CHAPTER 2

Models and Studies

Elements of a System Model

There are four main elements of a system model: the inputs, the outputs, the process model, and the model parameters. These are arranged as shown in Figure 2.1 to build a simulation of a system, for example, to simulate a business process. If the model employs discrete techniques it is termed a discrete system model.

Model Inputs

The inputs to a discrete system model are entities that arrive at the model for processing. These entities could arrive at a fixed rate, but usually there is some random variation in arrival times. Therefore, entity arrival times are usually described in terms of a standard statistical distribution. If a standard distribution cannot be found that fits the observed arrival data, a special distribution specific to the business system can be used for this description. For example, if truckloads of groceries arrive at a supermarket "every two hours, plus or minus 15 minutes," the arrival of these inputs might be described by a Normal distribution with a Mean of two hours (120 minutes) and Standard Deviation of 15 minutes. Other examples of such inputs might include customers, components, data frames, etc. In each case a representation of the arrival rate will be sought, to match direct observation in the field.

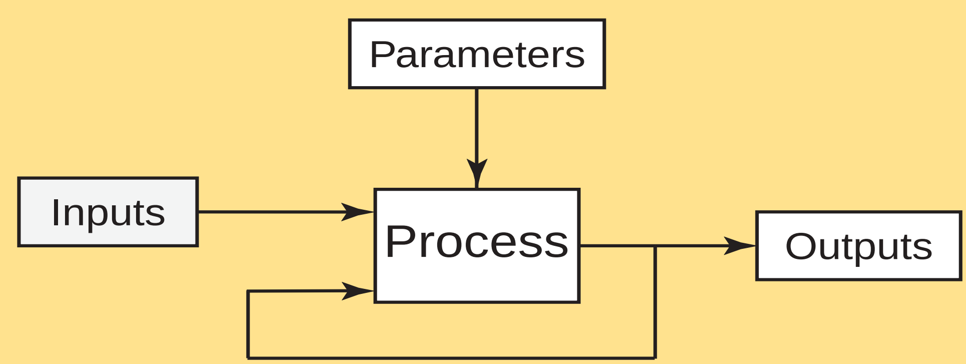

Figure 2.1 Structure of a simulation model

The arriving entities may also have characteristics that affect their later processing in the simulation model. These characteristics are called *Attributes* of the entities. Customers entering a coffee shop have their coffee preferences as attributes. When the simulated customers are generated in a simulation study, they must be generated with these preferences in approximately the same ratio as would be observed with the live customers. The simulation model will then split the customer stream into those with simple coffee preference and those who require special service, such as espresso preparation. Processing times for the two queues for these streams will differ just as it differs in the real coffee shop. Attributes such as these could be fixed in nature or could vary by standard or specialized statistical distributions. They are assumed to be determined at the time the input object arrives at the model. Thus all inputs are countable (discrete system) with known characteristics (attributes) upon arrival.

Model Parameters

The model is intended to act in a certain fashion. This behavior is controlled by the values of certain parameters, which are stored in tables. The model process references these tables when computing the simulation response to inputs. Typical of these parameters in a business process might be the statistical distribution to be used to represent service or the parameters governing the service time required to process an input. Other parameters might be used for cost accounting, such as the cost of resource use per unit of input or the expected outage characteristic of the resource. Parameters such as these are typically set by the model developer to reflect the observed behaviors of the business process modeled. They may be fixed for the duration of a simulation run, but successive simulation runs often are made with different parameters in the attempt to optimize a process. Having the simulation parameters tabulated and reported facilitates orderly analysis of these studies.

Model Process

The process is the heart of the model, and it applies the parameters to control response to the model inputs, producing the outputs. A process may be as simple as supplying a service to an arriving entity for an amount of time defined in the model parameters, and releasing the entity as an output to exit the system. Many processes are far more complex, of course, and respond to varying arrival rates and attributes, as well as the changing availability of resources.

The process may leave inputs unchanged, act on their attributes, or combine inputs into new entities. For example, a process might change the status of a truck from "loaded" to "empty," unbatch a truckload into a series of pallets, decrement the amount of money available to a customer, or combine foodstuffs in a shopping cart. Processes are controlled by the model parameters. Thus, if the statistical service time model is set by the parameters to Exponential with a given mean value, the discrete simulation system will apply these parameters, causing inputs to be processed accordingly. The process model and parameters thus act in concert to allow a given business process to be simulated under varying conditions.

Model Outputs

Entities exiting the model are termed outputs. They may carry attributes contributed by the process, or retain attributes acquired earlier. They may also acquire a time stamp or time of departure. As an example of an attribute acquired during processing, consider the security status of an airline passenger, “checked” or “unchecked.” Such attributes may influence downstream processing by later parts of the simulation model. Retained attributes may be unchanged by the process, as First Class Ticket status for instance, or they may be modified by the process, as exemplified by the available cash of a traveler who pays an airport tax. Attributes can be queried by the simulation system for reporting purposes.

More Complex Models

Models may have multiple processes, and entities or objects may take multiple and complex paths through the system. For example, truckloads may arrive at a fixed average rate with defined arrival time variations. A process can unload the trucks into a number of pallets defined by the characteristics of the truckload. Each pallet can then be viewed as an input, in its own right, to a downstream process and proceed to the appropriate conveyor designated by its attributes. The pallet can be further unbatched into a stream of entities, each of which then moves down a conveyor with its own quality attributes. These attributes may then be uncovered by an inspection process, causing further division of the stream. Streams can also be merged and batched into pallets and then loaded onto trucks, a reverse of the previous actions. The sequence of these actions may be complex, and the actions may be affected by attributes or modify the values of attributes.

The modeler can reduce this complexity by combining groups of operations into subsystems. The subsystem can then be combined into larger systems, and so on, until a comprehensible model evolves. By this process the modeler can readily simulate even very complex business systems. The outputs and reports from these simulations can guide the tuning and validation of the overall model and eventually provide insights into process improvements. New process ideas can be tested on the model first, allowing refinements and setting baselines for actual process operation expectations.

Model Development

Phases of Development

The goal of model development is to represent the actual process to the computer for the purposes of simulation. A system model can be considered organized as a hierarchy. At the top level there is the overall process, consisting of a number of interconnected operations. An example would be an entire manufacturing plant for a certain product. At the next level are the individual operations. To pursue the example, the model components might include a warehouse, preparation, assembly, testing, and packaging operations. These operations could easily be of such large scale that they have their own management structures. These operations, in turn, could consist of sub-operations, and so on, as suggested in Figure 2.2. A warehouse operation, for instance, might contain receiving, storage, retrieval, and dispensing sub-operations. A decision must be made regarding what portion of the hierarchy will be represented by the eventual model. If the concerns are with overall operation and the interface to customers and other organizations, perhaps only the top level will be modeled. If, however, there is concern over the impact of new production machinery on upstream and downstream processes, only those operations need be modeled. The general rule is to thoroughly model only what is needed and to avoid extending the model to more detail than is absolutely necessary.

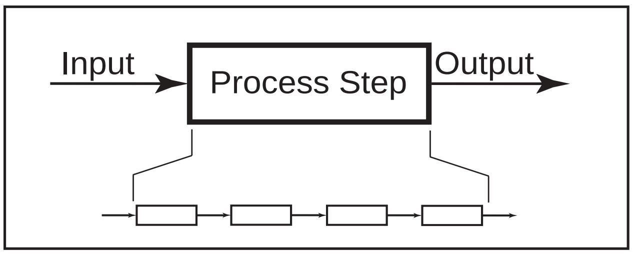

Figure 2.2 Model hierarchy

Each process step can be subdivided into component steps. This subdivision can continue until the desired level of modeling detail is reached

The development of the system models frequently proceeds through various steps, as suggested in Table 2.1. At each stage some elements of the model or data are held constant, while the others are allowed to vary. In a *model selection phase*, input and output data sets, gathered in the field, are held constant. Then the model and its parameters are adjusted to give a good fit of simulated to actual outputs, given the observed input data. This development is typically performed by the engineering scientist, because it frequently requires mathematical expertise as well as process know-how. The input and output data sets are gathered by observations in the field. The model is developed by postulating a form of model that might fit the data and then by adjusting the parameters of this model to see if it will produce outputs similar to those gathered in the field when exposed to inputs similar to those recorded. If there is a good fit between the model performance and the observations, then the model is selected. If the modeling results do not fit the observations within predefined confidence limits, then another form of model is tried and its parameters are adjusted similarly. Finally, this cycle of model refinement is repeated until the model is adequate.

Table 2.1 Model development studies

	Inputs	**Model**	**Parameters**	**Outputs**
Model selection	fixed	varied	varied	fixed
Parameter optimization	fixed	fixed	varied	fixed
Model limitations	varied	fixed	fixed	varied

In the *parameter optimization phase*, the model and observed data, both input and output data, are held constant. Model parameters are then adjusted to optimize the model fit to these data. The objective of this phase is to bring the model into agreement with the data as closely as possible, in some sense. The degree of closeness is usually measured objectively through the evaluation of a mathematical function, often called an *objective function*. The minimum sum of squared errors between model observed values and those calculated by the model would be an example of such a function. This cycle of development may have to be repeated as various models are tested and responses evaluated.

In a *model limitations evaluation*, an engineer further develops the model by varying inputs to give differing performance, under the conditions of constant model and parameters. The goal of such a study would be to determine the limits of validity of the process model. To perform this study, an increased loading on the process is often arranged, by varying the rate of arrival of inputs to be processed. At some arrival rate the ability of the model to deal with these inputs correctly will become saturated. Additional arrivals will only overflow the internal queue or cause some other form of conflict with reality. An example would be the arrival of cars at a car wash. Statistical models of this type assume that each arrival is independent of the previous ones, yet this condition for model validity is easily violated under actual conditions. First of all, the interarrival time between successive cars cannot be smaller than the time it takes to move the first car out of the way. Secondly, drivers can view the waiting line and could decide not to have the car washed. Thirdly, the average arrival rate of cars cannot exceed the average throughput through the car wash. It is often important to know such limitations. This type of study leads to assessment of the limitations of the system and its model in response to various inputs.

Gathering Field Data

Model development assumes the availability of statistical data about the inputs and outputs of a process. These inputs and outputs are events, such as customer arrivals or completions of customer service. The process will take action on the inputs, and these actions will produce output events. When recording these events, for model building purposes, the most effective method is to record observations with an accurate time stamp, as suggested in Table 2.2. Elapsed times and arrival rates can then be computed from these raw data. A fundamental assumption, which must be assured for accurate modeling, is that the process of observation does not affect the process observed. This may be difficult to insure, since observers with timers may affect performance of operations observed. One method the author has found effective, although time consuming, is to use a tape recorder to record arrivals and service completions. Times can then be read by digitizing the tape and processing it with a sound editor. This allows the observer to pay close attention to the process and insures consistent data.

Various supporting measures might be used to ease the burden of data collection. For instance, customers might be given a time stamped card upon entry into the queue. When these are submitted at time of service, the time in queue can be automatically and accurately

Table 2.2 Data to be gathered for building a statistical model

Arrivals (each sub-process)
- Entity name
- Quantity
- Time stamp
  - Date
  - Time
  - Location
- Attribute(s)

Path taken
- Length of path
- Travel velocity or time
- Attribute dependencies

Queue(s)
- Number
- Length
- Discipline (FIFO, Priority, etc.)

Service
- Number of servers
- Server properties (each server)
  - Service time
  - Quality of server
  - Resources used
  - Resource types
  - Resource downtimes expected
  - Costs
    - fixed
    - variable
- Process performed
- Enabling events (if any)
- Triggering events (if any)

Exit
- Entity name
- Quantity
- Time stamp
  - Date
  - Time
  - Location

captured. The cards could also capture customer attributes or preferences at time of issue, as well as record options performed at time of service. Such automatic data gathering could be conducted either with customer knowledge or without.

Entities arriving at a business process for service may be modeled statistically. Either the interarrival time or the number of arrivals per unit time, or both, can be modeled by fitting field data to standard statistical distributions. For purposes of documentation in the field, the name of the arriving entity and the number arriving are recorded with a time stamp. The time between successive time stamps will be used to calculate interarrival times. Since data may be gathered at multiple locations, in a complex process, this location will also be recorded. Any attributes of each arrival that could affect its processing time should be noted as well. This will be useful in later modeling of the service times. As will be discussed below, many data samples should be taken, and the data gathered should span the range of expected input conditions. This would imply data gathering during peak and off-peak times, various days of the week, various weather conditions, etc. Arrival times and numbers affect system loading, and an attempt should be made to gather enough data to permit formulation of an accurate model for each case to be considered.

Some processes and process delays are sensitive to the path taken. Warehouse operations, in which materials are transported across potentially conflicting paths are an example of this. In these cases, the data gathered should reference the path taken, and the velocity of travel or transit time should be recorded. There may also be attribute dependencies. An obvious case would be the transport of oversize loads on public highways. Knowing these facts, and knowing the total number of transport operations, the number of conflict cases, with their corresponding delays, can be modeled.

Discrete systems governed by statistical uncertainties generate queues of entities awaiting service. In general, the greater the uncertainty, and the greater the arrival rate, the longer the necessary queue will become. Gathering queue lengths in the field would be redundant, because these states can be found from arrival and service data. Nevertheless, recorded queue lengths can act as a check on the other data gathered. In any case, the nature of the queue discipline (FIFO, priority, or other) should be noted, since this will affect later processing. Note as well the nature of the queue. Do customers wait in a single queue for one of a number of servers to be free? Or, does each server have his or her own queue? Often, simulation studies will seek to determine the optimal type of customer queue for a given establishment and set of conditions.

Service systems are characterized by the number of servers, drawing entities from one or more queues, and by the properties of each server. Among the important properties, from the point of view of model development, is the service time that each server needs to perform the assigned tasks. Thus, the number of servers and their individual service times must be recorded in the field for later model development. It is often the case, however, that service times will vary according to certain properties of the arriving entities to be serviced. Field data should thus record factors needed to understand such dependencies.

Services also use resources, both human and material. Thus, when gathering field data, resource utilizations must also be documented, including both the quantities and the costs. Cost will be of both fixed and variable type, and these should be noted for subsequent development of a cost model. Since all services have associated downtimes, for maintenance or for human factors, these must also be researched and recorded. It will be found helpful to organize service data according to the operations performed. If these are very complex, and are significant to process costs, further hierarchic breakdown may be in order.

Services are sometimes conditional on certain enabling or triggering events. An enabling event is one that allows the service to proceed. The presence of an elevator operator may be required, for instance. There may be multiple enabling conditions, which must all be simultaneously present for the service action to occur. Once all enabling conditions are met, the triggering event will allow the serving process to proceed. Triggering events, such as the push of a button, are those that ultimately cause the service to take place. Even when the entities to be processed have arrived and all enabling events have taken place, the service must await the triggering event. Such events, if present, should also be noted in the field. One goal of a simulation study is to discover improvements resulting in reduced server idle time. Documentation of enabling and triggering events, in conjunction with service initiation and completion events, can serve this purpose.

Problems with Statistical Data

Discrete systems require statistical models, and the statistics of a process may change over time. Such cases are called non-stationary. Telephone calls and stock market trades, which tend to cluster at certain times of the day, would be examples of this. In such systems the mean arrival rate at one time of day may not be the same at some other time of day. This means that the statistical model is changing over the day, greatly complicating study of the process by simulation. In order to build a model to simulate such inputs and their system response, windows in time must be defined in which the statistics are *assumed* stationary. In other words, even though rates are changing, they may be *considered* constant over a short time interval. Unfortunately, the faster the mean rates are changing, the smaller such windows must be, with the result that fewer data samples are taken per time window. If this does not produce an adequate number of samples for analysis on a given day, then samples from corresponding windows at other times or days can be merged, assuming that the statistical distributions are equivalent.

Consider the car wash example, again. Customer arrivals will tend to be clustered around certain times of the day, and on certain days of the week. Suppose Saturday afternoons at 2 PM are peak times, with many customers per hour. If we wish to model the behavior of the system this time, we could not use data from other times that have lower mean arrival rates, such as Monday mornings. Instead we could pool recorded customer arrivals between 1 PM and 3 PM on a number of Saturdays, assuming approximately equivalent weather conditions and the absence of significant football games! It is easy to see that equivalency of the data between Saturdays will be in question. A more accurate approach might be to record arrivals over a wider window, say 10 AM to 4 PM on many Saturdays, and then combine data only in those windows for which the daily rates look similar. This merging of statistical data is a topic of research that will be avoided in the present discussion. For most models. It will suffice to take reasonable precautions to combine only data having similar statistical properties.

Gathering data to model processing times will be subject to similar problems, especially when human operators are observed. Processing times may vary with the number of hours of continuous operation. Workers tire, for instance, slowing down processing times. Machines may work faster when warmed up. Workers may work differently when observed. Worker productivity is also affected by seasonal and other influences. The solution is, again, to combine data from intervals considered to reflect similar behavior.

Validating the Model

The statistical model should be validated before it is used in a simulation study. Validation is the process of insuring agreement between model behavior and observed behavior under similar conditions of input. There are two cases in which model and observation must conform: (1) the model behavior and observed behavior must agree for the accumulated set of observations, and (2) the model behavior and observed behavior must agree for new observations under similar conditions. That is, the model must explain past behavior and it must predict future behavior. It should also be noted that the random number generators produced by computers are really not random but are a repeating sequence, beginning from a "seed" number. Validation should therefore be tested with alternative seed numbers, in order to remove any artifacts of the non-random generation process within the computer.

Simulation results should also make sense. Another important step in validating a computer model is to ask an expert in the problem domain if the results look reasonable. This should be checked under various operating regimes, such as time of day and degree of system loading. Through this, an attempt is made to bracket the region in which the model is valid. The model should then not be applied when inputs or operating regimes are outside this region.

In any set of observations of a physical process, "outlier" data occurs that does not fit the model. The validation process may uncover these deviations. In studying data network traffic between two cities, for instance, the low level traffic that keeps the network in operation can be safely ignored, because its intensity is far lower than the traffic due to institutional data flow. Similarly, car wash traffic after a rainstorm would not match normal patterns. Furthermore, the statistical model used in a simulation might not generate observed activity exactly. This can happen because an approximation was used, in order to use a standard statistical model, or the process may be non-stationary. A decision about how to handle these anomalies must be made. Whatever decisions are made, the validation process must insure that the final model accurately represents process behavior under the conditions to be simulated.

Simulation Study Design

A simulation study evaluates the performance of a system by using input and validated process models to predict the behavior of the system simulated. Studies can be of various types, as suggested in Table 2.3. The table shows various types of simulation study and the fixed or flexible nature of the corresponding data, parameters, or model.

Table 2.3 Simulation studies

	Inputs	Model	Parameters	Outputs
Process development	varied	fixed	fixed	varied
Process optimization	fixed	fixed	varied	varied
Process alternatives	fixed	varied	varied	varied

In a *process development study*, an engineer develops process know-how by varying inputs to give best performance, under the conditions of constant model and parameters. The goal of such a study would be to determine the processing limitations of a system, perhaps for marketing purposes. To perform this study, a hypothetical loading on the process is varied, most probably by varying the rate of arrival of inputs to be processed. At some arrival rate the ability of the process to deal with these inputs will be saturated. Additional arrivals will only overflow the internal queues. It is often important to know this limitation. This type of study has been used to justify the metering (throttling!) of traffic entering limited access highways, for instance. Studies have apparently shown that limiting highway entry rates can improve total highway throughput. This type of study leads to assessment of the robustness of the system to various inputs, and it can be a precursor to further product or process development.

In a *process optimization study*, the model and input data are held constant, and parameters are adjusted to optimize performance of the system. The cycle of development is frequently repeated as various combinations of parameters are tested and responses evaluated. This is a successive refinement procedure that should lead to optimal operation for the anticipated inputs. A study of this type might be conducted by a process manager or in a market study to define new product features. The result would be a set of parameters differing from the original model that would give better performance under the input conditions tested. This develops insight into new processing features, product specifications, or operating regimes that might be desired to deal with the contemplated input patterns. Such studies can also be used to minimize operating costs. This study can also give information on the probable behavior of a process under degraded conditions in which parameters have been altered.

A *process alternatives study* provides another study variant. In this case inputs are fixed, but there may be competing models. Parameters may be varied, for sensitivity study. The competing models are tested in response the given input data, and their merits are compared. In a simulation that tracks incremental processing costs, this type of study could also be used to reduce resource-related processing costs, the cost of in-process inventory, or the cost of processing delays. This can lead to an evolutionary strategy for process improvement.

Summary

Inputs, outputs, parameters, and process model are the main elements of a process model. Development of this model goes through model selection, parameter optimization, exploration of limitations, and exploration of alternatives. Each of these stages is characterized by holding constant one or more of the model elements and letting the remaining ones be free to settle to their natural results. Statistical data gathering is necessary for building the basic model, and techniques for gathering such data and validating the eventual model have been suggested. With a validated model, various simulation studies can be undertaken for the purposes of process documentation, development, optimization, and exploration of operational alternatives. Once again, constrained and free model elements are the organizing features of the studies.

Questions

For one or more of the following business processes, describe what kinds of simulation studies might be useful. Then identify the inputs, parameters, processes, and outputs for modeling and describe how data would be taken for building a model. Also identify any significant attributes and describe their possible influences on processing. Be sure to tell whether there will be daily, seasonal, or other non-stationary influences and how these will be handled in the modeling. Identify any expected risks and develop an approach to them.

(1) A car wash

(2) A bakery

(3) A bank branch office

(4) A “Seattle style” coffee shop

(5) A fast food restaurant

(6) A specialty ice cream parlor

(7) A subway or commuter train station

(8) An airline terminal

Design Homework

Develop a plan for simulating one of the businesses mentioned in the foregoing Questions, or a business process of your choice. Your simulation plans should include detailed plans for data gathering, model development, model selection, model validation, evaluation of processing alternatives, and process optimization. You will need to develop one or more data gathering aids, such as preprinted data sheets, that can be used in the field. Please discuss how you will study the behavior of the system under high customer loading.

Laboratory Homework

Try out your data gathering aids and methodology on a process of your choice.