Evaluating prediction uncertainty in simulation models

Michael D. McKay,[1] John D. Morrison, Stephen C. Upton

Los Alamos National Laboratory
Los Alamos, New Mexico 87545–0600 USA

Abstract

Input values are a source of uncertainty for model predictions. When input uncertainty is characterized by a probability distribution, prediction uncertainty is characterized by the induced prediction distribution. Comparison of a model predictor based on a subset of model inputs to the full model predictor leads to a natural decomposition of the prediction variance and the correlation ratio as a measure of importance. Because the variance decomposition does not depend on assumptions about the form of the relation between inputs and output, the analysis can be called nonparametric. Variance components can be estimated through designed computer experiments.

Key words: Model uncertainty; Uncertainty analysis; Sensitivity analysis; Nonparametric variance decomposition

[1] Los Alamos National Laboratory, Mail Stop F600, Los Alamos, NM 87545–0600 USA. Tel. No.: (1) 505.667.6227; Fax No.: (1) 505.667.4470; E-mail: mdm@lanl.gov

1 Background

In a general sense, uncertainty or variability in the prediction of a simulation model can arise from three sources. One source, called *simulation variability*, is an integral part of a stochastic simulation model and often corresponds to the stochastic variability one sees in the world or system being modeled. Simulation variability is an important model constituent in, for example, the probabilistic risk assessment exercises Helton [1][2][3] performed for the U.S. Nuclear Regulatory Commission. Another source of variability in prediction is called *input uncertainty*. It refers to incomplete knowledge of "correct" values of model inputs, including model parameters. Various methods related to assessing input uncertainty appear in the literature, including Cukier, Levine and Shuler [4], McKay [5], Cox [6], Iman and Hora [7], Krzykacz [8], Morris [9], Sobol' [10], and McKay [11]. Input uncertainty exists independently of any model. The final source of variability is called *structural uncertainty*, and exists because models are based on assumptions that are usually selected with some latitude. This source of uncertainty is associated with the mathematical form or structure of the model. The literature on structural uncertainty, in its infancy, includes Sacks, Welch, Mitchell, and Wynn [12], Atwood [13], McKay [14], Winkler [15], Draper [16], and Laskey [17].

This paper focuses its discussion on input-uncertainty analysis. Only one model is under investigation and its validity or structural uncertainty is not an issue. Simulation variability will not be addressed explicitly. A simple way to use the methods to be presented for true stochastic simulation models is to apply them to the simulation mean value of the model prediction. The added complication is that the precision of the estimator of the simulation mean should be considered. This consideration is outside the scope of this paper.

2 Mathematical abstraction

The ideas discussed in this section relate to input uncertainty, prediction uncertainty, and importance of inputs for a specified model. To set the stage, suppose that for a model $m(\cdot)$, the prediction y is determined by a vector of "input" variables x of length p. Input variables may define initial conditions or state of a system being modeled, as well as parameter values in the rules that determine y from the initial conditions. (When necessary, model parameters can be made explicit in a parameter vector θ.) We write the model prediction process as

$$y = m(x), \quad x \sim f_x(x), \quad y \sim f_y(y)\,. \tag{1}$$

The probability distribution f_x of the input variables x induces on y the probability distribution f_y, the *prediction distribution*. $V(y)$, the variance of the prediction distribution, is the *prediction variance*. The objective of our analysis will be to identify a subset of inputs x^s that "drives" the prediction variance.

Suppose that the model prediction based on the subset x^s of input variables is

$$\widetilde{y} = E(y \mid x^s)\,, \tag{2}$$

with the conditional expectation being over the complementary input subset $x^{\overline{s}}$. One can think of x^s as being "control" variables and $x^{\overline{s}}$ as "noise" variables. We can write the model prediction y to correspond to x^s and $x^{\overline{s}}$ as

$$\begin{aligned} y &= \widetilde{y} + (y - \widetilde{y}) \\ &= E(y \mid x^s) + e\Big(x^{\overline{s}} \mid x^s\Big) \ . \end{aligned} \tag{3}$$

The first term represents the average fixed value we expect for y due to the control variables x^s. The second term represents the random residual or error component due to the noise variables $x^{\overline{s}}$.

Our objective is to find a (small) subset of the input variables x^s for which their predictor $\widetilde{y}$ is a good approximation to the full model prediction y. One way to measure the quality of $\widetilde{y}$ as a predictor is by the quadratic loss function

$$\mathcal{L} = (y - \widetilde{y})^2 \ . \tag{4}$$

The expected value $E(\mathcal{L})$ is commonly called the mean squared error (MSE) of prediction. We relate the importance of the set x^s to its predictive ability. We measure predictive ability locally (at a specified value of x^s) by the loss function $\mathcal{L}$, and globally (averaged over values of x^s) by the expected value of $\mathcal{L}$. The global, mean squared error of prediction is a function of the difference in variances given by

$$\begin{aligned} E(\mathcal{L}) &= E(y - \widetilde{y})^2 \\ &= E(y - E[y \mid x^s])^2 \\ &= V(y) - V(E[y \mid x^s]) \\ &= V(y) - V(\widetilde{y}) \end{aligned} \tag{5}$$

because

$$\begin{aligned} \text{Cov}(y, E[y \mid x^s]) &= \int y E[y \mid x^s] f_{x^{\overline{s}}|x^s} f_{x^s} dx^{\overline{s}} dx^s - \mu_y^2 \\ &= V(E[y \mid x^s]) \ , \end{aligned} \tag{6}$$

where

$$\mu_y = \int y(x) f_x(x) dx \ . \tag{7}$$

Therefore, we see from

$$V(y) = V(\widetilde{y}) + E(\mathcal{L}) \tag{8}$$

that the prediction variance is "driven" by the subset x^s through the prediction variance $V(\widetilde{y})$ of the "restricted" prediction $\widetilde{y}$. The two variances approach equality as the prediction $\widetilde{y}$ approaches y. We use their ratio

$$\eta^2 = V(\widetilde{y}) \ / \ V(y) \ , \tag{9}$$

the correlation ratio of Pearson [18], as a measure of the importance of x^s relative to x.

3 Strategy for analysis

We now address the task of how one might find subsets x^s of important input variables that maximize the correlation ratio. We will construct an approach similar to subset selection in linear regression, which identifies "best" subsets of size 1, 2, and so forth. Following classical Analysis of Variance (ANOVA), we consider additive decompositions of the prediction variance with terms V_i that can be associated with the subsets of the inputs x. Symbolically, we write

$$V(y) = V_1 + V_2 + \cdots \ , \tag{10}$$

where the terms in the expansion represent contributions from subsets of inputs like, for example, individual inputs, pairs of inputs, triples, and so forth. When the components of the input vector x are statistically independent, several ANOVA-like decompositions have been used by, for example, Cukier, Levine and Shuler [4], Efron and Stein [19], and Sobol' [10]. Cox [6] relaxes the independence requirement, but only somewhat. For p independent inputs, there are $2^p - 1$ possible terms in the ANOVA-like decomposition. When no assumptions, i.e. y is approximately linear in x, are made about the form of the relationship between model inputs and output, we call the methods *nonparametric variance-based* methods.

An approach when the components are not independent follows from Panjer [20], who generalizes the well known variance formula

$$V(y) = V[E(y \mid w)] + E[V(y \mid w)] \tag{11}$$

from Parzen [21]. Suppose that the input vector x is partitioned into 3 disjoint subsets, denoted by $w_1, w_2,$ and w_3. (The number of such partitions may be enormous.) The prediction variance could be written

$$\begin{aligned} V(y) = & V_{w_3} E_{w_2|w_3} E_{w_1|w_2,w_3}(y \mid w_2, w_3) \\ & + E_{w_3} V_{w_2|w_3} E_{w_1|w_2,w_3}(y \mid w_2, w_3) \\ & + E_{w_2,w_3} V_{w_1|w_2,w_3}(y \mid w_2, w_3) \ . \end{aligned} \tag{12}$$

In the general case for k subsets (or random variables) $\{w_1, w_2, \cdots, w_k\}$ as presented by Panjer, there are k terms in the expansion and $k!$ possible expansions corresponding to the labeling of the set $\{w_1, w_2, \cdots, w_k\}$. When the random variables are independent, all decompositions are identical. In any event, Panjer's formula corresponds to ANOVA decompositions for nested or hierarchal models. Although the terms in Panjer's formula are nonnegative, they are not variances, in general, unless the components of x are independent.

Panjer's formula can be used in analysis of simulation models as follows. We note in Eq. 12 that

$$E_{w_2|w_3} E_{w_1|w_2,w_3}(y \mid w_2, w_3) = \widetilde{y}_3(w_3) \tag{13}$$

is the restricted predictor using w_3, and that

$$E_{w_1|w_2,w_3}(y \mid w_2, w_3) = \widetilde{y}_{2,3}(w_2, w_3) \tag{14}$$

is the restricted predictor using both w_2 and w_3. We interpret Eq. 12, written as

$$V(y) = V_{w_3}(\widetilde{y}_3) + E_{w_3} V_{w_2|w_3}(\widetilde{y}_{2,3}) + E_{w_2,w_3} V_{w_1|w_2,w_3}(y \mid w_2, w_3)\,, \tag{15}$$

to mean that combining the subset w_2 with the subset w_3 increases the restricted prediction variance for the combined subset by the amount

$$\Delta V = E_{w_3} V_{w_2|w_3}(\widetilde{y}_{2,3})\,. \tag{16}$$

Our strategy will be to build up x^s sequentially by examining estimates of ΔV for candidate input variables.

4 Sequential estimation of variance components

We now describe a screening procedure in the spirit of step-up regression, following McKay and Beckman [22]. The procedure uses estimates of elements in Eq. 15 based on Latin hypercube sampling (LHS) designs [23]. In the iterative procedure, x^s is the current set of variables selected to be important and plays the part of w_3. A candidate input variable x^* from $x^{\overline{s}}$ plays the part of w_2. x^* is under consideration for inclusion into x^s. The variable w_1 stands for all the other input variables, $x^- = \left\{x^{\overline{s}}\right\} - \{x^*\}$. In terms of the input variables, Eq. 15 becomes

$$V(y) = V_{x^s}(\widetilde{y}_{x^s}) + E_{x^s} V_{x^*|x^s}(\widetilde{y}_{x^*,x^s}) + E_{x^*,x^s} V_{x^-|x^*,x^s}(y \mid x^*, x^s)\,. \tag{17}$$

At each stage in the iteration, 1 or more variables x^* are moved from x^- into x^s. The iteration stops at the discretion of the analyst, and the selection of x^s is validated in a final step.

Estimators

To begin the iteration, we consider each input variable separately. Let x^* denote an input variable and x^- the remaining inputs in x. Suppose we have performed a computer experiment and generated the sample

$$\begin{gathered}\left\{y_{ij} = m\left(x_i^* \cup x_{ij}^-\right),\ i = 1, \cdots, I;\ j = 1, \cdots, J\right\} \\ x_i^* \sim f_{x^*} \text{ iid}, \quad x_{ij}^- \sim f_{x^-|x_i^*} \text{ c-iid}\,.\end{gathered} \tag{18}$$

That is, $\left\{x_1^*, \cdots, x_I^*\right\}$ is an independent and identically distributed (iid) sample from f_{x^*}, and for each value x_i^* in the sample, $\left\{x_{i1}^-, \cdots, x_{iJ}^-\right\}$ is a conditionally iid (c-iid) sample from $f_{x^-|x^*}$.

We calculate the sums of squares

$$\text{SST0} = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \overline{y})^2 \text{ and SSB0} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{y}_{i\cdot} - \overline{y})^2\,. \tag{19}$$

The quantities

$$\overline{y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} y_{ij} \text{ and } \overline{y} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij} \tag{20}$$

are sample means of the values of y_{ijk}. The correlation ratio for x^* is estimated with bias by

$$\widehat{\eta}^2_{x^*} = \text{ SSB0/SST0 }. \tag{21}$$

Typically, we select the input with the largest estimated correlation ratio to become the first element in x^s.

For subsequent stages or iterations, suppose we generate a sample of values y_{ijk} according to the design

$$\begin{gathered}\left\{y_{ijk} = m\left(x^s_k \cup x^*_{ik} \cup x^-_{ijk}\right),\ i = 1, \cdots, I;\ j = 1, \cdots, J;\ k = 1, \cdots, K\right\} \\ x^*_{ik} \sim f_{x^*|x^s_k} \text{ c-iid}, \quad x^-_{ijk} \sim f_{x^-|x^s_k, x^*_{ik}} \text{ c-iid}, \quad x^s_k \sim f_{x^s} \text{ iid }.\end{gathered} \tag{22}$$

We calculate the sums of squares

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk} - \overline{y}\right)^2 \text{ and SSB} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\overline{y}_{i\cdot k} - \overline{y}_{\cdot\cdot k}\right)^2 . \tag{23}$$

As before, $\overline{y}_{i\cdot k}$, $\overline{y}_{\cdot\cdot k}$ and $\overline{y}$ are the indicated sample means of the values of y_{ijk}. The estimate of the ΔV associated with x^* is proportional to SSB. The *partial correlation ratio* for x^*, "adjusted for" the subset x^s, is estimated with bias by

$$\widehat{\eta}^2_{x^*|x^s} = \text{ SSB/SST }. \tag{24}$$

The *incremental* correlation ratio for x^*, conditional on the subset x^s, is estimated with bias by using SST0 instead of SST in the denominator,

$$\Delta\widehat{\eta}^2_{x^*|x^s} = \text{ SSB/SST0 }. \tag{25}$$

It is the estimated amount by which the correlation ratio for the set x^s would increase with the addition of x^*. Both of these quantities are proportional to SSB and our estimate of ΔV. For sequential variable selection, we prefer to look at the partial correlation ratio together with the change in prediction variance because of the nature of the sample designs we use.

Sample designs

Independent samples for p input variables would require $N_0 = p \times I \times J$ model runs following Eq. 18 to begin the iterations. Subsequent stages would require $N = p \times I \times J \times K$ runs following Eq. 22. If LHS samples are used instead of simple random samples, both N_0 and N would be reduced by the factor p because the same y-values, albeit in a different order, would be used in the calculations for each input variable. Therefore, we generate first an LHS of size I for the p input variables. Then, we construct J pseudo replicates of that design matrix by independently applying new random permutations to each of its p columns. In this way, we generate a set of $N = I \times J$ input vectors x, containing I distinct values of each component of x, replicated J times in different pairings.

We further reduce N in the iterations by taking $I = 1$ and setting x^s in Eq. 22 to its nominal (median) value. This action is equivalent to approximating ΔV by

$$\widehat{\Delta V} = V_{x^*|x^s}\left(\widetilde{y}_{x^*|x^s=\text{nominal}}\right) \simeq E_{x^s} V_{x^*|x^s}\left(\widetilde{y}_{x^*|x^s}\right) = \Delta V \ . \tag{26}$$

The approximation to ΔV is equivalent to considering the additional importance of x^* not on average (globally) but only at the nominal value of x^s (locally). We then use the sample from the 0th iteration with Eq. 18 replacing the values of x^s with their nominal values. This is the design we use in place of Eq. 22 to obtain a new sample of y-values.

Properties of the estimator in Eq. 21 under LHS are not yet known. However, the authors have found the procedures to work very well in practice. Likewise, they have found that using $I = 1$ and setting x^s to its nominal value in the iteration steps has been adequate to identify important input subsets.

Validation

The final, validation step consists of estimating the correlation ratio for final x^s, the vector of the selected inputs. One can use the design of Eq. 18 (with x^s playing the role of x^* and $x^{\overline{s}}$ playing the role of x^-). Alternatively, one might use a fixed design, as will be done in the demonstration in the next section. We examine model responses at each of the x^s-design points to observe the variability caused by the "unimportant" inputs $x^{\overline{s}}$.

5 Demonstration application

The model in this demonstration is a discrete event simulation of the time-dependent movements of various cargos by various types of military aircraft. We study eight of the model's input variables (a small number): MOG, Max Wait, Use Rate, Enroute Time, Offload Time, Onload Time, Initial Hours and Fuel Flow. The model output predictions are cumulative hours (h) flown and tons (t) of cargo delivered for each aircraft type. Aircraft types are designated by C-141, C-17 and C-5A. The computer experiment is based on the form of replicated LHS described in the last section, with $I = 12$ distinct values of each input variable replicated $J = 4$ times for a total of 48 model runs at each stage. The "base case" or 0th iteration predictions for 6 outputs are presented in the bands in Figure 1. They show the prediction variability when all inputs are sampled from their joint uniform probability distribution f_x, which was taken to be joint uniform. The plots present a more complete description of the prediction distributions f_y that do prediction variances alone.

The model outputs y are not scalars but time series computed at 15 days. Therefore, we will apply the analysis calculations independently for each day. The prediction variances are estimated from the data in the Figure 1. The figure displays the object of the analysis, the spread in the plots, which is to be quantified by way of correlation ratios and attributed to various inputs. The correlation ratios for each of the 8 input variables are calculated from Eq. 21 for each day. The calculations are performed 8 times, with x^* corresponding to $x_1, x_2, \cdots, x_8$. Time trends in the

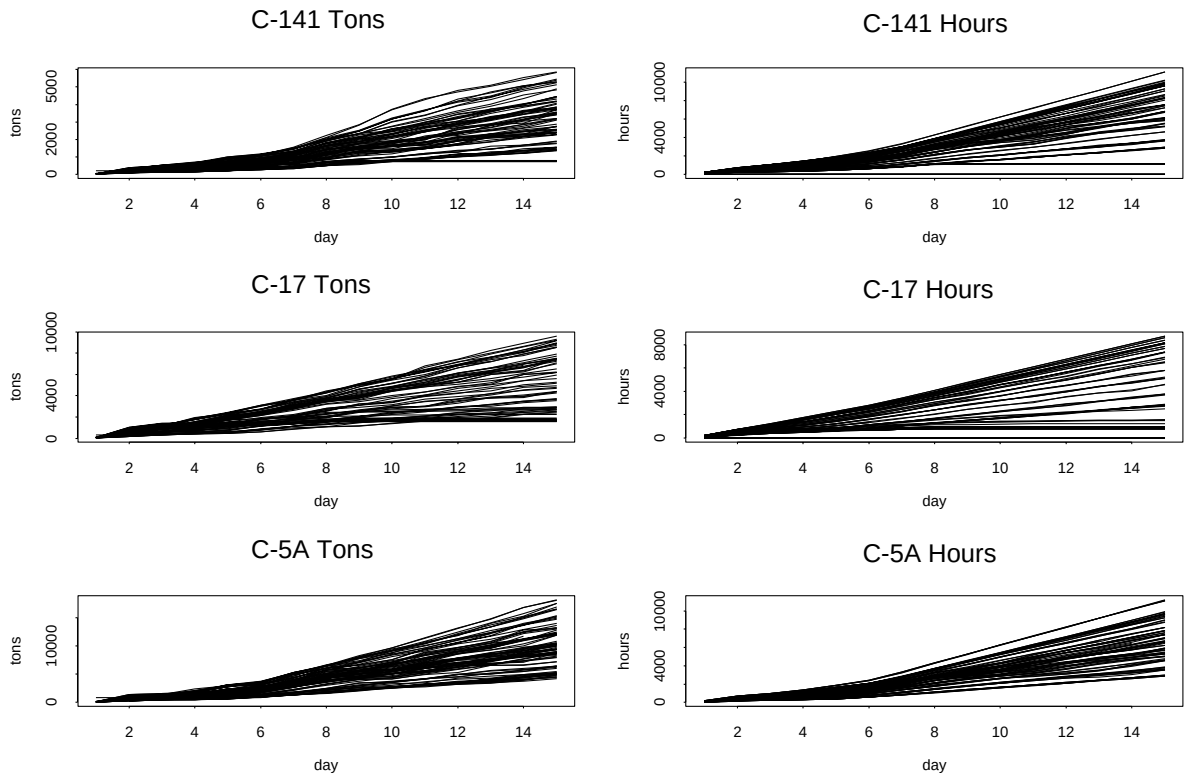

Figure 1. Base case, all 8 inputs vary

TABLE I
Correlation ratios from base case

Output ($\overline{\sigma}$)		MOG	Max Wait	Use Rate	Enroute Time	Offload Time	Onload Time	Initial Hours	Fuel Flow
C-5A.t (1227)	Avg $\widehat{\eta}^2$	0.28	0.23	0.45	0.14	0.27	0.28	0.17	0.40
	% days $\widehat{\eta}^2 \geq CV$	-	-	93	-	7	-	-	47
C-5A.h (1782)	Avg $\widehat{\eta}^2$	0.24	0.21	0.49	0.11	0.29	0.31	0.17	0.42
	% days $\widehat{\eta}^2 \geq CV$	-	-	100	-	-	-	-	53

CV is a critical value from normal theory under a null hypothesis of independence of x and y. It is used here only as a filter.

estimated correlation ratios can be very informative, pointing to regimes where different inputs are important. For this demonstration, however, we only present estimates averaged over days for each input. We point out that the estimates for each day and across outputs are not statistically independent. Sample results of the calculations for aircraft type C-5A[2] appear in Table I.

[2] Really for C-141

Average prediction variance is monitored by the average daily prediction variance in square root form. The average daily variability ($\overline{\sigma}$) in tons delivered by C-5A aircraft is 1227 from the first column of the table. It was computed as

$$\overline{\sigma} = \sqrt{\frac{1}{15}\sum_{d=1}^{15}\frac{1}{47}\sum_{i=1}^{48}(y_{id} - \overline{y}_d)^2} \tag{27}$$

where the subscript d indicates day. From the first row of numbers in the table we see that Use Rate, on average, accounts for 45% of the variability in y. Therefore, Use Rate is selected as the "x^s" at Stage 1. Because we choose to select variables one-at-a-time in this screening procedure, Stage 1 is complete. However, we see from Table I the indication that Fuel Flow is also important. Reduction in the bands of predicted values from setting Use Rate to its nominal value are seen in Figure 2. These runs constitute the data for Stage 2.

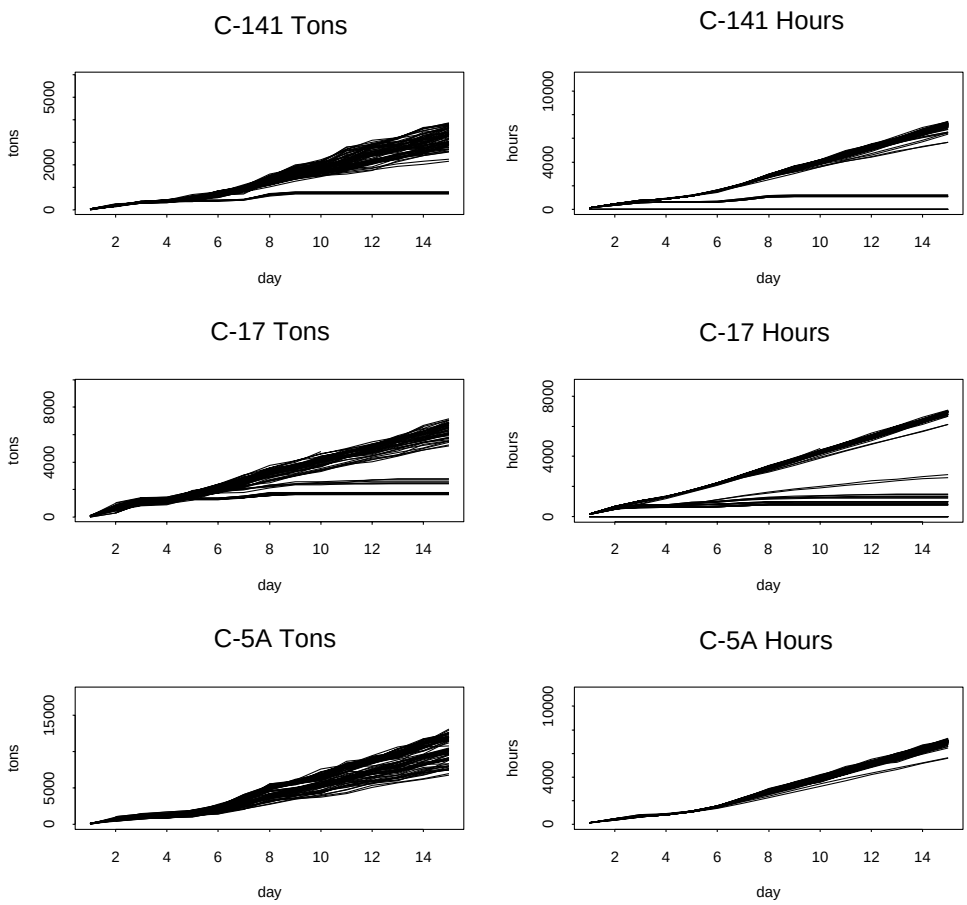

Figure 2. Use Rate set to nominal value and other 7 inputs vary

Validation

We continued selecting variables, one at a time, through three more screening stages but, finally, decided that only two (Use Rate and Fuel Flow) of the eight variables were the principal

contributors to prediction variance. We used a 2^2 factorial design on the extreme values of Use Rate and Fuel Flow, together with a small LHS on the other six inputs, to investigate contributions to prediction variance. Sample results for tons delivered by C-5A aircraft are displayed in Figure 3. The large differences among the four bands of predictions are due to the 4 combinations of values of Use Rate and Fuel Flow. The variations within each band are due to the other six inputs. These results indicate how well Use Rate and Fuel Flow account for the variability in the model predictions for C-5A tons depicted in Figure 1.

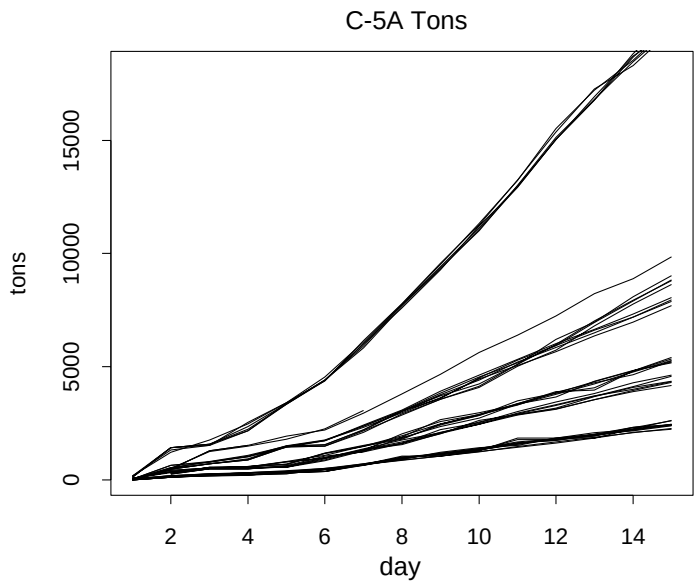

Figure 3. Bands of curves correspond to different combinations of 2 values of Use Rate (high in top 2 bands) and Fuel Flow (high within top 2 and bottom 2). Widths of bands correspond to variability attributable to the other 6 inputs

6 Conclusions

This paper presents techniques and suggests directions for research and development of methods for assessing effects of input uncertainty on model prediction. Variance-based measures of importance for input variables arise naturally when using the quadratic loss function of the difference between the full model prediction y and the restricted prediction $\widetilde{y}$. Practical limitations of the methods come from size of samples required to obtain adequate estimates of the variance components.

References

[1] J. C. Helton, Risk Analysis 14 (1994) 483–511.

[2] J. C. Helton, Reliability Engineering and System Safety 54 (1996) 145–164.

[3] J. C. Helton, Journal of Statistical Computation and Simulation 57 (1999) 3–76.

[4] R. I. Cukier, H. B. Levine, and K. E. Shuler, Journal of Computational Physics 26 (1978) 1–42.

[5] M. D. McKay, in: Proceedings of the Workshop on Validation of Computer-based Mathematical Models in Energy Related Research and Development, Texas Christian University, Fort Worth, TX, 1978, p. 36–.

[6] D. C. Cox, IEEE Transactions on Reliability R-31 (1982) 465–468.

[7] R. L. Iman and S. C. Hora, Risk Analysis 10 (1990) 401–406.

[8] B. Krzykacz, Tech. Rep. GRS-A-1700, Gesellschaft fur Reaktorsicherheit (GRS) mbH, Garching, Republic of Germany, 1990.

[9] M. D. Morris, Technometrics 33 (1991) 161–174.

[10] I. M. Sobol', Mathematical Modelling and Computational Experiment 1 (1993) 407–414.

[11] M. D. McKay, Reliability Engineering and System Safety 57 (1997) 267–279.

[12] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, Statistical Science 4 (1989) 409–435.

[13] C. L. Atwood, in: Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification, U.S. Nuclear Regulatory Commission, NUREG/CP–0138, Annapolis, MD, 1993, p. 99–106.

[14] M. D. McKay, in: Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification, U.S. Nuclear Regulatory Commission, NUREG/CP–0138, Annapolis, MD, 1993, p. 51–64.

[15] R. L. Winkler, in: Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification, U.S. Nuclear Regulatory Commission, NUREG/CP–0138, Annapolis, MD, 1993, p. 107–116.

[16] D. Draper, Journal of the Royal Statistical Society, B 57 (1995) 45–97.

[17] K. B. Laskey, IEEE Transactions on Systems, Man, and Cybernetics 26 (1996) 340–348.

[18] K. Pearson, Proceedings of the Royal Society of London 71 (1903) 288–313.

[19] B. Efron and C. Stein, Annals of Statistics 9 (1981) 586–596.

[20] H. H. Panjer, The American Statistician 27 (1973) 170–171.

[21] E. Parzen,, *Stochastic Processes*, page 55. San Francisco: Holden Day, 1962.

[22] M. D. McKay and R. J. Beckman, in: Proceedings of the American Statistical Association Section on Physical and Engineering Sciences, Toronto, 1994.

[23] M. D. McKay, W. J. Conover, and R. J. Beckman, Technometrics 21 (1979) 239–245.