Minicap
Multi Parameter Sensitivity Analysis

Toby Friend, Glasgow iGEM team

September 6, 2007

1 Abstract

The Multi-Parametric and Initial concentration analysis Sensitivity (Minicap) package is a Matlab function which executes a chosen Dynamic or Stochastic system for a defined number of different variable values across any desired range. The *subject* of analysis can either be constants in the user's system (eg. parameters in a biological system) usually termed MPSA, or initial values of the variables in the user's system (eg. initial substrate concentrations) termed ISCSA. The program will output a plot for each variable showing a comparison of *acceptable* and *unacceptable* samples across the subject's range with 3 calculated quantitative comparison figures: the *Correlation Coefficient*, the *area* between acceptable and unacceptable curves, and the *standard deviation of the gradient* of the acceptable plot. The comparative and intrinsic sensitivity of each chosen subject is thus highlighted. A plot showing the trend of the *Substrate-of-Interest* over time is also displayed.

2 Introduction

Multi-Parametric Sensitivity analysis is a general purpose technique that is often used to analyze how sensitive a system is with respect to the change of parameter values [3]. Minicap was developed specifically to aid the Molecular Biology industry in pre-experimental planning and post-experimental discussion with regards to these sensitivities. Generally, sensitivity analysis, in this way is not particularly new. However, the program developed and discussed in this paper hopes to provide an innovative user-friendly approach to carrying out the analysis.

2.1 Minicap's use in Pre-experiment

Once a mathematical model has been decided upon which is most suited to represent a particular biological system, it is often considered necessary to judge the importance of accurately determining the parameter values used. Parameters that are determined to be highly sensitive can be restricted more securely than those which are determined to be insensitive. In this way, preparation can be efficient and thorough.

2.2 Minicap's use in Post-experiment

Once experimental data has been retrieved, the data must be recorded in a satisfactory manner, so conclusions can be accurately made. No parameter can be absolutely recorded, only measured with a finite accuracy. It is therefore useful to have an idea of how much the possible variance of recorded parameters might affect the system.

3 Approach

3.1 General

In general, the Multi-Parameter Sensitivity Analysis (MPSA) approach can be broken down as follows [1]:

1. Select the parameters to be tested.

2. Set the range of each parameter to include the variations experienced in the field and laboratory measurement.

3. For each selected parameter, generate a series of, for example, 500 independent random numbers with a uniform distribution within the design range.

4. Run the model using the selected 500 parameter sets and [run the model for some observed or 'true' parameter values, then]calculate the objective function values.

5. Determine whether the 500 parameter sets are 'acceptable' or 'unacceptable' by comparing the objective function values to a given criterion (R).

6. Statistically evaluate parametric sensitivity. For each parameter, compare the distributions of the parameter values associated with the acceptable and unacceptable results. If the two distributions are not statistically different, that parameter is classified insensitive; otherwise the parameter is classified as sensitive. Relative importance can be evaluated statistically if desired.

3.2   Specific

The Minicap program designed by the current author varies in scope and ability on a number of points from the general approach(the Matlab variable names are given in brackets):

1. Minicap's Primary function is to analyse the sensitivity of parameters in substrate/enzyme systems (known as Multi-Parametric Sensitivity Analysis, MPSA); a similar method can, however, be applied to analyse the sensitivity over a range of initial substrate concentrations (ISCSA). The user's choice is denoted the *subject*.

2. The program's output plots can be cumulative or non cumualtive (Cum).

3. The program's output plots can be normalised or not ($Norm$).

4. If a normalised, cumulative output is requested by the user, the program calculates 3 quantitative comparison figures:

   (a) Correlation Coefficient between acceptable and unacceptable curves.
   (b) Area between the curves.
   (c) Standard deviation of the gradient of the acceptable plot.

5. The program also has two distinct methods of deducing the 'true' results curve from which to compare samples ($Compare$):

   (a) Creating a 'true' plot of the results by running the System Function with optimised or analytically determined parameters.
   (b) Inputting actual concentration vs time values (most likely gleaned from experiment).

4   Algorithm

For a given System Function, x with a defined number of parameters, p and initial concentration values, i, Minicap carries out the following steps.

4.1   LHSdesign

First a sample matrix is calculated which reads a user-defined number of samples ($numSamp$) from p/i-dimensioned hypercube (depending on the subject of analysis) eg. for parameter analysis of a 3-parameter model, a cube is produced with dimensions according to the specified range of each parameter (see fig. 1).

The specified number of samples are randomly distributed inside the cube. Each sample therefore has a unique coordinate with a specific value for each parameter.
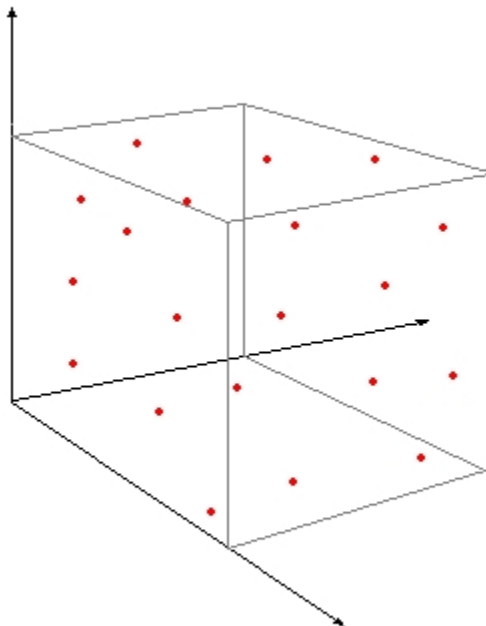
Figure 1: 3D sample cube with 20 samples

4.2   System Function

The System Function must suit either the **dynamic** ('*ode*') or **stochastic** ('*sto*') solver currently available in the 'simu' library file.

1. Dynamic simulation ('ode') uses **Ordinary Differential Equation** methods to solve a given system of a similar form to:

$$\frac{d[A]}{dt} = k_1[A] \tag{1}$$

   ie. the system is homogeneous. This approach should be used in systems with a satisfactorily large quantity of molecules, this implies the system can be treated as infinitely divisible, the '*molecules don't talk*'. [2]

2. Stochastic modelling ('sto') can be approached in a number of ways; Minicap invokes the **Gillespie Algorithm** to solve stochastic functions.

   In a given reaction system with v reactions, the hazard (probability) for a type i reaction is $h_i(x, c_i)$, so the hazard for a reaction of some type occurring is:

$$h_0(x, c) \equiv \sum_{i=1}^{v} h_i(x, c_i) \tag{2}$$

The time to the next reaction is $Exp(h_0(x, c))$, and also that this reaction will be a random type, picked with a probabilities proportional to to the $h_i(x, c_i)$, independent of the time to the next event. That is, the reaction type will be i with probability $(h_0(x, c))/h_i(x, c_i)$. Using the time to the next event and the event type, the state of the system can be updated, and simulation can continue. [4]

4.2.1 Random Sample Data and 'True' Data

The System Function is used in two ways, one to calculate **random sample data**, the other to calculate observed or **'true' data**. Data termed 'true' implies the values have been observed through experiment, retrieved form literature or calculated analytically. The 'true' data corresponds to the estimated or observed trend of output concentration of the *Substrate of Interest*. Minicap can read two types of 'true' data:

1. A vector of Parameters corresponding to the Parameters used in the System Function.

2. A vector of *Substrate-of-Interest* concentrations corresponding to specific time points throughout the system experiment.

Generally the 'true' value for a parameter or substrate concentration will fall within the variable range of the sensitivity analysis.

4.2.2 Computation

1. The System Function x for variable group j, will be solved. For each sample of variables, $p(j)$ or $i(j)$ depending on *subject*, $x(j)$ is found across the time range, t_0, and the *Substrate-of-interest's* concentration, $interestConc$, is recorded as x_0.

2. To simulate the 'true' trend of the System Function, x must use constant values c. Across the time range, t_0, and the *Substrate-of-interest's* concentration is recorded as x_c.

4.3 Objective function

The comparison method between the random sample simulation x_0 and the 'true' simulation x_c is a sum of the squared distances in the y-axis of the acceptable and unacceptable curves across t_0 for each sample point.

$$f = \sum_{i=1}^{n} [x_0(i) - x_c(i)]^2 \tag{3}$$

4.4 Nature

A sample's nature determines whether it is *acceptable* or *unacceptable* as a possible set of parameter or initial concentration values to correspond with the 'true' data.

4.4.1 Threshold

The nature of each sample is found via a comparison (criterion 'R') with a threshold value. The *threshold*, in the case of Minicap, is the mean value of the Object Function.

$$mean(OF) = threshold \tag{4}$$

5 Output

Minicap can return results in a number of formats.

5.1 Cumulative or Non-Cumulative frequency

Multi-Parametric Sensitivity Analysis is usually observed from a cumulative perspective. This allows for more quantitative analysis of curve comparison. If a normalised, cumulative frequency plot is requested from Minicap, it will display a plot for each variable with a format like that shown in fig. 2.
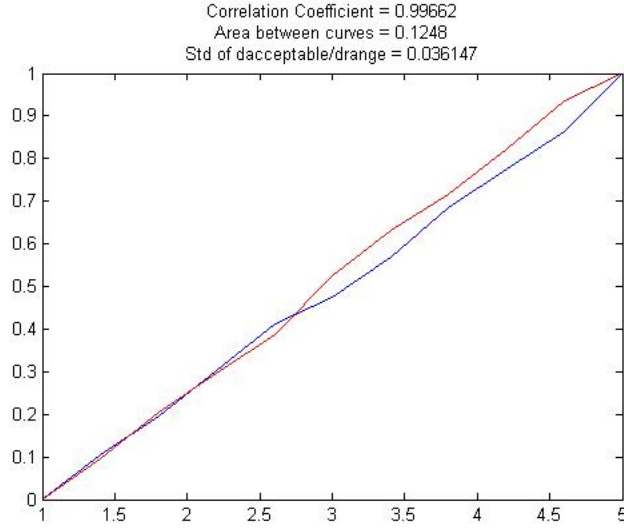
Figure 2: Cumulative example

As can be seen from fig. 2, the cumulative format returns 3 values for qualitative comparison of parameters.

1. The **Correlation Coefficient** (ρ) indicates the strength and direction of a linear relationship between the acceptable and unacceptable curves. It is calculated by dividing the *Coefficient of Variance* by the product of *Standard Deviations* of the two curves.

$$\rho_{a,u} = \frac{cov(a, u)}{\sigma_a \sigma_u} \tag{5}$$

2. The **Area between curves** also signifies how different the acceptable and unacceptable plots are from each other but in a less transferable comparison. It is calculated by first determining whether acceptable or unacceptable plots have maximum frequency for each interval across the variable range. At each interval, the minimum frequency is subtracted from the maximum and the difference multiplied by the width of the interval (similarly to the *Trapezium Rule*). Fig. 3 shows how the plots are divided into discrete intervals. The number of intervals corresponds to the user defined *definition*.

Correlation Coefficient = 0.99662
Area between curves = 0.1248
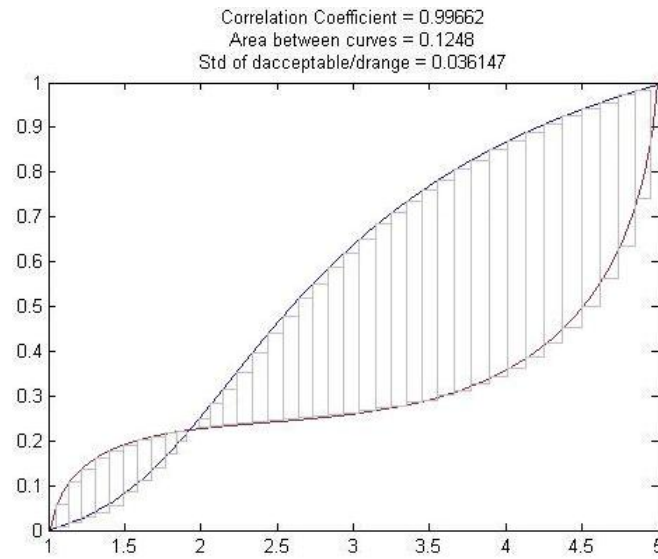Std of dacceptable/drange = 0.036147

Figure 3: Area calculation example

3. One way of highlighting the consistency of acceptable samples over the parameter range is to calculate the **Standard Deviation of the gradient of the acceptable results**; this is the third quantitative value presented by Minicap for normalised, cumulative frequency plots.

It may, however, be desirable to view the plots non-cumulatively to observe the data in a more qualitative manner (see fig. 4).
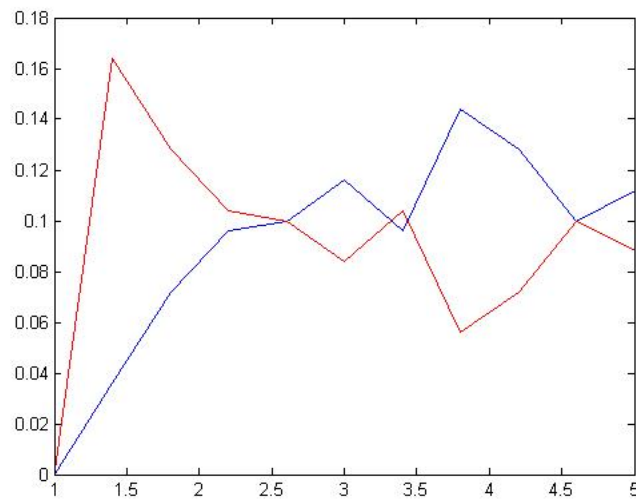

Figure 4: Non-Cumulative example

Here we see the same simulation as in fig. 2 but non-cumulatively distributed. Thus, sensitivity is much clearer across the parameter range as the gradient of acceptable and unacceptable plots can be hypothetically distributed between $-\infty$ and ∞ (as opposed to 0 and ∞ for cumulative).

5.2   Normalised and Non-Normalised Plots

Another characteristic of Multi-Parameter Sensitivity Analysis is the data is usually normalised. This pinpoints the trend of the samples' nature across the variable range rather than the actual quantity of samples. It may, however be useful to observe the data in a non-normailised plot. This displays a ratio of acceptable and unacceptable samples across the range, see fig. 5.
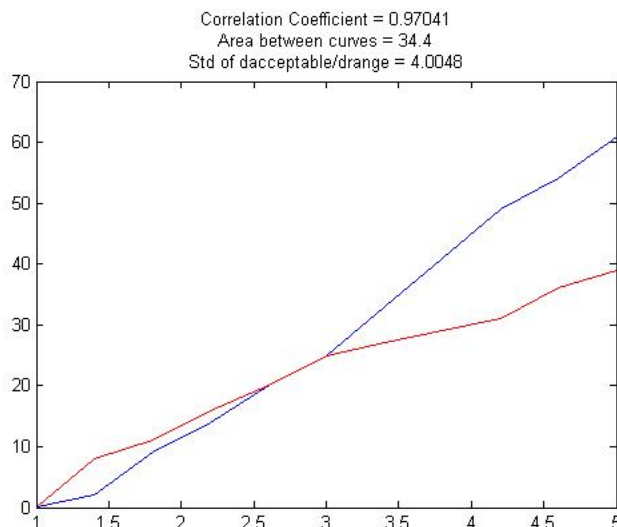

Figure 5: Non-Normalised example

In fig. 5 It is clear that a non-normalised graph can reveal some interesting information for example:

- There are more unacceptable samples than acceptable samples.

- From concentration values 1 to 2.5, for all ranges of defined variables, there are a greater proportion of acceptable samples than unacceptable.

6   Conclusion

Minicap is progressive analysis program which hopes to be accessible to modellers not necessarily acquainted with the MPSA technique. There are a number of different styles the results can be presented in, qualitative or quantitative, high or low definition etc. The introduction of ISCSA allows for another analysis of biological systems not formally documented much in the past.

7   The Future of Minicap

7.1   Simu Library

Although the library, Simu, currently only has two simulation techniques, the entire Minicap package has been designed to encorporate any other techniques should a user want to add them. Any new solver can be stored inside the directory similarly to *gillespie.m*. Simu must then be updated to incorporate this new solver.

Future solvers might be those involving

- modelling with partial differential equations or

- Cytoskeleton modelling

7.2 Critereon R

Currently the *threshold* is calculated from the **mean** of errors found in the *Objective Function*. There was much debate in the formulation of Minicap as to whether the **median** might be a better representation. In some respects, a 'median threshold' represents the information in a simpler manner. The number of samples falling into the *unacceptable* bracket will be equal to those in the *acceptable* bracket. This means the plots will have a somewhat symetric quality which. A 'mean threshold', will however, have different quantities in both brackets (it is more likely than not there will be many more *unacceptable* samples than *acceptable*). The symmetric quality is therefore lost; this affects the transferability of the 3 quantifiable comparison numbers.

It was decided that although there are arguments for a 'median threshold', poignant information is lost using this method. In the future a more satisfactory threshold calculation might be introduced which meets both arguments.

7.3 General Use

The program was designed with the Molecular Biology industry in mind. The program is, however, accessible to any user intending to carry out Sensitivity analysis for a System Function solvable by *ode15s.m*, *gillespie* or any new solver that can be called upon by the Simu library file with the input from the Minicap call-line.

References

[1] Jungyill Choi, Judson W. Harvey, and Martha H. Conklin. Use of multi-parameter sensitivity analysis to determine relative importance of factors influencing natural attenuation of mining contaminants. In *Proceedings of the Technical Meeting Charleston South Carolina March 8-12*, volume 1, 1999.

[2] David Gilbert. Modelling dynamic behaviour. http://www.brc.dcs.gla.ac.uk/d̃rg/courses/bioinformaticsHM/slides/modelling_dynamic.pdf, 2007.

[3] Kwang-HyanCho, Sung-Young Shin, Walter Kolch, and Olaf Wolkenhauer. Experimental design in systems biology based on parameter sensitivity analysis with monte carlo simulation: A case study for the TNFα mediated TNF-kB signal transdution pathway. *SIMULATION*, 79(12):726–739, 2003.

[4] Darren J. Wilkinson. The gillespie algorithm. In *Stochastic Modelling for Systems Biology*, 2006.