simulates fluctuations in water runoff or *discharge* over time at the Swiss Haute-Mentue research catchment. For this application, the model simulates 839 consecutive hours to compare to real-world data collected from the catchment between in August-September 1993. The real data includes (among other quantities), hourly rainfall and discharge measurements from the catchment.

The model is a "three compartment model with parallel transfer" in that water, which is either currently in the system or input by rain, may enter one of "three compartments",

labeled "ground water" (GW), "Acid Soil" (AS) and "Direct Precipitation" (DP), and eventually leave in "parallel" (from the compartments) via either "fast" or "slow" mechanisms. Per compartment, the model characterizes the fluid dynamics using 6 parameters which we denote as $p^{(l)},\ c_f^{(l)},\ c_s^{(l)},\ a^{(l)},\ b^{(l)},\ k^{(l)}$ for $l \in \{\text{GW}, \text{AS}, \text{DP}\}$. Parameter descriptions, ranges, and units are documented in Table I. of Iorgulescu, I. et al. (2005). Three parameters, $k^{(\text{AS})}$, $k^{(\text{DP})}$, and $k^{(\text{GW})}$, are constrained in the model to sum to one. Thus, the complete model requires only 17 input specifications, in addition to rainfall measurements (millimeters per hour), to predict water potential discharge.

Discharge is one of the three hourly times series output by the complete model. For this application however, we assess the water discharged at hour 620 which is shortly after a large rainfall. We explored different univariate summaries of the discharge-time series, and found that time 620 was particularly sensitive to the model inputs (more so than the mean discharge across time, the difference between the first and last discharge, and the maximum discharge across time). Given an approximate,[1] 17-dimensional sobol sequence experimental design with $n = 1040$ points ($\{p_i^{(l)},\ c_{f_i}^{(l)},\ c_{s_i}^{(l)},\ a_i^{(l)},\ b_i^{(l)},\ k_i^{(l')}\}$, for $l \in \{\text{GW}, \text{DP}, \text{AS}\},\ l' \in \{\text{DP}, \text{AS}\}$, and $i \in [1, ..., 1040]$), the outcomes are plotted in Figures 1 and 2.

We create the reduced model by removing compartment AS from the complete model. Now, the reduced model requires hourly precipitation measurements and the specification of only 11 input parameters. These parameters, except for $k^{(\text{GW})}$, have the same ranges listed in Table I. of Iorgulescu, I. et al. (2005). Since $k^{(\text{DP})}$ and $k^{(\text{GW})}$ must still sum to one, the range of possible values for $k^{(\text{GW})}$ shifts from $0.1 - 0.55$ to $0.45 - 0.90$.

Outcomes from the reduced model are plotted in Figures 1 and 2. The experimental design for the reduced model is an approximate, 11-dimensional sobol sequence that is a subset of the aforementioned 17-dimensional design. The appropriate dimensions of the 17-

[1]We consider the design to be "approximate" because we designed an experiment for 2500 runs and removed those that did not satisfy the summation constraint and the proposed ranges for $k^{(\text{AS})}$, $k^{(\text{DP})}$, and $k^{(\text{GW})}$; 1040 points in the experiment design remained.

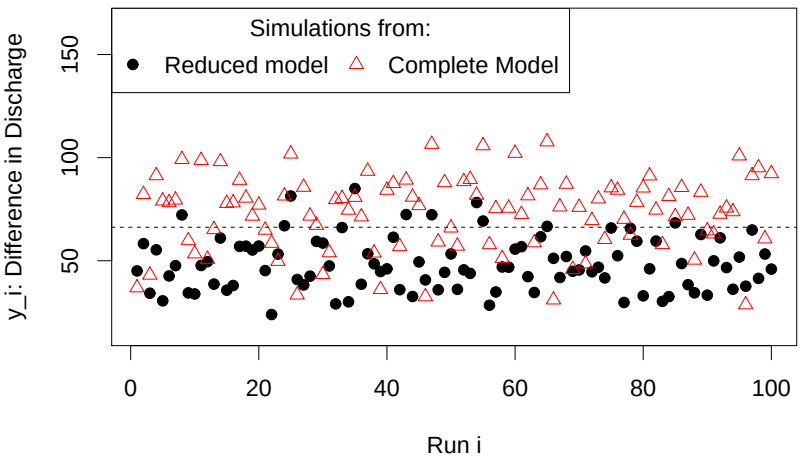

Figure 1: Typically, reified model runs are not available. However, for the purpose of displaying the effects of reification, we took a complete model and reduced its complexity. Thus, this plot displays both the reduced and complete model outcomes at time 620 for 100 runs of a 1040 experimental design. The dashed line marks the observed discharge at time 620.

dimensional design are removed to create the 11-dimensional design, so that the outcomes are paired; with the exception of $k^{(\text{GW})}$, all of the input values used for the parameters of the reduced model are also used for the complete model. Albeit, the pairwise comparisons are conditional on the experimental designs and would change for different specifications of the non-overlapping model parameters, $p^{(\text{AS})}$, $c_f^{(\text{AS})}$, $c_f^{(\text{AS})}$, $a^{(\text{AS})}$, $b^{(\text{AS})}$, $k^{(\text{AS})}$. However, since we specify legitimate designs for both models that would have likely been specified, regardless of the existence of the other, direct comparisons of the results are useful.

Since we are pretending that the reduced model is the realized model, let $f(x)$ represent the predicted model outcome for the reduced model given input x. In the next section, we emulate $f(x)$.

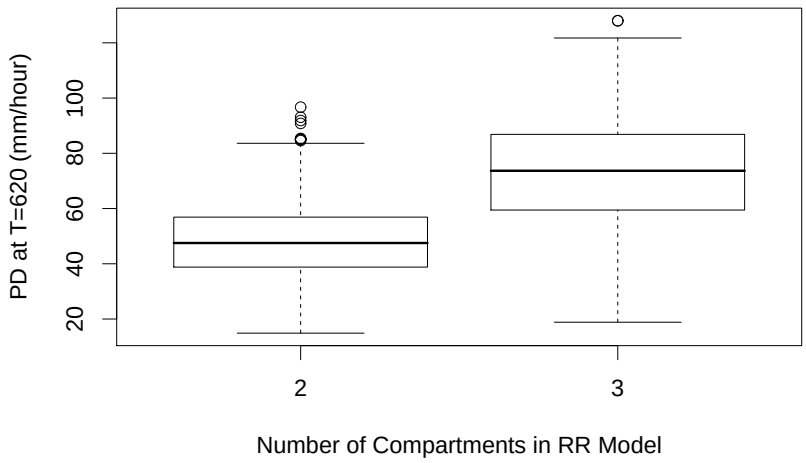

Figure 2: This boxplot displays the distribution of outcomes (potential discharge at time 620) for 1040 runs from either the reduced (2 compartment) and complete (3 compartment) models. Notice the increase discharge from the complete model.

6.2 Emulation

Statistical models that summarize simulated outcomes for the purpose of emulation should predict simulated outcomes with zero (or practically zero) error. Typical emulation approaches rely on Gaussian Process machinery similar to

$$f(x) = g(x)\beta + u(x), \quad u(x) \sim \mathrm{GP}(0, \sigma^2 K) \tag{16}$$

where $f = \{f(x_1), f(x_2), ..., f(x_n)\}$, x is an $n \times (p+1)$ matrix (includes a column of ones), $g_1()$ is a pre-specified function for the inputs variables, β is an $p \times 1$ matrix of coefficients, $u(x)$ is a vector of model residuals which are modeled jointly by a Gaussian Process centered at zero with covariance $\sigma^2 K$. Goldstein and Rougier (2008) on the other hand, chose to use linear regression (see GR-Section 6.1). With a predetermined, complex function of x

(denoted as $r()$), a linear model, such as

$$f(x) = r(x)\theta + e \qquad \text{where} \quad e \sim N(0, \sigma^2), \tag{17}$$

can fit any dataset almost perfectly in that $e = F - r(x) \approx 0$ and $\sigma^2 \approx 0$. Goldstein and Rougier (2008) specify $r()$ by first defining two orthogonal functions (with respect to a uniform weight function) $g()$ and $h()$. They then consider $r(x) = [g(x)\ h(x)]$ (the function $h()$ is not the same function defined on GR-Section 4.1) and $\theta = [\beta\ A]$, so that

$$f(x) = g(x)\beta + h(x)A + e, \quad \text{where} \quad u(x) = h(x)A + e. \tag{18}$$

In turn, this approach is similar in spirit to (16) in that $u(x) = h(x)A + e$, but the covariance of $h(x)A + e$ need not be stationery. Note, the expectation and variance estimates for both β and A are learned purely from GLS even though A is stated in GR-Section 6.1 to have a matrix normal distribution with expert elicited hyperparameters. These hyperparameters relate to the correlation between outputs, and in the univariate case, do not apply. In fact, for univariate model outcomes, the distribution for A reduces to a multivariate normal with a mean and variance determined simply from GLS fitted values; the expectation and variance for β are determined in a similar way and displayed in Goldstein and Rougier (2008)'s Table 2.

In the application section of Goldstein and Rougier (2008)(GR-Section), $g()$ is an identity function and $h()$ squares and pairwise multiplies the columns of x; we keep the same definition for $g()$ so that β corresponds to the intercept and main effects, but expand $h()$ so that A corresponds to quadratic, cubic, and 57 interaction terms. The dimensions for $g(x)$ and $h(x)$ are $n \times 12$ and $n \times 79$ respectively. For convenience, we will not apply GLS, but simple linear regression to assess the regression coefficients, but we will transform the ranges of the input parameters to $[-1, 1]$ for reasons explained in Section 6.4. Table 2 contains the estimates and

stand error's for β, as well as the correlation coefficient (R^2) and residual standard deviation (RSD).

Table 2: Table for parameters estimates $\hat{\beta}$ given observed and complete/reified model outcomes.

Index j	Coefficient	Reduced (Realized) Model Outcomes
0	Intercept	52.84
1	$p^{(\text{DP})}$	4.63
2	$p^{(\text{GW})}$	7.98
3	$c_f^{(\text{DP})}$	1.54
4	$c_f^{(\text{GW})}$	3.33
5	$c_s^{(\text{DP})}$	4.47
6	$c_f^{(\text{GW})}$	-1.19
7	$a^{(\text{DP})}$	-7.30
8	$a^{(\text{GW})}$	0.25
9	$b^{(\text{DP})}$	14.31
10	$b^{(\text{GW})}$	-1.76
11	$k^{(\text{DP})}$	9.93
	R2 (%)	98.37
	ARSD	1.65

The premise underlying the concept of reification is that this emulator is informative of an emulator for the reified simulator (Section 6.3). In the next three sections, we discuss how to adjust the two compartment model emulator.

6.3 Reify the two-compartment model

In both a reduced and reified model, runoff only generates from water that has entered a compartment. Thus, conditional on equal rainfalls (water input) and input parameters, simulators with more compartments will likely estimate higher amounts of discharge than models with fewer compartments. However, each compartment has different properties that affect the amount of water that leaves. For example, GW behaves differently from DP and AS, but AS and DP present similar patterns of discharge. In our case, the realized model

does not have compartment AS. Thus, we might expect a reified model to output 50% more than the realized model or double the output from DP.

For typical applications of reification, one may choose to implement general, structural, or both reification methods. To exemplify each approach, we will reify the reduced model using each method respectively in Sections 6.3.1-6.3.3.

6.3.1 General Reification

To reify generally, we start with model (18), and include β^{+*}, and A^{+*} in the following way,

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

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 (Section 6.3). Thus, we set c_1 and c_3 to 1.12^2 and

$$\begin{aligned} c_2 &= \left(\frac{0.5}{1.96}\right)^2 \mathrm{E}[\beta]\mathrm{E}[\beta]^T - c_1 \mathrm{Var}[\beta] \\ c_4 &= \left(\frac{0.5}{1.96}\right)^2 \mathrm{E}[A]\mathrm{E}[A]^T - c_3 \mathrm{Var}[A] \end{aligned}$$

(Appendix A justifies the choices for $c_1, ..., c_4$). Given $c_1, ..., c_4$, we may calculate the uncertainty associated with $f^*(x)$ (i.e., the inflated uncertainty of our realized model), and, if we

select a distribution for $f^*(x)$, we may calculate a 95% credible interval per x. For example, if we assume that $f^*(x)$ is distributed normally, then the credible intervals are center about $\mathrm{E}[f^*(x)]$ and span approximately four standard deviations. Assuming normality, 56% of the credible intervals contain the complete model outcomes as displayed in Figure 3.

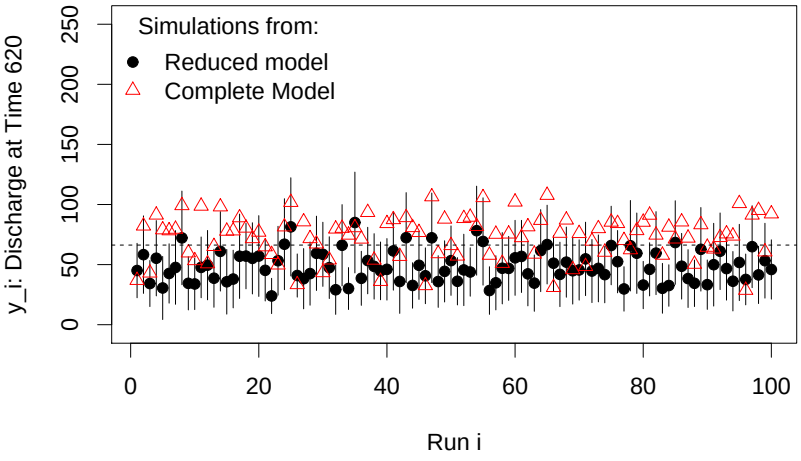

Figure 3: Model uncertainty learned from general reification.

6.3.2 Structural Reification

To reify a computer model structurally so that we consider the impact of an additional new compartment more carefully than in Section 6.3.1, we may take one of two courses of action. One, we could add six new parameters to emulator (18) which correspond to the following hypothetical inputs: $p^{(\mathrm{new})}$, $c_f^{(\mathrm{new})}$, $c_s^{(\mathrm{new})}$, $k^{(\mathrm{new})}$, $a^{(\mathrm{new})}$, $b^{(\mathrm{new})}$. Or two, we could exaggerate the effect of a compartment that is both currently in the realized model and expected to behave similarly to the new compartment. However, to implement the second approach, an additional k-parameter must be added to uphold the constraint that $k^{(\mathrm{DP})}$, $k^{(\mathrm{GW})}$, and $k^{(\mathrm{new})}$ sum to one. We opt to implement option two.

Expert judgments indicate that a new compartment would behave similarly to compartment DP. Thus,to adjust the effects of the DP-covariates and add $k^{(\text{new})}$ we take the approach described in GR-Section 6.2.1. We begin with

$$f'(x,\nu) = f(x) + g^{+\prime}(x,\nu)\beta^{+\prime} + e^{+\prime}(x,\nu)$$

and define

$$\begin{aligned} g^{+\prime}(x,\nu) &= [\nu_1 g(x),\ \nu_1\nu_2(1-k^{(\text{DP})})] && (19)\\ e^{+\prime}(x,\nu) &= \nu_1(\nu_3 e(x) + \delta^{+}) && (20)\\ \delta^{+} &\sim \pi(0,\nu_4) && (21) \end{aligned}$$

(where $\pi(0,\nu_4)$ represents any probability distribution function with mean zero and variance ν_4), so that $g^{+\prime}(x,\nu)$ transforms the 12 dimensional input matrix (11 main effects plus an intercept) to a 13 dimensional input matrix and $\beta^{+\prime} = [\beta_0^{+\prime}, \beta_1^{+\prime}, ..., \beta_{12}^{+\prime}]^T$. For $j \in [1, ..., 12]$,

$$\beta_j^{+\prime} = \begin{cases} 0 & \text{if } x_j \in (p^{(\text{GW})}, c_f^{(\text{GW})}, c_f^{(\text{GW})}, a^{(\text{GW})}, b^{(\text{GW})}) \\ \beta_j + s_j & \text{if } j = 0 \text{ or } x_j \in (p^{(\text{DP})}, c_f^{(\text{DP})}, c_f^{(\text{DP})}, a^{(\text{DP})}, b^{(\text{DP})}, k^{(\text{DP})}) \\ \beta_{k^{(\text{DP})}} + s_j & \text{if } j = 12, \end{cases} \quad (22)$$

where $\text{E}[s_j] = 0$ and $\text{Var}[s_j] = \nu_5$. With these definitions, $\text{E}[\beta_{(12)}^{+\prime}|\beta] = \beta$ (notation: $\theta_{(j)}$ represents the vector θ without the jth element, e.g., $\beta^{+\prime} = [\beta_{(12)}^{+\prime}, \beta_{12}^{+\prime}]$), $\text{E}[\beta_{12}^{+\prime}|\beta_{k^{(\text{DP})}}] = \beta_{k^{(\text{DP})}}$, $\text{Var}[\beta^{+\prime}|\beta] = \text{I}_{13}\nu_5$ (notation: I_{13} presents a 13-dimensional identity matrix), and

$f'(x,\nu)$ can be rewritten as

$$\begin{aligned} f'(x,\nu) &= g(x)\Big(\beta + \nu_1\beta^{+\prime}_{(12)}\Big) + \nu_2\Big(1 - k^{(\text{AS})}\Big)\nu_1\beta^{+\prime}_{12} + h(x)\Big(A + \nu_1\nu_3 A\Big) + \Big(e + \nu_1\nu_3 e\Big) + \nu_1\delta^{+} \\ &= g(x)\beta'_{(12)} + g(x,\nu)\beta'_{12} + h(x)A' + e' \\ &= g'(x,\nu)\beta' + e'(x,\nu) \end{aligned} \tag{23}$$

where

$$\beta'_{(12)} = \beta + \nu_1\beta^{+\prime}_{(12)}, \qquad \beta'_{12} = \nu_1\beta^{+\prime}_{12}, \qquad \beta' = [{\beta'_{12}}^T, \beta'_{12}]^T$$

$$A' = A + \nu_1\nu_3 A, \quad e' = e + \nu_1\nu_3 e + \nu_1\delta^{+}, \quad e'(x,\nu) = h(x)A' + e'.$$

Similar to constants $c_1, ...c_4$ in Section 6.3.1, we choose values for parameter $\nu = [\nu_1, ..., \nu_5]$ to reflect our judgment of the discrepancy between $f(x)$ and $f'(x)$. Thus, $\nu_1 \in [0,1]$ and, if $f(x) - f'(x) = 0$, $\nu_1 = 0$. In our case, we choose ν_1 based on the assumption that $\text{E}[f'(x)]$ equals $1.5\text{E}[f(x)]$ which, by Equation (18), is approximately $1.5g(x)\text{E}[\beta]$. Since our attempt to reify the reduced model structurally relies only on the discharge from the DP compartment (Equation (22)), we decompose $g(x)\beta$ by the compartments, $\beta_0 + g(x_{\text{GW}})\text{E}[\beta_{\text{GW}}] + g(x_{\text{DP}})\text{E}[\beta_{\text{DP}}]$, where x_{GW} and β_{GW} represent the GW$-$ covariates and slope coefficients, and x_{DP} and β_{DP} represent the DP$-$ covariates and coefficients. In doing so, the choice for ν_1 derives from

$$1.5g(x)\text{E}[\beta] = g(x)\text{E}[\beta] + \nu_1\frac{\beta_0 + g(x_{\text{DP}})\text{E}[\beta_{\text{DP}}]}{g(x)\text{E}[\beta]},$$

and solves to set $\nu_1 = 0.625$.

We select $\nu_2 \in (0,1)$ to segment $1 - k^{(\text{DP})}$ so that $k^{(\text{new})} = \nu_2(1 - k^{(\text{DP})})$ and $k^{(\text{GW})} = (1 - k^{(\text{DP})})(1 - \nu_2)$. We set $\nu_2 = 0.5$ which means that the reified computer model codes equal amounts of water into the new and GW compartments. Parameters ν_3 and ν_4 take on

the same meaning as c_i and the variance if δ_i^+ in Equation (24b) of Goldstein and Rougier (2008), and, for reasons explained in GR-Section 6.2.4, we set ν_3 to -0.1 and ν_4 to the expected variance of $u(x)$ ($u(x)$ for this application, is defined in Equation (18)). Finally, the parameter ν_5 allows us to account for our uncertainty about $\beta^{+\prime}$. Since our experts are very confident about the behavior of a third compartment, we want the standard deviation of $\nu_1\beta^{+\prime}$ to equal the standard error of β in emulator (18). Thus, we set ν_5 to $\text{Var}[\beta]/(\nu_1^2 + 0.00001)$. With these specifications for $\nu_1, ..., \nu_5$, we calculate the pointwise uncertainty bars for the structurally reified simulator (some are displayed in Figure 4), and 21% of the intervals contain the three-compartment model outcomes.

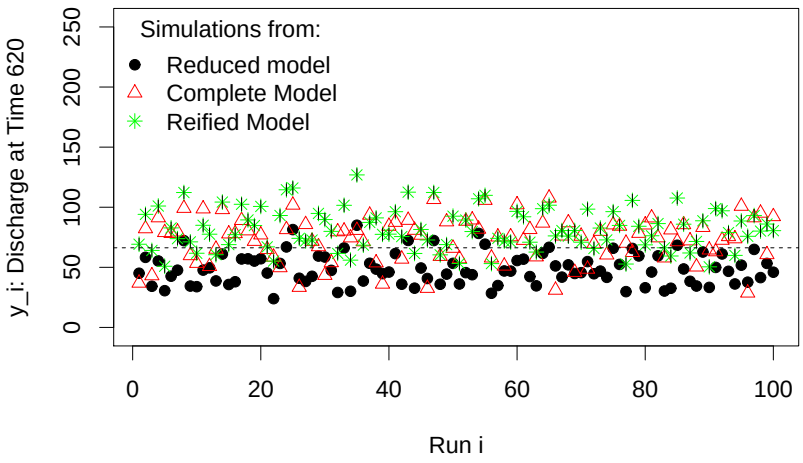

Figure 4: Model predictions and uncertainty learned from structure reification.

6.3.3 General and Structural Reification

The number of complete model outcomes covered using general reification is larger than the number covered using structural reification because the uncertainty intervals for the latter are too small. However, unlike the structurally reified model, the expectations of $f^*(x)$

(the generally reified model) per x never overlap the corresponding complete model outcomes. Thus, in this section, we apply both general and structural reification; we implement structural reification and expand the uncertainty intervals using general reification,

$$f(x) \rightarrow f'(x,\nu) \rightarrow f^*(x,\nu).$$

Starting with model (23), our final reified model $f^*(x)$ in this section is

$$\begin{aligned} f^*(x,\nu) &= g(x)\Big(\beta'_{(12)} + (\beta'_{(12)})^{+*}\Big) + g(x,\nu)\Big(\beta'_{12} + (\beta'_{12})^{+*}\Big) + h(x)\Big(A' + (A')^{+*}\Big) + e' \\ &= g'(x,\nu)\beta^* + h(x)A^* + e' \\ &= g'(x,\nu)\beta^* + e^*(x,\nu) \end{aligned} \tag{24}$$

We rely on the same specifications for $\nu_1, ..., \nu_5$ as stated in Section 6.3.2 to calculate the expected value and variance for β', A', and e'. We set the expectations of $(\beta')^{+*}$ and $(A')^{+*}$ to zero and calculate their variances based on specifications for $c_1, ..., c_4$. In Section 6.3.1, these specifications reflected our prior judgement that the distance between $f(x)$ and $f^*(x)$ was approximately $|0.5f(x)|$. Now, $c_1, ..., c_4$ reflect the distance between $f'(x,\nu)$ and $f^*(x,\nu)$ which we consider to be only $|0.25f'(x,\nu)|$. In turn, we set $c_1 = c_3 = 0.75^2$ and assess c_2 and c_4 as described in Section 6.3.1 and Appendix (A). With these specifications for ν and $c_1, ...c_4$, we calculate the 95% credible intervals. We cover 67% of the complete model outcomes as depicted in Figure 5.

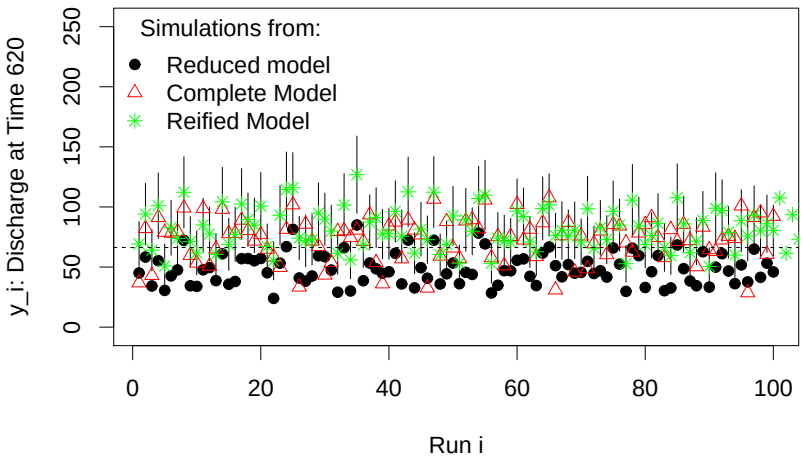

Figure 5: Model predictions and uncertainty learned from structure and general reification.

6.4 Inference

For this application, the observed measurement z is the measured discharge at time 620, and z =66.25 liters per second (lps). We model z as

$$z = y + \epsilon_z,$$

and characterize y using model (24) solved at the best inputs x^{**} and ν^{**},

$$f^*(x, \nu) \quad = \quad g'(x^{**}, \nu^{**})\beta^* + e^*(x^{**}, \nu^{**}) \tag{25}$$

Our goal is to assess the adjusted expectation and variance of y given z. To do so, we must specify the joint prior moments for y and z which rely on the prior moments of ϵ_z, ϵ_y^*, and the model components in our final, reified model (25): $g'(x^{**}, \nu^{**})$, β^*, and $e^*(x, \nu)$.

For ϵ_z, we assume that $\mathrm{E}[\epsilon_z] = 0$ and that the true discharge measurement is within 10% of recorded measurement; i.e., the standard deviation of ϵ_z is approximately $0.05z$. We also

assume that $\mathrm{E}[e^*(x,\nu)] = 0$, but we rely on the moment specifications for A^* and e' made in Section 6.3.3 to assess the variance of $e^*(x,\nu)$. Also in Section 6.3.3, we specified the moments for β^*.

For the remaining quantities, $g'(x^{**},\nu^{**})$ and ϵ_y^*, we follow the recommendations made by Goldstein and Rougier (2008). Since $g'()$ is a linear function of x and ν, $\mathrm{E}[g'(x^{**},\nu^{**})] = g(\mathrm{E}[x^{**}],\mathrm{E}[\nu^{**}])$ and $\mathrm{Var}[g'(x^{**},\nu^{**})]$ is straightforward to derive. Thus, we assume that the elements in x and ν are independent and uniformly distributed across their respective ranges. Since each x_l ($l \in [1,...,11]$) was transformed so that $x_l \in [-1,1]$ (Section 6.2), $\mathrm{E}[x_l^{**}] = 0$ and $\mathrm{Var}[x_l^{**}] = 1/3$, and the expectations and variances for ν^{**} are as follows: $\mathrm{E}[\nu_1^{**}] = 0.5$, $\mathrm{Var}[\nu_1^{**}] = 1/12$, $\mathrm{E}[\nu_2^{**}] = 0.5$, $\mathrm{Var}[\nu_2^{**}] = 1/12$, $\mathrm{E}[\nu_3^{**}] = 0$, $\mathrm{Var}[\nu_3^{**}] = 1/3$. For the expectation and variance of ϵ_y^*, we base our judgments on the distance between $f'(x)$ and $f^*(x)$; $\mathrm{E}[\epsilon_y^*] = 0$ and $\mathrm{Var}[\epsilon_y^*]$ equals the average variance of $f^*(x) - f'(x)$ across the inputs in the 11-dimensional experimental design.

Given these specifications, we have following prior quantities $\mathrm{E}[y] = \mathrm{E}[z] = 93.9$ lps, $\mathrm{Var}[y] = \mathrm{Cov}[y,z] = 25.7^2$ lps^2, and $\mathrm{Var}[z] = 26.0^2$ lps^2. Thus,

$$\mathrm{E}_{\tilde{z}}[y] = 66.7 \quad \text{and} \quad \mathrm{Var}_{\tilde{z}}[y] = 10.8.$$

Had we not reified and relied purely on the emulator of the realized model to assess the uncertainty of $f(x)$ (given the same moment specifications for ϵ_y^*), the posterior assessments would have been,

$$\mathrm{E}_{\tilde{z}}[y] = 65.9 \quad \text{and} \quad \mathrm{Var}_{\tilde{z}}[y] = 6.8.$$

The predictions and z are very similar, but the variance estimates differ. Both estimates are much larger than the variance of the measurement error, and the variance estimate based solely on the realized model is smaller than the variance estimate based on the reified model. For this example, had we assessed y without reification, we would have been overly confident

about our prediction.

7 Discussion

In the example we provided, our aim was to predict the true discharge of water at time 620 with reasonable ranges of uncertainty. To make the prediction, we had one real-world measurement with error and 1040 runs of an inadequate model computer model. Because of the model inadequacy, one approach is to ignore the model and simply use the one, observed measurement to estimate y. Another approach is to assume that the model is correct, and combine the simulated data with the real-world data as described in many current modeling approaches [Craig et al., 2001; Higdon et al., 2004; Goldstein and Rougier, 2004; Kennedy and O'Hagan, 2001]. Needless to say, both approaches are inadequate, and although reification isn't perfect, it offers an improved solution.

Goldstein and Rougier (2008)'s solution is to use another, untapped source of information: the experts. Given expert judgement and the simulated data, we can assess the prior expectation and variance for y by manipulating the prior moments for the realized model in justifiable ways. Reification formalizes this prior specification/manipulation process. However, validating the prior specifications is a challenge. Effectively, a prior sensitivity analysis of the posterior moments of y is the only available post-reification model-check. For this effort, we did not conduct a sensitivity analysis, but Goldstein and Rougier (2008) described an approach in GR-Section 7.2.

However, what we did do was justify the reification model choices by comparing the reified model to the complete model. The choices we made were independent of the complete model in that we did not use the complete model to guide our solution. Rather, we reified the reduced model legitimately and obtained reified model expectations $\mathrm{E}[f^*(x)]$ that can be compared to the complete model for accuracy. Specifically, 67% of the complete model

outcomes were contained in the interval, $\mathrm{E}[f^*(x)] \pm 2\mathrm{Std}[f^*(x)]$.

In practice, reducing a complete model might be impractical. However, information is gained when comparing two forms of the same model. Meaning, to reify a model a researcher may take one of two approaches: 1) hypothesize the ideal model and change the realized model accordingly, and 2) hypothesize the expected difference in outcomes from an ideal model and add the difference to the realized model. Outcomes from two forms of the same model provides analyzable data to assess the difference in outcomes from models of varying complexity. The results from the analysis might be useful when reifying via the second approach. For example, in Section 6, we reduced a complete model with three compartments to a model with two compartments. We might be able extrapolate from what we learned about the difference between the models to the difference between a three and four compartment model, and reify the complete model.

Appendix

A Specifying c_1, c_2, c_3 and c_4

The expected value for $f^*(x)$ in Section 6.3.1 equals $\mathrm{E}[f(x)]$, so we aim to make the predictions intervals about $\mathrm{E}[f(x)]$ large enough to contain $1.5f(x)$. This means that because we assume the distribution of $f^*(x)$ is symmetric, we have the following requirement:

$$\mathrm{E}[f(x)] \pm 1.96\sqrt{\mathrm{Var}[f^*(x)]} \subset 1.5\mathrm{E}[f(x)].$$

Thus,

$$\begin{aligned}\text{Var}[f^{*}(x)] &> \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[f(x)]\text{E}[f(x)]^T \\ \text{Var}[f^{+*}(x)] &> \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[f(x)]\text{E}[f(x)]^T - \text{Var}[f(x)].\end{aligned}$$

And, if we assume that the $\text{Var}[f^{*}(x)]$ is at minimum $1.5^2\text{Var}[f(x)]$, so that $\text{Var}[f^{*}(x)] > 1.5^2\text{Var}[f(x)]$ and $\text{Var}[f^{+*}(x)] > (1.5^2 - 1)\text{Var}[f(x)]$, we have

$$\begin{aligned}&(1.5^2 - 1)\Big[g(x)\text{Var}[\beta]g(x)^T + h(x)\text{Var}[A]h(x)^T + \text{Var}[e]\Big] + c \\ &\quad > \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[f(x)]\text{E}[f(x)]^T - \text{Var}[f(x)],\end{aligned}$$

by definition of $\text{Var}[f(x)]$ and for $c \geq 0$. This is equivalent to

$$\begin{aligned}&g(x)\Big(1.12^2\text{Var}[\beta] + c_2\Big)g(x)^T + h(x)\Big(1.12^2\text{Var}[A] + c_4\Big)h(x)^T + 1.12^2\text{Var}[e] \\ &\quad > \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[f(x)]\text{E}[f(x)]^T - \text{Var}[f(x)],\end{aligned}$$

and explains why c_1 and c_3 are set to 1.12^2. Assuming that the $\text{Var}[e] \approx 0$, we make the conservative choice to derive c_2 and c_4 from

$$g(x)\Big(1.12^2\text{Var}[\beta] + c_2\Big)g(x)^T + h(x)\Big(1.12^2\text{Var}[A] + c_4\Big)h(x)^T = \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[f(x)]\text{E}[f(x)]^T.$$

so that

$$\begin{aligned}c_2 &= \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[\beta]\text{E}[\beta]^T - 1.12^2\text{Var}[\beta] \\ c_4 &= \Big(\frac{0.5}{1.96}\Big)^2 \text{E}[A]\text{E}[A]^T - 1.12^2\text{Var}[A]\end{aligned}$$

(recall $g(x)$ is orthogonal to $h(x)$).

Acknowledgements

I thank Ian Vernon, Allan Seheult, and Michael Goldstein at the University of Durham for their helpful conversations and suggestions; and I thank Ian Iorgulescu and Keith Beven at the Swiss Federal Institute of Technology and Lancaster University respectively for their assistance with the rainfall-runoff model. I also acknowledge the financial support provided by EPSRC and MUCM.

References

Peter S Craig, Michael Goldstein, Jonathan C Rougier, and Allan H Seheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454):717–729, 2001.

Michael Goldstein and Jonathan Rougier. Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing*, 26(2):467–487, 2004.

Michael Goldstein and Jonathon Rougier. Reified bayesian modeling and inference for physical systems. *Journal of Statistical Planning and Inference*, 2008.

Dave Higdon, Marc Kennedy, James C. Cavendish, John A. Cafeo, and Robert D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, 2004.

I. Iorgulescu, K.J. Beven, and A. Musy. Data-based modelling of runoff and chemical tracer

concentrations in the haute-mentue research catchment (switzerland). *Hydrological Processes*, 19(13):2557–2573, 2005.

Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 63(3):425–464, 2001.