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

J.C. Helton\textsuperscript{a,}\textsuperscript{*}, F.J. Davis\textsuperscript{b}, J.D. Johnson\textsuperscript{c}

\textsuperscript{a}Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA
\textsuperscript{b}Sandia National Laboratories, Albuquerque, NM 87185-0779, USA
\textsuperscript{c}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.

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 \cite{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

\[ y = y(x) = f(x), \quad (1.1) \]

where

\[ x = [x_1, x_2, \ldots, x_n] \quad (1.2) \]

is a vector of uncertain analysis inputs and

\[ y = [y_1, y_2, \ldots, y_n] \quad (1.3) \]

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

\[ D_1, D_2, \ldots, D_n \quad (1.4) \]

is used to characterize the uncertainty associated with the elements of \( x \), where \( D_i \) is the distribution associated with \( x_i \) for \( i = 1, 2, \ldots, n \). Correlations and other restrictions involving the elements of \( x \) are also possible. The goal of uncertainty analysis is to determine the uncertainty in the elements of \( y \) that derives from the uncertainty in the elements of \( x \) characterized by the distributions \( D_1, D_2, \ldots, D_n \) 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: jhelton@sandia.gov (J.C. Helton).

0951-8320/$ - see front matter \copyright 2004 Elsevier Ltd. All rights reserved.
the uncertainty associated with individual elements of $x$ and the uncertainty associated with individual elements of $y$.

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 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 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

$$
\nabla \left[ \frac{\alpha_{bg} K_{bg} k_{bg}}{\mu_b} (\nabla p_g + \rho_g g \nabla h) \right] + \alpha q_{aq} + \alpha q_{aq} = \alpha \frac{\partial (\phi p_b S_b)}{\partial t}
$$

(2.1)

Brine conservation

$$
\nabla \left[ \frac{\alpha_{bo} K_{bo} k_{bo}}{\mu_b} (\nabla p_b + \rho_b g \nabla h) \right] + \alpha q_{ab} + \alpha q_{ab} = \alpha \frac{\partial (\phi p_b S_b)}{\partial t}
$$

(2.2)

Saturation constraint

$$
S_g + S_b = 1
$$

(2.3)

Capillary pressure constraint

$$
p_c = p_g - p_b = f(S_b)
$$

(2.4)

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

Brine density

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

(2.5)

Formation porosity

$$
\phi = \phi_0 \exp[\beta_{f} (p_b - p_{bo})]
$$

(2.6)
where $g =$ acceleration due to gravity (m/s$^2$), $h =$ vertical distance from a reference location (m), $K_l =$ permeability tensor (m$^2$) for fluid $l = l_g$ (gas), $l_b$ (brine), $k_{rl} =$ relative permeability (dimensionless) to fluid $l$, $p_{Cg} =$ capillary pressure (Pa), $p_l =$ pressure of fluid $l$ (Pa), $q_{rl} =$ rate of production (or consumption, if negative) of fluid $l$ due to chemical reaction (kg/m$^3$/s), $q_{wl} =$ rate of injection (or removal, if negative) of fluid $l$ (kg/m$^3$/s), $S_l =$ saturation of fluid $l$ (dimensionless), $t =$ time (s), $a =$ geometry factor (m in present analysis), $p_{l_i} =$ density of fluid $l$ (kg/m$^3$), $\mu_i =$ viscosity of fluid $l$ (Pa s), $\phi_0 =$ reference (i.e. initial) porosity (dimensionless), $p_{b0} =$ reference (i.e. initial) brine pressure (Pa) (constant in Eq. (2.5) and spatially variable in Eq. (2.6)), $\rho_0 =$ reference (i.e. initial) brine density (kg/m$^3$), $\beta_i =$ pore compressibility (Pa$^{-1}$), $\beta_b =$ brine compressibility (Pa$^{-1}$), and $f$ is defined by the model for capillary pressure in use (see the right hand sides of Eqs. (10), (19) and (20) in Ref. [103]). The conservation equations are valid in one (i.e. $\nabla = [\partial / \partial x]$), two (i.e. $\nabla = [\partial / \partial x \partial / \partial y]$) and three (i.e. $\nabla = [\partial / \partial x \partial / \partial y \partial / \partial z]$) dimensions. In the present analysis, the preceding system of equations is used to model two-phase fluid flow in a two-dimensional region (Fig. 1), with the result that the spatial scale factor $a$ in Eqs. (2.1) and (2.2) has units of meters (m).

In general, the individual terms in Eqs. (2.1)–(2.6) are functions of location and time (e.g. $p_g(x, y, t)$, $q_g(x, y, t)$, $t$, $p_{l_i}$, $\mu_i$, $\phi_0$, $\beta_i$, $\beta_b$, $\rho_{b0}$).
identified 31 uncertain inputs to the BRAGFLO program and other variables as well. A full description of how the individual terms in these equations are defined is beyond the scope of this presentation and is available elsewhere. The system of partial differential equations in Eqs. (2.1)–(2.6) is too complex to permit a closed form solution. In the present analysis, these equations were solved with finite difference procedures implemented by the BRAGFLO program on the computational grid in Fig. 1.

The problem under consideration involves a single drilling intrusion (regions 1A–C in Fig. 1) that passes through a waste disposal panel in the WIPP repository (region 23 in Fig. 1) 1000 yr after closure of the repository and also penetrates a region of pressurized brine beneath the repository (region 30 in Fig. 1). In the terminology of the 1996 WIPP PA, this designated an E1 intrusion. Due to regulatory requirements placed on the WIPP, the modeled period extends from slightly before closure of the repository \( t = -5 \text{ yr} \), through closure of the repository \( t = 0 \text{ yr} \), and out to \( t = 10,000 \text{ yr} \).

To assess the effects of uncertainty, the 1996 WIPP PA identified 31 uncertain inputs to the BRAGFLO program required in the formulation of the model in Eqs. (2.1)–(2.6) for an E1 intrusion (Table 1). The exact manner in which these inputs were used in the definition of the coefficients in Eqs. (2.1)–(2.6) is described in Table 1 of Ref. [109].

The analysis was structured to require a single value for each of the variables in Table 1. However, the exact values to use for these variables were felt to be poorly known. Therefore, ranges of possible values for these variables were developed, and distributions were assigned to these ranges to characterize a degree of belief with respect to the location of the appropriate values to use in the 1996 WIPP PA. Thus, the distributions indicated in Table 1 are characterizing subjective (i.e. epistemic) uncertainty.

The 1996 WIPP PA used Latin hypercube sampling to investigate the effects of the uncertain variables in Table 1 on predictions of two-phase flow in the vicinity of the repository. In particular, three replicated LHSs of size 100 each were generated with use of the Iman and Conover restricted pairing technique to control correlations. The results in Fig. 2 indicate the result at 10,000 yr minus the result at 1000 yr, which is the end of the simulation period.

The solution of Eqs. (2.1)–(2.6) yields time-dependent results for each of the dependent variables in Table 2 (Fig. 2), and also for many additional dependent variables as previously indicated. The results in Fig. 2 are for the first of the three replicates (i.e. R1) used in the 1996 WIPP PA and illustrate both the spectrum of behaviors and the complexity of behavior that solutions to Eqs. (2.1)–(2.6) can display. In particular, a significant change in behavior occurs subsequent to the drilling intrusion at 10,000 yr.

### 3. Uncertainty analysis results

The time-dependent results in Fig. 2 display the uncertainty in solutions to Eqs. (2.1)–(2.6) that results from uncertainty in the 31 variables in Table 1. The goal of this presentation is to illustrate the robustness of such uncertainty representations with respect to the type and size of the sample in use. As previously indicated, results at 1000, 10,000–10000, and 10,000 yr will be used for illustration.

One way to compare uncertainty analysis results is to present cumulative distributions functions (CDFs)
Table 1

<table>
<thead>
<tr>
<th>Uncertain variables used as input to BRAFLO in the 1996 WIPP PA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANHBCEXP</strong>—Brooks–Corey pore distribution parameter for anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 0.491–0.842. Mean, median: 0.644, 0.644</td>
</tr>
<tr>
<td><strong>ANHBCGVP</strong>—Point factor for selection of relative permeability model for use in anhydrite. Distribution: discrete with 60% 0, 40% 1. Value of 0 implies Brooks–Corey model; value of 1 implies van Genuchten–Parker model</td>
</tr>
<tr>
<td><strong>ANHICOMP</strong>—Bulk compressibility of anhydrite (Pa⁻¹). Distribution: Student's with 3 degrees of freedom. Range: 1.09×10⁻¹¹–2.75×10⁻¹⁰ Pa⁻¹. Mean, median: 8.26×10⁻¹⁴, 8.26×10⁻¹⁴ Pa⁻¹. Correlation: −0.99 rank correlation with <strong>ANHPRM</strong></td>
</tr>
<tr>
<td><strong>ANHPRM</strong>—Logarithm of anhydrite permeability (m²). Distribution: Student's with 5 degrees of freedom. Range: 1.0×10⁻¹⁴–1×10⁻¹⁷ m². Mean, median: −18.9, −18.9. Correlation: −0.99 rank correlation with <strong>ANHICOMP</strong></td>
</tr>
<tr>
<td><strong>ANRBRSAT</strong>—Residual brine saturation in anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 7.85×10⁻³–1.74×10⁻¹. Mean, median: 8.36×10⁻², 8.36×10⁻²</td>
</tr>
<tr>
<td><strong>ANRGSSAT</strong>—Residual gas saturation in anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 1.39×10⁻²–1.79×10⁻¹. Mean, median: 7.71×10⁻², 7.71×10⁻²</td>
</tr>
<tr>
<td><strong>BHPRM</strong>—Logarithm of borehole permeability (m²). Distribution: uniform. Range: −14 to −11 (i.e. permeability range is 1×10⁻¹⁴–1×10⁻¹¹ m²). Mean, median: −12.5</td>
</tr>
<tr>
<td><strong>BPCOMP</strong>—Logarithm of bulk compressibility of brine pocket (Pa⁻¹). Distribution: triangular. Range: −11.3 to −8.00 (i.e. bulk compressibility range is 1×10⁻¹³–1×10⁻⁸ Pa⁻¹). Mean, mode: −9.80, −10.0. Correlation: −0.75 rank correlation with <strong>BPRPM</strong></td>
</tr>
<tr>
<td><strong>BPINPRS</strong>—Initial pressure in brine pocket (Pa). Distribution: triangular. Range: 1.11×10⁻¹–1.70×10³ Pa. Mean, median: 1.36×10³, 1.27×10³ Pa</td>
</tr>
<tr>
<td><strong>BPPRM</strong>—Logarithm of intrinsic brine pocket permeability (m²). Distribution: triangular. Range: −14.7 to −9.80 (i.e. permeability range is 1×10⁻¹⁴–1×10⁻⁸ m²). Mean, mode: −12.1, −11.8. Correlation: −0.75 rank correlation with <strong>BPCOMP</strong></td>
</tr>
<tr>
<td><strong>BPRPVOL</strong>—Point factor for selection of brine pocket volume. Distribution: discrete, with integer values 1, 2, ..., 32 equally likely</td>
</tr>
<tr>
<td><strong>HALCOMP</strong>—Bulk compressibility of halite (Pa⁻¹). Distribution: uniform. Range: 2.94×10⁻¹²–1.92×10⁻¹⁰ Pa⁻¹. Mean, median: 9.75×10⁻¹¹, 9.75×10⁻¹¹ Pa⁻¹. Correlation: −0.99 rank correlation with <strong>HALPRM</strong></td>
</tr>
<tr>
<td><strong>HALP</strong>—Halite porosity (dimensionless). Distribution: piecewise uniform. Range: 1.0×10⁻³–3×10⁻². Mean, median: 1.28×10⁻², 1.00×10⁻²</td>
</tr>
<tr>
<td><strong>HALPRM</strong>—Logarithm of halite permeability (m²). Distribution: uniform. Range: −24 to −21 (i.e. permeability range is 1×10⁻¹⁰–1×10⁻¹⁹ m²). Mean, median: −22.5, −22.5. Correlation: −0.99 rank correlation with <strong>HALCOMP</strong></td>
</tr>
<tr>
<td><strong>HALS</strong>—Initial brine pressure, without the repository being present, at a reference point located in the center of the combined shafts at the elevation of the midpoint of Marker Bed (MB) 139 (Pa). Distribution: uniform. Range: 1.10×10⁻⁸–1.38×10⁻⁷ Pa. Mean, median: 1.247×10⁻⁷, 1.247×10⁻⁷ Pa</td>
</tr>
<tr>
<td><strong>SHBCEXP</strong>—Brooks–Corey pore distribution parameter for shaft (dimensionless). Distribution: piecewise uniform. Range: 0.11–8.10. Mean, median: 2.52, 0.94</td>
</tr>
<tr>
<td><strong>SHPRMASC</strong>—Logarithm of permeability (m²) of asphalt component of shaft seal (m²). Distribution: triangular. Range: −21 to −18 (i.e. permeability range is 1×10⁻¹⁵–1×10⁻¹² m²). Mean, mode: −19.7, −20.0</td>
</tr>
<tr>
<td><strong>SHPRMCCL</strong>—Logarithm of permeability (m²) for clay components of shaft seal. Distribution: triangular. Range: −21 to −17.3 (i.e. permeability range is 1×10⁻¹⁵–1×10⁻¹² m²). Mean, mode: −18.9, −18.3</td>
</tr>
<tr>
<td><strong>SHPRMCDAS</strong>—Same as <strong>SHPRMASC</strong>, but for concrete component of shaft seal for 0–400 yr. Distribution: triangular. Range: −17.0 to −14.0 (i.e. permeability range is 1×10⁻¹⁰–1×10⁻¹⁷ m²). Mean, mode: −15.3, −15.0</td>
</tr>
<tr>
<td><strong>SHPRMCLZ</strong>—Logarithm of permeability (m²) of DRZ surrounding shaft seal. Distribution: triangular. Range: −17.0 to −14.0 (i.e. permeability range is 1×10⁻¹⁰–1×10⁻¹⁷ m²). Mean, mode: −15.3, −15.0</td>
</tr>
<tr>
<td><strong>SHPRMHAL</strong>—Point factor (dimensionless) used to select permeability in crushed salt component of shaft seal at different times. Distribution: uniform. Range: 0–1. Mean, mode: 0.5, 0.5. A distribution of permeability (m²) in the crushed salt component of the shaft seal is defined for each of the following time intervals: [0, 10 yr], [10, 25 yr], [25, 50 yr], [50, 100 yr], [100, 200 yr], [200, 10,000 yr]. <strong>SHPRMHAL</strong> is used to select a permeability value from the cumulative distribution function for permeability for each of the preceding time intervals with result that a rank correlation of 1 exists between the permeabilities used for the individual time intervals</td>
</tr>
<tr>
<td><strong>SHRBSAT</strong>—Residual brine saturation in shaft (dimensionless). Distribution: uniform. Range: 0–0.4. Mean, median: 0.2, 0.2</td>
</tr>
<tr>
<td><strong>SHRGSAT</strong>—Residual gas saturation in shaft (dimensionless). Distribution: uniform. Range: 0–0.4. Mean, median: 0.2, 0.2</td>
</tr>
<tr>
<td><strong>WAST</strong>—Increase in brine saturation of waste due to capillary forces (dimensionless). Distribution: uniform. Range: 0–1. Mean, median: 0.5, 0.5</td>
</tr>
<tr>
<td><strong>WFBET</strong>—Scale factor used in definition of stoichiometric coefficient for microbial gas generation (dimensionless). Distribution: uniform. Range: 0–1. Mean, median: 0.5, 0.5</td>
</tr>
<tr>
<td><strong>WGCRO</strong>—Corrosion rate for steel under inundated conditions in the absence of CO₂ (m/s). Distribution: uniform. Range: 0–1.58×10⁻¹⁴ m/s. Mean, median: 7.94×10⁻¹⁵, 7.94×10⁻¹⁵ m/s</td>
</tr>
<tr>
<td><strong>WGXMICH</strong>—Microbial degradation rate for cellulose under humid conditions (mol/kg s). Distribution: uniform. Range: 0–1.27×10⁻⁹ mol/kg s. Mean, median: 6.34×10⁻¹⁰, 6.34×10⁻¹⁰ mol/kg s</td>
</tr>
<tr>
<td><strong>WGXMICL</strong>—Microbial degradation rate for cellulose under inundated conditions (mol/kg s). Distribution: uniform. Range: 3.17×10⁻¹⁰–5.91×10⁻⁹ mol/kg s. Mean, median: 4.92×10⁻⁹, 4.92×10⁻⁹ mol/kg s</td>
</tr>
</tbody>
</table>

constructed from the individual replicated random and LHSs. For each of the 12 analysis outcomes under consideration (i.e. three values for each of BRNREPTC, GAS_MOLE, REP_SATB, and WAS_PRES), three CDFs resulting from three random samples of size 100 and also three CDFs resulting from three LHSs of size 100 are available. These CDFs are generally very similar, both when compared within sampling procedures (i.e. random or Latin
hypercube) and when compared across sampling procedures. The greatest variability occurred for WAS_PRES (Fig. 3), with Latin hypercube sampling producing noticeably more stable CDFs than random sampling. For the other results, visual inspection indicated little difference between the CDFs obtained with random and Latin hypercube sampling.

Plots of CDFs are too bulky to permit their presentation for all 12 analysis outcomes under consideration. In particular, 12 analysis outcomes, two sampling procedures, and three replicates result in 72 (i.e. $12 \times 2 \times 3$) CDFs. However, box plots provide a compact representation of the information contained in the 72 CDFs under consideration that can be presented in a single figure (Fig. 4). Further, the flattened structure of box plots facilitates the comparison of CDFs both within and across sampling procedures.

As inspection of Fig. 4 shows, the distributions of results obtained with the two sampling techniques are quite stable, both within and across the two techniques. Visual inspection suggests that the results obtained with Latin hypercube sampling are slightly more stable than those obtained with random sampling, but the difference is not very large. If desired, the $t$-test can be used to determine confidence intervals for the estimated means for the two sampling procedures [109,116].

In typical uncertainty analyses dealing with subjective (i.e. state of knowledge or epistemic) uncertainty, the primary goal is to obtain a general assessment of

Table 2
Dependent variables arising from the solution of Eqs. (2.1)–(2.6) for an E1 intrusion at 1000 yr selected for consideration

| BRNREPTC | total brine flow (m$^3$) into repository, which, in the context of Fig. 1, corresponds to regions 23 and 24, the part of region 1 (i.e. the borehole) between the two parts of region 23, and the part of region 25 (i.e. the panel closure) between regions 23 and 24 |
| GAS_MOLE | total gas generation (moles) in repository due to corrosion and microbial degradation of cellulose |
| REP_SATB | brine saturation (dimensionless) in part of repository not penetrated by a drilling intrusion, which corresponds to region 24 in Fig. 1 |
| WAS_PRES | pressure (Pa) in part of the repository penetrated by a drilling intrusion, which corresponds to region 23 in Fig. 1. A capillary pressure of zero is assumed within regions 23 and 24 in Fig. 1, with the result that gas and brine pressure are equal within these regions |

Fig. 2. Time-dependent solutions to Eqs. (2.1)–(2.6) obtained for the first replicated LHS (i.e. replicate R1) of size 100 used in the 1996 WIPP PA for BRNREPTC, GAS_MOLE, REP_SATB and WAS_PRES.
the uncertainty in analysis outcomes of interest. In particular, there is neither need nor justification for the estimation of very small or very large quantiles of distributions characterizing subjective uncertainty. This is in contrast to risk studies where much emphasis is placed on the determination of the effects of stochastic (i.e. random or aleatory) uncertainty due to the need to determine the likelihood of rare, high consequence events [110–112,117].

The present analysis is concerned with the effects of subjective uncertainty. In this context, the use of any of the individual random or LHSs would have led to operationally similar assessments of the uncertainty in analysis outcomes. The word operational is used because the individual assessments of uncertainty are sufficiently similar that it is difficult to envision that the individual assessments would have led to different courses of action being chosen (e.g. whether or not to fund additional research to reduce the indicated state of uncertainty).

4. Stepwise results

A sensitivity analysis based on stepwise regression analysis with rank-transformed data [118] was carried...
out for the replicated samples summarized in Fig. 4 (Tables 3–6). This analysis required \( \alpha \)-values of 0.02 and 0.05 for variables to enter and to be retained in a given analysis, respectively, and was carried out with the STEPWISE program [119]. The summary tables (Tables 3–6) present results for both the individual replicates and for the three replicates of a given type (i.e. random or Latin hypercube) pooled. The standardized rank regression coefficient (SRRC) is used as a measure of variable importance.

Inspection of Tables 3–6 shows that the results obtained with the individual replicates are very consistent. In particular, the results obtained for a given dependent variable for the three replicated random samples are very similar to each other and also to the results obtained for the three replicated LHSs. This similarity includes the order in which variables are selected in the stepwise process, the SRRCs associated with individual variables, and the \( R^2 \) value of the final regression model.

The results obtained with the pooled replicates tend to include a few more variables than the results obtained with the individual replicates. However, the effects associated with the addition of these variables are
Table 3
Sensitivity analysis results based on stepwise rank regression for replicated random and Latin hypercube samples of size 100 for \textit{BRNREPTC} at 1000, 10,000–1000, and 10,000 yr

<table>
<thead>
<tr>
<th>Step</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicates 1, 2, 3 pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variableb</td>
<td>SRRCc</td>
<td>$R^2$d</td>
<td>Variable</td>
</tr>
<tr>
<td>Random: $BRNREPTC$, 1000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>HALPOR</td>
<td>0.99 0.97</td>
<td>HALPOR</td>
<td>0.97 0.94</td>
</tr>
<tr>
<td>2</td>
<td>WMICDFLG 0.01 0.98</td>
<td>ANHPRM 0.13 0.95</td>
<td>WMICDFLG 0.07 0.96</td>
<td>ANHPRM 0.10 0.97</td>
</tr>
<tr>
<td>3</td>
<td>ANHPRM 0.07 0.99</td>
<td>WMICDFLG 0.07 0.96</td>
<td>ANHPRM 0.07 0.96</td>
<td>BPPRM 0.05 0.97</td>
</tr>
<tr>
<td>4</td>
<td>SALPRES 0.04 0.99</td>
<td>HALPRM 0.07 0.96</td>
<td>SALPRES 0.05 0.97</td>
<td>HALPRM 0.07 0.97</td>
</tr>
<tr>
<td>5</td>
<td>WRBNSAT 0.03 0.99</td>
<td>WGRCOR 0.06 0.97</td>
<td>HALPRM 0.04 0.98</td>
<td>WGRCOR 0.02 0.99</td>
</tr>
<tr>
<td>6</td>
<td>HALPRM 0.03 0.99</td>
<td>SHBCEXP 0.05 0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>WASTWICK 0.03 0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>SHPRMORZ 0.02 0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LHS: $BRNREPTC$, 1000 yr

<table>
<thead>
<tr>
<th>Step</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicates 1, 2, 3 pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variableb</td>
<td>SRRCc</td>
<td>$R^2$d</td>
<td>Variable</td>
</tr>
<tr>
<td>Random: $BRNREPTC$, 10,000–1000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>BHPRM 0.70 0.51</td>
<td>BHPRM 0.68 0.45</td>
<td>BHPRM 0.59 0.34</td>
<td>BHPRM 0.66 0.44</td>
</tr>
<tr>
<td>2</td>
<td>BPCOMP 0.37 0.64</td>
<td>BPCOMP 0.40 0.63</td>
<td>BPCOMP 0.37 0.49</td>
<td>BPCOMP 0.40 0.60</td>
</tr>
<tr>
<td>3</td>
<td>WMICDFLG 0.26 0.70</td>
<td>ANHPRM 0.21 0.68</td>
<td>ANHPRM 0.29 0.57</td>
<td>WMICDFLG 0.22 0.65</td>
</tr>
<tr>
<td>4</td>
<td>BPINTPRS 0.14 0.72</td>
<td>WMICDFLG 0.21 0.72</td>
<td>WMICDFLG 0.23 0.63</td>
<td>BPINTPRS 0.11 0.70</td>
</tr>
<tr>
<td>5</td>
<td>HALPOR 0.13 0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LHS: $BRNREPTC$, 10,000–1000 yr

<table>
<thead>
<tr>
<th>Step</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicates 1, 2, 3 pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variableb</td>
<td>SRRCc</td>
<td>$R^2$d</td>
<td>Variable</td>
</tr>
<tr>
<td>Random: $BRNREPTC$, 10,000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>BHPRM 0.67 0.43</td>
<td>BHPRM 0.66 0.43</td>
<td>BHPRM 0.64 0.44</td>
<td>BHPRM 0.66 0.43</td>
</tr>
<tr>
<td>2</td>
<td>BPCOMP 0.43 0.59</td>
<td>BPCOMP 0.46 0.66</td>
<td>BPCOMP 0.36 0.58</td>
<td>BPCOMP 0.42 0.60</td>
</tr>
<tr>
<td>3</td>
<td>WMICDFLG 0.25 0.64</td>
<td>WMICDFLG 0.24 0.71</td>
<td>WMICDFLG 0.29 0.67</td>
<td>WMICDFLG 0.27 0.67</td>
</tr>
<tr>
<td>4</td>
<td>ANHPRM 0.16 0.67</td>
<td>BPVOL 0.14 0.74</td>
<td>ANHPRM 0.17 0.70</td>
<td>BPVOL 0.16 0.70</td>
</tr>
<tr>
<td>5</td>
<td>BPVOL 0.16 0.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>WGRCOR 0.14 0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|         |         |         |         |         |         |         |         |         |         |             |             |         |         |

a Steps in stepwise rank regression analysis with $a$-values of 0.02 and 0.05 required for a variable to enter and to be retained in an analysis, respectively.

b Variables listed in order of selection in regression analysis with ANHCOMP and HALCOMP excluded from entry into regression model because of $-0.99$ rank correlation with the pairs (ANHPRM, ANHCOMP) and (HALPRM, HALCOMP).
Table 4
Sensitivity analysis results based on stepwise rank regression for replicated random and Latin hypercube samples of size 100 for GAS_MOLE at 1000, 10,000–1000, and 10,000 yr

<table>
<thead>
<tr>
<th>Stepa</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicas 1, 2, 3 pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variable</td>
<td>SRCC</td>
<td>R^2</td>
<td>Variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Random: GAS_MOLE, 1000 yr</td>
</tr>
<tr>
<td>1</td>
<td>WMICDFLG</td>
<td>0.88</td>
<td>0.74</td>
<td>WMICDFLG</td>
</tr>
<tr>
<td>2</td>
<td>WGRCOR</td>
<td>0.37</td>
<td>0.88</td>
<td>WGRCOR</td>
</tr>
<tr>
<td>3</td>
<td>WASTWICK</td>
<td>0.21</td>
<td>0.92</td>
<td>WASTWICK</td>
</tr>
<tr>
<td>4</td>
<td>HALPOR</td>
<td>0.11</td>
<td>0.93</td>
<td>HALPOR</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>WGRMICI</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>SHPRMDRZ</td>
</tr>
<tr>
<td>LHS:</td>
<td></td>
<td></td>
<td></td>
<td>Random: GAS_MOLE, 10,000–1000 yr</td>
</tr>
<tr>
<td>ANHPRM</td>
<td>0.53</td>
<td>0.31</td>
<td>HALPOR</td>
<td>0.52</td>
</tr>
<tr>
<td>WRBRNSAT</td>
<td>0.43</td>
<td>0.51</td>
<td>WGRCOR</td>
<td>0.43</td>
</tr>
<tr>
<td>BPVOL</td>
<td>0.24</td>
<td>0.63</td>
<td>BHPRM</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHS:</td>
<td></td>
<td></td>
<td></td>
<td>Random: GAS_MOLE, 10,000 yr</td>
</tr>
<tr>
<td>ANHPRM</td>
<td>0.52</td>
<td>0.27</td>
<td>HALPOR</td>
<td>0.50</td>
</tr>
<tr>
<td>BPVOL</td>
<td>0.33</td>
<td>0.40</td>
<td>BHPRM</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>0.50</td>
<td>WGRCOR</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.59</td>
<td>SHRGSSAT</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>0.13</td>
<td>0.78</td>
<td>BPVOL</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHS:</td>
<td></td>
<td></td>
<td></td>
<td>LHS: GAS_MOLE, 10,000 yr</td>
</tr>
<tr>
<td>BPVOL</td>
<td>0.53</td>
<td>0.30</td>
<td>HALPOR</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>0.43</td>
<td>0.48</td>
<td>WGRCOR</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>0.63</td>
<td>WMICDFLG</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>0.26</td>
<td>0.70</td>
<td>BHPRM</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>0.74</td>
<td>SHRGSSAT</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Steps in stepwise rank regression analysis with α-values of 0.02 and 0.05 required for a variable to enter and to be retained in an analysis, respectively.

b Variables listed in order of selection in regression analysis with ANHCOMP and HALCOMP excluded from entry into regression model because of −0.99 rank correlation with the pairs (ANHPRM, ANHCOMP) and (HALPRM, HALCOMP).

c Standardized rank regression coefficients (SRRCs) in final regression model.

d Cumulative R^2 value with entry of each variable into regression model.
small, and the $R^2$ values for the pooled analyses are not much larger than the $R^2$ values obtained for the individual replicates. Further, the results obtained with the pooled random and LHSs are very similar.

The comparisons of random and Latin hypercube sampling in this section are based on nonquantitative impressions gained from inspecting the results in Tables 3–6. Section 5 introduces quantitative procedures.
for comparing the results in Tables 3–6 obtained with random and Latin hypercube sampling.

5. Coefficients of concordance

Inspection of the results in Tables 3–6 suggests that the individual replicates are producing similar results. Kendall’s coefficient of concordance (KCC) provides a way to formally assess this similarity (p. 305, Ref. [120]). This coefficient is based on the consideration of arrays of the form

\[
\begin{align*}
R_1 & \quad R_2 & \cdots & \quad R_{nR} \\
O_{11} & \quad r(O_{11}) & \cdots & \quad r(O_{11},nR) \\
O_{12} & \quad r(O_{12}) & \cdots & \quad r(O_{12},nR) \\
\vdots & \quad \vdots & \cdots & \quad \vdots \\
O_{nX} & \quad r(O_{nX,1}) & \cdots & \quad r(O_{nX,nR})
\end{align*}
\]

\( (5.1) \)
where \(x_1, x_2, \ldots, x_{nX}\) are the variables under consideration (i.e. \(nX = 29\) with the exclusion of ANHCOMP and HALCOMP from the analysis; see Footnote b, Table 3), \(R_1, R_2, \ldots, R_{nR}\) designate the replicates (i.e. \(nR = 3\)), \(O_{ij}\) is the outcome (i.e. sensitivity measure) for variable \(x_i\) and replicate \(R_j\), and \(r(O_{ij})\), \(i = 1, 2, \ldots, nX\), are the ranks assigned to the outcomes associated with replicate \(R_j\). In the assigning of ranks, (i) a rank of 1 is assigned to the outcome \(O_{ij}\) with the largest value for \(j\) (or \(O_{ij}\), (ii) a rank of 2 is assigned the outcome \(O_{ij}\) with the second largest value for \(O_{ij}\), and so on, and (iii) averaged ranks are assigned to equal values of \(O_{ij}\). This is the reverse of the procedure used to assign ranks for use in rank regression.

Kendall’s coefficient of concordance (KCC) is defined by

\[
W = \frac{12}{nR^2 nX(nX + 1)(nX - 1)} \times \sum_{i=1}^{nX} \left[ \sum_{j=1}^{nR} r(O_{ij}) - \frac{nR(nX + 1)}{2} \right]^2
\]

(5.2)

(see Eq. (23), p. 305, Ref. [120]). The coefficient \(W\) is related to the average \(\rho_{as}\) of the \(nR(nR - 1)/2 \) correlations (i.e. rank or Spearman correlations due to the indicated rank transformation) between the columns in Eq. (5.1) by

\[
W = [(nR - 1)\rho_{as} + 1]/nR.
\]

(5.3)

The preceding equality follows from a rewriting of Eq. (29), p. 307, of Ref. [120] in the form \(\rho_{as} = (nR W - 1)/(nR - 1)\) with \(\rho_{as}\) corresponding to \(\rho_{as}\) in the indicated equation from Ref. [120]. Under repeated random assignment of the integers in the columns of Eq. (5.1),

\[
T = nR(nX - 1)W
\]

(5.4)

approximately follows a \(\chi^2\)-distribution with \(nX - 1\) degrees of freedom (see Eq. (24), p. 304, Ref. [120]; Iman and Davenport [121] recommend using an \(F\)-distribution with \(k_1 = nX - 1\) and \(k_2 = (nR - 1)(nX - 1)\) degrees of freedom rather than the indicated \(\chi^2\)-distribution).

Kendall’s coefficient of concordance (KCC) places equal weight on agreement of rankings for both important variables (i.e. variables with ranks close to 1) and unimportant variables (i.e. variables with ranks close to \(nX\)). In practice, only a few variables typically have significant effects on a given model prediction, with the remaining variables having no discernable effects and rankings that are either unassigned or meaningless. The stopping of the regressions in Tables 3–6 at an \(\alpha\)-value of 0.02 is an example of only the important variables being assigned ranks, with the remaining variables (i.e. the variables not selected in the stepwise regression) assigned no rank. Alternatively, the regression could be forced to include all variables, which would result in the assignment of ranks to all variables, but with most of these ranks having no meaning. As a result, KCC can be a poor indicator of agreement when only a few variables have significant effects.

As an alternative to KCC, Iman and Conover [122] proposed the top down coefficient of concordance (TDCC) as a measure of agreement between multiple rankings for use when it is desired to emphasize agreement between rankings assigned to important variables and to deemphasize disagreement between rankings assigned to less important/unimportant variables. For the TDCC, the ranks \(r(O_{ij})\) in Eq. (5.1) are replaced by the corresponding Savage scores \(ss(O_{ij})\), where

\[
ss(O_{ij}) = \sum_{i=1}^{nX} 1/i
\]

and average Savage scores are assigned in the event of ties. The result is an array of the form

\[
\begin{array}{cccc}
R_1 & R_2 & \ldots & R_{nR} \\
x_1 & ss(O_{11}) & ss(O_{12}) & \ldots & ss(O_{1,nR}) \\
x_2 & ss(O_{21}) & ss(O_{22}) & \ldots & ss(O_{2,nR}) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
x_{nX} & ss(O_{nX,1}) & ss(O_{nX,2}) & \ldots & ss(O_{nX,nR}) \\
\end{array}
\]

(5.6)

which has the same form as the array in Eq. (5.1) except that the ranks \(r(O_{ij})\) have been replaced by the corresponding Savage scores \(ss(O_{ij})\).

The TDCC is defined by

\[
C_T = \left\{ \left[ \sum_{i=1}^{nX} \left[ \sum_{j=1}^{nR} ss(O_{ij}) \right]^2 \right] - nR^2 nX \right\} / \left\{ nR^2 \left( nX - \sum_{i=1}^{nX} 1/i \right) \right\}
\]

(5.7)

and is equivalent to KCC calculated with Savage scores rather than ranks. In particular,

\[
C_T = [(nR - 1)\rho_{as} + 1]/nR
\]

(5.8)

where \(\rho_{as}\) is the average of the \(nR(nR - 1)/2\) correlations (i.e. ordinary or Pearson correlations involving Savage scores) between the columns in Eq. (5.6). Under repeated random assignment of the integers in the columns of Eq. (5.1),

\[
T = nR(nX - 1)C_T
\]

(5.9)

approximately follows a \(\chi^2\)-distribution with \(nX - 1\) degrees of freedom (see Sect. 4, Ref. [122]).

Sensitivity analysis results obtained with the random and LHSs were compared with both KCC and the TDCC (Table 7). For this comparison, the associated rank regression models were forced to include all 29 variables under consideration (i.e. all variables in Table 1 except ANHCOMP and HALCOMP as indicated in Footnote b of Table 3), and the ranking was done on the basis of the absolute values of the SRRCs for the regression model containing all variables. An alternative would be to rank the variables included in the stepwise regressions in Tables 3–6.
Table 7  
Consistency of variable rankings with stepwise rank regression for three replicated random samples of size 100 and three replicated Latin hypercube samples of size 100

<table>
<thead>
<tr>
<th>Variablea</th>
<th>Random sampling</th>
<th>Latin hypercube sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KCCb  p-valuec</td>
<td>TDCCd  p-valuee</td>
</tr>
<tr>
<td>BRNREPTC1</td>
<td>0.58 8.2×10^{-3}</td>
<td>0.80 5.2×10^{-5}</td>
</tr>
<tr>
<td>BRNREPTC2</td>
<td>0.55 1.7×10^{-2}</td>
<td>0.79 6.4×10^{-5}</td>
</tr>
<tr>
<td>BRNREPTC3</td>
<td>0.59 7.4×10^{-3}</td>
<td>0.83 2.0×10^{-5}</td>
</tr>
<tr>
<td>GAS_MOLE1</td>
<td>0.55 1.6×10^{-2}</td>
<td>0.81 3.4×10^{-5}</td>
</tr>
<tr>
<td>GAS_MOLE2</td>
<td>0.53 2.5×10^{-2}</td>
<td>0.76 1.4×10^{-4}</td>
</tr>
<tr>
<td>GAS_MOLE3</td>
<td>0.58 8.6×10^{-3}</td>
<td>0.84 1.5×10^{-5}</td>
</tr>
<tr>
<td>REP_SATB1</td>
<td>0.61 5.2×10^{-3}</td>
<td>0.83 2.0×10^{-5}</td>
</tr>
<tr>
<td>REP_SATB2</td>
<td>0.61 4.4×10^{-3}</td>
<td>0.85 1.1×10^{-5}</td>
</tr>
<tr>
<td>REP_SATB3</td>
<td>0.73 2.5×10^{-4}</td>
<td>0.88 4.6×10^{-6}</td>
</tr>
<tr>
<td>WAS_PRES1</td>
<td>0.52 3.2×10^{-2}</td>
<td>0.78 7.3×10^{-5}</td>
</tr>
<tr>
<td>WAS_PRES2</td>
<td>0.55 1.8×10^{-2}</td>
<td>0.80 5.0×10^{-5}</td>
</tr>
<tr>
<td>WAS_PRES3</td>
<td>0.46 9.3×10^{-2}</td>
<td>0.58 9.5×10^{-3}</td>
</tr>
</tbody>
</table>

a Dependent variables (Table 2) with 1–3 designating results at 1000, 10,000–1000, and 10,000 yr, respectively.
b Kendall’s coefficient of concordance (KCC).
c p-value for KCC.
d Top down coefficient of concordance (TDCC).
e p-value for TDCC.

and then to assign tied ranks to the variables not selected in a particular regression. This approach was not used.

The TDCC values in Table 7 provide more insightful indications of analysis consistency than the KCC values. In particular, the numerical values for the TDCC are larger than those for KCC, and more importantly, the corresponding p-values are more significant (i.e. the TDCC is producing smaller p-values than KCC). For example, \textit{BRNREPTC1} for random sampling has a KCC of 0.58 with a p-value of 8.2×10^{-3} and a TDCC of 0.80 with a p-value of 5.2×10^{-5}; similar comparisons also exist for the other analyses in Table 7. This behavior results because the TDCC emphasizes agreement on important variables and deemphasizes disagreement on unimportant variables. In contrast, KCC tends to weight agreement/disagreement on the rankings assigned to all variables equally.

As indicated by the TDCC, random and Latin hypercube sampling show similar levels of consistency in rankings of variable importance for the three replicated samples. In particular, both approaches have similar TDCC values for a given variable, and neither approach has TDCC values across all variables that are consistently higher than the values for the other approach. Thus, at least in this example, neither sampling approach appears to have an advantage in the consistent identification of important variables with a sample size of 100.

6. Sensitivity analysis with the TDCC

Replicated samples and the TDCC provide the basis for a sensitivity analysis procedure to identify important sets of variables that does not depend on direct testing of the statistical significance of sensitivity measures (e.g. the significance of the coefficients in a stepwise regression model as defined by an α-value for entry into the model). Rather, important variables are identified by the similarity of outcomes in analyses performed for the individual replicated samples.

The procedure operates in the following manner: (i) The sensitivity analysis technique in use (e.g. stepwise regression analysis) is applied to each replicate to rank variable importance. (ii) The TDCC is applied to the variable rankings obtained with each replicate to determine if there is a significant agreement between the replicates (e.g. as defined by a specified p-value for the TDCC). (iii) If there is significant agreement, the top ranked variable (i.e. rank 1) for each replicate is removed from consideration for all replicates; this results in the removal of one variable if all replicates assign the same variable a rank of 1 and more than one variable if different variables are assigned a rank of 1 in different replicates. (iv) A new sensitivity analysis is then performed for each replicate with the remaining variables, the remaining variables are reranked for each replicate, and Steps (ii) and (iii) are repeated with the reduced set of variables. (v) The process is continued until the deleted variable result in the analysis reaches a point at which the TDCC indicates that there is no significant agreement between the variable rankings obtained with the individual replicates. (vi) At this point, the analysis ends, and the significant set of variables are those deleted before the TDCC indicated no significant agreement between the variable rankings obtained with the individual replicates.

This procedure is illustrated for rank regression analysis with the three random samples for \textit{BRNREPTC1} at 1000 yr (i.e. \textit{BRNREPTC1}). The individual regression analyses all
Table 8
Sensitivity analysis results based on SRRCs for three replicated random samples (RS1 RS2, RS3) and three replicated Latin hypercube samples (LS1, LS2, LS3) of Size 100 for BRNREPTC at 1000 yr

<table>
<thead>
<tr>
<th>Variable</th>
<th>RS1 (10^{-4})</th>
<th>RS2 (10^{-4})</th>
<th>RS3 (10^{-4})</th>
<th>LS1 (10^{-4})</th>
<th>LS2 (10^{-4})</th>
<th>LS3 (10^{-4})</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMICDFLG</td>
<td>-9.72</td>
<td>-6.92</td>
<td>-1.13</td>
<td>-9.53</td>
<td>-8.37</td>
<td>-1.03</td>
</tr>
<tr>
<td>ANHPRM</td>
<td>6.49</td>
<td>1.33</td>
<td>9.84</td>
<td>7.62</td>
<td>1.04</td>
<td>7.20</td>
</tr>
<tr>
<td>SALPRES</td>
<td>-4.00</td>
<td>-2.70</td>
<td>-1.41</td>
<td>-1.90</td>
<td>-3.17</td>
<td>-4.57</td>
</tr>
<tr>
<td>HALPRM</td>
<td>3.53</td>
<td>7.67</td>
<td>4.05</td>
<td>2.92</td>
<td>4.99</td>
<td>6.83</td>
</tr>
<tr>
<td>WRBRNSAT</td>
<td>-3.08</td>
<td>-1.79</td>
<td>9.13</td>
<td>-1.59</td>
<td>-5.15</td>
<td>-5.07</td>
</tr>
<tr>
<td>WASTWICK</td>
<td>-2.82</td>
<td>-2.27</td>
<td>-4.47</td>
<td>-3.78</td>
<td>-2.74</td>
<td>-10.14</td>
</tr>
<tr>
<td>BPCOMP</td>
<td>2.61</td>
<td>3.56</td>
<td>8.05</td>
<td>-1.04</td>
<td>9.96</td>
<td>-1.29</td>
</tr>
<tr>
<td>SHPRMRRZ</td>
<td>2.29</td>
<td>-1.37</td>
<td>2.58</td>
<td>1.05</td>
<td>-2.45</td>
<td>9.97</td>
</tr>
<tr>
<td>BPPRM</td>
<td>-1.85</td>
<td>1.27</td>
<td>5.08</td>
<td>-2.46</td>
<td>2.51</td>
<td>-1.20</td>
</tr>
<tr>
<td>WFBETCEL</td>
<td>-1.60</td>
<td>1.89</td>
<td>1.46</td>
<td>-1.39</td>
<td>-5.76</td>
<td>-2.70</td>
</tr>
<tr>
<td>SHPRMSP</td>
<td>-1.30</td>
<td>1.42</td>
<td>-2.36</td>
<td>-5.85</td>
<td>2.83</td>
<td>-9.23</td>
</tr>
<tr>
<td>BPINTPRS</td>
<td>1.27</td>
<td>2.07</td>
<td>-1.27</td>
<td>-7.75</td>
<td>1.24</td>
<td>2.54</td>
</tr>
<tr>
<td>WGRMIC</td>
<td>-1.14</td>
<td>7.36</td>
<td>1.32</td>
<td>-1.34</td>
<td>-1.64</td>
<td>-1.84</td>
</tr>
<tr>
<td>SHRGSAT</td>
<td>-1.13</td>
<td>-1.64</td>
<td>5.93</td>
<td>-6.60</td>
<td>1.89</td>
<td>-1.31</td>
</tr>
<tr>
<td>SHPRMHAL</td>
<td>1.11</td>
<td>-5.05</td>
<td>-3.49</td>
<td>-7.12</td>
<td>4.97</td>
<td>1.10</td>
</tr>
<tr>
<td>WGRCOR</td>
<td>-1.03</td>
<td>-6.13</td>
<td>-2.15</td>
<td>-2.74</td>
<td>-3.95</td>
<td>-2.45</td>
</tr>
<tr>
<td>SHBICEP</td>
<td>9.95</td>
<td>4.62</td>
<td>1.35</td>
<td>2.21</td>
<td>-6.49</td>
<td>1.45</td>
</tr>
<tr>
<td>SHPRMCN</td>
<td>-9.62</td>
<td>4.86</td>
<td>-1.09</td>
<td>1.21</td>
<td>-1.17</td>
<td>1.45</td>
</tr>
<tr>
<td>BHPRM</td>
<td>7.42</td>
<td>1.37</td>
<td>1.77</td>
<td>1.67</td>
<td>2.03</td>
<td>1.11</td>
</tr>
<tr>
<td>SHPRMCLEY</td>
<td>-6.59</td>
<td>8.78</td>
<td>-1.44</td>
<td>-4.35</td>
<td>1.19</td>
<td>-2.11</td>
</tr>
<tr>
<td>ANRGSAT</td>
<td>-6.11</td>
<td>-1.75</td>
<td>7.94</td>
<td>8.88</td>
<td>1.77</td>
<td>-13.12</td>
</tr>
<tr>
<td>ANBCCVGP</td>
<td>-4.40</td>
<td>-2.40</td>
<td>-1.80</td>
<td>1.29</td>
<td>2.74</td>
<td>-18.88</td>
</tr>
<tr>
<td>WGRCIC</td>
<td>-3.33</td>
<td>-1.48</td>
<td>-3.35</td>
<td>-8.81</td>
<td>-1.44</td>
<td>-1.60</td>
</tr>
<tr>
<td>SHRBRBSAT</td>
<td>2.75</td>
<td>-2.23</td>
<td>2.87</td>
<td>-4.44</td>
<td>3.09</td>
<td>1.73</td>
</tr>
<tr>
<td>ANRBRBSAT</td>
<td>1.95</td>
<td>-4.05</td>
<td>1.64</td>
<td>6.56</td>
<td>-1.54</td>
<td>4.31</td>
</tr>
<tr>
<td>BPVOL</td>
<td>-1.58</td>
<td>6.54</td>
<td>2.12</td>
<td>2.37</td>
<td>-2.18</td>
<td>-10.28</td>
</tr>
<tr>
<td>ANHBCEXP</td>
<td>-1.30</td>
<td>4.32</td>
<td>-2.88</td>
<td>-2.15</td>
<td>-1.30</td>
<td>1.69</td>
</tr>
<tr>
<td>WRGSSAT</td>
<td>-1.19</td>
<td>1.32</td>
<td>-5.33</td>
<td>1.43</td>
<td>2.45</td>
<td>1.29</td>
</tr>
</tbody>
</table>

\* Variables included in regression model (i.e. all variables in Table 1, except for ANHCOMP and HALCOMP which are not included because of -0.99 rank correlation with the pairs (ANHPRM, ANHCOMP) and (HALPRM, HALCOMP)).

\* SRRC in model containing all variables for indicated sample.

\* Variable rank based on absolute value of SRRC for indicated sample.

rank HALPOR as the most important variable (see left three columns of results in Table 8) and have a TDCC of 0.80 with a \(p\)-value of \(5.2 \times 10^{-5}\) (Table 9). As a result, HALPOR is removed from consideration, which reduces the number of independent variables from 29 to 28. A new rank regression is then performed for each replicate with the remaining 28 variables, and the variables are reranked (i.e. from 1 to 28) on the basis of their SRRCs, with ANHPRM having a rank of 1 in one replicate and WMICDFLG having a rank of 1 in two replicates. For this new ranking (i.e. without HALPOR), the TDCC has a value of 0.71 with a \(p\)-value of \(5.0 \times 10^{-4}\) (Table 9). As this is considered to be significant agreement, ANHPRM and WMICDFLG are dropped; the remaining 26 variables are reranked; new regressions are performed for each replicate; and a resultant TDCC of 0.46 with a \(p\)-value of \(9.8 \times 10^{-2}\) is calculated (Table 9). If a \(p\)-value of \(9.8 \times 10^{-2}\) is considered to be insignificant, then the analysis ends, and the set of significant variables is taken to be \{HALPOR, ANHPRM, WMICDFLG\}.

If a \(p\)-value of \(9.8 \times 10^{-2}\) is considered to be significant (e.g. if the analysis was using 0.1 as the \(p\)-value above which the analysis stopped), then the analysis would continue with the top ranked variables in the individual replicates being dropped (i.e. SALPRES, HALPRM, BPPRM) and the TDCC recalculated for the remaining 23 variables. This process would continue until either an insignificant value for the TDCC was obtained or all variables were dropped, with the latter being an unlikely outcome.

For perspective, the process is also illustrated for BRNREPTC, 10,000–1000 yr (i.e. BRNREPTC2) and BRNREPTC at 10,000 yr (i.e. BRNREPTC3) in Table 9. If a \(p\)-value of 0.02 was being used to determine significance for the TDCC, then the analyses for BRNREPTC2 and BRNREPTC3 would identify \{BHPRM, BPCOMP, WMICDFLG, ANHPRM\} and \{BHPRM, BPCOMP, HALPOR, ANHPRM, WMICDFLG\}, respectively, as the important sets of variables.

Sensitivity analysis results based on the TDCC as described in this section are presented in Table 10 for the 18 dependent variables considered in Tables 3–6. There is little difference between the sets of important variables identified with random and Latin hypercube sampling.

The sensitivity analysis procedure presented in this section is analogous to forward stepwise regression analysis in the sense that the procedure operates by finding the most
Table 9  
Sensitivity analysis with the TDCC for three replicated random samples of size 100 for BRNREPTC at 1000, 10,000–1000, and 10,000 yr

<table>
<thead>
<tr>
<th>Step</th>
<th>TDCC</th>
<th>p-value</th>
<th>Variable(s) removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random: BRNREPTC, 1000 yr</td>
<td>1</td>
<td>0.80</td>
<td>5.2 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.71</td>
<td>5.0 × 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.46</td>
<td>9.8 × 10^{-2}</td>
</tr>
<tr>
<td>Random: BRNREPTC, 10,000–1000 yr</td>
<td>1</td>
<td>0.79</td>
<td>6.4 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.72</td>
<td>4.3 × 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.60</td>
<td>8.1 × 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.30</td>
<td>5.9 × 10^{-1}</td>
</tr>
<tr>
<td>Random: BRNREPTC, 10,000 yr</td>
<td>1</td>
<td>0.83</td>
<td>2.0 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.77</td>
<td>1.4 × 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.64</td>
<td>3.9 × 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.28</td>
<td>6.8 × 10^{-1}</td>
</tr>
</tbody>
</table>

a Steps in analysis.  
b TDCC at beginning of step.  
c p-value for TDCC at beginning of step.  
d Variable(s) removed at end of step.

7. Sensitivity analysis with small samples

The sensitivity analysis results obtained with random and LHSs of size 100 are very similar and thus indicate that a sample size of 100 is adequate for the problem under consideration. The question naturally arises if smaller sample sizes would also be adequate.

To partially address this question, the random samples were pooled to produce 300 observations, and then three important variable(s), then the next most important variable(s), and so on until no more variables having identifiable effects can be found. However, the procedure differs from forward stepwise regression analysis in that a variable is removed from further consideration once it is identified as being important. In contrast, forward stepwise regression analysis retains those variables identified as being important at previous steps as it moves forward to identify additional important variables. At a certain operational level, the sensitivity analysis procedure presented in this section is analogous to backward stepwise regression analysis in which unimportant variables are sequentially eliminated from inclusion in the regression model. However, there is a very important difference. Backward stepwise regression analysis eliminates variables from further consideration on the basis of being unimportant; in contrast, the presented sensitivity procedure eliminates variables on the basis of being important.

Table 10  
Sensitivity analysis results with the TDCC for three replicated random samples of size 100 and three replicated Latin hypercube samples of size 100

<table>
<thead>
<tr>
<th>Variable</th>
<th>Random sampling</th>
<th>Latin hypercube sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRNREPTC1</td>
<td>HALPOR(1), WMICDFLG(2), ANHPRM(2)</td>
<td>HALPOR(1), WMICDFLG(2), ANHPRM(2), WASTWICK(3), WBRRNBSAT(3), HALPRM(3)</td>
</tr>
<tr>
<td>BRNREPTC2</td>
<td>BHPRM(1), BPVOL(2), WMICDFLG(3), ANHPRM(3)</td>
<td>BHPRM(1), BPVOL(2), WMICDFLG(3), ANHPRM(4), BPVOL(4)</td>
</tr>
<tr>
<td>BRNREPTC3</td>
<td>BHPRM(1), HALPOR(2), BPVOL(2), WMICDFLG(3), ANHPRM(3)</td>
<td>BHPRM(1), HALPOR(2), BPVOL(2), WMICDFLG(3), ANHPRM(4), BPVOL(4)</td>
</tr>
<tr>
<td>GAS_MOLE1</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3), HALPOR(4)</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3), HALPOR(4)</td>
</tr>
<tr>
<td>GAS_MOLE2</td>
<td>HALPOR(1), WGRCOR(2)</td>
<td>HALPOR(1), BHPRM(1), WGRCOR(2)</td>
</tr>
<tr>
<td>GAS_MOLE3</td>
<td>WMICDFLG(1), WGRCOR(2), HALPOR(3), BHPRM(4), BPVOL(4)</td>
<td>WMICDFLG(1), WGRCOR(1), HALPOR(2), BHPRM(3)</td>
</tr>
<tr>
<td>REP_SATB1</td>
<td>HALPOR(1), WGRCOR(2), WMICDFLG(3), WASTWICK(4)</td>
<td>HALPOR(1), WGRCOR(2), WASTWICK(3), WMICDFLG(3)</td>
</tr>
<tr>
<td>REP_SATB2</td>
<td>BHPRM(1), BPVOL(2), HALPOR(3), HALPOR(4)</td>
<td>BHPRM(1), BPVOL(2), HALPOR(2), HALPOR(4)</td>
</tr>
<tr>
<td>REP_SATB3</td>
<td>BHPRM(1), WGRCOR(1), HALPOR(2), ANHPRM(2), BPVOL(2), WMICDFLG(3)</td>
<td>BHPRM(1), WGRCOR(1), HALPOR(2), BPVOL(2), WASTWICK(3), ANHPRM(4), BPVOL(5)</td>
</tr>
<tr>
<td>WAS_PRES1</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3)</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3), ANHPRM(4), HALPOR(4)</td>
</tr>
<tr>
<td>WAS_PRES2</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3), BPVOL(3)</td>
<td>WMICDFLG(1), WGRCOR(2), WASTWICK(3), ANHPRM(4), HALPOR(4)</td>
</tr>
<tr>
<td>WAS_PRES3</td>
<td>HALPOR(1), BPVOL(2), ANHPRM(2), WGRCOR(2)</td>
<td>HALPOR(1), BPVOL(2), ANHPRM(2), WGRCOR(2)</td>
</tr>
</tbody>
</table>

a Dependent variables (Table 2) with 1–3 designating results at 1000, 10,000–1000, and 10,000 yr, respectively.  
b Significant variables identified with replicated random sampling with a p-value cutoff of 0.02 for the TDCC.  
c Significant variables identified with replicated Latin hypercube sampling with a p-value cutoff of 0.02 for the TDCC.  
d Step at which variable is identified as being significant.
Table 11
Sensitivity analysis results based on stepwise rank regression for replicated random samples of size 50 for BRNREPTC, GAS_MOLE, WAS_SATB and WAS_PRES at 1000, 10,000–1000, and 10,000 yr

<table>
<thead>
<tr>
<th>Step</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variableb</td>
<td>SRRCc</td>
<td>R2d</td>
</tr>
<tr>
<td>Random 50: BRNREPTC, 1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>HALPOR</td>
<td>1.00 0.96</td>
<td>HALPOR 0.97 0.94</td>
</tr>
<tr>
<td>2</td>
<td>WMICDFLG</td>
<td>−0.09 0.97</td>
<td>ANHPRM 0.11 0.95</td>
</tr>
<tr>
<td>3</td>
<td>ANHPRM</td>
<td>0.08 0.98</td>
<td>WMICDFLG −0.08 0.96</td>
</tr>
<tr>
<td>4</td>
<td>SHRGSSAT</td>
<td>−0.05 0.98</td>
<td>HALPOR 0.10 0.96</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: BRNREPTC, 10,000–1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: GAS_MOLE, 1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: GAS_MOLE, 10,000–1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: REP_SATB, 1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: REP_SATB, 10,000–1000 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
samples of size 50 were obtained by randomly sampling from these 300 observations. Each new sample of size 50 was produced by sampling without replacement from the 300 observations (i.e. each sample of size 50 was generated without replacement from the original 300 observations). This resampling process is used instead of generating entirely new analysis results because BRAGFLO is time consuming to run for the problem under consideration (i.e. 2–4 h of CPU time on a VAX Alpha per model evaluation). As a result, it is desirable to reuse the available results rather than generate entirely new results. This process was not performed for the LHSs because the stratification associated with Latin hypercube sampling would not be preserved in the resampling, with the result that the new samples of reduced size would not be LHSs.

The new random samples of size 50 were analyzed with stepwise rank regression in the same manner as the random samples of size 100 in Section 4 (Table 11). For a given dependent variable, the results for the samples of size 50 were similar to each other and also similar to the corresponding results for random samples of size 100 in Tables 3–6.

Although the results with random samples of size 50 and 100 are generally similar, the impression emerges that the results with samples of size 100 are somewhat better in the sense of being more consistent and having more variables identified as being significant. To test this, rank regression models containing all independent variables were constructed for the samples of size 50 and 100, and variable importance was ranked on the basis of the resultant SRRCs. The general impression that the analyses with samples of size 100 are somewhat better than the analyses with samples of size 50 is confirmed by the resultant TDCC values (Table 12). In particular, the values obtained for the three

### Table 11 (continued)

<table>
<thead>
<tr>
<th>Step</th>
<th>Replicate 1</th>
<th></th>
<th>Replicate 2</th>
<th></th>
<th>Replicate 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variable</td>
<td>SRRC</td>
<td>( R^2 )</td>
<td>Variable</td>
<td>SRRC</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>3</td>
<td>HALPOR</td>
<td>0.31</td>
<td>0.70</td>
<td>HALPOR</td>
<td>0.27</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>ANHPRM</td>
<td>0.26</td>
<td>0.77</td>
<td>ANHPRM</td>
<td>0.19</td>
<td>0.77</td>
</tr>
<tr>
<td>5</td>
<td>SHPRMHAL</td>
<td>-0.25</td>
<td>0.82</td>
<td>WGRCOR</td>
<td>0.19</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>BPCOMP</td>
<td>0.17</td>
<td>0.84</td>
<td>WASTWICK</td>
<td>0.14</td>
<td>0.94</td>
</tr>
<tr>
<td>Random 50: WAS_PRES, 1000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>WMICDFLG</td>
<td>0.82</td>
<td>0.78</td>
<td>WMICDFLG</td>
<td>0.94</td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>WGRCOR</td>
<td>0.35</td>
<td>0.88</td>
<td>WGRCOR</td>
<td>0.33</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>WFBETCEL</td>
<td>-0.12</td>
<td>0.90</td>
<td>WASTWICK</td>
<td>0.19</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>SHBCEXP</td>
<td>0.11</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SHPRMRLY</td>
<td>-0.09</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SHPRMDRZ</td>
<td>0.08</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ANHBCEXP</td>
<td>0.07</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: WAS_PRES, 10,000–1000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>WMICDFLG</td>
<td>-0.75</td>
<td>0.67</td>
<td>WMICDFLG</td>
<td>-0.88</td>
<td>0.67</td>
</tr>
<tr>
<td>2</td>
<td>WGRCOR</td>
<td>-0.20</td>
<td>0.73</td>
<td>WGRCOR</td>
<td>-0.23</td>
<td>0.77</td>
</tr>
<tr>
<td>3</td>
<td>ANRGSSAT</td>
<td>0.20</td>
<td>0.77</td>
<td>WASTWICK</td>
<td>-0.22</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>SHBCEXP</td>
<td>-0.21</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 50: WAS_PRES, 10,000 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ANRGSSAT</td>
<td>0.39</td>
<td>0.15</td>
<td>BPCOMP</td>
<td>0.33</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>SHBCEXP</td>
<td>-0.33</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

a Steps in stepwise rank regression analysis with \( \alpha \)-values of 0.02 and 0.05 required for a variable to enter and to be retained in an analysis, respectively.

b Variables listed in order of selection in regression analysis with ANHCOMP and HALCOMP excluded from entry into regression model because of \(-0.99\) rank correlation with the pairs (ANHPRM, ANHCOMP) and (HALPORM, HALCOMP).

c Standardized rank regression coefficients (SRRCs) in final regression model.

d Cumulative \( R^2 \) value with entry of each variable into regression model.
samples of size 100 are consistently larger and more significant than those obtained for the three samples of size 50.

The analysis was also tried for random samples of size 25. At this sample size, considerable deterioration in the results was observed (i.e. few or no variables identified as being significant, and considerable variation in identified variables from replicate to replicate). However, even at this small sample size, the sensitivity analysis was typically successful in identifying the important independent variables for the dependent variables that had high $R^2$ values in the analyses for samples of size 50 and 100. The TDCC was not calculated for the samples of size 25 because a regression model containing all 29 variables cannot be constructed when only 25 observations are available, and a TDCC based on a regression model containing a fewer number of variables would not be directly comparable to a TDCC based on regression models containing all variables obtained from samples of size 50 or 100.

8. Sensitivity analysis without regression

The regression analyses summarized in Tables 3–6 exhibit various levels of success. Some analyses are quite good, with $R^2$ values above 0.9. Other analyses are not quite so good, with $R^2$ values in the range from 0.6 to 0.8. The analyses for WAS_PRES at 10,000 yr are effectively failures, with $R^2$ values in the vicinity of 0.2.

An important aspect of the analyses in Tables 3–6 is that the identification of dominant variables tends to remain the same across replicates for both random and Latin hypercube sampling. This consistency holds for regression models with both high and low $R^2$ values. This implies that the failure to account for uncertainty as measured by $R^2$ values probably derives from the sensitivity analysis technique in use (i.e. stepwise regression analysis with rank-transformed data) rather than from an overly small sample size.

When regression-based approaches to sensitivity analysis do not yield satisfactory insights, important variables can be searched for by attempting to identify patterns in scatterplots between sampled and predicted variables with techniques that are not predicated on searches for linear or monotonic relationships. For a sampled variable $x$ (i.e. one of the variables in Table 1) and a predicted variable $y$ (i.e. one of the variables in Table 2 at a specific point in time), possibilities include use of (i) the $F$-statistic to identify changes in the mean value of $y$ across the range of $x$, (ii) the $\chi^2$-statistic to identify changes in the median value of $y$ across the range of $x$, (iii) the Kruskal–Wallis statistic to identify changes in the distribution of $y$ across the range of $x$, and (iv) the $\chi^2$-statistic to identify a nonrandom joint distribution involving $y$ and $x$ [70]. For convenience, the preceding will be referred to as tests for (i) common means (CMNs), (ii) common medians (CMDs), (iii) common locations (CLs), and (iv) statistical independence (SI), respectively.

The indicated statistics are based on dividing the values of $x$ into intervals (Fig. 5). Typically, these intervals contain equal numbers of values for $x$ (i.e. the intervals are of equal probability); however, this is not always the case (e.g. when the sample space for $x$ has a finite number of values of unequal probability). The calculation of the $F$-statistic for CMNs and the Kruskal–Wallis statistic for CLs involves only the division of $x$ into intervals. The $F$-statistic and the Kruskal–Wallis statistic are then used to indicate if the $y$ values associated with these intervals appear to have different means and distributions, respectively. The $\chi^2$-statistic for CMDs involves a further division of the predicted $y$ values into values above and below their median (i.e. the horizontal line in Fig. 5a), with the corresponding significance test used to indicate if the $y$ values associated with the individual intervals defined for $x$ appear to have medians that are different from the median for all values of $y$. The $\chi^2$-statistic for SI involves a division of the $y$ values into intervals of equal probability analogous to the division of the values of $x$ (i.e. the horizontal lines in Fig. 5b), with the corresponding significance test used to indicate if the observed distribution of the ($x, y$) pairs over the cells in Fig. 5b appears to be different from what would be expected if there was no relationship between $x$ and $y$. For each statistic, a $p$-value can be calculated which corresponds to the probability of observing a stronger pattern than the one actually observed if there is no relationship between $x$ and $y$. An ordering of $p$-values then provides a ranking of variable importance (i.e. the smaller the $p$-value, the stronger the effect of $x$ on $y$ appears to be).

Owing to the poor resolution of the regression analyses in Table 6, WAS_PRES at 10,000 yr was analyzed with the tests for CMs, CMDs, CLs and SI (Table 13). For both random and Latin hypercube sampling, BHPRM was identified as the dominant variable by all four tests. In contrast, BHPRM was not identified as being significant by any of the corresponding regression analyses in Table 6. Basically, although there is a strong relationship between BHPRM and WAS_PRES, the nonmonotonic, nonlinear nature of this relationship (Fig. 5) prevents it from being identified in a regression analysis with rank-transformed data.

After the identification of BHPRM as the most important variable, the individual replicates and also the individual analysis techniques show considerable variability in the second and subsequent variables selected for both random and Latin hypercube sampling. The small $p$-values for the pooled replicates for random and Latin hypercube sampling indicate more significant variables than is the case for the individual replicates. This is in contrast to most of the results in Tables 3–6, where the individual replicates and associated pooled replicates produced similar results. This may indicate that the significance tests for nonrandomness require larger sample sizes to be effective than
the significance tests used in conjunction with the regression analyses in Tables 3–6. However, more investigation is needed before any conclusions can be safely drawn.

In addition to the four tests illustrated in this section, many other procedures also exist that might be effective in the identification of patterns in sampling-based sensitivity analyses. For example, the two-dimensional Kolmogorov–Smirnov test has the potential to be a useful technique for the identification of nonrandom patterns [123–126]. As another example, techniques developed to identify randomness in spatial point patterns also have potential for use in the identification of nonrandom patterns in sampling-based sensitivity analyses [127–140]. Finally, nonparametric regression procedures are likely to be effective sensitivity analysis tools in the presence of significant nonlinear relationships between analysis inputs and analysis results [141–145].

A complete variance decomposition of model predictions is a very appealing approach to sensitivity analysis in the presence of nonmonotonic relationships between model inputs and model predictions [36–55]. However, this approach is complicated to implement and is likely to require substantially more model evaluations than a sampling-based approach. Fortunately, publicly available procedures to implement variance decomposition calculations are provided as part of the SIMLAB program [146].
9. Discussion

Uncertainty and sensitivity analysis results obtained with replicated random and LHSs are compared. In particular, uncertainty and sensitivity analyses were performed for a large model for two-phase fluid flow with three independently generated random samples of size 100 each and also three independently generated LHSs of size 100 each.

For the outcomes under consideration, analyses with random and LHSs produced similar results. Specifically, there is little difference in the uncertainty and sensitivity analysis results obtained with random and LHSs of size 100. Further, the results obtained with samples of size 100 are similar to the results obtained for the samples of size 300 that result from pooling the three replicated samples for each sampling procedure. The results obtained with random

<table>
<thead>
<tr>
<th>Variable name</th>
<th>CMN(^b)</th>
<th>CL(^c)</th>
<th>CMD(^d)</th>
<th>SI(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHPRM</td>
<td>1.0</td>
<td>0.0000</td>
<td>1.0</td>
<td>0.0000</td>
</tr>
<tr>
<td>BPCOMP</td>
<td>2.0</td>
<td>0.0114</td>
<td>2.0</td>
<td>0.0119</td>
</tr>
<tr>
<td>BPVOL</td>
<td>3.0</td>
<td>0.0157</td>
<td>4.0</td>
<td>0.0249</td>
</tr>
<tr>
<td>WGRCOR</td>
<td>4.0</td>
<td>0.0166</td>
<td>3.0</td>
<td>0.0231</td>
</tr>
<tr>
<td>HALPOR</td>
<td>5.0</td>
<td>0.0202</td>
<td>5.0</td>
<td>0.0315</td>
</tr>
<tr>
<td>LHPRM</td>
<td>2.0</td>
<td>0.0019</td>
<td>2.0</td>
<td>0.0051</td>
</tr>
<tr>
<td>BPVOL</td>
<td>3.0</td>
<td>0.0042</td>
<td>3.0</td>
<td>0.0069</td>
</tr>
<tr>
<td>ANHBCEXP</td>
<td>4.0</td>
<td>0.0143</td>
<td>2.0</td>
<td>0.0052</td>
</tr>
<tr>
<td>WGRCOR</td>
<td>5.0</td>
<td>0.0585</td>
<td>5.0</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable name</th>
<th>CMN(^b)</th>
<th>CL(^c)</th>
<th>CMD(^d)</th>
<th>SI(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHPRM</td>
<td>1.0</td>
<td>0.0000</td>
<td>1.0</td>
<td>0.0000</td>
</tr>
<tr>
<td>ANHBCEXP</td>
<td>4.0</td>
<td>0.0166</td>
<td>3.0</td>
<td>0.0249</td>
</tr>
<tr>
<td>WGRCOR</td>
<td>5.0</td>
<td>0.0143</td>
<td>2.0</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

For the outcomes under consideration, analyses with random and LHSs produced similar results. Specifically, there is little difference in the uncertainty and sensitivity analysis results obtained with random and LHSs of size 100. Further, the results obtained with samples of size 100 are similar to the results obtained for the samples of size 300 that result from pooling the three replicated samples for each sampling procedure. The results obtained with random

\(^a\) Variables for which at least one of the tests (i.e. CMN, CL, CMD, SI) has a \(p\)-value less than 0.02; variables ordered by \(p\)-values for CMNs.
\(^b\) Ranks and \(p\)-values for CMNs test with \(1 \times 5\) grid.
\(^c\) Ranks and \(p\)-values for CLs (Kruskal–Wallis) test with \(1 \times 5\) grid.
\(^d\) Ranks and \(p\)-values for CMDs test with \(2 \times 5\) grid.
\(^e\) Ranks and \(p\)-values for SI test with \(5 \times 5\) grid.
and LHS in this study are more similar than what has been observed in several other comparisons [63,72–74]. An important implication of this study is that large sample sizes are often unnecessary to develop an understanding of a complex system. This has also been demonstrated in several other studies with relatively small, replicated samples [147,148]. A possible analysis strategy is to initially carry out an analysis with a relatively small sample size, and then add additional sample elements and associated model evaluations only if the initial analysis is found to be inadequate.

In considering appropriate sample sizes, it is important to recognize the distinction between analyses carried out to assess the effects of subjective (i.e. epistemic) uncertainty, and analyses carried out to assess the effects of stochastic (i.e. aleatory) uncertainty. In assessing the effects of stochastic uncertainty, it is often desired to determine the likelihood of low probability, but high consequence, events. The determination of such likelihoods in a naïve sampling-based analysis requires a very large sample size. However, in assessing the effects of subjective uncertainty, the goal is usually to determine general patterns of behavior rather than likelihoods for specific, low probability behaviors. As a result, analyses to assess the effects of subjective uncertainty can be carried out with much smaller sample sizes than analyses carried out to assess the effects of stochastic uncertainty. In practice, analyses that must estimate the likelihood of low probability events typically use some type of importance sampling procedure (e.g. an event tree) rather than an unstructured random sampling procedure [149–157].

Extensive regression-based sensitivity analyses were carried out for the individual replicated samples. When these analyses performed poorly, this performance was due to the inappropriateness of the regression model for the patterns in the mapping between model input and a model output rather than to an inadequate sample size. In particular, employing a more appropriate sensitivity analysis procedure is more effective than simply increasing the sample size. Fortunately, a number of procedures exist that can be used to identify nonrandom patterns in the mapping between model input and a model output.

The TDCC was found to be an effective procedure for comparing variable rankings obtained with replicated samples. Owing to its emphasis on agreement between the rankings assigned to important variables and deemphasis on disagreement between the rankings assigned to unimportant variables, the TDCC was more effective in comparing variable rankings than KCC. Further, when replicated samples are available, the TDCC provides the basis for a sensitivity analysis procedure predicated on the agreement of importance measures obtained for the individual replicates.

Although random and Latin hypercube sampling performed similarly in this analysis, the authors' preference remains Latin hypercube sampling for use in analyses of complex systems with small sample sizes. On the whole, the enforced stratification over the range of each sampled variable gives Latin hypercube sampling a desirable property that should not be given up. In a large analysis with many inputs and even more outputs, this stratification should decrease the likelihood of being mislead in assessing the relationships between individual inputs and outputs.

**Acknowledgements**

Work performed for Sandia National Laboratories (SNL), which is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000. Review provided at SNL by M. Chavez, J. Garner, and S. Halliday. Editorial support provided by F. Puffer, J. Ripple, and K. Best of Tech Reps, Inc.

**References**


[105] Bean JE, Lord ME, McArthur DA, MacKinnon RJ, Miller JD, Schreiber JD. Analysis package for the salado flow calculations (task 1) of the performance assessment analysis supporting


