Reproduced from the journal Health Physics with permission from the Health Physics Society

A COMPARISON OF SENSITIVITY ANALYSIS TECHNIQUES

D. M. Hamby

Savannah River Technology Center, Bldg. 773-A

Westinghouse Savannah River Company

Aiken, SC 29808

ABSTRACT

Modeling the movement and consequence of radioactive pollutants is critical for environmental protection and control of nuclear facilities. Sensitivity analysis is an integral part of model development and involves analytical examination of input parameters to aid in model validation and provide guidance for future research. Sensitivities of 21 input parameters have been analyzed for a specific-activity tritium dose model using fourteen methods of parameter sensitivity analysis. This report demonstrates, for each sensitivity method, the required calculational effort, the sensitivity ranking of parameters, and the relative method performance. The sensitivity measures include the following: partial derivatives, variation of inputs by one standard deviation and by 20%, a sensitivity index, an importance index, a relative deviation of the output distribution, a relative deviation ratio, partial rank correlation coefficients, standardized regression coefficients, rank regression coefficients, the Smirnov test, the Cramer-von Mises test, the Mann-Whitney test, and the squared-ranks test.

INTRODUCTION

Engineering and scientific phenomena are often studied with the aid of mathematical models designed to simulate complex physical processes. In the nuclear industry, modeling the movement and consequence of radioactive pollutants is extremely important for environmental protection and facility control. One of the steps in model development is the determination of the parameters most influential on model results. A "sensitivity analysis" of these parameters is not only critical to model validation but also serves to guide future research.

Many of the methods available for conducting sensitivity analyses have been summarized by the author (Hamby 1994). The current paper is a comparative assessment of several methods and is intended to demonstrate calculational rigor and compare parameter sensitivity rankings resulting from various sensitivity analysis techniques. An atmospheric tritium dosimetry model (Hamby 1993) is used here as an example, but the techniques described can be applied to many different modeling problems. Other investigators (Rose 1983; Dalrymple and Broyd 1987; Iman and Helton 1988; Saltelli and Marivoet 1990; Helton 1993) present comparisons of a few of the sensitivity analyses methodologies discussed here.

SENSITIVITY ANALYSIS METHODS

The sensitivity of a tritium dose model (Hamby 1993) to 21 input parameters has been analyzed using fourteen methods of sensitivity analysis. A comprehensive review of these methods is given by Hamby (1994). A Latin hypercube sampling procedure was used to generate input to the dose model with a sample size of 1000 (Hamby 1993).

The sensitivity methods include the utilization of the following one-at-a-time sensitivity measures: partial derivatives (PD), one standard deviation increase and decrease of inputs (±SD), a 20% increase and decrease of inputs (±20%), and a sensitivity index (SI). The sensitivity measures investigated that utilize an array of input and output values generated through random sampling include: an importance index (II), a relative deviation of the output distribution (RD), a relative deviation ratio (RDR), partial rank correlation coefficients (PRCC), standardized regression coefficients (SRC), and rank regression coefficients (RRC). Four additional techniques have been used to estimate parameter sensitivity rankings based on the partitioning of input data (Crick et al. 1987). These methods include the Smirnov test (S), the Cramer-von Mises test (CM), the Mann-Whitney test (MW), and the squared-ranks test (SR). Sensitivity analyses are often referred to as either "local" or "global". A local analysis addresses sensitivity relative to point estimates of parameter values while a global analysis examines sensitivity with regard to the entire parameter distribution.

A simplistic and qualitative approach to determining parameter sensitivity is achieved by aggregating the mathematical model, i.e., algebraically combining exposure pathway models, evaluating the resulting equation using best-estimate parameter values, and assessing the relative contribution to dose via each pathway component. It is a simple task to aggregate the tritium model utilized here (Hamby 1993). Total atmospheric tritium dose to a receptor is the sum of the inhalation and ingestion pathway doses and is given by,

$$D = \left\{ \frac{4.84 \mathrm{x} 10^{-9}\ T_e\ f_w\ C^a\ R_{pa}}{M\ H} \right\} \bullet \left\{ (2.74\ U_m\ f_m\ f_{pm}\ I_m\ e^{-(\lambda t_m)}) + (2.74\ U_b\ f_b\ f_{pb}\ I_b\ e^{-(\lambda t_b)}) + (1000\ U_v\ f_v) + (1000\ U_l\ f_l) + \left(\frac{(1.5)\ BR\ H}{f_w\ R_{pa}} \right) \right\} \quad (1)$$

where the constants account for unit conversions. Distribution definitions and characteristics of parameters are given in Tables 1 and 2. The five components in the right set of brackets represent the five exposure pathways: 1) milk consumption, 2) beef consumption, 3) produce consumption, 4) leafy vegetable consumption, and 5) inhalation, respectively. It is immediately apparent that the model will be sensitive to some degree to three of the parameters in the left set of brackets (T_e, C^a, and M) since their values influence all pathway dose estimates. The three remaining parameters in the left brackets (f_w, R_{pa}, and H) cancel in the inhalation portion of the equation, therefore, they are expected to be sensitive, but to have less influence than T_e, C^a, and M, since all pathway dose estimates are not affected by their values.

Differential Sensitivity Analysis (PD). Differential analyses, also referred to as the direct method, are structured on the behavior of the model for a base-case scenario, e.g., all parameters set equal to their mean value. Differential sensitivity analysis is based on partial differentiation of the aggregated model. When an explicit algebraic equation describes the modeled relationship, the sensitivity coefficient for a particular independent variable is calculated from the partial derivative of the dependent variable with respect to the independent variable. Partial derivatives of the aggregated equation describing the atmospheric tritium model were calculated for each input parameter. Table 3 shows the numerical results of these calculations. Derivatives are multiplied by the ratio of the parameter value to the model result for the base-case scenario to normalize the results by removing the effects of units.

One-at-a-Time Sensitivity Measures (±20%, ±SD). This type of sensitivity analysis only addresses parameter sensitivity relative to the point estimates chosen for the parameters held constant. One test was conducted where the sensitivity measure was determined by adjusting parameter values by a percentage of their base-case value. The sensitivity measure (±20%) was determined by calculating the ratio of model results while the input parameter was varied by ±20%.

A more powerful test of local sensitivity examines the change in output as each parameter is individually increased by a factor of its standard deviation (±SD). This type of sensitivity measure takes into account the parameter's variability and the associated influence on model output. This test is similar to that described above except that parameters were varied by one standard deviation of their input distribution rather than 20% of their base-case value.

Factorial design (Box et al. 1978) is another one-at-a-time analyses, however, this method requires a great deal of effort when dealing with large models. A simple factorial design for the twenty-one parameters used here and only two levels requires more than two million model runs. In some cases, a fractional factorial method can be implemented to reduce the number of trials to a manageable size.

The Sensitivty Index (SI). Another of the simple methods of determining parameter sensitivity is to calculate the output percent difference when varying one input parameter from its minimum value to its maximum value. The "sensitivity index" was introduced by Hoffman and Gardner (1983) to account for all possible values when determining parameter sensitivity.

The Importance Index (II). Hoffman and Gardner (1983) introduced an "importance index" which is equal to the variance of the parameter value divided by the variance of the dependent values. For additive models the variance is of the raw data, whereas for simple multiplicative models the variance is of the log-transformed data. The model under consideration here is not a simple additive nor a simple multiplicative model. Importance indices were calculated, however, using the input and output data following a log transformation, as would be carried out for a purely multiplicative model.

The "Relative Deviation" Method (RD). This sensitivity ranking method measures the amount of variability in the model output while varying each input parameter, one-at-a-time, according to its probability density function. This method is similar to local sensitivity methods with the exception that a larger sampling is made of the input distribution. The sensitivity figure-of-merit is the "relative deviation" (RD), the ratio of the output density function's standard deviation to its mean, and is similar to a coefficient of variation.

The "Relative Deviation Ratio" (RDR). Given two model input distributions, one narrow and one wide, producing identical output distributions, the model is more sensitive to the input parameter of the narrow distribution. This one-at-a-time measure of sensitivity, therefore, is the ratio of the output distribution's relative deviation to the input distribution's relative deviation. In principle, this sensitivity statistic is very similar to the importance index calculated above. A large value of the RDR indicates that the model is sensitive to the parameter and that either the output distribution varies widely or that the variability of the input parameter is small.

The Partial Rank Correlation Coefficient (PRCC). Correlation can be determined qualitatively by a scatter plot of the independent and the dependent variable (Fig. 1), or quantitatively by calculation of a correlation coefficient, r. The larger the absolute value of r, the stronger the degree of linear relationship between the input and output values (IAEA 1989). One of the problems encountered in calculating test statistics from raw data is that the data are not necessarily linear. For this reason, parameter sensitivity was not calculated based on simple correlation.

A method of reducing the effects of nonlinear data is to use the rank transformation. If the input/output associations are monotonic then rank transformations of the input and output values (i.e., replacing the values with their ranks) will result in linear relationships and the rank correlation coefficient (RCC) will indicate the degree of monotonicity between the input and output sample values (IAEA 1989). The RCC can be calculated using the equation for simple correlations with the exception of operating on the rank transformed data (Iman and Conover 1979).

RCCs have been calculated for the correlations between each input parameter and the output. Strong correlations between input parameters, however, may influence these input/output correlations. Partial correlation coefficients (PCC) can be calculated to account for correlations among other input variables. Again, since the raw data may not necessarily be linear, sensitivity rankings using partial correlations were calculated using ranked data.

The rank transformation also can be applied to partial correlation as a test of monotonicity between input and output variables while accounting for relationships between input parameters. The partial rank correlation coefficient (PRCC) is widely utilized for sensitivity studies. Because of the difficulty in determining correlations between input parameters, many investigators assume that input correlations do not exist in model evaluations. Therefore, the use of RCCs or PRCCs will result in identical sensitivity rankings, although actual values of the correlation coefficients will differ (Table 4). The implications of various rank correlations are presented by Iman and Davenport (1980).

Standardized Regression Coefficients and Rank Regression Coefficients (SRC & RRC). The use of the regression technique allows the sensitivity ranking to be determined based on the relative magnitude of the regression coefficients. The coefficients are indicative of the amount of influence the parameter has on the model as a whole. Because of units and the relative magnitudes of parameter values, a standardization process is sometimes warranted. Standardization in regression analysis takes place in the form of a transformation by ranks or by the ratio of the parameter's standard deviation to its mean. The rank regression coefficient (RRC) is calculated by performing regression analysis on rank-transformed data rather than the raw data. The RRC is often referred to as a standardized rank regression coefficient.

Sensitivity Tests Involving Segmented Input Distributions

These statistical tests involve some form of dividing or segmenting input parameters into two or more empirical distributions based on an associated partitioning of the output distribution (Crick et al. 1987). In one of the two examples presented here, the median value of the dose distribution is chosen as the partitioning point. For a given parameter, all input data associated with a dose below the median are said to belong to one random sample while input data associated with a dose above the median belong to a second random sample. These two random samples are used to generate the empirical distributions. Means, medians, variances, and other characteristics of these empirical distributions are compared to determine whether the distributions are statistically identical. In the second numerical example, input data are partitioned according to the 90th percentile dose. Since their results are specific to the partitioning point, the sensitivity tests performed on the segmented data are not compared to the tests discussed above.

The Smirnov Test (S). The Smirnov test operates on the two empirical distributions, $S_1(x)$ and $S_2(x)$, generated as a result of partitioning a given input parameter distribution. The degree of similarity, measured as the greatest absolute difference in the vertical direction between distributions, is used to indicate the sensitivity between the input and output values (Fig. 2). The resulting sensitivity ranking is based on the distribution partitioning; ranks resulting from data partitioned based on the output median value may be different than ranks resulting from data partitioned based on the 90th percentile.

The Cramer-von Mises Test (CM). The Cramer-von Mises test is very similar to the Smirnov test in that its purpose is to determine whether two empirical distributions are statistically identical. The test statistic is the sum of all squared vertical distances between the two empirical distributions.

The Mann-Whitney Test (MW). The Mann-Whitney test, also known as the Wilcoxon test, is utilized to compare the means of two independent samples (Conover 1980). Two distribution functions are ordered as a single sample and ranks are assigned based on the ordering. The test statistic is the sum of the resulting ranks of data from one of the distributions. Since the Mann-Whitney test is two-tailed (the mean of X could be larger or smaller than the mean of Y), sensitivity ranks are based on a normalized value of T (Hamby 1994). After normalization, the smaller values of T indicate the more sensitive parameters since the means of the distributions show a greater difference based on the partitioning of input data.

The Squared-Ranks Test (SR). The variances of two independent samples can be compared using the squared-ranks test. Ranks are not based on the raw data, rather on the absolute difference between the random sample and the sample mean. For parameter ranking purposes, the normalization procedure executed on the Mann-Whitney statistic also is necessary with the squared-ranks statistic.

RESULTS

The sensitivity results for each test are given in Tables 5 and 6. Tables 7 and 8 present the sensitivity measures in terms of ranks. Since one sensitivity method does not stand out as being universally accepted as the "correct" method, a "composite" sensitivity ranking has been determined. For the sake of comparing methods, the composite sensitivity ranking is based on the sum of ranks over all methods shown in Table 7. The parameter with the lowest total rank is considered to have the greatest sensitivity. Iman and Conover (1987) have presented a measure of "top-down correlation" for similar problems.

The relative performance of each method was determined by comparing the method-specific sensitivity ranking to the composite ranking. A "performance index" was calculated for this comparison. The performance index is a test of trend and is the sum of the squared-differences of the compared ranks, the T statistic in Spearman's ρ (Conover 1980). Table 9 shows an example of the calculation of the performance index. A smaller index indicates a better trending of the method-specific and composite rank orders. The composite sensitivity ranking and the method performance ranking are shown in Table 10. Parameters are listed in decreasing order of sensitivity and the sensitivity techniques are listed in order of increasing performance index. Sensitivity ranks of the top ten parameters for each method are given in the table.

Table 11 shows the ten most sensitive parameters based on the data-partitioned sensitivity methods. Because of the nature of these sensitivity tests, their results are not compared quantitatively to the composite sensitivity ranking. For convenience and a qualitative view of their performance, however, the table lists the model parameters in the same order as the composite ranking.

The test of trend using Spearman's ρ also was used to calculate a performance index and compare sensitivity ranks between methods. These comparisons show which tests behave similarly and which tests appear to be inappropriate for sensitivity analysis, at least for the type of model considered in this work. The sums of squared-differences are given in Tables 12 and 13. Again, smaller values indicate better trending of ranks and greater parity between methods. As an example, the performance index for the comparison between the ±20% and PD methods is 1.5, indicating remarkable agreement between the two rank orders. Again, the techniques involving partitioning of input distributions have not been compared with the other methods.

DISCUSSION

As stated earlier, the performance of each method is measured by how closely the method-specific sensitivity rank compares to the composite rank. The performance index (PI) indicates that the SI and RD methods produce ranking results that are most similar to the composite rank (refer to Table 10). It is encouraging to see that all methods (except the importance index) produce the same general ranking of parameter sensitivity. The importance index is meant to be used with simple additive or multiplicative models; it is apparently not appropriate as a sensitivity measure for the model used in this example. The SI method chooses all of the top ten sensitive parameters while the RD method chooses the top six parameters in the composite order. The first five methods choose the top six parameters, but not necessarily in the composite order.

The top ten parameter sensitivity rankings determined by the partitioned data methods are given in Table 11. The ranks based on the Smirnov, Cramer-von Mises, and Mann-Whitney tests appear to be very similar to the composite rank. The three methods produce the three most sensitive parameters regardless of the partitioning point. Ranking results from the squared-ranks test are much different and show greater sensitivity to the selection of partitioning point than the others.

Table 12 provides a comparison between sensitivity methods; the PI is calculated for each combination of ten sensitivity techniques discussed. Small values of PI indicate similar sensitivity rankings. The partial derivative method is the most fundamental of the "local" sensitivity analysis techniques. It is appropriate only for relatively small changes (on the order of several percent) in the input parameter. It is not surprising, therefore, that sensitivity ranks based on the PD and ±20% methods result in very similar orders. The standard deviation increments (±SD) can at times be quite large, therefore, the ±SD ranks are not as similar. The RDR method acts globally, yet produces rankings similar to PD and ±20%. As suggested by Table 10 and confirmed in Table 12, rankings obtained from the sensitivity index (SI) and the relative deviation (RD) are quite similar. And, to a lesser degree, the SI and RD methods produce results similar to the ±SD method. Parameter sensitivity ranks based on the rank regression coefficient (RRC) are similar to the rankings from the SI, ±SD, and PRCC techniques. The importance index (II), meant for simple multiplicative models, produces results unlike any of the other methods; its utility is questionable.

The Smirnov (S), Cramer-von Mises (CM), Mann-Whitney (MW) tests all produce very similar parameter sensitivity rankings when partitioned based on the median and 90th percentile model results (Table 13). It is not surprising that the rankings based on the Smirnov and Cramer-von Mises tests are similar since the two show little difference in their statistical power (Conover 1980). Parameters can be ranked entirely different for tests operating on a partitioning point at the median and at the 90th percentile. For example, the performance indices show that the rankings between the CM test at the median and 90th percentiles are markedly different than the similaries between the CM test and the MW test, at both partitioning points. Ranking results from the squared-ranks (SR) tests are quite different than the other tests operating on partitioned data sets. The variances of the two empirical distributions generated by the partitioning process apparently are not good indicators of parameter sensitivity.

CONCLUSIONS

A number of sensitivity analysis techniques have been presented. The majority of the techniques result in similar rankings of the top several sensitive parameters. Since the actual ranking is not as important as the general ranking, most of the techniques would be appropriate for sensitivity analysis for the type of model considered in this report. The criterion most important, therefore, is the ease with which the sensitivity method can be performed. With the proper software, all methods presented here are relatively easy to execute. Given a moderate number of parameters and a hand calculator, however, the sensitivity index is the easiest and most reliable sensitivity measure. The SI can be calculated without detailed knowledge of the parameter distribution and without the use of random sampling schemes or large computer programs.

The relative deviation (RD) is a reliable measure of parameter sensitivity. Calculation of the RD is quite simple if a sampling technique is employed and the output values are stored for the statistical analysis. This analysis requires a one-at-a-time approach, however, and can be labor intensive. Estimating sensitivity based on the relative deviation ratio (RDR) is not recommended since its results are less reliable and it requires more calculational rigor than the RD.

Rank regression coefficients are easily obtained with the use of commerically available software. An electronic spreadsheet and the SAS* statistical package were utilized for this analysis. The calculation of sensitivity rankings by varying the parameter over its standard deviation (±SD) is as simple as calculating the sensitivity index with the exception that some knowledge of the parameter distribution must be available. Varying the input parameter by a standard amount (±20%) is an easy test to perform, but its reliability is less desirable than the simpler SI method.

The simplest approach to conceptualize is the one-at-a-time method where sensitivity measures are determined by varying each parameter independently while all others are held constant. These sensitivity techniques, however, become rather time intensive with large numbers of parameters. The most fundamental of sensitivity techniques is the direct method of using partial differentials to calculate the rate of change in the model output with respect to a given input parameter. The one-at-a-time techniques are valid only for small variability in parameter values and the partials must be recalculated for each change in the base-case scenario.

Crick et al. (1987) utilize statistical analysis of selected input vectors by segmenting input values based on their relationship to some critical output value (e.g., the mean, median, or 90th percentile of the output distribution). This type of analysis provides detailed information on parameter sensitivity based on the calculated output. The non-parametric tests on partitioned data sets are very labor intensive yet are not necessarily beneficial unless a particular question is to be answered regarding the sensitivity of a parameter with respect to model output.

Acknowledgement. The author would like to thank Drs. L.R. Bauer and W.H. Carlton for their critical reviews of this manuscript. This work was supported by the U.S. Department of Energy under Contract No. DE-AC09-89SR18035 with the Westinghouse Savannah River Company.

FOOTNOTES

* SAS Institute Inc., Cary, North Carolina. 27511

REFERENCES

Box, G.E.P.; Hunter, W.G.; Hunter, J.S. Statistics for experimenters: an introduction to design, data analysis, and model building. New York: John Wiley & Sons; 1978.

Conover, W.J. Practical nonparametric statistics. 2nd Ed. New York: John Wiley & Sons; 1980.

Crick, M.J.; Hill, M.D.; Charles, D. The role of sensitivity analysis in assessing uncertainty. In: Proceedings of an NEA workshop on uncertainty analysis for performance assessments of radioactive waste disposal systems. Paris: OECD; 1987: 1-258.

Dalrymple, G.J.; Broyd, T.W. The development and use of parametric sampling techniques for probabilistic risk assessment. IAEA-SR-111/54P. In: Cullingford, M.C.; Shah, S.M.; Gittus, J.H. eds. Implications of probabilistic risk assessment. London: Elsevier; 1987.

Hamby, D.M. A probabilistic estimation of atmospheric tritium dose. Health Physics. 65:33-40; 1993.

Hamby, D.M. A review of techniques for parameter sensitivity analysis. Accepted for publication in Environmental Monitoring and Assessment. 1994.

Helton, J.C. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering and System Safety. 42:327-367; 1993.

Hoffman, F.O.; Gardner, R.H. Evaluation of uncertainties in environmental radiological assessment models. In: Till, J.E.; Meyer, H.R. eds. Radiological assessments: a textbook on environmental dose assessment. Washington, DC: U.S. Nuclear Regulatory Commission; Report No. NUREG/CR-3332; 1983.

Iman, R.L.; Conover, W.J. The use of the rank transform in regression. Technometrics. 21:499-509; 1979.

Iman, R.L.; Conover, W.J. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. Communications in Statistics - Theory and Methods. A9:1749-1842; 1980.

Iman, R.L.; Conover, W.J. A measure of top-down correlation. Technometrics. 29:351-357; 1987.

Iman, R.L.; Davenport, J.M. Rank correlation plots for use with correlated input variables in simulation studies. Sandia National Laboratories. Albuquerque, NM: Report No. SAND80-1903; 1980.

Iman, R.L.; Helton, J.C. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis. 8:71-90; 1988.

International Atomic Energy Agency. Evaluating the reliability of predictions made using environmental transfer models. Vienna: Safety Series No. 100. Report No. STI/PUB/835; 1989, 1-106.

Rose, K.A. A simulation comparison and evaluation of parameter sensitivity methods applicable to large models. In: Lauenroth; Skogerboe; Flug eds. Analysis of ecological systems: State-of-the-art in ecological modelling. New York: Elsevier; 1983.

Saltelli, A.; Marivoet, J. Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques. Reliability Engineering and System Safety. 28:229-253; 1990.

CAPTIONS

Fig. 1 Sample scatter plot showing the relationship between one input parameter and the model output. Plot is of individual dose from tritium exposure (all pathways) versus tritium oxide concentration in air (Hamby 1993).

Fig. 2 Example of the graphical determination of the Smirnov statistic. The value of T1 is a measure of the disparity of the two empirical distributions.

Sensitivity analysis is an analysis technique that works on the basis of what-if analysis like how independent factors can affect the dependent factor and is used to predict the outcome when analysis is performed under certain conditions. It is commonly used by investors who takes into consideration the conditions that affect their potential investment to test, predict and evaluate result. Sensitivity Analysis Formula. The formula for sensitivity analysis is basically a financial model in excel where the analyst is required to identify the key variables for the output formula and then assess t