

TEACHING SIMULATION WITH SPREADSHEETS

Jelena Pecherska and Yuri Merkurjev
Department of Modelling and Simulation
Riga Technical University
1, Kalku Street, LV-1658 Riga, Latvia
E-mail: merkur@itl.rtu.lv, jelena@itl.rtu.lv

KEYWORDS

Spreadsheets, discrete-event simulation, System Dynamics, education

ABSTRACT

This paper focuses on a spreadsheet-based approach to teaching simulation. The objective is to introduce spreadsheets as a powerful simulation tool for educational purposes. The spreadsheet simulation facilitates the familiarisation with basic simulation concepts apart from simulation software. The paper deals with basic types of simulation models suitable for spreadsheets. The static stochastic or Monte Carlo simulation models are introduced through a sample model for lead-time demand distribution evaluation. The two main types of dynamic simulation spreadsheets models are overviewed as well. The discrete-event system simulation basic concepts are introduced by simulating a two-server system with a single queue. A population dynamics simulation model presents the System Dynamics concepts. The basic simulation model types that are available on spreadsheets make it possible to study simulation concepts, methodology, and techniques as well as to use spreadsheets for on-hand simulation in business games, case studies, engineering, etc. The overviewed experiences in using spreadsheets for teaching simulation result in indication of spreadsheets as an alternative simulation software.

INTRODUCTION

Simulation studies as well as simulation techniques studies can be complicated by unreasonable usage of professional simulation software. This statement is especially valid for overview or introductory courses. Simulation tools vary from universal programming languages to special simulation software. Model creation using simulation software requires experiences and/or training as well as a good knowledge of simulation theory. However, there is an alternative approach to learn the basic simulation concepts and methodology. Spreadsheets provide almost complete set of tools for simulation. Students' prior experience using spreadsheets is both advantage and disadvantage. The advantage of using a familiar tool is obvious. The disadvantage is due to the use of spreadsheets mostly as an illustrative and calculating tool. Most users are not acquainted with spreadsheets as rather powerful software. Of course, the use of spreadsheets as a simulation tool requires specific skills and the

effectiveness of simulation depends upon the developer's efficiency. Curriculum simulation studies can be carried out by using spreadsheets. Such a possibility is due to the development of electronic spreadsheets in recent years. Spreadsheets can also facilitate on-hand simulations. This paper describes the usage of spreadsheets at various stages of simulation studies such as input data analysis, random variate generation, and simulation model creation. Microsoft Excel is used throughout the paper, although most functions and models can be transformed into other spreadsheet formats.

TYPES OF SIMULATION MODELS ON SPREADSHEETS

After formulating a simulation task, an appropriate logical model should be developed. The choice of simulation model type can be prescribed by characteristics of the system and an objective of the study. One can create either a static or dynamic simulation model. The created model can be deterministic or stochastic as well as discrete or continuous. This paper is related to some types of simulation models such as Monte Carlo, discrete-event simulation, and System Dynamics models.

MONTE CARLO SIMULATION ON SPREADSHEETS

Monte Carlo methods generally are stochastic ones. Simulation models that contain uncertainties and are static are called Monte Carlo simulations (Banks et al. 1996). Monte Carlo simulation is based on the use of random numbers and probability statistics to investigate a variety of problems. Monte Carlo simulations are often used in such areas as economics, logistics, management, etc. Sometimes Monte Carlo simulation is defined as a sampling experiment (Evans and Olson, 1998). Actually, to call something a "Monte Carlo" experiment or simulation, all you need to do is to use random numbers to examine a problem.

Spreadsheets provide a variety of built-in statistical functions. For example, in Excel such functions as BETADIST, CHIDIST, EXPONDIST and other popular distribution functions are available. Built-in functions facilitate Monte Carlo simulation on spreadsheets, but an experienced user is not restricted by built-in distributions.

As an example of Monte Carlo simulation on spreadsheets let's consider solving of the lead-time

demand evaluation (Banks et al. 1996). In this case bulk rolls of newsprint are the inventory items. The daily demand for rolls and the lead-time are described by the probability distributions. Both distributions are used for definition of uncertain inputs of the model. In this simple example a Monte Carlo simulation represents 100 random trials. In each trial one value of a random number is used for lead-time value generation. The obtained lead-time value is used for demand during lead-time generation using different random numbers. Performing a predefined number of trials gives a possibility to summarize results and assess lead time demand. The problem is solved as the lead-time demand distribution is determined. The histogram of lead-time demand after 100 trials is shown in Figure 1.

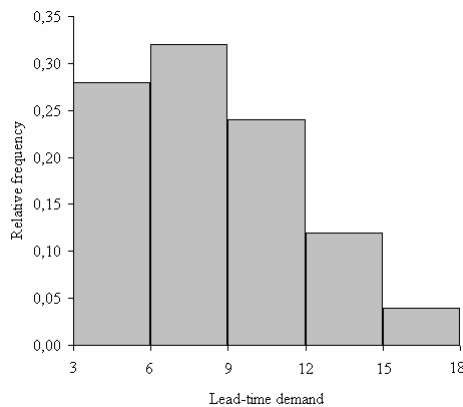


Figure 1: Histogram of Lead-Time Demand for Monte Carlo Simulation.

Monte Carlo simulation allows building in the uncertainty (present in most real-life problems) into spreadsheet models. This example is developed using only standard Excel statistical, lookup, and reference functions. A spreadsheet model contains cells which have random values generated from the given probability distributions. Every time the spreadsheet is recalculated, these random values change. After recalculation the different set of simulation results is drawn. Replications using the same parameters help to figure out the most likely behaviour of the unknown random lead-time demand. The use of Monte Carlo method for teaching simulation results in better understanding of the probabilistic nature of real-life occurrences as well as better insight into different problem solving with uncertainties.

SYSTEMS SIMULATION ON SPREADSHEETS

Unlike Monte Carlo simulation, systems simulation models sequences of events over time. This section introduces a common approach to dynamic systems simulation on spreadsheets. Actually, the range of systems under consideration is restricted to discrete-event systems and continuous systems.

Spreadsheet Discrete-Event Simulation Models

In discrete-event systems the state variables change only at discrete time moments. An occurrence that changes the system state is an event. Thus, discrete-event dynamic simulation models, independently of the means of realization, always use some kind of time advance algorithm. The simulation clock value recalculation in spreadsheet models is performed using one of two basic approaches: time advance using fixed increment or time advance using event scheduling. Time advance using fixed increment in discrete-event simulation differs from genuine event scheduling. However, the event planning within both approaches is rather close. The discrete event system simulation with time advance using fixed increment is not considered separately in this paper.

Traditionally, simulation basics of discrete-event systems are introduced via a single-server queuing system. A single server queuing system model on a spreadsheet gives a possibility to introduce almost all basic discrete-event simulation concepts. The sample model under consideration is rather close to the basic one.

The simulated system consists of an independent entity source, a common FIFO queue, and two service devices. The source and devices use random variables during the simulation process. At simulation time $t=0$ the system is "empty". Random variates for arrival times and cycle times are generated using individual probability mass functions. Entities interarrival time distribution is shown in Table 1.

Table 1: Interarrival Time Distribution of Entities

Time between Arrivals (Minutes)	Probability
1	0.25
2	0.40
3	0.20
4	0.15

The distributions of service times for both service devices are given in the same way. Sample data is used for interarrival time and service time generation as it is described in **Random Variate Generation** section. Arrival times and service times are unique for each replication. Data tables are located on the separate model sheet and can be easily changed without affecting the model logic.

The considered system spreadsheet model is worked out using the event-scheduling approach. The model algorithm concentrates on events and their effect on system state. Any time moments between events are ignored. Each model spreadsheet row keeps a record necessary to advance simulation time and activate future events in time order. Records are put into spreadsheet cells and contain formulas and conditions for future changes. The current system state provides the possibility to produce the next system snapshot after

advancing simulation time up to the moment of the simulation End event.

During simulation the appearance of the model spreadsheet changes automatically. Further we will concentrate on a single replication.

The results of a single replication are available via the Excel spreadsheet. By adding some extra calculations we can obtain necessary statistics. During each simulation trial the system state is updated and statistical performance measures are recalculated according to scheduled events' times. In this case the system status variables are recalculated at discrete time moments $t_{i+1}=t_i+dt$ (dt = random time interval between scheduled events). A fragment of the spreadsheet simulation screen is shown in Figure 2. The system simulation model consists of system snapshot series at discrete points of time. A given snapshot includes system state at time t_i , entity attributes, future event list and necessary cumulative statistics and counters. The future event list contains event notices that have been scheduled to occur at future time. The system snapshot at time 0 is additionally defined by initial conditions as well as by the generation of exogenous events. For this study the initial state is an "empty" queue, free service devices. The exogenous event is the first entity arrival and is scheduled at time 0. Service completion events for both servers are generated and scheduled in the future event list when the entity arrives or upon the system state.

	A	B	C		Event	States	Attributes			
	Customer Arrives	Customer Service	Service Process	Time	Type	Q	S1	S2	Arr	SS1
5	Initial	0	0	0	0	0	0	0	0	0
6		1		0	3	1	0	0	0	1
7		2		2	3	0	1	1	2	
8			1	2	4	2	0	0	1	
9		3		3	4	3	0	1	1	4
10			2	3	5	1	0	1	0	
11			3	3	6	2	0	0	0	
12		4		4	6	3	0	1	0	6
13		5		5	7	3	0	1	1	7
14		6		5	8	3	1	1	1	8
15			4	6	10	1	0	1	1	
16		7		6	10	3	1	1	1	10
17			5	7	11	2	0	1	1	1

Figure 2: A Fragment of the Single Queue System Model

A stopping event, here called End, defines the length of a simulation run. This event is scheduled at time 0 for a specified future time.

This example introduces discrete-event concepts using Excel lookup and reference functions. Such concepts as entity, entity attributes, system state, event, event notice and event list, simulation clock are dealt with while creating a spreadsheet simulation model. In contradistinction to simulation software application, model building on spreadsheets is practically impossible without clear and deep perception of basic modelling concepts.

System Dynamics Spreadsheet Models

System dynamics (SD) is another approach to systems analysis using simulation. In system dynamics various complex systems (e.g. population, ecological, and economic, etc.) are simulated. Actually, these simulations mean substitution of a continuous system with a discrete simulation model.

The basic model structure in system dynamics includes a number of elements such as levels and flows, decision functions and information links. Elements of the SD model have specific features and form a complete set of equations for simulation. The simulation trial can be interpreted as a sequence of solutions of these equations calculated with a fixed time interval. Each solution is unambiguously defined by the previous solution and the length of fixed time interval. Actually, the model logical base can be formulated by means of spreadsheets very effectively and briefly. Two main components completely describe the SD spreadsheet model. They are the initial state variable values and the set of equations.

As a sample SD spreadsheet model the simplified fish population system is considered (that corresponds to the well-known Fish Banks resource management game). The system structure is shown in Figure 3.

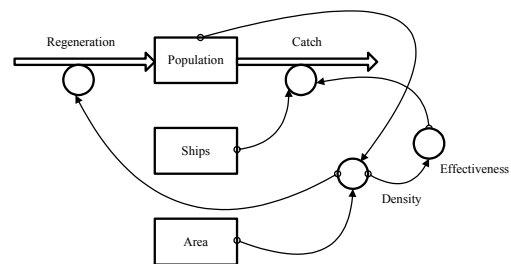


Figure 3: The Structure of Simulated Dynamic System

The structure in Figure 3 is drawn using the symbols that are evident to anyone familiar with system dynamics approach. The rectangles represent levels; the double arrows represent flows; the circles – decision functions; the thin arrows – information links. Model creation assumes fixed initial values for all levels, equations defined for all flows, links and decisions. Some specific information is needed, such as functions defining the relationships between fish density, and both ship effectiveness and fish regeneration. The functions are shown in Figures 4 and 5.

Simulation results of 3 trials are shown in Figure 6. The dynamic simulation results provide information about population dynamics and support decision making in management of renewable resources.

STEPS OF SIMULATION STUDY USING SPREADSHEETS

Simulation studies independently of means of realization (by hand, using simulation software, etc.) include a typical set of steps. In this section only some of those steps are considered, such as input data

analysis, random variate generation and output data analysis. The steps are selected because they require special techniques in order to be realised on spreadsheets.

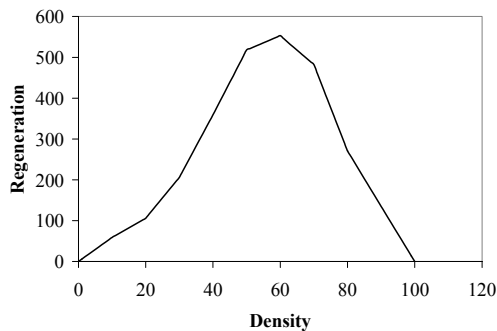


Figure 4: Fish Regeneration Function

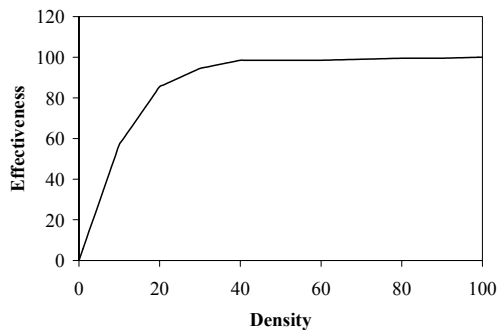


Figure 5: Ship Effectiveness Function

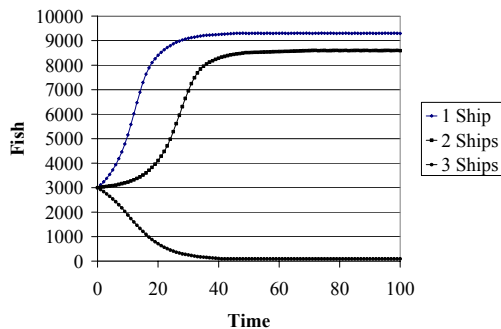


Figure 6: Simulated Population Dynamics

Input Data Analysis

Input data for simulation models can be obtained from different sources. Measurements, observations and historical data that provide quantitative information for simulation should be “processed” before becoming a driving force of a simulation model. The tabular nature of empirical data nicely fits spreadsheets. There are alternative approaches to data processing. Standard spreadsheet functions, toolpacks, external data processing (e.g. with the help of Arena Input Analyzer or with special software), add-ins such as Crystal Ball or @Risk, and manual data processing provide different

possibilities. The standard spreadsheet functions assume reasonable input data processing.

Data processing includes identifying the distribution, distribution parameter estimation, and goodness-of-fit testing as the main stages. All these activities and calculations can be effectively performed on spreadsheets. It should be mentioned that in all cases the data is discrete.

As an example, an empirical distribution function for customer demand is obtained. The data about demand for mineral water from a small shop was collected. Data collection period was four months. Measurement units are bottles per day. Excel standard functions MIN, MAX, AVERAGE, STDEV and FREQUENCY were used. The resulting histogram is shown in Figure 7, sample parameter estimations are shown in Table 2.

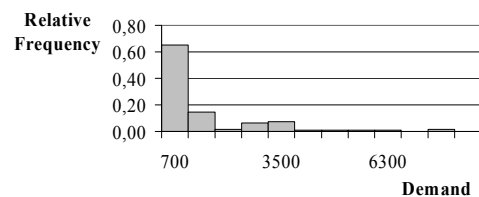


Figure 7: Histogram of the Daily Demand for Mineral Water

Table 2: Parameter Estimation for the Daily Demand for Mineral Water

Minimum	0.00
Maximum	7324.00
Average	907.23
Standard Deviation	1477.58

The chi-square test for distribution fitting for Weibull distribution was performed and the hypothesis about the random variable Weibull distribution was rejected. The Excel functions WEIBULL and CHINV were used for testing. The selection of distribution type based on the shape of the histogram was unsuccessful. Therefore, an empirical distribution was used for simulation.

Random Variate Generation

A variety of probability distributions is available to generate random data for simulation. Three basic generation approaches in spreadsheets are applied. One can generate random variates using a number of standard distribution functions. The generated values are sampled from continuous or discrete distributions. The second approach is generating random variates from empirical distributions. Using only Excel built-ins gives us a possibility to get a new sample for each replication.

The third approach assumes the usage of Data Analysis Toolpack. Each reference to this tool will provide a sample. For spreadsheet simulation purposes this approach does not suit dynamic simulations.

The most specific for simulation purposes is random variate generation from continuous distributions. In this

case the distribution function should be presented in a tabulated form. In Excel we use the VLOOKUP function to generate random variates from empirical distributions. Using this function we can generate discrete outcomes defined by almost any probability mass function (using inverse transformation method). An example of this method application for generating and analysing data using a discrete triangular distribution function is considered. Figure 8 shows the Excel spreadsheet snapshot with distribution parameters and generation results, while in Figure 9 both appropriate cumulative distribution function and probability mass function are shown.

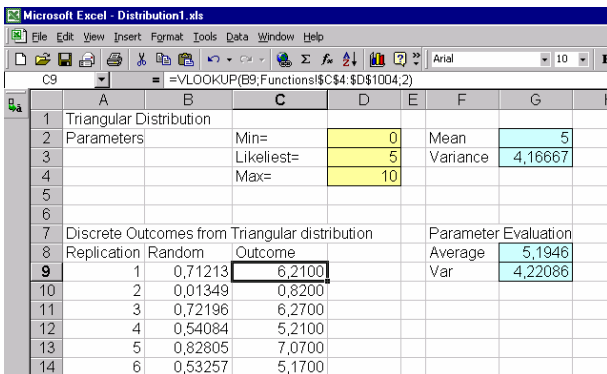


Figure 8: Discrete Distribution Parameters and Generated Outcomes

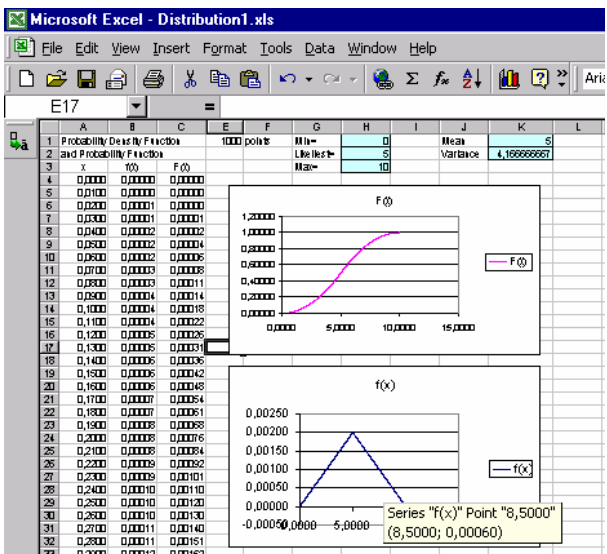


Figure 9: Cumulative Distribution Function F(x) and Probability Mass Function f(x)

The fitting Chi-square test for the obtained thousand points gives a result of $\chi^2=0.673744$, while the critical value for this sample is $\chi_{crit}(0.05,26)=38.88513$, where probability of rejecting a valid distribution is 0.05 and degrees of freedom correspond to the number of histogram cells. As χ^2 is less than a critical value, we fail to reject the hypothesis that the data originates from the hypothesized distribution.

The inverse transformation method applied to generate random variates from discrete probability distributions provides usage of empirical distributions for simulation purposes in spreadsheets, as well as gives an insight of using any distribution in simulation.

Output Data Analysis

The simulation models that use random input data produce random results. Each simulation experiment or trial gives a unique set of random variables. To interpret results correctly the simulation experiments should be replicated, results should be accumulated and after performing a number of replications the statistical performance measures should be calculated. To obtain a necessary set of results the simulation model should be supplemented with an appropriate section. The realization and the appearance of such a section may vary. In Excel it is possible to perform replications not only using macros or add-ins. The choice of Tools\Options\Calculations\Manual option allows calculations using cross-references. Such calculation mode is useful for accumulating simulation results. Actually, the data analysis actions in spreadsheets are almost manual if not using additional tools. The estimation of the basic measures of performance supposes the application of standard functions and goodness-of-fit tests. These calculations are already overviewed in the above sections. The features of simulation model define the performance measure types. For example, the simulation results for simulation model from section **Spreadsheet discrete-event simulation models** shown in Table 3 are obtained after 100 replications.

Table 3: The 55-minute Simulation Results for Queuing System Model

Performance Measure	Resulting value
Total production	21
Average delay in queue	0.5
Server1 Utilization	0.8545
Server2 Utilization	0.6364
Maximum number of entities in queue	1
Maximum delay in queue	2.0
Maximum service time	6.0

Using add-ins for replication of simulation runs can sufficiently facilitate not only experimentation with the model. Add-ins provide convenient input and output data analysis as well as simplify random number generation for different purposes. The only disadvantage for teaching purposes is their “hidden” algorithms for performing replications and performance measure calculations.

MODEL CONSTRUCTION AND SIMULATION RUNS

Model construction using spreadsheets substantially differs from the same process using simulation software. Actually, it looks more like a kind of programming. The flexibility and complexity of spreadsheet simulation models depends on author's skills. Some stages are typical for spreadsheet calculations:

- assignment of spreadsheet areas for data and parameters;
- programming of cell formulas;
- copying cell formulas to provide a number of trials or a necessary simulation length;
- visualisation by using charts.

The model implementation requires data accumulation for calculations of statistical performance measures. This goal is specific to spreadsheet simulation model construction. Using manual calculations as well as cross-references, it is possible to accumulate the results for further calculations. Advanced spreadsheet users can implement more complicated techniques by creating macros, using Add-ins e.g. Analysis ToolPack, Crystal Ball, etc. Crystal Ball Add-in can be used as an engine of dynamic simulation providing a necessary number of simulation experiments both for static and dynamic models.

CONCLUSIONS

Spreadsheets provide broad possibilities of creating simulation models for educational purposes. They are useful for on-hand simulations in business games, case studies, engineering, etc. Spreadsheet models are suitable for logistics, business management, operations research and other studies as well as for implementing at small and medium-size enterprises. The basic spreadsheet features facilitate static simulations, and dynamic simulations can be rather effective and helpful in different areas of application. The spreadsheet simulation models can be used as a platform for understanding the mechanisms behind the discrete-event, as well as System Dynamics approaches. Simulation models developed on spreadsheets are not very flexible. But, from a teaching point of view, they have advantages of getting software skills in a short period of time, broad availability, ease of use, ease of validation, and low price (Robinson 2003). It is possible to implement both static and dynamic simulation using spreadsheet standard functions only. Macros and add-ins as simulation engines give an additional support to model construction and replication of simulation experiments. Understanding of simulation concepts by using spreadsheet simulation moulds a basis for application of simulation for different studies, for practical investigations, and for further shifting to special simulation software.

REFERENCES

- Banks, J.; J. S. Carson; and B. L. Nelson. 1996. *Discrete-Event System Simulation (2nd Edition)*. Prentice Hall, N.J.
- Evans J. and D.L. Olson. 1998. *Introduction to Simulation and Risk Analysis*. Prentice Hall, N.J.
- Pecherska J. and Y. Merkurjev. 2001. "Monte Carlo Simulation on Spreadsheets". In *Scientific Proceedings of Riga Technical University*. RTU, Riga, 122-129.
- Robinson S. 2003. *Simulation. The Practice of Model Development and Use*. John Wiley & Sons, West Sussex, England.

AUTHOR BIOGRAPHIES

JELENA PECHERSKA comes from Riga, Latvia. She works for Riga Technical University from 1979 and obtained her master degree there. At the moment she is working at the Department of Modelling and Simulation of Riga Technical University as a lecturer. Professional interests are in the fields of methodology of discrete-event simulation, supply chains modelling, practical applications of discrete-event simulation and discrete-event simulation in education.

YURI MERKURYEV is Habilitated Doctor of Engineering, Professor of the Institute of Information technology at Riga Technical University, Head of the Department of Modelling and Simulation. His professional interests include a methodology and practical implementation of discrete-event simulation, supply chain modelling and management, and education in the areas of modelling, simulation and logistics management. He is Programme Director of the Master-level curriculum "Industrial Logistics Management" at Riga Technical University. Prof Merkurjev has wide experiences in performing research and educational projects in the simulation area, at both national and European levels. He regularly participates in organising international conferences in the simulation area. He is the General Chair of the 2005 European Simulation Multiconference, SCS – ESM2005, to be held in June, 2005 in Riga, Latvia. Prof. Merkurjev has about 180 scientific publications, including 2 books. He is a Board member of the European Council of the Society for Modeling and Simulation International, President of the Latvian Simulation Society, Board member of the Latvian Transport Development and Education Association, and corresponding member of the Latvian Academy of Sciences.