JAERI-M
89-111

AN IMPORTANCE QUANTIFICATION TECHNIQUE IN UNCERTAINTY ANALYSIS FOR COMPUTER MODELS

September 1989

Tsutomu ISHIGAMI and Toshimitsu HOMMA

日本原子力研究所
Japan Atomic Energy Research Institute

JAERI-Mレポートは，日本原子力研究所が不定期に公刊している研究報告書です。
入手の問合わせは，日本原子力研究所技術情報部情報資料課（〒319－11茨城県那珂郡東海村）あて，お申しこしください。なお，このほかに財団法人原子力弘済会資料センター（〒319-11茨城県那珂郡東海村日本原子力研究所内）で複写による実費頒布をおこなっております。

JAERI-M reports are issued irregularly.
Inquiries about availability of the reports should be addressed to Information Division, Department of Technical Information, Japan Atomic Energy Research Institute, Tokai-mura, Naka-gun, Ibaraki-ken 319-11, Japan.

編集兼発行　日本原子力研究所
印　　刷　日立高速印刷株式会社

JAERI-M 89-111

An Importance Quantification Technique in Uncertainty Analysis for Computer Models

Tsutomu ISHIGAMI and Toshimitsu HOMMA+

Department of Reactor Safety Research
Tokai Research Establishment
Japan Atomic Energy Research Institute
Tokai-mura, Naka-gun, Ibaraki-ken

(Received August 3, 1989)

We have developed a new technique to numerically quantify importance of input variables including uncertainties to the output uncertainty. The technique makes it practically possible to estimate the importance measure, proposed by Hora and Iman, which is based on the concept of uncertainty reduction. The technique requires a limited number of calculations based on the original model using the Monte Carlo or the Latin hypercube sampling. Effectiveness of the technique is demonstrated in a comparative study by applying the technique and a conventional regression method to two computer models, the TERFOC (Terrestrial Food Chain) and the PSA (Probabilistic Systems Assessment) models, for dose prediction.

Keywords : Uncertainty Analysis, Importance Analysis, Sensitivity Analysis, Monte Carlo Method, Latin Hypercube Sampling, Regression Method, Uncertainty Reduction Method, Radioactive Waste Disposal

+ Department of Environmental Safety Research

計算コードの不確実さ解析における重要度評価手法

日本原子力研究所東海研究所原子炉安全工学部
石神　努・本間　俊充[+]

（1989年8月3日受理）

計算モデルの入力値に不確実さが含まれているとき，その出力値（計算結果）は不確実さを有する。出力値の不確実さに寄与する入力変数を識別するための重要度評価について，新しい計算手法を開発した。同手法は，HoraおよびIman の提案した不確実さ減少の概念に基づいた重要度の尺度を，元の計算モデルに基づきモンテカルロ法あるいは改良型モンテカルロ法を用いて計算するものである。同手法ならびに従来の回帰分析法を，線量予測用計算コード，TERFOC（Terrestrial Food Chain）およびPSA（Probabilistic Systems Assessment），に適用し，これらの手法の有用性についての比較検討を行った。

東海研究所：〒319-11　茨城県那珂郡東海村白方字白根２－４
+　環境安全研究部

Contents

目　　次

1. Introduction

In order to analyze complex systems, computer models are widely used in many fields such as probabilistic safety assessment of nuclear power plants,[1]-[2] dose assessment for disposal of radioactive waste[3] and so on. However, the results (outputs) calculated include uncertainties due to uncertainties included in both the complex computer models themselves and values of input variables or parameters used in the analysis. It is important not only to quantify the output uncertainties but also to identify dominant input variables which contribute to the output uncertainties.

Several techniques have been developed and reviewed to quantify the output uncertainties caused by propagation of uncertainties in inputs.[4]-[6] The Monte Carlo method is known to be most useful when the number of input variables including uncertainties are small and the cost of calculations with the computer model of concern is inexpensive. However, the Monte Carlo method loses its effectiveness for very long-running computer codes due to time consuming and high cost of calculations. Today, two methods are often used alternatively. One is the stratified Monte Carlo sampling method known as the Latin hypercube sampling (LHS) method,[7],[8] which is considered to be one of the most powerful methods. The other is the response surface method[9] that replaces the original complex computer model by a simplified surrogate model, that is a response surface model. Its usefulness depends on the degree of fitness of the response surface model to the original computer model.

After the output uncertainties have been estimated, an importance analysis is useful to determine rankings of the input variables with respect to the output uncertainties for the computer model.[10] Two approaches may be considered as the importance analysis techniques. One is based on replacing the original complex computer model by a simplified surrogate model such as response surface models (Approach A). The other does not rely on such a simplified surrogate model but relies on a limited number of Monte Carlo calculations based on the original computer model (Approach B). As to the approach A, Downing et al.[5] have discussed and reviewed several importance analysis techniques including subjective, differential sensitivity analysis, one-at-a-time design, rank order correlation and the adjoint method. In application to a model

of dose assessment, their conclusion was that the partial rank-order correlation measure was the most appropriate, because it accounted for the nonlinearity and correctly incorporated the monotonicity. Although the partial rank-order correlation measure seems to be promising, its reliability would become worse as the regressibility to the original computer model is poorer.

On the other hand, regarding approach B, Hora and Iman[11] introduced the uncertainty importance measure which is based upon the concept of uncertainty reduction by ascertaining the value of the input variable of concern. Their introduced measure seems to be theoretically satisfactory, because no approximation to the original computer model is made and because the importance measure does not depend upon the specified value of the input variable of concern. However, it has had drawback in its practical use except for application to specic problems, since it seems to be very difficult and costly to calculate the measure numerically or analytically.

The purpose of this report is to present a calculational technique to estimate the importance measure proposed by Hora and Iman, and to demonstrate its usefulness by applying to two computer models, the TERFOC(Terrestrial Food Chain)[12] and the PSA (Probabilistic Systems Assessment)[13] models, for dose prediction. The technique needs a limited number of calculations based on the original computer model using the Monte Carlo or Latin hypercube (LH) sampling, and hence it is expected that the technique gives more accurate results as the size of samples becomes larger. The results of input variable rankings obtained by the technique are compared with those by the regression method of partial (rank) correlation coefficients (P(R)CCs) and standardized (rank) regression coefficients (S(R)RCs).[14]

Chapter 2 describes a calculational technique for the importance measure proposed by Hora and Iman. In chapter 3, the technique is applied to two computer models, the TERFOC and the PSA models, to demonstrate its practical use. Concluding remarks are given in chapter 4.

2. Calculational Technique of Importance Measure

Let the output variable Y be a function of variables X_i's (i=1,2,...,n) as given by

$$Y = h(X_1, X_2, \ldots, X_n). \tag{2.1}$$

We hereafter assume that variables X_i's (random variables) are independent. Denoting a probability density function (PDF) of X_i as $f_i(X_i)$, the PDF of a set of random variables $(X_1, X_2, \ldots, X_n)$ is expressed as the product of the PDF of X_i:[15]

$$f(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f_i(X_i), \tag{2.2}$$

owing to independency of random variables X_i's. Then the mean and variance of Y are given by

$$\langle Y \rangle = \int\!\!\ldots\!\int h(X_1, X_2, \ldots, X_n) \prod_{i=1}^{n} f_i(X_i)dX_i, \tag{2.3}$$

$$V_Y = \int\!\!\ldots\!\int \{h(X_1, X_2, \ldots, X_n) - \langle Y \rangle\}^2 \prod_{i=1}^{n} f_i(X_i)dX_i$$

$$= \int\!\!\ldots\!\int h(X_1, X_2, \ldots, X_n)^2 \prod_{i=1}^{n} f_i(X_i)dX_i - \langle Y \rangle^2. \tag{2.4}$$

2.1 Importance Measure Based on Uncertainty Reduction

Before we introduce the importance measure of input variables X_i's to the output variable Y, we first consider uncertainty reduction of Y by ascertaining the specified value of the variable X_j. Let us introduce a variance of Y, $V_Y(\underline{x}_j)$, in the case where the value of X_j is ascertained to the specified value $\underline{x}_j$:

$$V_Y(\underline{x}_j) = \int \cdots \int \{ h(X_1, \ldots, X_{j-1}, \underline{x}_j, X_{j+1}, \ldots, X_n) - \langle h(\underline{x}_j) \rangle \}^2 \cdot$$

$$\prod_{\substack{i=1 \\ (i \neq j)}}^{n} f_i(X_i) dX_i$$

$$= \int \cdots \int \{ h(X_1, \ldots, X_{j-1}, \underline{x}_j, X_{j+1}, \ldots, X_n) \}^2 \cdot$$

$$\prod_{\substack{i=1 \\ (i \neq j)}}^{n} f_i(X_i) dX_i - \langle h(\underline{x}_j) \rangle^2, \qquad (2.5)$$

where $\langle h(\underline{x}_j) \rangle$ denotes a mean of Y with the input variable X_j ascertained by the specified value $\underline{x}_j$ as given by

$$\langle h(\underline{x}_j) \rangle = \int \cdots \int h(X_1, \ldots, X_{j-1}, \underline{x}_j, X_{j+1}, \ldots, X_n) \cdot$$

$$\prod_{\substack{i=1 \\ (i \neq j)}}^{n} f_i(X_i) dX_i. \qquad (2.6)$$

Since the value of $\underline{x}_j$ has been ascertained, it is expected that the resulting variance of Y, $V_Y(\underline{x}_j)$, is smaller than the variance V_Y, Eq.(2.4), where all input variable uncertainties have been taken into account. In other words, uncertainty in Y has been somewhat reduced.

The dependency of the variance $V_Y(\underline{x}_j)$ on the specified value of the variable X_j can be eliminated by averaging $V_Y(\underline{x}_j)$ according to the PDF of X_j as

$$V_Y{}^j = \int V_Y(\underline{x}_j) f_j(\underline{x}_j) d\underline{x}_j. \qquad (2.7)$$

Substitution of Eq.(2.5) into Eq.(2.7) gives

$$V_Y{}^j = \int \cdots \int h(X_1, X_2, \ldots, X_n)^2 \prod_{i=1}^{n} f_i(X_i) dX_i$$

$$- \int \langle h(\underline{x}_j) \rangle^2 f_j(\underline{x}_j) d\underline{x}_j. \qquad (2.8)$$

Comparison of Eqs.(2.4) and (2.8) leads to the relation:

$$V_Y - V_Y{}^j = U_j - \langle Y \rangle^2, \tag{2.9}$$

with

$$U_j = \int \langle h(\underline{x}_j) \rangle^2 \, f_j(\underline{x}_j) d\underline{x}_j. \tag{2.10}$$

Eq.(2.9) is equivalent to the expected reduction in the variance of Y proposed by Hora and Iman. According to them, we refer to the square root of this reduction as a measure of uncertainty importance for X_j:

$$I_j = \sqrt{V_Y - V_Y{}^j}$$

$$= \sqrt{U_j - \langle Y \rangle^2} \, . \tag{2.11}$$

This importance measure, Eq.(2.11), is general and thus favorable because no approximation has been made in the above procedure except for the assumption of independency of input variables. Since the quantity $\langle Y \rangle^2$ is constant, the importance rankings are determined by the values of U_j. In other words, the input variable X_j is more important than X_i if $U_j > U_i$. Hence the importance analysis to determine the importance rankings of input variables is reduced how to practically estimate the quantity U_j.

2.2 Calculational Technique

If the function $h(X_1, X_2, \ldots, X_n)$ has a simple form, we may estimate the quantity U_j analytically. However it is not possible in general to analytically estimate U_j for a complex computer model. Two approaches to estimation of U_j may be considered. The first approach is to approximate $h(X_1, X_2, \ldots, X_n)$ to a simplified surrogate model such as a response surface. If a linear regression model is selected as a response surface, it turns out that the resulting I_j for the standardized Y (defined by $(Y-\langle Y \rangle)/\sigma(Y)$, where $\sigma(Y)$ is the standard deviation of Y) coincides with the absolute value of the standardized regression coefficient for X_j (see Appendix A). Although this approach seems to be promising, it should be noted that the importance measure is calculated

with regard to the response surface model and not on the original model. Hence the obtained results are only as valid as the response surface model itself.

The second approach is to numerically calculate the quantity U_j in Eq.(2.10) based on the original computer model without relying on the regression or the response surface method. Using the Monte Carlo method, we may be able to estimate the quantity $\langle h(\underline{x}_j)\rangle$ which is the mean of the output variable Y where all input variable uncertainties other than uncertainty in the input variable X_j are taken into account. Since the quantity U_j is expressed as an integral form, direct estimate of U_j needs a large number of Monte Carlo calculations by varying the specified value $\underline{x}_j$. Namely, the calculation number is MxN, where M is a mesh number to calculate the integral in Eq.(2.10) and N is a number presenting Monte Carlo sample size to estimate $\langle h(\underline{x}_j)\rangle$ for the specified value $\underline{x}_j$. This calculational method is too costly and thus impractical.

The proposed technique to numerically calculate the quantity U_j is based on the second approach with an allowable number of Monte Carlo calculations. In order to overcome the drawback mentioned above, we rewrite Eq.(2.10) as

$$U_j = \int \{ \int \cdots \int h(X_1, \ldots, X_{j-1}, \underline{x}_j, X_{j+1}, \ldots, X_n) \cdot \prod_{\substack{i=1 \\ (i\neq j)}}^{n} f_i(X_i)\}^2 f_j(\underline{x}_j) d\underline{x}_j$$

$$= \int \cdots \int h(X_1, \ldots, X_{j-1}, \underline{x}_j, X_{j+1}, \ldots, X_n) \, h(X'_1, \ldots, X'_{j-1}, \underline{x}_j, X'_{j+1}, \ldots, X'_n) \left(\prod_{\substack{i=1 \\ (i\neq j)}}^{n} f_i(X_i) dX_i \right) \left(\prod_{\substack{i=1 \\ (i\neq j)}}^{n} f_i(X'_i) dX'_i \right) f_j(\underline{x}_j) d\underline{x}_j$$

$$= \int \cdots \int h(X_1, \ldots, X_n) h(X'_1, \ldots, X'_{j-1}, X_j, X'_{j+1}, \ldots, X'_n) \left(\prod_{i=1}^{n} f_i(X_i) dX_i \right) \left(\prod_{\substack{i=1 \\ (i\neq j)}}^{n} f_i(X'_i) dX'_i \right). \qquad (2.12)$$

The above equation, Eq.(2.12), shows that the quantity U_j is nothing more than an expectation value of the function defined by

$$H(X_1, \ldots, X_n, X'_1, \ldots, X'_{j-1}, X'_{j+1}, \ldots, X'_n)$$

$$= h(X_1, \ldots, X_n)\, h(X'_1, \ldots, X'_{j-1}, X_j, X'_{j+1}, \ldots, X'_n), \qquad (2.13)$$

with a set of (2n-1) independent random variables $(X_1, \ldots, X_n, X'_1, \ldots, X'_{j-1}, X'_{j+1}, \ldots, X'_n)$ whose PDF is given by $\left(\prod_{i=1}^{n} f_i(X_i)\right)\left(\prod_{\substack{i=1 \\ (i\neq j)}}^{n} f_i(X'_i)\right)$. Thus the quantity U_j would be estimated numerically with allowable Monte Carlo calculations. It is noted that to estimate U_j the required computer run number for the original model function $h(X_1, \ldots, X_n)$ is 2xN, where N is a number presenting Monte Carlo sample size, because the function H is expressed by twofold product of the original model function h. This number is much smaller than the number MxN based upon the direct use of Eq.(2.10).

If Monte Carlo calculations for the original model $h(X_1, X_2, \ldots, X_n)$ is costly and impractical, the Latin hypercube sampling (LHS) method (see Appendix B) would be applicable, which requires much smaller number of samples than the Monte Carlo method. Now the computer run number to estimate U_j is reduced to 2xL and thus the computer run number needed to determine the importance rankings of input variables is reduced to 2xLxK, where L is a number presenting LH sample size and K is the number of input variables of concern. Hence the calculational technique for the importance analysis which does not rely on a simplified surrogate model but uses directly the original computer model has been established. The technique will be referred to the uncertainty reduction method hereafter. It should be emphasized that this technique is expected to give more accurate results as the size of the Monte Carlo or the LH samples becomes larger, owing to reduced error in statistics.

3. Application to Computer Models for Dose Prediction

3.1 The TERFOC Model

3.1.1 Description of the TERFOC Model

The purpose of the TERFOC (Terrestrial Food Chain) computer code developed by Japan Atomic Energy Research Institute is to assess the potential radiological impact of routine releases of radioactive effluents from nuclear facilities.[12] The TERFOC model calculates the concentration of radionuclides in vegetation and animal products as the result of deposition of radionuclides on agricultural land.

To simplify the problem for the present analysis an average long-term concentration of ^{131}I [Bq/m^3] in the above-ground atmosphere is assumed, and the physico-chemical form of iodine is assumed to be 10% elemental, 50% organic and 40% particulate. The equation used for predicting the i-th radionuclide concentration, C_m [Bq/ℓ], in milk is given by

$$C_m = F_m \left(Q_{FF} C_{VF} + Q_{FS} C_{VS} \right) \exp(-\lambda_i t_f), \tag{3.1}$$

where

F_m = feed to milk transfer factor [d/ℓ]
Q_{FF} = daily dry intake of fresh forage by dairy cows [kg/d]
Q_{FS} = daily dry intake of stored forage by dairy cows [kg/d]
t_f = time delay from production to consumption of milk [d]
C_{VF} = radionuclide concentration in fresh forage [Bq/kg]
C_{VS} = radionuclide concentration in stored forage [Bq/kg]
λ_i = physical decay constant of the i-th radionuclide [1/d].

The radionuclide concentrations in fresh forage (C_{VF}) and stored forage (C_{SF}) are calculated as

$$C_{VF} = d \left[\frac{r}{Y_V(\lambda_i + \lambda_w)} \{1 - \exp\left(-(\lambda_i + \lambda_w)t_e \right)\} + \frac{B_V}{P(\lambda_i + \lambda_s)} \{1 - \exp\left(-(\lambda_i + \lambda_s)t_b \right)\} \right], \tag{3.2}$$

$$C_{VS} = C_{VF} \exp(- \lambda_i t_h), \quad (3.3)$$

where

d = deposition rate of the radionuclide from both wet and dry processes [Bq/(m^2 d)]
r = interception fraction retained of the edible portion of pasture forage [-]
Y_V = agricultural productivity of pasture forage [kg/m^2]
λ_w = weathering removal rate constant [1/d]
t_e = time period that vegetation is exposed to contamination during the growing season [d]
B_v = soil to plant concentration factor [(Bq/kg dry forage)/(Bq/kg dry soil)]
P = effective surface density of soil [kg/m^2]
λ_s = rate constant for reduction of radionuclide deposited in root zone of soil [1/d]
t_b = period of long-term deposition for activity in soil [d]
t_h = time delay between harvest and consumption of stored forage [d].

The total deposition rate, d, on ground and vegetation is given by the sum of processes of dry deposition and wet deposition:

$$d = V_g \chi + \omega I \chi , \quad (3.4)$$

where

χ = average long-term concentration of a radionuclide in the above-ground atmosphere [Bq/m^3]
V_g = dry deposition velocity [m/d]
ω = washout ratio [-]
I = long-term average of the precipitation intensity [m/d].

Here, the washout ratio approach is adapted to calculate the wet deposition rate. Both the dry deposition velocity and the washout ratio depend on the physico-chemical form of radionuclide considered.

3.1.2 Variables Considered in the Importance Analysis for the TERFOC Model

Of many output variables provided by the TERFOC model, one output variable has been selected in the present importance analysis for simplicity. The output variable selected is the ^{131}I concentration in milk, C_m given by Eq. (3.1), at the time of 30 years. The TERFOC input variables selected are given in Table 3.1 together with their assigned ranges and distribution patterns as used for the present importance analysis. The primary information on the ranges and distribution patterns of these variables was taken from an extensive review[16] of the literature describing the statistical properties of environmental parameters. Some of the ranges of input variables such as Q_{FF} and Q_{FS}, however, were derived from a review of site-specific information in Japan. Triangular distributions were assumed when detailed information was not available. These twelve input variables will be sampled independently using the LHS technique in accordance with the ranges and distribution patterns of the input variables.

3.1.3 Importance Analysis for the TERFOC Model

(1) Partial Correlation and Standardized Regression Coefficients[14]

The partial correlation coefficient (PCC) is a measure of the unique linear relationship between two variables that cannot be explained in terms of the relationships of these two variables with any other variables. Thus, it provides an importance measure with which to identify the variables which should be accounted for in a regression model.

When nonlinear relationships are involved, it is often more appropriate to calculate standardized regression coefficients and partial correlation coefficients on variable ranks rather than on the actual values for the variables: such coefficients are known as standardized rank regression coefficients (SRRCs) and partial rank correlation coefficients (PRCCs). Specially, the smallest value of a variable is assigned the rank 1, the next smallest value is assigned the rank 2, and so on up to the largest value which is assigned the rank L, where L denotes the number of observations. The standardized regression coefficients and/or partial correlation coefficients are then calculated

on these ranks than on the raw variables. The rank transformation permits a better fit of the regression model to the actual model since the weaker assumption of monotonicity between raw outputs and inputs replaces the linearity requirement.

Based on the TERFOC-run results for two hundreds LHS input vectors, the evaluation of PCCs, PRCCs, SRCs, and SRRCs was performed using the computer program in Ref. 14. Table 3.2 shows these coefficients and coefficients of determination, R^2, for the ^{131}I concentration in milk, C_m, at the time of 30 years. The results show that the fit of the regression model to the TERFOC model is almost satisfactory ($R^2 > 0.8$). It is found that the three input variables, F_m, RP and ω_p, as revealed by their high correlation coefficients, predominantly govern the uncertainty in the output C_m.

(2) Uncertainty Reduction Method

In calculating the quantity U_j relating to the importance measure I_j for the j-th input variable, the TERFOC computer runs for four hundreds LHS input vectors have been performed to obtain the expectation value of the function H, Eq. (2.13), corresponding to the output variable C_m. Here the number of "input" variables in the function H is not twelve but twenty three (= 2x12 - 1), and this is why the LH sample size is larger by two times than the LH sample size used in the regression analysis. The calculated values of U_j for twelve input variables are shown in Table 3.3. The results indicate that the input variables, F_m, ω_p and RP, make dominant contribution to the uncertainty in the output variable C_m.

The rankings of input variables obtained from the uncertainty reduction method are also shown in Table 3.3 together with the results by the regression analysis of PCCs, PRCCs, SRCs and SRRCs. The results show the following:

(a) Both the uncertainty reduction method and the regression method identify the three dominant input variables, F_m, ω_p and RP, that contribute to the output uncertainty.

(b) There can be seen some differences between the results by the uncertainty reduction method and those by the regression method. This may reveal that the fit of the regression model to the TERFOC model is not completely adequate in spite of high value of R^2.

3.2 The PSA Model

3.2.1 Description of the PSA Model

The PSA (Probabilistic Systems Assessment) model was used here, which was introduced in the level 0 intercomparison PSA exercise of the NEA Codes User Group. The detailed model specifications are described in PSA level 0 report.[13]

The problem is to calculate the statistics of the annual effective dose equivalent to the maximum exposed individual due to a certain specified nuclide. The dose arises after leaching from a repository, transport through a buffer and the geosphere, and ingestion by drinking from a contaminated well.

The waste-form is assumed to dissolve at a constant rate for a finite time. The leach rate $R_{wf}(t)$ is given by

$$R_{wf}(t) = \begin{cases} R^o & \text{if } t < \tau_D \\ & \qquad [kg/(m^2 \cdot a)] \\ 0 & \text{if } t > \tau_D \end{cases} \tag{3.5}$$

where

$$\tau_D = Q/(R^o\ S), \quad [a] \tag{3.6}$$

R^o = initial leach rate [kg/(m^2·a)]
Q = the total mass of waste-form [kg]
S = the surface area of waste-form [m^2]

The flow of radionuclide i from the waste-form is given by:

$$F_i(t) = R_{wf}(t)\ I_i^o\ \exp(-\lambda_i t)\ S, \ [mol/a] \tag{3.7}$$

where

I_i^o = initial inventory of radionuclide i [mol/kg]
λ_i = the decay constant of radionuclide i [mol/kg].

A buffer of thickness X_B acts to delay the flow of radionuclides from the waste form to the geosphere. The delay time depending on the

radionuclide is given by

$$\tau_i^B = X_B^2 \, R_i^B \, / \, (4 \, D_B), \, [a] \tag{3.8}$$

where

$$R_i^B = 1 + \rho_B \, K_{Di}^B \, (1 - \varepsilon_B) / \varepsilon_B, \, [-] \tag{3.9}$$

ρ_B = buffer density [kg/m^3]
ε_B = buffer porosity [-]
K_{Di}^B = buffer sorption constant for radionuclide i [m^3/kg]
D_B = diffusion coefficient of the buffer [m^2/a].

A geosphere path of length X_G acts to delay and spread contaminant over time. The contaminant is assumed to spread uniformly between times τ_i^L and τ_i^H which are the low and high roots, respectively, for τ_i^G of the equation:

$$X_G = 2 \left(\frac{D_G \, \tau_i^G}{R_i^G} \right)^{1/2} + \frac{V_G \, \tau_i^G}{R_i^G}, \, [m] \tag{3.10}$$

where

$$R_i^G = 1 + \rho_G \, K_{Di}^G \, (1 - \varepsilon_G) \, / \, \varepsilon_G, \, [-] \tag{3.11}$$

ρ_G = geosphere density [kg/m^3]
ε_G = geosphere porosity [-]
K_{Di}^G = geosphere sorption constant for radionuclide i [m^3/kg]

$$D_G = D_G^o + \alpha_G \, V_G, \, [m^2/a] \tag{3.12}$$

D_G^o = geosphere diffusion coefficient [m^2/a]
V_G = geosphere groundwater velocity [m/a]
α_G = geosphere dispersion length [m].

Only the drinking water scenario is considered in the biosphere model. The annual effective dose equivalent, $H_i(t)$, is given by

$$H_i(t) = A_i\ W_m\ D_i\ F_i^G(t)\ /\ W,\ [Sv/a] \tag{3.13}$$

where

A_i = molar specific activity of radionuclide i [Bq/mol]
W_m = consumption rate of water [m^3/a]
W = water extraction rate [m^3/a]
D_i = dose conversion factor of radionuclide i by ingestion [Sv/Bq].

The flow $F_i^G(t)$ entering the biosphere from the geosphere is given by

$$F_i^G(t) \begin{cases} = \dfrac{I_i^o\ Q\ \exp(-\lambda_i t)}{\tau_i^H - \tau_i^L + \tau_D} & \text{if } \tau_i^B + \tau_i^L < \tau_i^B + \tau_i^H + \tau_D \\ & [\text{mol/a}] \\ = 0 & \text{otherwise} \end{cases} \tag{3.14}$$

3.2.2 Variables Considered in the Importance Analysis for the PSA Model

The PSA model provides predicted values of dose as a function of time. The output variable selected in the present importance analysis is the annual effective dose equivalent to the maximum exposed individual due to ^{129}I, $H_{I\text{-}129}(t)$ given by Eq. (3.13), at the time of 10^5 years and 10^6 years. The PSA input variables selected are given in Table 3.4 together with their assigned ranges and distribution patterns. These ranges and distribution patterns of the variables were taken from the level 0 inter comparison exercise of the NEA probabilistic systems assessment codes user group. These ten input variables will be sampled independently using the LHS techniques in accordance with their ranges and distribution patterns.

3.2.3 Importance Analysis for the PSA Model

In the regression analysis for the PSA model, one thousand LHS input vectors have been prepared and PSA computer runs according to the LHS

input vectors have been performed. Table 3.5 shows PCCs, PRCCs, SRCs, SRRCs and coefficients of determination, R^2, for the annual effective dose equivalent to the maximum exposed individual due to ^{129}I, $H_{I-129}(t)$ at the time of 10^5 years and 10^6 years. As shown in the Eqs (3.13) and (3.14) the annual effective dose equivalent for a given input vector is non-zero only over a limited time interval. The position of this time interval varies over millions of years, and hence only a few input vectors give non-zero values at a particular time. The percentage of non-zero values is below nine percent of one thousand runs. The existence of many ties (same values) in the output data leads to the very poor fit of the regression model to the PSA model ($R^2 < 0.15$).

In calculating the quantity U_j relating to the importance measure I_j with the uncertainty reduction method, five thousands LHS input vectors have been prepared, and the PSA computer runs according to the LHS input vectors have been performed to obtain the expectation value of the function H, Eq.(2.13), corresponding to the output variable $H_{I-129}(t)$. The calculated values of the importance measure of the ten input variables for the output variables are shown in Table 3.6 together with rankings of input variables obtained by the uncertainty reduction method and the regression method. The results show the following:

(a) Both the uncertainty reduction method and the regression method identify the same dominant input variable, V_G, which contributes uncertainty in both the output variables, $H_{I-129}(10^5$ years) and $H_{I-129}(10^6$ years).

(b) However, the resulting rankings of input variables are different between the two methods. It is noted that there are large difference even between the results by PCC/SRC and PRCC/SRRC, probably due to poor regression fit to the original model.

Although we used a relatively large number of LH samples of five thousands, the resulting values of uncertainty reduction given by Eq.(2.9) for some input variables were negative. This may reveal that the size of the LH samples is not enough to give reasonable results by making the numerical calculation error in statistics negligibly small, or that there is a limitation of the Monte Carlo or LH sampling approach in the importance analysis for the PSA model. However, we may say that those

variables for which the uncertainty reduction values are negative do not much contribute to the output uncertainty. We, therefore, expect that the calculational technique presented here based on the uncertainty reduction method would not miss identifying several important input variables that contribute to the output uncertainty.

Table 3.1 Input Variables, their Ranges and Distributions Considered in the Importance Analysis for the TERFOC Model.

Variable	Description	Probability Distribution	Unit	Range		
				Lower Bound		Upper Bound
ω_E	washout ratio for elemental iodine	lognormal	-	2×10^4		2×10^6
ω_P	washout ratio for particulate iodine	lognormal	-	3×10^4		3×10^6
V_{gE}	deposition velocity for elemental iodine	lognormal	cm/s	0.02		26
V_{gP}	deposition velocity for particulate iodine	lognormal	cm/s	0.03		3
RP	Mass interception factor for pasture forage (r/Y_V)	lognormal	m^2/kg (wet)	0.47		8.5
λ_W	weathering removal rate constant	lognormal	1/d	5.9×10^{-3}		0.42
F_m	feed to milk transfer factor	lognormal	d/ℓ	1.9×10^{-3}		5.2×10^{-2}
Q_{FF}	daily dry intake of fresh forage by daily cows	triangular	kg/d	3	4*	5
Q_{FS}	daily dry intake of stored forage by daily cows	triangular	kg/d	9	13*	18
t_e	time period that vegetation is exposed to contamination during the growing season	triangular	d	15	30*	60
t_h	time delay between harvest and consumption of stored forage	triangular	d	15	90*	180
t_f	time delay from production to consumption of milk	triangular	d	0	2*	7

* denotes the most probable value in the triangular distribution.

Table 3.2 PCCs, PRCCs, SRCs, SRRCs and R^2 Values for the TERFOC Output Variable C_m.

Input Variable	PCC	PRCC	SRC	SRRC
ω_E	0.25	0.32	0.11	0.10
ω_P	0.64	0.82	0.36	0.43
V_{gE}	0.36	0.51	0.17	0.18
V_{gP}	0.30	0.55	0.14	0.19
RP	0.71	0.86	0.44	0.49
λ_W	-0.52	-0.70	-0.27	-0.29
F_m	0.80	0.88	0.59	0.54
Q_{FF}	0.27	0.41	0.12	0.13
Q_{FS}	0.06	0.02	0.02	0.01
t_e	0.07	0.10	0.03	0.03
t_h	-0.02	-0.07	-0.01	-0.02
t_f	-0.16	-0.40	-0.07	-0.13
R^2	0.81	0.91	0.81	0.91

Table 3.3 Values of Quantity U_j with the Uncertainty Reduction Method (URM) and Rankings of Input Varaibles for the TERFOC Output Variable C_m.

Input Variable	U_j	Ranking				
		URM	PCC	PRCC	SRC	SRRC
ω_E	2.00×10^5	5	8	9	8	9
ω_P	2.38×10^5	2	3	3	3	3
V_{gE}	2.04×10^5	4	5	6	5	6
V_{gP}	1.97×10^5	6	6	5	6	5
RP	2.29×10^5	3	2	2	2	2
λ_W	1.90×10^5	10	4	4	4	4
F_m	2.56×10^5	1	1	1	1	1
Q_{FF}	1.91×10^5	9	7	7	7	7
Q_{FS}	1.88×10^5	12	11	12	11	12
t_e	1.88×10^5	11	10	10	10	10
t_h	1.93×10^5	8	12	11	12	11
t_f	1.93×10^5	7	9	8	9	8

Table 3.4 Input Variables, their Ranges and Distributions Considered in the Importance Analysis for the PSA Model.

Variable	Description	Probability Distribution	Unit	Range	
				Lower Bound	Upper Bound
R^o	time-invariant waste form leach rate	loguniform	$kg/(cm^2\ a)$	2.69×10^{-3}	12.9
X_B	buffer thickness	uniform	m	0.5	5
$K_{DI}{}^B$	sorption constant in the buffer for iodine	lognormal	m^3/kg	6.18×10^{-10}	0.118
X_G	geosphere pathlength	uniform	m	1×10^3	1×10^4
V_G	geosphere groundwater velocity	loguniform	m/a	0.001	0.1
$D_G{}^o$	geosphere diffusion coefficient	normal	m^2/a	0.0369	0.0431
α_G	dispersivity in the geosphere	loguniform	m	2.0	200
$K_{DI}{}^G$	sorption constant in the geosphere for iodine	lognormal	m^3/kg	7.87×10^{-15}	92.0
W	water extraction rate	uniform	m^3/a	5×10^5	5×10^6
W_m	water ingestion rate	uniform	m^3/a	0.7	0.9

Table 3.5 PCCs, PRCCs, SRCs, SRRCs and R^2 Values for the PSA Output Variables, $H_{I-129}(10^5$ years) and $H_{I-129}(10^6$ years).

Input Variable	$H_{I-129}(10^5$ years)				$H_{I-129}(10^6$ years)			
	PCC	PRCC	SRC	SRRC	PCC	PRCC	SRC	SRRC
R^o	-0.04	-0.05	-0.04	-0.04	0.04	-0.00	0.04	-0.00
X_B	-0.05	-0.04	-0.04	-0.03	-0.02	0.01	-0.02	0.01
$K_{DI}{}^B$	-0.01	-0.03	-0.01	-0.03	-0.01	0.02	-0.01	0.02
X_G	-0.07	-0.19	-0.06	-0.18	0.01	-0.08	0.01	-0.08
V_G	0.21	0.28	0.20	0.27	-0.12	-0.14	-0.11	-0.14
$D_G{}^o$	-0.04	-0.06	-0.04	-0.06	0.01	-0.01	0.01	-0.01
α_G	-0.04	0.08	-0.03	0.08	0.01	0.10	0.01	0.10
$K_{DI}{}^G$	-0.00	-0.13	-0.00	-0.12	-0.01	-0.04	-0.01	-0.04
W	-0.11	0.02	-0.10	0.02	-0.14	-0.04	-0.14	-0.04
W_m	-0.00	-0.00	-0.00	-0.00	0.03	0.01	0.03	0.01
R^2	0.06	0.13	0.06	0.13	0.04	0.04	0.04	0.04

Table 3.6 Values of Quantity U_j with the Uncertain Reduction Method (URM) and Rankings of Input Varaibles for the PSA Output Variables, $H_{I-129}(10^5$ years) and $H_{I-129}(10^6$ years).

Input Variable	$H_{I-129}(10^5$ years)						$H_{I-129}(10^6$ years)					
		Ranking						Ranking				
	U_j	URM	PCC	PRCC	SRC	SRRC	U_j	URM	PCC	PRCC	SRC	SRRC
R^o	1.12×10^{-11}	10	6	6	6	6	1.43×10^{-13}	10	3	10	3	10
X_B	1.19×10^{-11}	9	4	7	4	7	1.44×10^{-13}	8	5	7	5	7
K_{DI}^B	1.24×10^{-11}	8	8	8	8	8	1.44×10^{-13}	8	8	6	8	6
X_G	2.19×10^{-11}	4	3	2	3	2	2.66×10^{-13}	5	10	3	10	3
V_G	5.54×10^{-11}	1	1	1	1	1	8.75×10^{-13}	1	2	1	2	1
D_G^o	2.26×10^{-11}	3	5	5	5	5	3.23×10^{-13}	2	9	8	9	8
α_G	1.54×10^{-11}	6	7	4	7	4	3.01×10^{-13}	4	6	2	6	2
K_{DI}^G	2.12×10^{-11}	5	10	3	10	3	2.27×10^{-13}	6	7	4	7	4
W	3.19×10^{-11}	2	2	9	2	9	3.22×10^{-13}	3	1	5	1	5
W_m	1.32×10^{-11}	7	9	10	9	10	1.76×10^{-13}	7	4	9	4	9

4. Concluding Remarks

We proposed a new technique to numerically quantify importance of input variables including uncertainties to the output variable uncertainty, and applied the technique to the two computer models, the TERFOC and the PSA models, for dose prediction. The application demonstrated effectiveness of the technique in comparing the results with those by the conventional regression method. Since the technique does not replace the original computer model to a simplified surrogate model but relies on a limited number of calculations based on the original model, the technique is expected to be widely applicable to the importance analysis for computer models when the regression and/or response surface methods could not provide sufficiently reliable results.

Finally we would point out a couple of problems in the technique:

(1) The technique still needs 2xLxK computer runs based on the original computer model, where L and K are numbers of LH samples and input variables of concern, respectively. The run number is larger by 2xK times than that needed in the conventional regression analysis based on the LHS approach. This fact makes the importance analysis using the technique proposed relatively expensive.

(2) Independency among input variables of concern has been assumed in developing the technique to estimate the importance measure based on the uncertainty reduction method. It is desired to extend the new technique to the case where there exists dependency among input variables.

Acknowledgment

The authors are grateful to K. Namatame, K. Kobayashi and H. Matsuzuru of Japan Atomic Energy Research Institute for reviewing and making many helpful remarks on the manuscript.

4. Concluding Remarks

We proposed a new technique to numerically quantify importance of input variables including uncertainties to the output variable uncertainty, and applied the technique to the two computer models, the TERFOC and the PSA models, for dose prediction. The application demonstrated effectiveness of the technique in comparing the results with those by the conventional regression method. Since the technique does not replace the original computer model to a simplified surrogate model but relies on a limited number of calculations based on the original model, the technique is expected to be widely applicable to the importance analysis for computer models when the regression and/or response surface methods could not provide sufficiently reliable results.

Finally we would point out a couple of problems in the technique:

(1) The technique still needs 2xLxK computer runs based on the original computer model, where L and K are numbers of LH samples and input variables of concern, respectively. The run number is larger by 2xK times than that needed in the conventional regression analysis based on the LHS approach. This fact makes the importance analysis using the technique proposed relatively expensive.

(2) Independency among input variables of concern has been assumed in developing the technique to estimate the importance measure based on the uncertainty reduction method. It is desired to extend the new technique to the case where there exists dependency among input variables.

Acknowledgment

The authors are grateful to K. Namatame, K. Kobayashi and H. Matsuzuru of Japan Atomic Energy Research Institute for reviewing and making many helpful remarks on the manuscript.

References

1. U.S. Nuclear Regulatory Commission, "Reactor Safety Study, An Assessment of Accident Risk in U.S. Commercial Nuclear Power Plants," WASH-1400 (1975).

2. U.S. Nuclear Regulatory Commission, "Reactor Risk Reference Document," NUREG-1150 (Draft) (1987).

3. D.M. Wuschke, P.A. Gillespie, K.K. Mehta, W.F. Heinrich, D.M. Leneveu, V.M. Guvanasen, G.R. Sherman, D.C. Donahue, B.W. Goodwin, T.H. Andres, and R.B. Lyon, "Second Interim Assessment of the Canadian Concept for Nuclear Fuel Waste Disposal - Volume 4: Post Closure Assessment," Atomic Energy of Canada Ltd Report AECL-8373-4, (1985).

4. D.C. Cox and P. Baybutt, "Methods for Uncertainty Analysis: A Comparative Survey," Risk Analysis **1**, 251-258 (1981).

5. D.J. Downing, R.H. Gardner and F.O. Hoffman, "An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models," Technometrics **27**, 151-163 (1985).

6. R.L. Iman and J.C. Helton, "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models," Risk Analysis **8**, 71-90 (1988)NUREG/TM-282 (1979).

7. M.D. Mckay, W.J. Conover, and R.J. Beckman, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics **21**, 239-245 (1979).

8. R.H. Iman and M.J. Shortencarier, "A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models," NUREG/CR-3624, SAND83-2365 (1984).

9. R.H. Myers, "Response Surface Methodology," Allyn and Bacon, Boston, MA (1975).

10. C. Park and M. Khatib-Rahbar, "Quantification and Uncertainty Analysis of Source Terms for Severe Accidents in Light Water Reactors (QUASAR), Part I - Methodology and Program Plan," NUREG/CR-4688, BNL/NUREG-52008, Volume 1 (1986).

11. S.C. Hora and R.L. Iman, "A Comparison of Maximus/Bounding and Bayes/Monte Carlo for Fault Tree Uncertainty Analysis," SAND85-2839 (1986).

12. O. Togawa, "A Computer Code TERFOC-N to Calculate Doses to the Public due to Atmospheric Releases to Radionuclides in Normal Operations of Nuclear Facilities," to be published.

13. A. Saltelli, E. Sartori, T.H. Andres, B.W. Goodwin and S.G. Carlyle, "PSACOIN Level 0 Intercomparison, An International Code Intercomparison Exersise on a Hypothetical Assessment Case Study for Raldioactive Waste Disposal Systems," OECD-NEA publication, Paris (1987).

14. R.L. Iman, M.J. Shortencarier and J.D. Johnson, "A FORTRAN 77 Program and User's Guide for the calculation of Partial Correlation and Standardized Regression Coefficients," NUREG/CR-4122, SAND85-0044 (1985).

15. M. Fisz, "Probability Theory and Mathematical Statistics," John Wiley & Sons, Inc. (1963).

16. F.O. Hoffman and C.F. Bases III (eds.), "A Statistical Analyusis of Selected Parameters for Predicting Food Chain Transport and Internal Dose of Radionuclides," ORNL/NUREG/TM-282 (1979).

Appendix A Standardized Regression Coefficient

We start with a linear regression model defined as

$$Y = a_o + \sum_{i=1}^{n} a_i X_i, \tag{A.1}$$

where a_o and a_i are regression coefficients. We assume that X_i's are independent each other. Since the coefficients a_i depend on the unit for the variables X_i, they are meaningless as a measure of importance. In order to remove the problem of different unit of measurement in the variables, let us standardize all variables as:

$$X_i \longrightarrow X_i^* = (X_i - \langle X_i \rangle) / \sigma(X_i), \tag{A.2}$$

$$Y \longrightarrow Y^* = (Y - \langle Y \rangle) / \sigma(Y). \tag{A.3}$$

Here $\langle X_i \rangle$ and $\langle Y \rangle$ are the means of variables X_i and Y, and $\sigma(X_i)$ and $\sigma(Y)$ are their standard deviations. Then Eq.(A.1) is rewritten in the following standardized form:

$$Y^* = \sum_{i=1}^{n} a_i^* X_i^*, \tag{A.4}$$

where a_i^* is called the standardized regression coefficient (SRC).

Using Eq.(A.4), the variance of Y^* is given by

$$V_{Y^*} = \sum_{i=1}^{n} a_i^{*2}. \tag{A.5}$$

The variance $V_{Y^*}(\underline{x}_j^*)$ defined by Eq.(2.5) in the text is given by

$$V_{Y^*}(\underline{x}_j^*) = \sum_{\substack{i=1 \\ (i\neq j)}}^{n} a_i^{*2} + a_j^{*2}\underline{x}_j^{*2} - (a_j^*\underline{x}_j^*)^2$$

$$= \sum_{\substack{i=1 \\ (i\neq j)}}^{n} a_i^{*2}. \tag{A.6}$$

Thus the resulting importance measure I_j is given by

$$I_j = \sqrt{V_Y^* - \int V_Y^*(\underline{x}_j^*)\, f(\underline{x}_j^*)\, d\underline{x}_j^*}$$

$$= |a_j^*|, \qquad (A.7)$$

which is nothing more than the absolute value of SRC for the variable X_j.

Appendix B Latin Hypercube Sampling Method[8]

The Latin hypercube sampling (LHS) technique provides an alternative to the conventional Monte Carlo approach. The LHS method can provide a precise estimate of the model's response to input variability with a smaller total number of calculations. This is accomplished by a constrained sampling scheme. LHS selects L different values from each of n variables X_1, ..., X_n in the following manner. The range of each variable is divided into L intervals on the basis of equal probability. One value from each interval is selected at random with respect to the probability density in the interval. The L values thus obtained for X_1 are paired in a random manner (equally likely combinations) with the L values of X_2. These L pairs are combined in a random manner with the L values of X_3 to form L triplets, and so on, until L n-tuplets are formed. This is the Latin Hypercube Sample. Thus, for given values of L and n, there exist $(L!)^{n-1}$ possible interval combinations for an LHS. It is convenient to think of the LHS as forming an L x n matrix of input where the i-th row contains specific values of each of the n input variables to be used on the i-th run of the computer model. A particular row is often referred to as an LHS input vector.