StatRes – A tool for statistical analysis of stochastic StochSD models

Leif Gustafsson and Erik Gustafsson.

1. Introduction

StatRes is a tool for statistical analysis of stochastic models created for StochSD.

A stochastic model produces different results for different simulations. Therefore, it is necessary to perform many simulation runs of the model to see what can happen, and to collect these results and present them in statistical terms. This is what StatRes does.

2. The user's interface of StatRes and its features

The StatRes form is opened by clicking the **StatRes** button in the **Tools** menu in StochSD. StatRes is then shown to the left of the Modelling Window.

StatRes user's interface is shown in Figure 1. The natural order to work with StatRes is from left to right and from top to bottom.

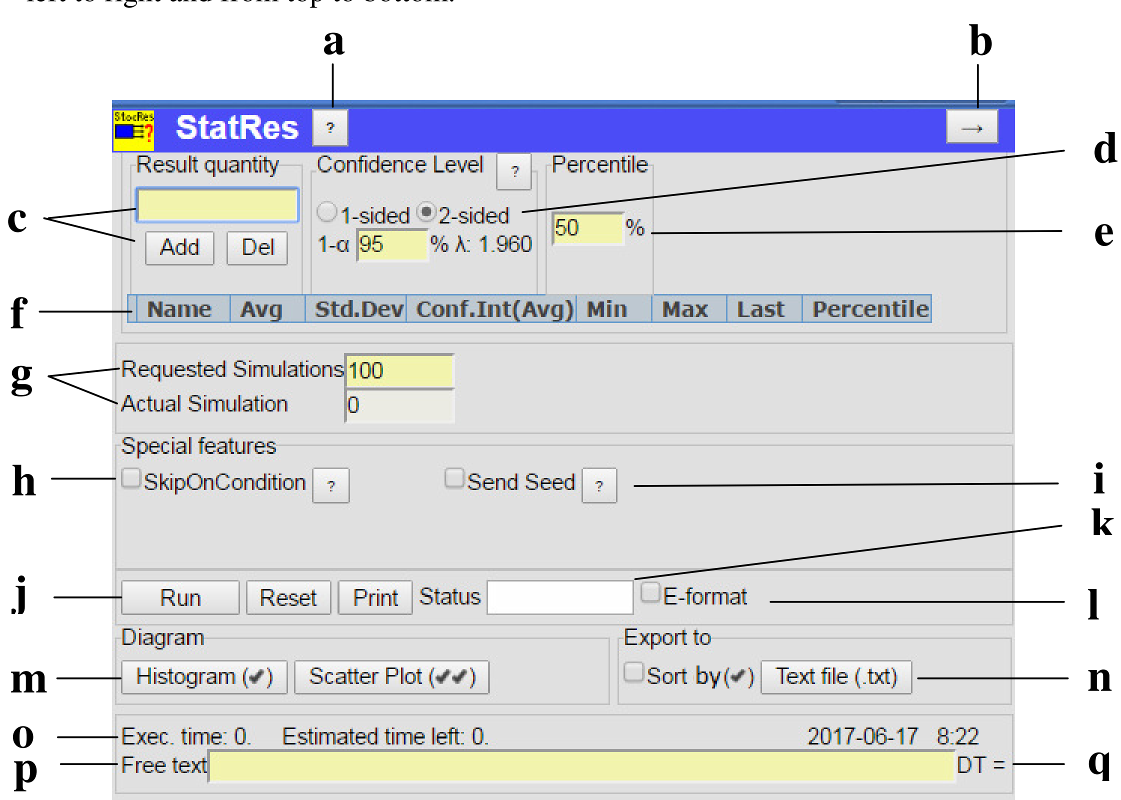

Figure 1. The user's interface of StatRes.

StatRes is handled by pressing buttons, checking boxes, and filling in text fields. As seen from the figure, the user's interface is composed of the following parts:

a) At the StatRes headline a help button (**?**) gives instructions about how to run StatRes.

b) An arrow button (→) minimises the StatRes form. Reopening StatRes restores it.

c) Specify Result quantities (without the brackets [ ]) to be analysed. When the **Add** button is pressed, a check if the specified quantity exists in the model is done. If so the quantity gets a new row in the Results Grid. (Note that StochSD and StatRes are not case-sensitive!) The **Del** button removes checked quantities in the Result Grid.

d) The 'Confidence Level' frame allows you to specify 1- or 2-sided Confidence Interval for the estimates, and the Confidence Level '1-α' can be specified in per cent (Default 2-sided interval and 95 %.). The corresponding λ-value is automatically calculated. (Also after the simulation you may change these options.) A help button (**?**) gives the theoretical background.

e) In the 'Percentile' frame you can specify any percentile between 0 and 100 %. Default is 50 % which means the median value. (Also after the simulation you may change the percentile.)

f) The 'Results Grid' shows the names of the specified variables. Here, estimates of Average, Standard Deviation, Confidence Interval for the average, Min value, Max value, Last value, and Percentile for each specified variable are presented. When a quantity is added, a row for it is created. (In Figure 4, the StatRes form is shown after 1000 runs.)

The very first column of the Results Grid holds a checkbox that can be checked or unchecked. The checkboxes are needed for several purposes: The **Del** button operates on checked rows only. The Histogram shows the distribution of a single checked variable. The Scatter Plot shows a xy-plot of two checked variables. Exported results can be unsorted or sorted by a checked variable.

g) Here you specify the number of 'Requested Simulations' to be done (default 100). Below the textbox you can see the 'Actual Simulation' number.

h) Checking the 'SkipOnCondition' box enables you to use a variable with the name [SkipOnCondition] that must be included in the model. This variable should define a condition (e.g. '[R]<1.5'). If this condition is TRUE at the end of a simulation run, this run will be skipped and replaced by a new run. This is handy if you for some reason want to exclude certain outcomes e.g. runs where nothing of interest happens. The number of skipped runs is also shown. (See Section 3 and Figure 4.)

i) By the 'Send Seeds' device you will make the study of a stochastic model reproducible. Checking the 'Send Seed' box opens a sub-frame that contains two text fields. In the first, named 'Seed-of-Seeds', you specify an integer number (default 123) that will generate a new seed ('Actual Seed') that is sent to the model before each new simulation run.

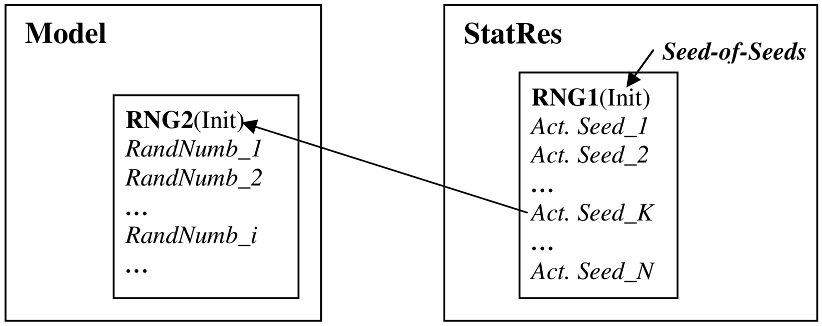

Figure 2. How the Send Seeds mechanism works.

j) The buttons **Run**, **Reset**, and **Print** do the following:

- The **Run** button (actually **Run/Halt/Continue**) saves the specifications done and starts the run sequence.
  - Pressing this button again during the runs will halt the sequence after the actual run is completed.
  - After the run sequence you may increase the number 'of 'Requested Simulations' and press this button (which is now labelled 'Continue') again.
- **Reset** resets the specifications to the state before the last run session.
- **Print** prints the form as shown.

k) To the right of the row of buttons the 'Status' textbox shows the current status of the session: RUNNING, HALTED or DONE.

l) Checking the 'E-format' box gives the results in exponential form. Default is unchecked which gives the normal numerical presentation. If 'E-format' is checked the results are presented as: ±X.XX E±XX meaning $\pm X.XX \cdot 10^{\pm XX}$.

m) In the Diagram frame there are two buttons: **Histogram** (✓) and **Scatter Plot** (✓✓) The check marks tells how many variables that have to be selected to display a Histogram or a Scatter Plot.

n) With the **Text file** button you can send all collected data (not the statistical estimates of it) to a specified text file to view them or to perform further analysis by an external program. Data from this session such as: Model name, Date and Time are then sent to the text file followed by the data results of all variables in columns.

With the 'Sort on' box unchecked the 'Text file' button will create a text field with the crude results of the *N* simulation runs *unsorted*. By marking *exactly one* variable in the Result Grid and checking the 'Sort on' box the data will be displayed in ascending order of the specified variable. All variables from each run are, of course, presented together on the same row in the text file.

o) At the bottom frame you will see the 'Execution time' for the session of the *N* simulation runs. The 'Expected time left' is also updated during the session. To the right the Date and Time when **Run** was pressed are shown.

p) At the bottom of the form is a 'Free text' field where you can write your own comments.

q) To the right of the ‘Free text’, the time-step, DT, used in the simulations is shown.

3. An StochSD model example (A stochastic SIR model)

A simple SIR model is presented below. SIR models are used in epidemiology to study the process of an epidemic. Such a model is composed of three stages, here represented by the single compartments: [S] (<u>S</u>usceptibles), [I] (<u>I</u>nfectious) and [R] (<u>R</u>ecovered).

An Infectious person has a certain probability, [c], to infect a Susceptible one at each time unit, who then becomes Infectious. The sojourn time in the infectious stage is on average [T] time units; thereafter the person becomes recovered (and immune).

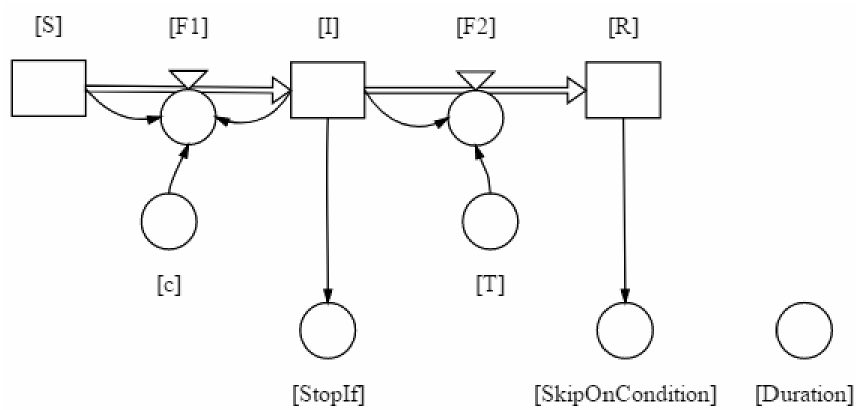

Figure 3. A stochastic SIR model in StochSD.

The main code of the SIR model could *symbolically* be written (where ‘←’ means ‘assigned to’, and ΔX replaces [X](t+Δt) – [X](t)):

[c] ← 0.0003	Infectious constant
[T] ← 4	Constant for expected time in I-stage
[S] ← 1000	(Initial values of S, I and R.)
[I] ← 1	
[R] ← 0	
ΔS ← −[F1]*DT()	(The three Stock equations are implic-)
ΔI ← ([F1] − [F2])*DT()	(itly specified by the model structure.)
ΔR ← [F2]*DT()	
[F1] ← RandPoisson(DT()*[c]*[S]*[I])/DT()	(Flow equ. Or: PoFlow([c]*[S]*[I]).)
[F2] ← RandPoisson(DT()*[I]/[T])/DT()	(Or: PoFlow([I]/[T].)

The stochastic SIR model requires Poisson distributed random numbers in the transitions (flows) [F1] and [F2].

A variable named [Duration] is set to T() to record the length of the epidemic. (Don’t confuse the time constant [T] and the time function T().)

To reduce the execution time, it is better to make it sufficiently large for a worst-case simulation run, but end each simulation run as soon as the I-compartment becomes empty – because nothing further will happen when there are no Infectious persons left. This is done with a variable [StopIf} having the algorithm:

```
IF [I] < 0.5 Then      ('IF [I] = 0' is avoided. Although [S], [I] and [R] stay integers, they)
  Stop()               (are stored as real numbers - and real numbers are approximate.)
End If                 (Stop() acts instantly – before presenting the current values in a)
                       (graph or table. To see these values, replace [I] by Delay([I], DT()).)
```

To also demonstrate the 'SkipOnCondition' facility in StatRes we assume that we are not interested in runs where 'nothing happens' because the single infectious person at start will recover before he or she has infected another person (so in this trivial case R will end equal to 1). To activate the 'SkipOnCondition' we check its box in the StatRes form. A Boolean variable named [SkipOnCondition] must also be included in the SIR model above, which only contains the condition:

```
[R] < 1.5              (The condition '[R] = 1' is avoided of the reason described above.)
```

If this condition is TRUE when the run terminates, the current run is skipped.

In Figure 4, the results of 1000 (approved) simulation runs of the SIR model are shown.

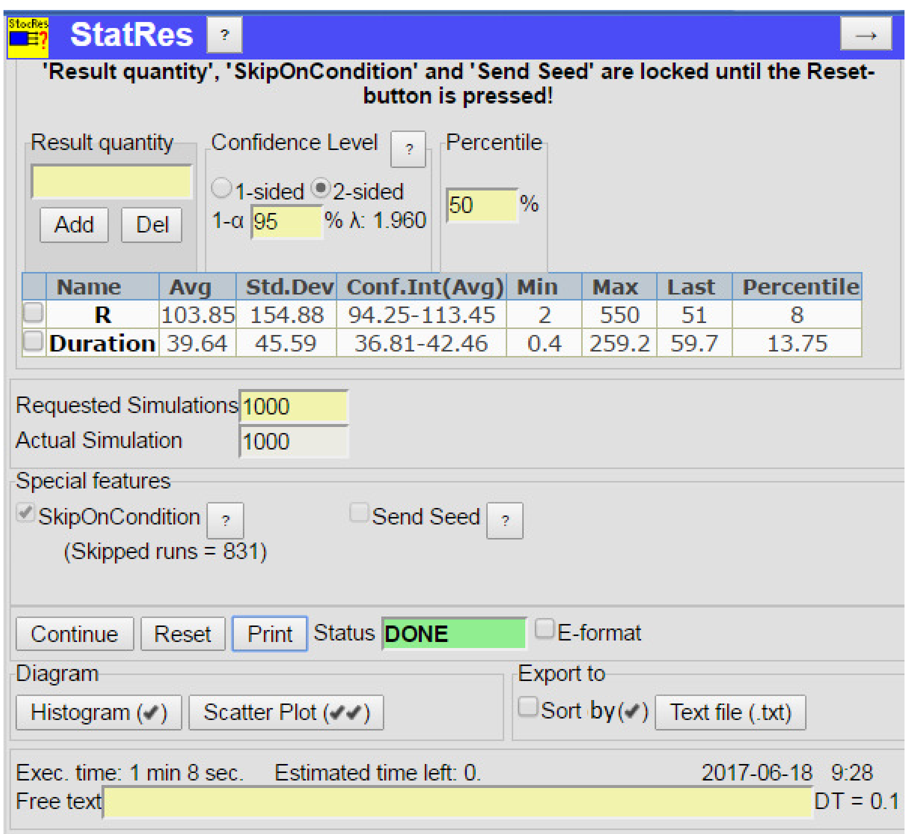

Figure 4. The results after approved 1000 simulation runs (where 831 additional runs were skipped because of no epidemic.)

The results are shown in the Result Grid of StatRes. Here you get the statistics (Average, Standard Deviation, Confidence Interval, Min, Max and Last values, and the Percentile) of the selected variables: [R], and [Duration] of the epidemic.

The execution time of the session of 1000 (+831 skiped) runs where 68 seconds when the time step, DT, was 0.1 time units.

By checking appropriate quantities and clicking the Histogram or Scatter Plot buttons you can see how the number of Recovered from disease, [R], is distributed over the 1000 runs, and how [R] covariates with [Duration] of the epidemic. See Figure 5.

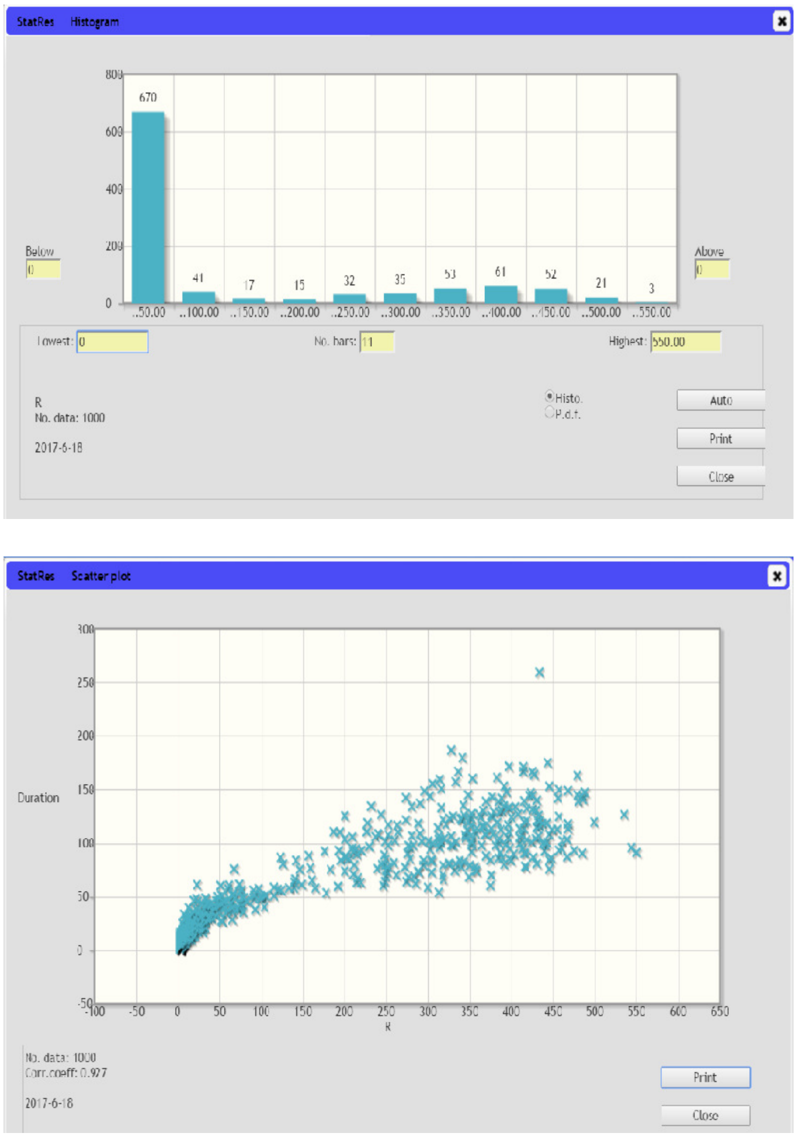

Figure 5. A histogram of [R], and a scatter plot of [R] vs. [Duration] of the epidemic.

4. References

- Devore, J.L. (2004) *Probability and Statistics for Engineering and Sciences*, sixth edition, Thomson Learning, Inc., Toronto, Canada.

- Gustafsson, L. (2000) Poisson Simulation – A method for generating stochastic variations in Continuous System Simulation. *Simulation*, **74/5**, 264-274.

- Gustafsson, L. and Sternad, M. (2016) A guide to population modelling for simulation, *Open Journal of Modelling and Simulation*, 4, 55-92. http://file.scirp.org/pdf/OJMSi_2016042717425486.pdf

- Gustafsson, L. and Sternad, M. (2017) The full potential of Continuous System Simulation modelling, OJMSi, 2017. http://xxx ???

5. Responsibility

The user is fully responsible for the use of this product. The producer and the supplier of this code take no responsibility for the use or functioning of StatRes.