A Formal but Non-Automated Method to Test the Sensitivity of System Dynamics Models

Jonathan D. Moizer[1], Dan Arthur[2] and Ian Moffatt[3]

[1]Plymouth Business School, University of Plymouth, England,
Telephone: 0044 1752 232834
E-mail: Jonathan.Moizer@pbs.plym.ac.uk;

[2]Department for Business Development, University of Plymouth, England,
Telephone: 0044 1752 233522
E-mail: D.Arthur@plymouth.ac.uk;

[3]Department of Environmental Science, University of Stirling, Scotland
Telephone: 0044 1786 467854
E-mail: Ian.Moffatt@stir.ac.uk.

Abstract

Sensitivity testing of parameters can add greatly to the validity of a system dynamics model. Most model builders view parameter sensitivity tests as confirming whether a small perturbation to a parameter's numerical value results in a significant change in the model's behaviour. The results of these tests can indicate the level of accuracy that is required when assigning numerical values to a model's parameters, and also narrow down the search for improved policy.

It can be impractical to run a sensitivity analysis on a trial and error basis because of the large number of permutations that exist. There are various strategies for approaching the sensitivity testing task and these are reviewed. A formal and straightforward process for analysing the sensitivity of system dynamics models is proposed. A range of single parameter sensitivity tests is performed on all model parameters. Static and behavioural performance measures are compared using Spearman's Rank Correlation Coefficient to measure the congruence between the results of the separate tests.

Keywords

system dynamics; sensitivity testing, formal, non-automated

Introduction to the Study of Sensitivity Testing

System dynamics has had criticism levelled at it because of its relatively informal, subjective and qualitative validation procedures. They are more relativistic and take multiple approaches to confidence building in comparison with traditional operational research methods. The criticisms have been levelled by people more familiar with hard input-output models where statistical measurement of model output is the principal determinant of model confidence. Building confidence in system dynamics models requires a range of on-going tests to be performed on a model to examine its structure, behaviour and policy. Sensitivity testing is one aspect of establishing validity or confidence building in a model. It is concerned with examining the behaviour of a model. Normally, this involves searching for

instances where a small numerical change to a parameter results in a significant change in a model's behaviour.

This paper will introduce the background on the use of sensitivity testing as a means of building confidence in system dynamics models. The method of developing and testing a formal and non-automated method for sensitivity testing is outlined. Finally, the perceived benefits and also limitations of this exposition are raised.

The Scope of Sensitivity Testing of System Dynamics Models

Sensitivity testing of the parameters of a system dynamics model has a number of uses:

➢ It can help to narrow down those areas where more data gathering would be useful. It can be used to set a priority for data collection and the associated level of accuracy required.
➢ It can assist with improving understanding of complex problems being modelled, in particular help the modeller understand the structure-orientated behaviour of a model.
➢ It can be used to identify the pressure points in a model where the potential for improved behaviour lies.

Sensitivity testing of the parameters of a system dynamics model is essential for a number of reasons:

➢ As system dynamics models are populated by feedback loops and non-linearity, the relationship between a model's structure and behaviour is complex. It is not always obvious prior to running a simulation which parameters the model is actually sensitive to. This can only be determined by inspection of model outputs post-simulation. A proportional change in an input is unlikely to lead to the same proportional change in the output.
➢ Many system dynamics models use soft variables and associated parameters. These parameters represent softer, less easily measured factors which are not precisely known and are hard to measure. Therefore, the effects of numerical changes to these parameters may have to be more fully examined.
➢ A well constructed and robust system dynamics model exhibits behaviour that is often insensitive to most parameter changes. If a model is robust or stable, i.e. where the sources of instability are reduced through the introduction of negative feedback loops, the model behaviour is often insensitive to most but not all parameter changes. It is vital though to locate these leverage points where sensitivity exists for designing system improvements.

Sensitivity testing allows an exhaustive analysis of the effects of parameter change(s) on model behaviour and performance. These measures can be dynamic or static, and of course tests can be continued until time, money, effort and even sanity are expended.

Developments in Sensitivity Testing of System Dynamics Models

Sensitivity testing of system dynamics models has been a subject addressed by a number of popular authors (e.g. Forrester and Senge, 1980; Tank-Nielson, 1980, Richardson and Pugh, 1981). These earlier authors have emphasised the purpose and importance of sensitivity testing. They discuss non-automated or manual methods for analysis of sensitivity. The awareness of sensitivity appears to be built up through less formal, and more experimental or

intuitive means. Learning about sensitivity through experimentation appears to be most important. Raiswell (1978) developed a formal but non-automated method of sensitivity testing. Monte-Carlo sampling is used to select single parameter values from a predefined probability distribution. Formal automated techniques have also been developed. These include the use of Latin Hypercube Sampling (Clemson *et al*, 1995) and Taguchi methods (Ford *et al*, 1983) which allow multiple parameter sensitivity tests through structured sampling strategies. The strength of such automated sensitivity techniques lies in their ability to identify a range of sensitivity values through simulating combinations of parameter changes. Taguchi involves a different parameter sampling method which can be more efficient than Latin Hypercube Sampling in instances where there is no strong interdependencies or non-linearities. Kleijen (1995) developed a formal approach to the design of experiments, using regression analysis for looking at interactions between variables. The regression analysis is used to design sensitivity experiments by selecting a partial factorial set of parameter combinations.

Performance Metrics and Indices

A performance index can be used as a relative measure outputs from a system dynamics model. Coyle (1978) sets out a method which uses a weighted combination of final values of a run to be taken, less instability penalties. This is a convenient way to compare one simulation run with another. The performance index is a single number which summarises the whole performance of a model run. The measure of a whole simulation run is condensed into a very simple form. It can be a useful approach, particularly where the difference between behavioural outputs are not visually significant. The idea of a performance index could be taken and used to assist with measuring the sensitivity of a system dynamics model.

Method Developed to Test the Sensitivity of a System Dynamics Model

Scholl's (1995) benchmarking survey of the system dynamics community suggested that there was inconsistent use of confidence tests. Given that Forrester and Senge (1980) suggested that behaviour and policy sensitivity tests are 'core' tests, less than 60 percent of respondents indicated that they use sensitivity analysis as a confidence building test. Is it possible that this test is not universally applied because there is no method or set of methods commonly accepted and applied across the system dynamics community? Given this notion, it was worth investing some time in developing a simple and transferable method of analysis. A formal and straightforward process for analysing the sensitivity of system dynamics models is proposed in which a range of single parameter sensitivity tests are performed on all model parameters. The results of static and behavioural performance measures are compared using Spearman's Rank Correlation Coefficient. This statistical test is applied in order to measure the congruence between the results of the separate tests.

The method employs a formal manual means of identifying model sensitivity to parameter change. Given that the Ithink (High Performance Systems 1994) software used was not available to support sensitivity testing using the system dynamics package used, then this method was developed. The purpose of the sensitivity test was two-fold. Firstly, to discover which parameters needed to be accurately validated with real world data when a model is empirically tested; and secondly to obtain an idea of where policy improvement may lie in a model.

The testing is applied to single parameter values. It is not impossible but rather impractical to apply this method to table functions, as these are collections of parameters. Most system dynamics models usually contain many parameters. It would be deemed impractical to conduct multiple parameter tests manually, given the huge range of permutations. Using single parameter testing, the effects of each parameter change could be precisely measured. A base run was set to replicate a state of equilibrium. This would allow more precise comparison to be made between alternative simulation runs.

Multiple simulation performance measures were employed. The results of a range of behavioural and point sensitivity measures were collated for each run. A number of model outputs were selected as performance metrics. Each output assumes equal weighting when used to analyse overall performance or sensitivity. A unified index was then used to compare the variability in performance of any given run against the base run. Performance was measured by comparing the change in outputs over the change in inputs. This measurement was referred to as the 'gearing ratio'. It was used as a normalised measure of sensitivity or performance. Changes in output were measured against the base run for the model. Finally, Spearman's Rank Correlation Coefficient[1] was applied to the results, in order to test the level of congruence between the range of sensitivity measures.

Using different sensitivity tests and sensitivity performance measures the method should help to identify whether a pattern emerges amongst the parameter sensitivities, i.e. is the model sensitive to the same parameters, despite different sensitivity tests? Spearman's Rank Correlation Coefficient test helps to answer this question by comparing each set of results against each other.

A Straightforward Manual Method of Sensitivity Analysis

Two sensitivity tests are conducted which result in three measures of sensitivity (see Coyle, 1977 for range of different measures of model performance). The first test is a 'final value test', where a fixed change is made to a parameter at the outset of a simulation run, and the final value of the output noted. This measure is represented in Figure 1. The second test is an 'equilibrium disturbance test'. Two measures of sensitivity are taken: the time for the output to settle within x percent of its final value following a disturbance, and the maximum deflection from equilibrium.
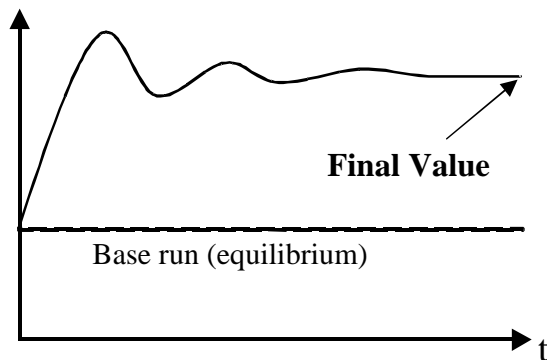

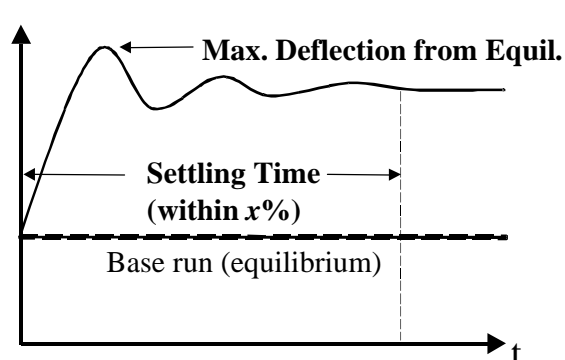**Figure 1:** Final value test   **Figure 2:** Equilibrium disturbance test

The desire is to test each parameter over a wide range of values. A specific proportional change to each parameter is introduced for both sets of tests. A range is set for the change to

the parameter. Within that range, a gradation is specified and this is named the 'adjustment fraction'. The results of each model run are compared against the base or equilibrium run. The sensitivity for all parameters tested are ranked. Therefore, three sets of ranked data exist. Each set of ranked data is compared against each other set to identify the strength of correlation between the results of the tests. Strong correlation should indicate more robust sensitivity findings.

Application Case Study of the Method

A generic occupational safety model had been developed using the system dynamics method. This work contributed towards a doctoral thesis (Moizer 1999; Moizer and Moffatt, 2000). It was populated with synthetic data and was purported to represent a safety management system across a variety of workplaces. The model was to be presented to a potential host firm. The intention was to subsequently validate the model with real world data from the firm and calibrate it to represent the typical safety system behaviour it experienced. The model has been capable of simulating a number of modes of behaviour but for comparative purposes was set to replicate a state of equilibrium for the duration of the sensitivity testing exercise.

Sensitivity testing would help in translating the generic model into a real world model in two ways. The tests would identify the parameters which needed to be accurately validated in the subsequent empirical study. Also, an early idea of the range of future scenario tests could be gained.

For the sensitivity tests a range of +/-100% was set around the base run values of the parameters. This incorporates a strong measure of extreme behaviour testing. In instances where division by zero would be evident, the parameter was taken down to one percent of its base run value. A moderate level of granularity was used for proportional changes to parameters. Each parameter under test had its numerical value varied by 25% for each new simulation run. The percentage change from the base run value was termed the 'adjustment fraction'. Eight simulation runs were performed to test each parameter, with six performance metrics used, one from each sector of the model. Using a range of metrics from various parts of the model allowed both upstream (e.g. employee safety awareness) and downstream (e.g. accidents) measures of performance to be made. Therefore, eight sensitivity runs were performed on each parameter, with six output metrics; this produces 48 final value outputs. An example of the test results for one parameter is shown in Table 1.

Metric	Adjustment Fraction							
	-100%	-75%	-50%	-25%	+25%	+50%	+75%	+100%
Cumulative Accidents	45	56	71	87	121	153	197	245
Average KSA	4.60	4.43	4.28	4.14	3.87	3.75	3.64	3.53
Actual Length of Employment	120	120	120	120	120	120	120	120
Cumulative Accident Reports	46	57	72	87	120	123	123	124
RBAAIH	0.04	0.04	0.04	0.04	0.05	0.08	0.12	0.16
Cumulative Safety Costs	249502	250612	252108	253681	257095	260286	264729	269547

Table 1: Raw values for a parameter *x* tested across the range of values

The three measures of parameter sensitivity are:

1. final value (FV);
2. maximum deflection from equilibrium (MDFE); and
3. settling time following disturbance (STFD).

For 1. and 2. above, 'gearings' are produced from the results through change in output divided by change in input:

$$\frac{\Delta \text{Output}}{\Delta \text{Input}} \quad \text{where;}$$

$$\Delta \text{Output} = \left(\frac{|\text{New Run Final Value - Base Run Value}|}{\text{Base Run Value}} \right)$$

$$\Delta \text{Input} = \text{Adjustment Fraction}$$

$$\frac{\Delta \text{Output}}{\Delta \text{Input}} \quad \text{where;}$$

$$\Delta \text{Output} = \left(\frac{|\text{Maximum Deflection from Equilibrium - Base Run Value}|}{\text{Base Run Value}} \right)$$

$$\Delta \text{Input} = \text{Adjustment Fraction}$$

Table 2 shows the raw values converted into geared values.

Metric	Adjustment Fraction							
	-100%	-75%	-50%	-25%	+25%	+50%	+75%	+100%
Cumulative Accidents	0.56	0.61	0.62	0.63	0.70	0.97	1.22	1.38
Average KSA	0.15	0.14	0.14	0.14	0.13	0.13	0.12	0.12
Actual Length of Employment	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cumulative Accident Reports	0.55	0.59	0.60	0.61	0.66	0.38	0.26	0.20
RBAAIH	0.20	0.27	0.40	0.80	0.00	1.20	1.87	2.20
Cumulative Safety Costs	0.02	0.02	0.03	0.03	0.03	0.04	0.05	0.06

Table 2: Geared values for a parameter *x* tested across the range of values

In the measurement of settling time, gearings were not necessary, as no comparison was being made with the base run. The magnitude of the gearing is a good indicator of the model's sensitivity to parameter change.

Sixteen parameters in total were tested for sensitivity. Their mean overall sensitivity was determined (i.e. for the 48 recorded values), and then these means were ranked in order of sensitivity for each set of results. The Spearman's Rank Correlation Coefficient Test was applied so as to determine whether the same parameters were sensitive across the three sets of measures. The final results are shown in Table 3.

Parameter	Mean FV Rank	Mean STFD Rank	Mean MDFE Rank	Grand Mean Rank	Overall Rank Order
Accident Reporting Policy	9	12	2	7.67	7
Accident Reporting Time	15	16	13	14.67	14
Base Length of Employment	1	1	1	1.00	1
Fixed Proportion of Knowledge Lost	12	6	12	10.00	11
Full Hazard Regulation Policy	4	10	6	6.67	6
Full Hazard Regulation Time	8	13	10	10.33	12
Full Hazard Regulation Weighting	10	9	5	8.00	8=
Intermediate Hazard Regulation Policy	6	11	7	8.00	8=
Intermediate Hazard Regulation Time	14	17	15.5	15.50	15
Intermediate Hazard Regulation Weighting	7	7	4	6.00	5
Learning Delay	18	18.5	18	18.17	18=
Perceived Accident Incidence Smooth	18	14	18	16.67	17
Ratio Between Hires and Average KSA	13	8	14	11.67	13
Ratio Between Quits and Average KSA	11	5	11	9.00	10
Safety Monitoring Policy	16	15	15.5	15.50	15
Staff Adjustment Time	18	18.5	18	18.17	18=
Training Effectiveness	3	3	9	5.00	4
Training Policy	2	4	8	4.67	3
Unregulated Hazard Regulation Weighting	5	2	3	3.33	2

Table 3: Spearman's rank correlation coefficient test summary rankings

The levels of significance of the correlation coefficients were then tested to further establish the reliability of the results.

Summary of the Case Study Results

The test method established that a minority of parameters have a significant effect upon model behaviour, with the majority of parameters having little or no effect. A pattern emerged amongst the results. The same parameters were generally sensitive or insensitive for all three sets of measures. The significance of the Spearman's Rank Correlation results further confirmed these patterns.

Benefits of the Method

There is logic in this method with clear steps involved. It is simple to perform the tests and collate the results. Like for like comparisons of parameter sensitivities can be applied. As the results are normalised, then all parameter sensitivities can be compared and subsequently ranked. A range of test results produces more comprehensive data on model performance. Dynamic and point measurements of sensitivity should be preferential to one single measure, as in some instances a model may be behaviourally insensitive but numerically sensitive to parameter change. Through using a range of tests it is easier to identify where the search for data is important, and of course offers an idea of where policy improvement may lie. This method reduces some of the monotony, yet retains learning through simulation. The formalised approach takes some of the drudgery and most of the intuition out of the testing. A spreadsheet is set to task in performing the analysis of the results. Yet because the modeller is still very much engaged in the process of sensitivity testing, then learning through experimentation is still retained.

Limitations of the Method

A number of limitations are associated with this method of sensitivity testing. The base run selected will have an effect on the sensitivity results. It is set to simulate an equilibrium state but still runs at an arbitrary level. This could indicate results which are misleading. The drawback with using Spearman's Rank Correlation to compare multiple results of parameter sensitivity tests is the fact that the results are classified ordinally. As a result, it can not be suggested which parameters are very insensitive or sensitive. Only the order can be ranked. Comparing absolute values would be more informative. This may though be a good reason for ensuring that the method is not fully automated to avoid making erroneous conclusions over sensitivity. This method could be seen as somewhat cumbersome. The method is easy but has a large element of repetition. In the case example, six metrics or output measures of sensitivity were selected. Was this too many or too few? Was the test range too extensive or narrow? Were the proportional changes to parameters between runs too coarse or too fine? It was not as likely to have been too narrow, as the incorporation of some extreme behaviour testing would require a wide range. The sensitivity of the granularity could be tested at a future point. The introduction of some automation would reduce time and effort, albeit potentially reduce learning and understanding. Powersim (1998) Application Programmer's Interface may assist with partially automating this method.

Summary of the Sensitivity Method

These behavioural and point sensitivity tests have been used to discover which parameters might have a bearing on the overall model sensitivity. The tests were able to identify a number of sensitive parameters. The range of sensitivities exhibited by the parameters appears to be plausible, as they fit a definite pattern. These test results could assist with

building further confidence in the model. Effort could be concentrated on carefully setting the numerical parameters which have been shown to be most significant. The policies most likely to offer greatest leverage over the problem under study are now also better known. This should aid the search for effective policy decisions.

References

Clemson, B., Yongming, T. Pyne, J. and Unal R. 1995. Efficient methods for sensitivity analysis. System Dynamics Review 11 (1): 31-50.

Coyle, R.G. 1977. Management System Dynamics. Wiley: London.

Coyle, R.G. 1978. An approach to the formulation of equations for performance indices. Dynamica 4 (2): 62-81.

Ford, A., Amlin, J.S. and Backus, G.A. 1983 A practical approach to sensitivity testing of system dynamics models. International System Dynamics Conference, Chestnut Hill, MA; 261-280.

Forrester, J.W. and Senge, P. 1980. Tests for building confidence in system dynamics models. In Studies in the Management Sciences: System Dynamics 14, Legasto A.A., Forrester J.W. and Lyneis J.M. (eds.). North-Holland Publishing: Amsterdam; 209-228.

High Performance Systems. 1994. Ithink 3.0 technical documentation. High Performance Systems: Hannover NH.

Kleijnen, J.P. 1995. Sensitivity analysis and optimisation of system dynamics models: regression analysis and statistical design of experiments. System Dynamics Review 11 (4): 275-288.

Moizer, J.D. 1999. System dynamics modelling of occupational safety: a case study approach. Doctoral thesis. University of Stirling: Stirling UK.

Moizer, J.D. 2000. Learning and policy making in occupational safety using a dynamic simulation. International Conference on Systems Thinking in Management, Deakin, Australia; 450-455.

Powersim. 1998. Reference Manual. Powersim Press: Reston VA.

Raiswell, J.E. 1978. Sensitivity analysis revisited. Dynamica 4 (2): 82-88.

Richardson, G.P, and Pugh A.L. 1981. Introduction to System Dynamics Modeling. Productivity Press: Portland OR.

Scholl, G.J. 1995. Benchmarking the System Dynamics Community: Research Results. System Dynamics Review 11 (2): 139-155.

Tank-Neilsen, C. 1980. Sensitivity analysis in system dynamics. In Elements of the System Dynamics Method, J. Randers (ed.). Productivity Press: Cambridge MA; 185-201.

---

[1] This coefficient is also known as the rank correlation coefficient. It is a measure of the extent of an association between two variables when the variables are ranked.