

# A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling

J.C. Helton<sup>a,\*</sup>, F.J. Davis<sup>b</sup>, J.D. Johnson<sup>c</sup>

<sup>a</sup>Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA

<sup>b</sup>Sandia National Laboratories, Albuquerque, NM 87185-0779, USA

<sup>c</sup>ProStat, Mesa, AZ 85204-5326, USA

Received 16 October 2003; accepted 3 September 2004

Available online 21 November 2004

## Abstract

Uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling are compared. The comparison uses results from a model for two-phase fluid flow obtained with three independent random samples of size 100 each and three independent Latin hypercube samples (LHSs) of size 100 each. Uncertainty and sensitivity analysis results with the two sampling procedures are similar and stable across the three replicated samples. Poor performance of regression-based sensitivity analysis procedures for some analysis outcomes results more from the inappropriateness of the procedure for the nonlinear relationships between model input and model results than from an inadequate sample size. Kendall's coefficient of concordance (KCC) and the top down coefficient of concordance (TDCC) are used to assess the stability of sensitivity analysis results across replicated samples, with the TDCC providing a more informative measure of analysis stability than KCC. A new sensitivity analysis procedure based on replicated samples and the TDCC is introduced.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Epistemic uncertainty; Kendall's coefficient of concordance; Latin hypercube sampling; Monte Carlo analysis; Random sampling; Replicated sampling; Sensitivity analysis; Stability; Subjective uncertainty; Top down coefficient of concordance; Two-phase fluid flow; Uncertainty analysis

## 1. Introduction

The identification and representation of the implications of uncertainty is widely recognized as a fundamental component of analyses of complex systems [1–10]. The study of uncertainty is usually subdivided into two closely related activities referred to as uncertainty analysis and sensitivity analysis, where (i) uncertainty analysis involves the determination of the uncertainty in analysis results that derives from uncertainty in analysis inputs and (ii) sensitivity analysis involves the determination of relationships between the uncertainty in analysis results and the uncertainty in individual analysis inputs.

At an abstract level, the analysis or model under consideration can be represented as a function of the form

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) = f(\mathbf{x}), \quad (1.1)$$

where

$$\mathbf{x} = [x_1, x_2, \dots, x_{nX}] \quad (1.2)$$

is a vector of uncertain analysis inputs and

$$\mathbf{y} = [y_1, y_2, \dots, y_{nY}] \quad (1.3)$$

is a vector of analysis results. Further, a sequence of distributions

$$D_1, D_2, \dots, D_{nX} \quad (1.4)$$

is used to characterize the uncertainty associated with the elements of  $\mathbf{x}$ , where  $D_i$  is the distribution associated with  $x_i$  for  $i=1, 2, \dots, nX$ . Correlations and other restrictions involving the elements of  $\mathbf{x}$  are also possible. The goal of uncertainty analysis is to determine the uncertainty in the elements of  $\mathbf{y}$  that derives from the uncertainty in the elements of  $\mathbf{x}$  characterized by the distributions  $D_1, D_2, \dots, D_{nX}$  and any associated restrictions. The goal of sensitivity analysis is to determine relationships between

\* Corresponding author. Address: Department 6874, MS 0779, Sandia National Laboratories, Albuquerque, NM 87185-0779, USA. Tel.: +1-505-284-4808; fax: +1-505-844-2348.

E-mail address: [jhelto@sandia.gov](mailto:jhelto@sandia.gov) (J.C. Helton).

the uncertainty associated with individual elements of  $\mathbf{x}$  and the uncertainty associated with individual elements of  $\mathbf{y}$ .

A variety of approaches to uncertainty and sensitivity analysis are in use, including (i) differential analysis, which involves approximating a model with a Taylor series and then using variance propagation formulas to obtain uncertainty and sensitivity analysis results [11–24], (ii) response surface methodology, which is based on using classical experimental designs to select points for use in developing a response surface replacement for a model and then using this replacement model in subsequent uncertainty and sensitivity analyses based on Monte Carlo simulation and variance propagation [25–35], (iii) the Fourier amplitude sensitivity test (FAST) and other variance decomposition procedures, which involve the determination of uncertainty and sensitivity analysis results on the basis of the variance of model predictions and the contributions of individual variables to this variance [36–55], (iv) fast probability integration, which is primarily an uncertainty analysis procedure used to estimate the tails of uncertainty distributions for model predictions [56–62], and (v) sampling-based (i.e. Monte Carlo) procedures, which involve the generation and exploration of a probabilistically based mapping from analysis inputs to analysis results [63–73]. Additional information on uncertainty and sensitivity analysis is available in a number of reviews [69,70,74–80]. The primary focus of this presentation is on sampling-based methods for uncertainty and sensitivity analysis.

Sampling-based approaches for uncertainty and sensitivity analysis are very popular [81–96]. Desirable properties of these approaches include conceptual simplicity, ease of implementation, generation of uncertainty analysis results without the use of intermediate models, and availability of a variety of sensitivity analysis procedures [67,69,76,97,98]. Despite these positive properties, concern is often expressed about using these approaches because of the computational cost involved. In particular, the concern is that the sample sizes required to obtain meaningful results will be so large that analyses will be computationally impracticable for all but the most simple models. At times, statements are made that 1000 to 10,000s of model evaluations are required in a sampling-based uncertainty/sensitivity analysis.

In this presentation, results obtained with a computationally demanding model for two-phase fluid flow are used to illustrate that robust uncertainty and sensitivity analysis results can be obtained with relatively small sample sizes. Further, results are obtained and compared for replicated random and Latin hypercube samples (LHSs) [63,73]. For the problem under consideration, random and LHSs of size 100 produce similar, stable results.

The presentation is organized as follows. The analysis problem is described in Section 2. Then, the following topics are considered: stability of uncertainty analysis results (Section 3), stability of sensitivity analysis results based on stepwise rank regression (Section 4), use of

coefficients of concordance in comparing replicated sensitivity analyses (Section 5), sensitivity analysis based on replicated samples and the top down coefficient concordance (Section 6), sensitivity analysis with reduced sample sizes (Section 7), and sensitivity analysis without regression analysis (Section 8). Finally, the presentation ends with a concluding discussion (Section 9).

## 2. Analysis problem

The analysis problem under consideration comes from the 1996 performance assessment (PA) for the Waste Isolation Pilot Plant (WIPP) [99,100]. This PA was the core analysis that supported the successful Compliance Certification Application (CCA) by the US Department of Energy (DOE) to the US Environmental Protection Agency (EPA) for the operation of the WIPP [101]. With the certification of the WIPP by the EPA for the disposal of transuranic waste in May 1998 [102], the WIPP became the first operational facility in the United States for the deep geologic disposal of radioactive waste. Thus, the example used to illustrate properties of sampling-based approaches to uncertainty and sensitivity analysis in this presentation is part of a real analysis rather than a hypothetical example constructed solely for illustrative purposes.

The analysis problem involves the model for two-phase fluid flow that is at the center of the 1996 WIPP PA. This model is based on the following system of nonlinear partial differential equations:

Gas conservation

$$\nabla \cdot \left[ \frac{\alpha \rho_g \mathbf{K}_g k_{rg}}{\mu_g} (\nabla p_g + \rho_g g \nabla h) \right] + \alpha q_{wg} + \alpha q_{rg} = \alpha \frac{\partial(\phi \rho_g S_g)}{\partial t} \quad (2.1)$$

Brine conservation

$$\nabla \cdot \left[ \frac{\alpha \rho_b \mathbf{K}_b k_{rb}}{\mu_b} (\nabla p_b + \rho_b g \nabla h) \right] + \alpha q_{wb} + \alpha q_{rb} = \alpha \frac{\partial(\phi \rho_b S_b)}{\partial t} \quad (2.2)$$

Saturation constraint

$$S_g + S_b = 1 \quad (2.3)$$

Capillary pressure constraint

$$p_c = p_g - p_b = f(S_b) \quad (2.4)$$

Gas density

$\rho_g$  determined by Redlich–Kwong–Soave equation of state (see Eqs. (31) and (32), Ref. [103]).

Brine density

$$\rho_b = \rho_0 \exp[\beta_b(p_b - p_{b0})] \quad (2.5)$$

Formation porosity

$$\phi = \phi_0 \exp[\beta_f(p_b - p_{b0})] \quad (2.6)$$

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات