What Every Engineer Should Know About Modeling and Simulation

Raymond Madachy

Daniel Houston

Dedication

To our families who allowed us the time to write this book.

1 Introduction

1.1 BOOK SCOPE

The title of this book indicates that it is directed toward engineers, although it is actually suited for a broader audience of engineers, scientists, and managers of technical organizations. These are people who perform or manage technical work but have not studied modeling and simulation (M&S).

These days, virtually all complex engineering projects involve some M&S. The increasing use of simulation modeling has been enabled by the growth of computing and is expected to continue. Consequently, more people must be conversant in the concepts of M&S.

Some M&S concepts are generally not well understood. These include the role of simulation in human learning and problem solving, representations of uncertainty in modeling, and degrees of necessity for verification and validation. This book addresses these concepts to provide a basic understanding of simulation modeling for those who have not studied the discipline.

This book is also intended for those who would like to learn simulation modeling with a general purpose simulation software program. While this book cannot provide thorough treatment of any particular simulation program, it does seek to make general purpose simulation programs more accessible to the beginner in simulation. To this end, several general purpose simulation programs are shown and explained to draw out the general methods underlying each simulation paradigm.

1.2 SYSTEMS AND MODELS

A *system* is a set of interacting parts that form a connected whole. A more thorough definition is:

> A *system* is a construct or collection of different elements that together produce results not obtainable by the elements alone. The elements, or parts, can include people, hardware, software, facilities, policies, and documents; that is, all things required to produce systems-level results. The results include system level qualities, properties, characteristics, functions, behavior and performance. The value added by the system as a whole, beyond that contributed independently by the parts, is primarily created by the relationship among the parts; that is, how they are interconnected [14].

A system must be represented in some form in order to analyze it and communicate about it. A *model* in the broadest sense is a representation of reality, ranging from physical mockups to graphical descriptions to abstract symbolic models. A model in the context of this book is a logical, quantitative description of how a system or process behaves. Though both *engineered* systems and natural systems are modeled in simulations, this book focuses on human-designed engineered systems. The models are abstractions of real or conceptual systems used as surrogates for low-cost experimentation and study. Models allow us to understand a system by dividing it into parts and looking at how they are related.

1.3 WHAT IS SIMULATION MODELING AND WHY MODEL?

Engineers design systems with many different kinds of models including mental models, diagrams and schematics, scaled physical models, mathematical models, and computational models. All are representations that enable understanding, design, and building of processes or systems. Many models are static representations where time plays no role, such as a scale model of a building, an equation for estimating the cost of an airplane, or a Bayesian network of contributing factors to an industrial accident. This book discusses modeling for simulation, which usually, but not always, implies dynamic modeling of process or system changes over time.

Simulation is the numerical evaluation of a mathematical model describing a system of interest. Many systems are too complex for closed-form analytical solutions, hence simulation is used to exercise models with given inputs to see how the system performs. Simulation can be used to explain system behavior, improve existing systems, or to design new systems too complex to be analyzed by spreadsheets or flowcharts.

A simulation model is a representation of a system that consists of variables related through formulas. Simulation software provides a means of specifying the structure of a model through formulas and performs the task of automatically updating variable values. At any particular moment in a time-based simulation, a model's state consists of its structure and the values of its variables.

Engineered systems are often too complex to represent in a closed form solution, an equation that solves a given problem in terms of functions and mathematical operations. For complex systems where analytical solutions aren't possible, one can use a simulation model consisting of linked equations that are executed whereby the outputs of some are inputs to others.

Simulation models are used for many reasons. One of these is to understand how a process or system works, particularly when inputs may vary and a model's structure contains feedback loops and delays. The effects of such complexity can be very difficult to understand without a model that allows one to represent and trace behavior.

Simulation modeling can play an important role in training and learning. Flight simulators are routinely used to train pilots to reduce risk by practicing in scenarios before actual flying. Simulators have also been used to teach project management decision-making skills, such as how much a schedule can be reduced before product quality is undermined.

Another reason for simulation modeling is supporting system or process design or redesign. Modeling can be used to compare designs, identify trade-offs, and choose among alternatives that provide the best support for the desired outcomes. Along with design, modeling can support planning the use of resources and forecasting outcomes using objective criteria.

When designing or planning, the cost of simulation modeling may appear high. At those times, one must determine how much is unknown about a system or process and then ask whether the cost of learning is higher with a model or with a trial implementation. One will often find that a program cannot afford not to model.

1.4 OVERVIEW OF SIMULATION MODEL TYPES

Static models do not involve time in their computations. There is no change of variable values over time. *Dynamic* models are used to simulate the time-varying behavior of a system. In time-based simulation the system is considered a collection of interacting elements, the properties of which change with time. These elements are typically referred to as *entities* and their properties as *attributes*. The interactions are often called *activities*. Continuous and discrete simulation models must be distinguished however. The decision as to which type of simulation model to use depends upon the nature of the system and the purpose of the simulation study (see Chapters 2 and 3).

Continuous models compute differential or algebraic equations at regular intervals to update state variable values to represent continual change over time. A continuous model shows how values of the attributes change as functions of time, computed at equidistant small time steps. These changes are typically represented by smooth continuous curves. Continuous models are useful in such areas as engineering design where well-established mathematical relationships give rise to models consisting of differential or algebraic equations. An example is computing the level of water in a dam based on continuous flow rates. Other areas of application address long-range trends, or growth and decay behavior. In these studies, many variables can be conveniently aggregated and local discrete fluctuations ignored. Continuous simulation models may also be applied to systems that are discrete in real life but where reasonably accurate solutions can be obtained by averaging values of the model variables. Examples of such approximations are traffic pattern studies or analysis of mass production by assembly lines. In these cases, one is not interested in individual cars or machine parts, but desires to know general traffic trends or the overall efficiency of a manufacturing process.

Continuous models solving differential equations are not all time-based. Integration may be across spatial distances (e.g., modeling thermodynamics or material stresses) instead of time progression. The underlying modeling procedures and numerical methods are structurally similar to time-based models.

Discrete or *discrete event* models generally consist of entities and their attributes that change during the simulation runs at event times. These changes occur instantaneously as the simulated time lapses, and are reflected in the output as discontinuous fluctuations. There are many systems that should be modeled as discrete because no continuous approximations are valid. An example situation is a model built to study the effects of different machine maintenance scheduling policies. In such a model one is concerned with queue lengths, waiting times, sequencing of related jobs, and availability of maintenance resources where one cannot ignore the discrete character of the system by aggregation or averaging.

Agent-based models can be considered forms of discrete models. They also model individual entities and use triggers to cause events. Agent-based models are distinguished by their encapsulation of behavioral rules within individual agents. Agents operate autonomously, interacting with an environment and with each other. The states of agents can change with an event like entity characteristics change in discrete event modeling. A similarity with continuous systems is that transitions move agents between states like continuous flows move entities between levels (continuous state variables). Agents may interact in discrete, continous, or combined environments.

A simulation model can be *deterministic*, *stochastic*, or *hybrid* (mixed). In the deterministic case, all input parameters are specified as single values. Stochastic modeling recognizes the inherent uncertainty in many parameters and relationships using random numbers drawn from a specified probability distributions. Mixed modeling employs both deterministic and stochastic parameters.

In a purely deterministic model, only one simulation run is needed for a given set of parameters. However, with stochastic or mixed modeling the result variables differ from one run to another because the random numbers drawn differ from run to run. In this case the result variables are best analyzed statistically (e.g., mean, standard deviation, distribution type) across a batch of simulation runs. This is termed *Monte Carlo simulation* or Monte Carlo analysis (Chapter 4).

1.5 APPLICATIONS

To illustrate the broad spectrum in which M&S can be used, representative engineering applications are shown in Table 1.1 for general engineering disciplines. The list is not exhaustive across disciplines, inclusive of all applications, or static. Not shown are many subfields or specialties such as aerospace engineering, automotive engineering, environmental engineering, engineering management, robotic engineering, manufacturing engineering, and countless others. These examples are common traditional applications and more areas will come into vogue as the fields advance. The applications also have subcategories not shown. A good number of applications are shared across disciplines (e.g., materials science in mechanical and/or chemical engineering) so the traditional closest fits are shown.

TABLE 1.1: Representative Engineering Applications

Discipline	**Applications**
Biomedical	Biomechanics
	Biomedicine
	Medical devices
	Bioinformatics
	Biomaterials

(*continued*)

TABLE 1.1: Representative Engineering Applications (*continued*)

Discipline	**Applications**
Chemical	Biochemical
	Corrosion
	Genetics
	Pharmaceuticals
	Transport phenomena
	Waste management
	Plant design
	Process design
	Chemical synthesis
Civil	Structural design
	Building materials
	Hydraulics
	Transportation systems
	Environmental
	Municipal infrastructure
	Construction management
Electrical and Electronic	Circuit analysis and design
	Power systems and grids
	Signal processing and communications
	Networks
	Computer engineering
Mechanical	Materials
	Thermodynamics
	Fluid dynamics
	Vehicle dynamics
	Kinematics
	Nanotechnology
	Control system design
Software	Software development processes
	Software evolution and maintenance
	Defect modeling
	Business value
Industrial and Systems	Systems of Systems
	Sustainment
	System Architecture

(*continued*)

TABLE 1.1: Representative Engineering Applications (*continued*)

Discipline	**Applications**
	Manufacturing
	Operations research
	Economic and financial analysis

1.6 INTRODUCTORY EXAMPLE WITH BASIC MEASURES

The methods of modeling and simulation will be demonstrated first with a running example for electric car charging scenarios starting with a simple scenario that can be manually calculated. In subsequent chapters we will extend it with increasingly complex working models and simulations to demonstrate the expanding concepts.

The case study example is for a company that develops, manufactures, and installs electric car charging station equipment. The charging stations are for customers "on-the-go" for short-term charges. The company needs to decide where to install charging stations, how many charging bays to include at each location, and what options for charging power capability need to be provided. They cannot address these questions with a simple mathematical solution because of the complexity of operations and system uncertainties, thus they will simulate the operations to support the decisions.

To gain insight they will analyze standard metrics from the simulations including waiting time, queue length, and resource utilization. This is because excessive wait time leads to unhappy customers, drives customers to competitors' services, and reduces profit. Queues require physical space. If the queue cannot accommodate enough cars safely, then cars won't even be able to enter the queue. If queues are too long, customers may also leave. For resource utilization, underused resources are wasteful and reduce profit. An overused resource indicates that another resource may be warranted, and it leads to excessive waits and queue lengths. Optimizing the waiting time, queue length, and resource utilization will be key to success or failure.

1.6.1 QUEUE-SERVER SYSTEM

The most common discrete system is the queue/server system. Queue/server systems exist whenever a line forms in front of some processing mechanism. The line is called a *queue* and the processing mechanism is called the *server*. In this example the cars form in queues to be served by the charging bay resource. Modeling the scenario entails quantifying the car arrival times and their respective charging (service) times.

In this first manual example we'll assume some random values for car arrivals and charging times. For simplicity we'll use integer values for easy calculation and assess the first one hour of operations using basic measures for simulation output shown in Table 1.2. These are the per-car calculations.

Assume the first eight cars arrive at a single charging bay with the following inter-arrival times in minutes (denoting the times between successive arrivals) starting at $time = 0$:

$$[2, 8, 7, 2, 11, 3, 15, 9].$$

Their respective charging times in minutes at the station are

$$[11, 8, 5, 8, 8, 5, 10, 12].$$

These are the inputs to compute through a simple discrete event simulation as illustrated.

For the car charging discrete event model we'll compute basic measures to quantify the charging service operations. We will calculate the minimum and maximum queue length, number of customers served at the one-hour mark, the total idle time for the charging bay, the total queue time, and the average waiting time in queue. Table 1.2 shows the resulting events and measures using the given arrival and charging times.

TABLE 1.2
Example Discrete Event Calculations

Car #	Inter-arrival Time	Charging Time	Clock Time	Charging Start	End	Wait Time	Charger Idle Time	Queue Start	Length End
1	2	11	2	2	13	0	2	0	1
2	8	8	10	13	21	3	0	1	2
3	7	5	17	21	26	4	0	2	1
4	2	8	19	26	34	7	0	1	2
5	11	8	30	34	42	4	0	2	2
6	3	5	33	42	47	9	0	2	0
7	15	10	48	48	58	0	1	0	1
8	9	12	57	58	70	1	0	1	0

Table 1.3 shows system snapshots of all the events and statistics simulating this scenario as time progresses. It includes the event list and updated statistics at each event. The upcoming event list consists of arrival and departure events, of which the closest one in time is scheduled next for processing. The clock is updated at each event invocation and the ongoing statistics re-calculated. Additional mathematical description of this next event time advance approach is in Section 3.2.1.

Resource Utilization

An important measure of system performance is a measure of how busy the server is. The utilization of the server is the proportion of time during a simulation that the server is busy (not idle). The quantity can be measured as a continuous time average (similar to average queue length) by defining a binary busy function:

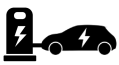
$$B(t) = \begin{cases} 0, & \text{if } x = 1 \\ 1, & \text{if } x = 2. \end{cases} \tag{1.1}$$

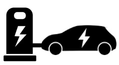
The utilization $u(n)$ is the proportion of time that $B(t)$ is equal to 1 as in Equation 1.2. In the equation, $u(n)$ is the continuous average of the $B(t)$ function, which matches intuition about utilization.

TABLE 1.3: System Snapshots with Events and Statistics

Clock	**Server**	**Queue**	**Statistics**
00:00			Cars Entered = 0 Cars Served = 0 Total Time in Queue = 0
Initialization	**Next Arrival** Car 1 at t=2	**Next Departure** -	Mean Waiting Time = N/A Utilization = N/A

Clock	**Server**	**Queue**	**Statistics**
00:02			Cars Entered = 1 Cars Served = 0 Total Time in Queue = 0
Event Car 1 Arrival	**Next Arrival** Car 2 at t=10	**Next Departure** Car 1 at t=13	Mean Waiting Time = 0 Utilization = 0

Clock	**Server**	**Queue**	**Statistics**
00:10			Cars Entered = 2 Cars Served = 0 Total Time in Queue = 0
Event Car 2 Arrival	**Next Arrival** Car 3 at t=17	**Next Departure** Car 1 at t=13	Mean Waiting Time = 0 Utilization = 0.8

Clock	**Server**	**Queue**	**Statistics**
00:13			Cars Entered = 2 Cars Served = 1 Total Time in Queue = 3
Event Car 1 Departure	**Next Arrival** Car 3 at t=17	**Next Departure** Car 2 at t=21	Mean Waiting Time = 1.5 Utilization = 0.85

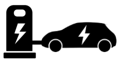
(*continued*)

TABLE 1.3: System Snapshots with Events and Statistics (*continued*)

Clock	**Server**	**Queue**	**Statistics**
00:17			Cars Entered = 3 Cars Served = 1 Total Time in Queue = 3
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 1.5
Car 3 Arrival	Car 4 at t=19	Car 2 at t=21	Utilization = 0.88

Clock	**Server**	**Queue**	**Statistics**
00:19			Cars Entered = 4 Cars Served = 1 Total Time in Queue = 3
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 1.5
Car 4 Arrival	Car 5 at t=30	Car 2 at t=21	Utilization = 0.89

Clock	**Server**	**Queue**	**Statistics**
00:21			Cars Entered = 4 Cars Served = 2 Total Time in Queue = 7
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 2.3
Car 2 Departure	Car 5 at t=30	Car 3 at t=26	Utilization = 0.90

Clock	**Server**	**Queue**	**Statistics**
00:26			Cars Entered = 4 Cars Served = 3 Total Time in Queue = 14
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.5
Car 3 Departure	Car 5 at t=30	Car 4 at t=34	Utilization = 0.92

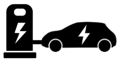
Clock	**Server**	**Queue**	**Statistics**
00:30			Cars Entered = 5 Cars Served = 3 Total Time in Queue = 14
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.5
Car 5 Arriving	Car 6 at t=33	Car 4 at t=34	Utilization = 0.93

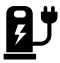
(*continued*)

TABLE 1.3: System Snapshots with Events and Statistics (*continued*)

Clock	**Server**	**Queue**	**Statistics**
00:33			Cars Entered = 6 Cars Served = 3 Total Time in Queue = 14
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.5
Car 6 Arrival	Car 7 at t=48	Car 4 at t=34	Utilization = 0.94

Clock	**Server**	**Queue**	**Statistics**
00:34			Cars Entered = 6 Cars Served = 4 Total Time in Queue = 18
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.6
Car 4 Departure	Car 7 at t=48	Car 5 at t=42	Utilization = 0.94

Clock	**Server**	**Queue**	**Statistics**
00:42			Cars Entered = 6 Cars Served = 5 Total Time in Queue = 18
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.6
Car 5 Departure	Car 7 at t=48	Car 6 at t=47	Utilization = 0.95

Clock	**Server**	**Queue**	**Statistics**
00:47			Cars Entered = 6 Cars Served = 6 Total Time in Queue = 27
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 4.5
Car 6 Departure	Car 7 at t=48	-	Utilization = 0.96

Clock	**Server**	**Queue**	**Statistics**
00:48			Cars Entered = 7 Cars Served = 6 Total Time in Queue = 27
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 4.5
Car 7 Arrival	Car 8 at t=57	Car 7 at t=58	Utilization = 0.94

(*continued*)

TABLE 1.3: System Snapshots with Events and Statistics (*continued*)

Clock	**Server**	**Queue**	**Statistics**
00:57			Cars Entered = 8 Cars Served = 6 Total Time in Queue = 27
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 4.5
Car 8 Arrival	-	Car 7 at t=58	Utilization = 0.95

Clock	**Server**	**Queue**	**Statistics**
00:58			Cars Entered = 8 Cars Served = 7 Total Time in Queue = 27
Event	**Next Arrival**	**Next Departure**	Mean Waiting Time = 3.86
Car 7 Departure	-	Car 8 at t=70	Utilization = 0.95

$$u(n) = \frac{\int_0^{T(n)} B(t)}{T(n)} \tag{1.2}$$

Waiting Time

The basic statistics for waiting time are the mean (average) waiting time in Equation 1.3 and sample variance in Equation 1.4.

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i \tag{1.3}$$

$$Var_W = \frac{1}{n-1}\sum_{i=1}^{n} (W_i - \overline{W})^2 \tag{1.4}$$

The mean waiting time for the first hour of operation (seven cars completing) is

Mean Waiting Time ($\overline{W}$) = (0 + 3 + 4 + 7 + 4 + 9 + 0) / 7
= 3.86 minutes.

Queue Length

The mean and variance of the queue length are in Equations 1.5 and 1.6 where T is the total simulation time and t_i is the length of time between the $(i-1)^{\text{th}}$ and i^{th} events. A graph of queue length is shown in Figure 1.1 and continued in Figure 1.2 (which can also be derived from Table 1.2).

$$\overline{L} = \frac{1}{T}\sum_{i=1}^{n} L_i t_i \tag{1.5}$$

$$Var_L = \frac{1}{T}\sum_{i=1}^{n} (L_i - \overline{L})^2 t_i \tag{1.6}$$

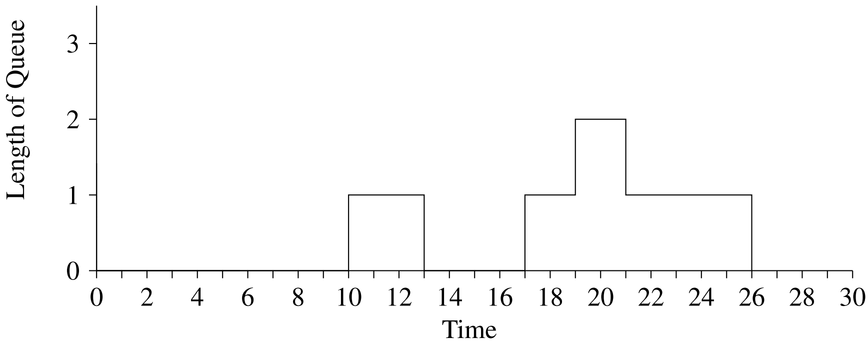


FIGURE 1.1: Length of Queue (0–30 Minutes)

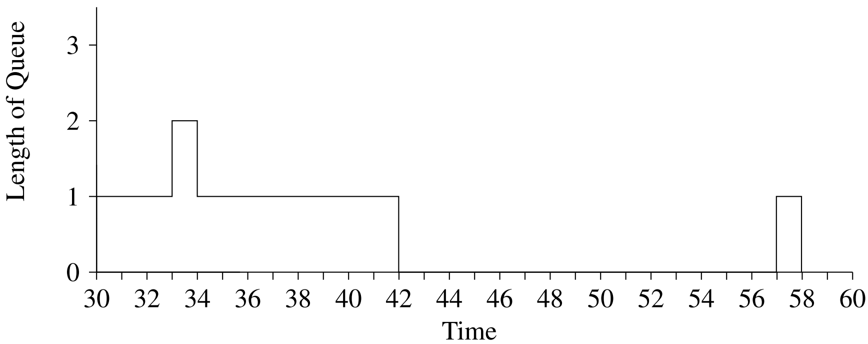


FIGURE 1.2: Length of Queue (30–60 Minutes)

$$\begin{aligned} \text{Average queue length } (\overline{L}) &= [(10-0)0+(13-10)1 \\ &\quad +(17-13)0+(19-17)1+(21-19)2+(26-21)1 \\ &\quad +(30-26)0+(33-30)1+(34-33)2+(42-34)1 \\ &\quad +(57-42)0+(58-57)1+(60-58)0]/60 \\ &= [0+3+0+2+4+5+0+3+2+8+0+1+0]/60 \\ &= .47 \text{ cars} \end{aligned}$$

A graph of the utilization for the charging resource is shown in Figure 1.3 and continued in Figure 1.4 (which can also be derived from Table 1.2).

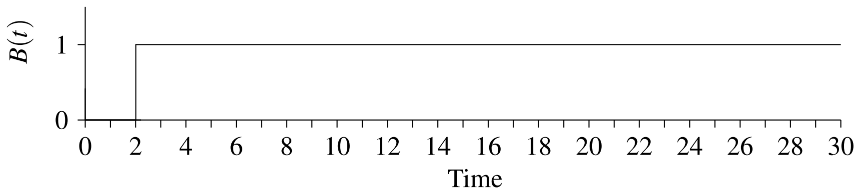


FIGURE 1.3: Resource Utilization (0–30 Minutes)

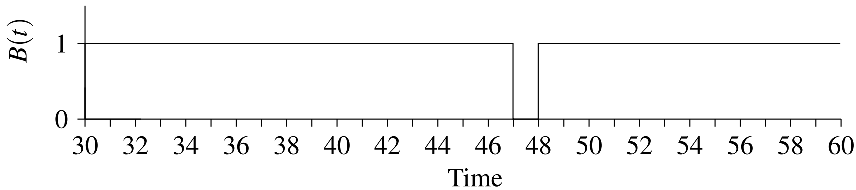


FIGURE 1.4: Resource Utilization (30–60 Minutes)

The resource utilization of the charging server for the simple example would be calculated as:

$$\begin{aligned}\text{Utilization } u(n) &= [(2-0)0+(47-2)1\\ &\quad +(48-47)0+(60-48)1]/60\\ &= [0+45+0+12]/60\\ &= .95.\end{aligned}$$

Thus the server is busy 95% of the time during the 60 minutes of simulation time. This car charging station example is the basis for further elaboration in subsequent chapters to illustrate the modeling process and concepts. It will be constructed as a stochastic model with randomness, inputs will be varied, it will run with various scenarios, outputs will be tested statistically, and conclusions drawn from the simulation experiments.

1.7 SUMMARY

Engineers, scientists, and managers should be well informed about M&S concepts. Virtually all complex engineering projects involve some modeling and simulation. Purely analytic methods seldom apply to complex, engineered systems with inherent uncertainties. Thus simulation models are used to understand how a system or process works, to design or redesign an engineering process, and to support training and learning.

A system is defined as a set of interacting parts that form a connected whole producing results not achievable by the elements alone. It must be represented in some form in order to analyze it and communicate about it. A model in this book's context is a logical, quantitative description of how a system or process behaves. Simulation is the numerical evaluation of such a mathematical model describing a system of interest.

The models are abstractions of real or conceptual systems used as surrogates for low-cost experimentation and study. Models allow us to understand a system or process by dividing it into parts and looking at how they are related. This book focuses on dynamic models used to simulate the time-varying behavior of a system. In time-based simulation the system is considered a collection of interacting elements, the properties of which change with time. These elements are referred to as entities and their properties as attributes.

The most popular types of dynamic simulation models are continuous, discrete event and agent based. A continuous model shows how values of the attributes change continuously as functions of time computed at equidistant small time steps. These changes are typically represented by smooth continuous curves. Continuous models may also be applied to systems that are actually discrete when the outputs have sufficient accuracy.

Discrete or discrete event models consist of entities and their attributes that change during simulation runs at aperiodic event times. These changes occur instantaneously as the simulated time lapses, and are reflected in the output as discontinuous fluctuations.

Agent-based models share common aspects with discrete models. They also model individual entities and use triggers to cause events, but the entities are autonomous agents that encapsulate their own rules of behavior. They operate within an external environment and interact with each other. At event times they may change their internal state.

Furthermore, simulation models can be deterministic, stochastic, or a hybrid. Most systems and models of concern are stochastic whereby the result variables differ from one run to another because of inherent randomness. For this the results must be analyzed statistically across multiple simulation runs to determine probabilistic means, standard deviations, distributions, etc. Monte Carlo simulation is commonly used to run models many times with random inputs for this purpose.

There is a very broad spectrum of applications in which modeling and simulation can be used across all engineering disciplines. In all these areas, simulation models are ultimately used to support engineering decision making based on the simulation results.

The most common discrete system is a queue/server system. In these systems, entities form lines (queues) in front of processing mechanisms called servers that provide resources. The entities use the resources to perform activities in the system. To gain insight for decisions, standard metrics from discrete event simulations include waiting time, queue length, and resource utilization. These measurements have impact in real-world operations and thus answer questions to address modeling goals.