MATH 5131 Lecture 1

A better description of applied mathematics had been given by John L. Synge. He noted that applying mathematics to a real problem involves three stages. The first stage is a dive from the world of reality into world of mathematics. The second stage is a swim in the world of mathematics. The third stage is a climb back from the world of mathematics into the world of reality and, importantly, with a prediction in your teeth.

Suppose we want to understand some behavior or phenomenon in the real world. We may wish to make predictions about that behavior in the future and analyze the effects various situations have on that behavior. For instance, when studying the populations of two interacting species, we may wish to know if the species can coexist within their environment or if one species will eventually dominate and drive the other to extinction. Or in the management of a fishery, it may be important to determine the optimal sustainable yield of a harvest and the sensitivity of the species to population fluctuation caused by harvesting. How can we construct and use models in the mathematical world to help us better understand real-world systems?

System

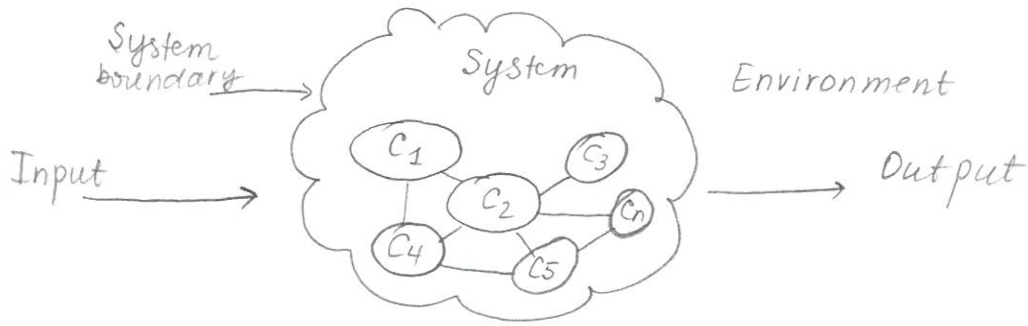
Let us consider what we mean by a real-world system.

Definition A system is an assemblage of objects joined in some regular interaction or interdependence.

Most systems that surround us are multidimensional, extremely complex, time varying, and nonlinear in nature as they are comprised of large varieties of actively or passively interacting subsystems. These systems consist of interacted subsystems, which have separate and conflicting objectives. The term system is derived from the Greek word systema, which means an organized relationship among functioning units or components. It is used to describe almost any orderly arrangement of ideas or construct.

FIGURE 1

System as collection of interconnected components

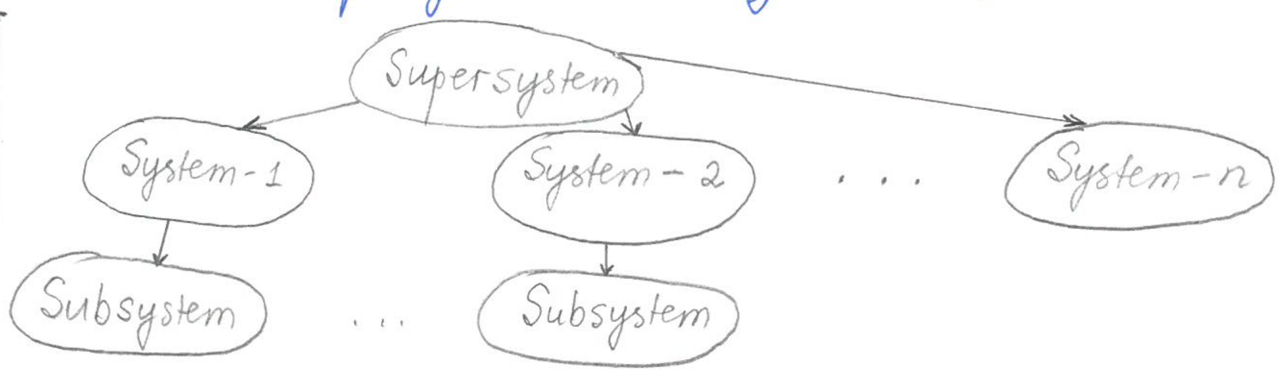


Some examples of the systems are:

- Esoteric systems
- Medical / biological systems
- Socioeconomic systems
- Communication and information systems
- Planning systems
- Manufacturing systems
- Management systems
- Transportation systems
- Physical systems – electrical, mechanical, thermal, hydraulic systems, and combinations of them.

Every system consists of subsystems or components at lower levels and supersystems at higher levels.

FIGURE 2



Hierarchically nested set of systems.

One needs to be extremely careful to define such a hierarchically nested system because this will determine the kind of results one will obtain.

A system is characterized by the following attributes:

- System boundary
- System components and their interactions
- Environment

System boundary

To study a given system, it is necessary to determine what comprises (falls inside and what falls outside) a system. For this demarcation is required to differentiate entities from the environment. Such a partition is called a system boundary. The system boundaries are observer-dependent, time-dependent, and most importantly system-dependent. The different observers may draw different boundaries for the same system. Also, the same observer may draw the system boundaries differently for different times. Finally, they may also be drawn differently with respect to the nature of the study, that is, steady state and transient.

Some salient points about the system boundary are:

- It is a partitioning line between the environment and the system.
- System is inside the boundary and environment is outside the system.
- A real or imaginary boundary separates the system from the rest of the universe, which is referred to as the environment or surroundings
- System exchanges input - output from its environment
- This boundary might be material boundary (like skin of a human body) or immaterial boundary (like the membership to a certain social group).
- Considering a system boundary in system's analysis and evaluation is of immense importance as it helps in identifying the system and its components. The interaction between a system and its environment takes place mainly at the boundaries. It determines what can enter or leave a system (input and output).

- System boundary may be crisp (clearly defined) or fuzzy (ill defined). In crisp boundaries, it is quite clear that what is inside the boundary (i.e., part of system) and what is outside the boundary (i.e., part of environment). In fuzzy boundaries, it is not very clear whether a particular component belongs to the environment or the system.

System Components and Their Interactions

System component is a fundamental building block. It is quite easy to find the input-output relations for the system components with the help of some fundamental laws of physics, which is called the mathematical model for components. It may be written in the form of difference of differential equations. They are pretty simple and easily understandable.

Business system environment includes customers, suppliers, other industries, and government. Its inputs include materials, services, new employees, new equipment, facilities, etc. Output includes product, waste materials, money, etc.

- It is static or dynamically changing with time, input, or state of the system.
- Interaction may be constrained or nonconstrained type.
- The component interaction may be unidirectional or bidirectional.
- Interaction strength may be 0, 1, or between 0 and 1.
  a. If interaction strength is zero then there is no interaction.
  b. If interction strength is one it means full interaction and if the interaction strength lies between zero and one, then the interaction is partial interaction.

Environment

A living organism is a system. Organisms are open systems: they cannot survive without continuously exchanging matter and energy with their environment.

When we separate a living organism from its surraending, it will die shortly due to lack of oxygen, water, and food. The peculiarity of open systems is that they interact with other systems outside of themselves. This interaction has two components: input, that is, what enters the system from outside the boundary, and output, that is, what leaves the system boundary to the environment. In order to speak about the inside and the outside of a system, we need to be able to distinguish between the system and its environment, which is in general separated by a boundary (for example, living systems, skin is the boundary). The output of a system is generally a direct or indirect result to given input. For example, the food, drink, and oxygen we consume are generally separated by a boundary and discharged as urine, excrements, and carbon dioxide. The transformation of input into output by the system is usually called throughput.

A system is intended to "absorb" inputs and process them in some way in order to produce outputs. Outputs are defined by goals, objectives, or common purposes. In order to understand the relationship between inputs, outputs, and processes, we need to understand the environment in which all of this occurs. The environment represents everything that is important to understand the functioning of the system, but it is not part of the system. The environment is that part of the world that can be ignored in the analysis except for its interaction with the system. It includes competition, people, technology, capital, raw materials, data, regulation, and opportunities.

When we are concerned only with the input and correspon-ding output of a system, while undermining the internal intricacies of component-level dynamics of the system, such study may be called a black box study. For example, if we consider a city, we may safely measure the total fuel consumption (input) of the city and the level of emmisions (output) out of such consumptions without actually bothering about trivial details like who/what consumed more and who/what emitted or polluted the most.

Such point of view considers the system as a "black box" that is, something that takes input, and produces output, without looking at what happens inside the system during process. Contrary to the former, when we are equally concerned about the internal details of the system and its processes besides the input and output variable, such an approach of system is considered as white box. For example, when we model a city as a pollution production system, regardless of which chimney emitted a particular plume of smoke, it is sufficient to know the total amount of fuel that enters the city to estimate the total amount of carbon dioxide and other gases produced. The "black box" view of the city will be much simpler and easier to use for the calculation of overall pollution levels than the more detailed "white box" view, where we trace the movement of every fuel tank to every particular building in the city.

The system as a whole is more than the sum of its parts. For example, if person A alone is too short to reach an apple on a tree and person B is too short as well, once person B sits on the shoulders of person A, they are more than tall enough to reach the apple. In this example, the product of their synergy would be one apple.

Classifications of Systems

Systems can be classified on the basis of time frame, type of measurements taken, type of interactions, nature, type of components, etc.

- According to the Time Frame

Systems can be categorized on the basis of time frame as

- Discrete
- Continuous
- Hybrid

A *discrete* system is one in which the state variables change instantaneously at separated points in time, for example, queuing systems (bank, telephone network, traffic lights, machine breakdowns), card games, and cricket match. In a bank system, state variables are the number of customers in the bank, whose value changes only when a customer arrives or when a customer finishes being served and departs.

A *continuous* system is one in which the state variables change continuously with respect to time, for example, solar system, spread of pollutants, charging a battery. An airplane moving through the air is an example of a continuous system, since state variables such as position and velocity can change continuously with respect to time.

Few systems in practice are wholly discrete or wholly continuous, but since one type of change predominates for most systems, it will usually be possible to classify a system as being either discrete or continuous.

A *hybrid* system is a combination of continuous and discrete dynamic system behavior. A hybrid system has the benefit of encompassing a larger class of systems within its structure, allowing more flexibility in modeling continuous and discrete dynamic phenomena, for example, traffic along a road with traffic lights.

- According to the Complexity of the System

Systems can be classified on the basis of complexity

- Physical systems
- Conceptual systems
- Esoteric systems

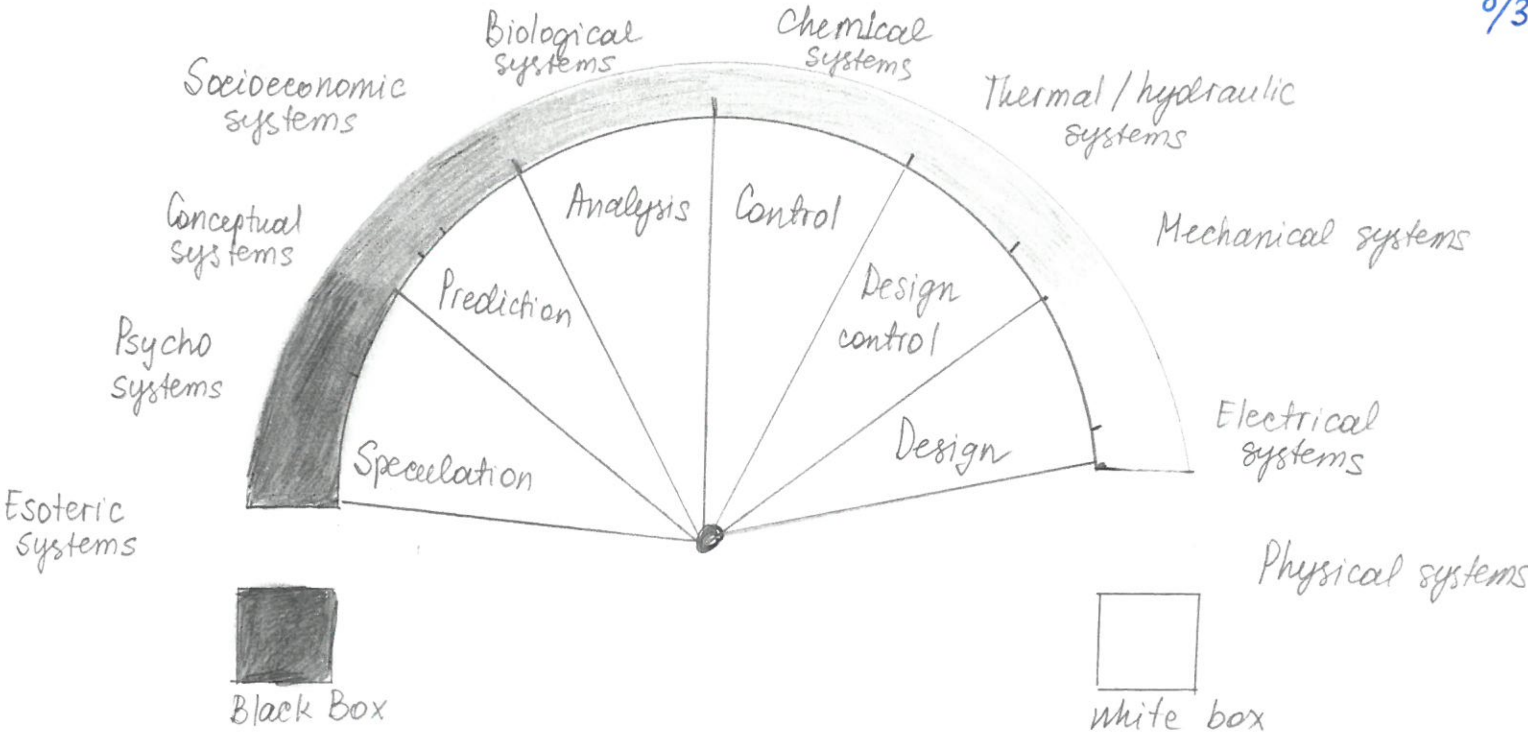


Figure 3 Classification of system based on complexity

Physical systems can be defined as systems whose variables can be measured with physical devices that are quantitative such as electrical systems, mechanical systems, computer systems, hydraulic systems, thermal systems, or a combination of these systems. Physical system is a collection of components, in which each component has its own behavior, used for some purpose. These systems are relatively less complex.

Conceptual systems are those systems in which all the measurements are conceptual or imaginary and in qualitative form as in psychological systems, social systems, health care systems, and economic systems. (Ex. the transportation system). Conceptual systems are those systems in which the quantity of interest cannot be measured directly with physical devices. These are complex systems.

Esoteric systems are the systems in which the measurements are not possible with physical measuring devices. The complexity of these systems is of highest order.

According to the Interaction

Interactions may be unidirectional or bidirectional, crisp or fuzzy, static or dynamic, etc. Classification of systems also depends upon the degree of interconnection of events from none to total. Systems will be divided into three classes according to the degree of interconnection of events.

1. Independent – If the events have no effect upon one another, then the system is classified as independent.
2. Cascaded - If the effects of the events are unilateral (that is, part A affects part B, B affects C, C affects D and not vice versa), the system is classified as cascaded.
3. Coupled – If the events mutually affect each other, the system is classified as coupled.

According to the Nature and Type of Components

1. Static or dynamic components
2. Linear or nonlinear components
3. Time-invariant or time-variant components
4. Deterministic or stochastic components
5. Lumped parametric component or distributed parametric component.
6. Continuous-time and discrete-type systems.

According to the Uncertainties Involved

Deterministic – No uncertainty in any variables, for example, model of pendulum.

Stochastic – Some variables are random, for example, airplane flight with random wind gusts, mineral-processing plant with random grade ore, and phone network with random arrival times and call lengths.

Fuzzy systems – The variables in such type of systems are fuzzy in nature. The fuzzy variables are quantified with linguistic terms.

Static vs. Dynamic Systems

Normally, the system output depends upon the past inputs and system states. However, there are certain systems whose output does not depend on the past inputs called static or memoryless systems. On the other hand, if the system output depends on the past inputs and earlier system states which essentially implied that the system has some memory elements, it is called a dynamic system. For example, if an electrical system contains inductor or capacitor elements, which have some finite memory, due to which the system response at any time instant is determined by their present and past inputs.

Linear vs. Nonlinear Systems

The study of linear systems is important for two reasons:

1. Majority of engineering situations are linear at least within specified range.

2. Exact solutions of behavior of linear systems can usually be found by standard techniques.

Except, a handful special types, there are no standard methods for analysing nonlinear systems. Solving nonlinear problems practically involves graphical or experimental approaches. Approximations are often necessary, and each situation usually requires special handling. The present state of art is such that there is neither a standard technique which can be used to solve nonlinear problems exactly, nor is there any assurance that a good solution can be obtained at all for a given nonlinear system.

The Ohm's law governs the relation between the voltage across and the current through a resistor. It is a linear relationship because voltage across a resistor is linear preportional to the current through it.

$$V = RI.$$ (The resistor is held at a constant temperature)

But even for this simple situation, the linear relationship does not hold good for all conditions. Practically all substances show a variation of resistance with change in temperature. All metals and most alloys used in electrical engineering increase in resistance with increase in temperature; the resistance of nonmetalic conductors such as carbon and electrolytes, also of most dielectrics, decreases with increase in temperature. In general, the relation between the resistance R_t of a given mass of a substance at any temperature t may be expressed in terms of its resistance R_0 at zero degrees by the following power series:

$$R_t = R_0 (1 + at + bt^2 + ct^3 + \text{etc.})$$

where R_t and R_0 are in ohms, t is in degrees centigrade, and a, b, c, etc., are constants. With metalic conductors such as copper and aluminium the following expression is sufficiently accurate for most practical purposes:

$$R_t = R_0 (1 + \alpha_0 t)$$ [For dielectrics this linear relationship is not sufficiently accurate.]

where R_t and R_0 are as above and α_0 is called the zero degree temperature coefficient of resistance. α_0 is the change in ohms per ohm per degree in the neighborhood of zero centigrade.

The amount of change in resistance is being dependent upon the magnitude of the current, and it is no longer correct to say that the voltage across the resistor bears a linear relationship to current through it.

Similarly, the Hooke's law states that the stress is linearly proportional to the strain (deformation) in a spring. But this linear relationship breaks down when the stress on the spring is too great. When the stress exceeds the elastic limit of the material of which the spring is made, stress and strain are no longer linear related. The actual relationship is much more complicated than the Hooke's law situation, that is,

$$\text{Stress}(\sigma) \propto \text{Strain}(\varepsilon) \quad \left[\frac{\text{About symbol } \propto:}{\sigma \text{ varies directly as } \varepsilon} \right]$$

Therefore, we can say that restrictions always exist for linear physical situation, saturation, breakdown, or material changes with ultimate set in and destroy linearity. Under ordinary circumstances physical conditions in many engineering problems stay well within the restrictions and the linear relationship holds good.

Linear Systems

An engineer's interest in a physical situation is very frequently the determination of the response of a system to a given excitation. Both the excitation and the response may be any physically measurable quantity, depending upon the particular problem.

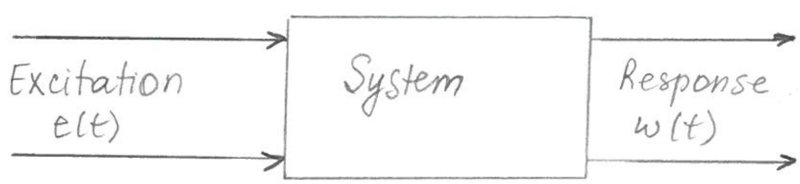


Fig. 4 Representation of two-port system

The linear system obeys superposition and homogeneity theorems.

- ### Superposition Theorem

Suppose that an excitation function, $e_1(t)$ which varies with the time in a specified manner, produces a response function, $w_1(t)$, and a second excitation function, $e_2(t)$ produces a response function $w_2(t)$.

Symbolically, the above-mentioned observation may be written as

$$e_1(t) \rightarrow w_1(t) \quad \text{and} \quad e_2(t) \rightarrow w_2(t)$$

For the linear system, if these excitations applied simultaneously

$$e_1(t) + e_2(t) \rightarrow w_1(t) + w_2(t)$$

The above equation shows the superposition theorem, which can be described as a superposition of excitation functions results in a response which is the superposition of the individual response functions.

- Homogeneity

If there are n identical excitations applied to the same part of the system, that is, if

$$e_1(t) = e_2(t) = e_3(t) = \cdots = e_n(t)$$

then for a linear system

$$\sum_{k=1}^{n} e_k(t) = n e_1(t) \rightarrow \sum_{k=1}^{n} w_k(t) = n w_1(t)$$

[For functions, which are homogeneous of degree 1: $f(kx) = k^1 f(x)$]

Hence, a characteristic of linear systems is that the magnitude is preserved. This characteristic is referred to as the property of homogeneity.

⚠ A system is said to be linear if and only if both the properties of superposition and homogeneity are satisfied.

Time-Varying vs. Time-Invariant Systems

A <u>time-variant</u> <u>system</u> is a system whose output response depends on moment of observation as well as moment of input signal application. In other words, a time delay or time advance of input not only shifts the output signal in time but also changes other parameters and behavior. Time-variant systems respond differently to the same input at different times.

If system parameters do not change with time then such systems are called <u>time-invariant</u> (or constant parameter) systems. If the excitation function $e(t)$ applied to such a system is an alternating function of time with frequency f, then the steady-state response $w(t)$, after the initial transient has died out, appearing at any part of the system will also be alternating with the frequency f. In other words, time-invariant nonlinear systems create no new frequencies

Lumped vs. Distributed Parameter Systems

A lumped system is one in which the components are considered to be concentrated at a point. For example, the mass of a pendulum in simple harmonic motion is considered to be concentrated at a point in space. This is a lumped parametric system because the mass is a point mass. This assumption is justified at lower frequencies (higher wavelengths). Therefore, in lumped parameter models, the output can be assumed to be functions of time only. Hence, the system can be expressed with ordinary differential equations.

In contrast, distributed parametric systems such as the mass or stiffness of mechanical power transmission shaft cannot be assumed to concentrate at a point; thus the lumped parameter assumption breaks down. Therefore, the system output is a function of time and one or more spatial variables (space), which results in a mathematical model consisting of partial differential equations.

Continuous-Time and Discrete-Time Systems

Systems whose inputs and outputs are defined over a continuous range of time (i.e. continuous-time signals) are continuous-time systems. On the other hand, the systems whose inputs and outputs are signals defined only at discrete instants of time $t_0, t_1, t_2, \ldots, t_k$ are called discrete systems, as shown in Fig. 5. The digital computer is a familiar example of this type of systems.

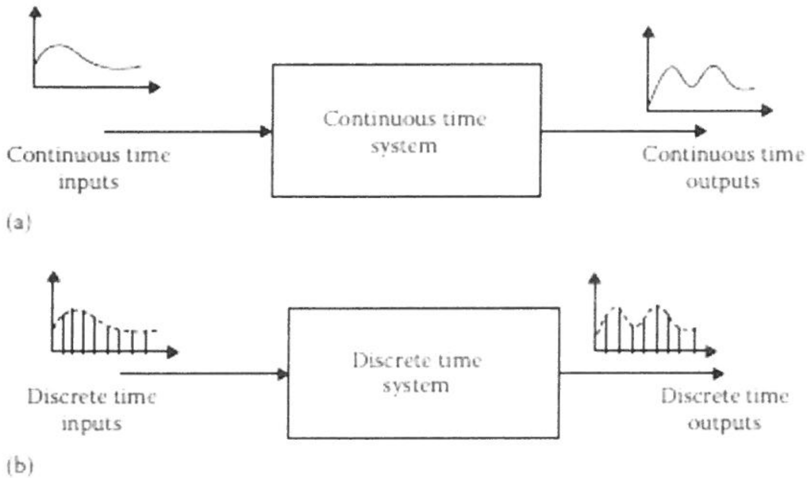


Fig. 5 Analog and digital systems. (a) Continuous time system and (b) discrete time system

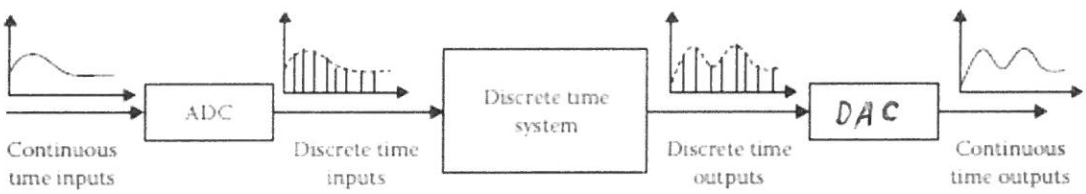
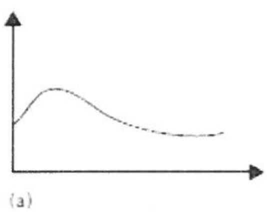
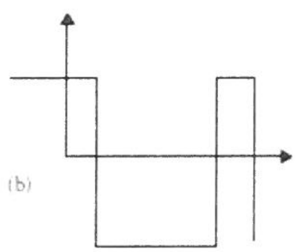


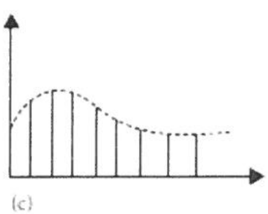
Fig. 6 Processing continuous-time signals by discrete-time systems



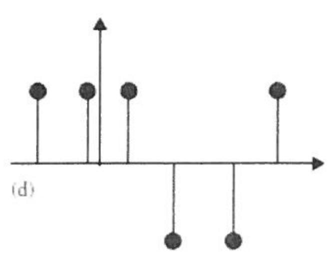
(a) Analog and continuous signal



(b) digital and continuous signal



(c) analog and discrete signal



(d) digital and discrete signal.

Fig. 7 Different types of signals.

The discrete - time signals arise naturally in situations which are inherently discrete time such as population in a particular town and customers served at ATM counter. Sometimes, we want to process continuous - time signals with discrete - time systems. In such situations it is necessary to convert continuous- time signals to discrete signals using analog - to - digital converters (ADC) and process the discrete signals with discrete systems. The output of discrete - time system is again converted back into continuous - time signals using digital - to - analog converters (DAC), as show in Fig. 6

The terms discrete - time and continuous - time signals qualify the nature of signal along the time axis (x-axis) and the terms analog and digital, on the other hand, qualify the nature of signal amplitude (y-axis), as shown in Fig. 7.

Deterministic vs. Stochastic Systems

A system that will always produce the same output for a given input is said to be deterministic. Determinism is the philosophical proposition that every event, including human cognition and behavior, decision, and action, is causally determined by an unbroken chain of prior occurences.

A system that will produce different outputs for a given input is said to be stochastic. A stochastic process is one whose behavior is nondeterministic and is determined both by the process's predictable actions and by a random element. Classical examples of this are medicine: a doctor can administer the same treatment to multiple patients suffering from the same symptoms, however, the patients may not all react to the treatment the same way. This makes medicine a stochastic process. Additional examples are warfare, meteorology, and rhetoric, where success and failure are so difficult to predict that explicit allowances are made for uncertainty.

Complexity of Systems

Another basic issue is the complexity of a system. Complexity of a system depends on the following factors:

- Number of interconnected components
- Type / nature of component
- Number of interactions
- Strength of the interaction
- Type / nature of interactions
  a. Static or dynamic
  b. Unidirectional or bidirectional
  c. Constrained or nonconstrained interaction

Systems Modeling

The modeler is interested in understanding how a particular system works, what causes changes in the system, and the sensitivity of the system to certain changes. He or she is also interested in predicting what changes might occur and when they occur. How might such information be obtained?

First, we make some specific observations about the behavior being studied and identify the factors that seem to be involved. Usually we cannot consider or even identify, all the factors involved in the behavior, so we make simplifying assumptions that eliminate some factors. Next, we conjecture tentative relationships among the factors we have selected, thereby creating a rough "model" of the behavior. Having constructed a model, we then apply appropriate mathematical analysis leading to conclusions about the model. Note that these conclusions pertain only to the model, not to the actual real-world system under investigation. Since we made some simplifications in constructing the model, and since the observations upon which the model is based invariably contain errors and limitations, we must carefully account for these anomalies before drawing any inferences about the real-world behavior.

In summary we have the following "rough" modeling procedure:

1. Through observation, identify the primary factors involved in the real-world behavior, possibly making simplifications.
2. Conjecture tentative relationships among the factors.
3. Apply mathematical analysis to the resultant "model".
4. Interpret mathematical conclusions in terms of the real-world problem.

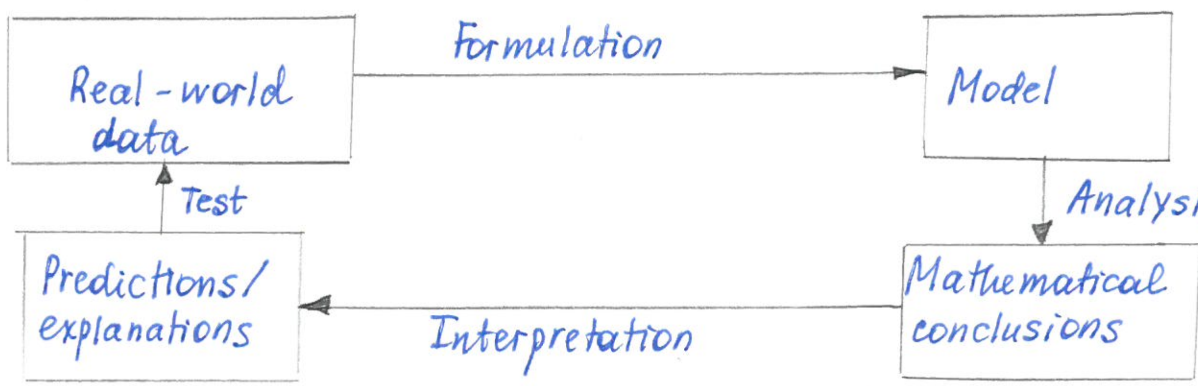


Fig. 8 The modeling process as a closed system/organization

Figure 8 portrays the entire modeling process as a closed system. Given some real-world system, we gather sufficient data to formulate a model. Next we analyze the model and reach mathematical conclusions about it. Then we interpret the model and make predictions or offer explanations. Finally we test our conclusions about the real-world system against new observations and data. We may then find we need to go back and refine the model in order to improve its predictive or descriptive capabilities. Or perhaps we will discover that the model really doesn't "fit" the real world at all accurately, so we must formulate a new model.

Mathematical Models

The models used to represent or approximate a real-world system can be very different in both appearance and purpose. For instance, one kind of model is a miniature replication of a real-world object of interest, like a model spacecraft that might be used to study certain design features under experimental conditions.

Another kind of model is a mathematical model. For our purpose we define a mathematical model as a mathematical construct designed to study a particular real-world system or phenomenon.

Models have been widely accepted as a means for studying complex phenomena for experimental investigations at a lower cost and in less time, than trying changes in actual systems. Knowledge can be obtained more quickly, and for conditions not observable in real life. Models tell us about our ignorance and give better insights into the system.

There are many modeling approaches when formulating a mathematical model.

Emperical models

The modeler desires to construct an emperical model based on the collected data rather than select a model based on certain assumptions. In such cases the modeler is strongly influenced by the data that have been carefully collected and analyzed, so he or she seeks a curve that captures the trend of the data in order to predict in between the data points.

Stochastic models ("stochastic" comes from the Greek word to guess)

Using this method we try to estimate the probability of certain outcomes based on the available data.

These models can be extremely complicated. They do have the advantage of incorporating a degree of uncertainty within them, and ideally should be used when there is a high degree of variability in the problem. They have valuable application in many areas such as economic fluctuations, insurance problems, telecommunications and traffic theory, and biological models.

Simulation Models

In a simulation model one writes a computer program which applies a set of rules, or possibly even physically builds a scale model. It is intended to produce a set of data which mimic a real outcome including extreme events. Usually a simulation will involve some random components. The computer program can be run many times and statistical information gained in the process.

Typically, such models are used in engineering applications as an aid to identifying problems which may arise during use or construction. They also provide a useful means by which data sets can be generated. An example of a simulation is a computer program that generates 1000 integers randomly from 1 to 6 as representative of 1000 tosses of a six-sided die. Another example is using a scaled-down model to simulate the drag force on a proposed design for a submarine.

It might be argued that this simulation approach provides the most realistic models, but this does not mean it provides the best models. The best models are usually those which are simplest yet still provide results which are useful.

Deterministic models

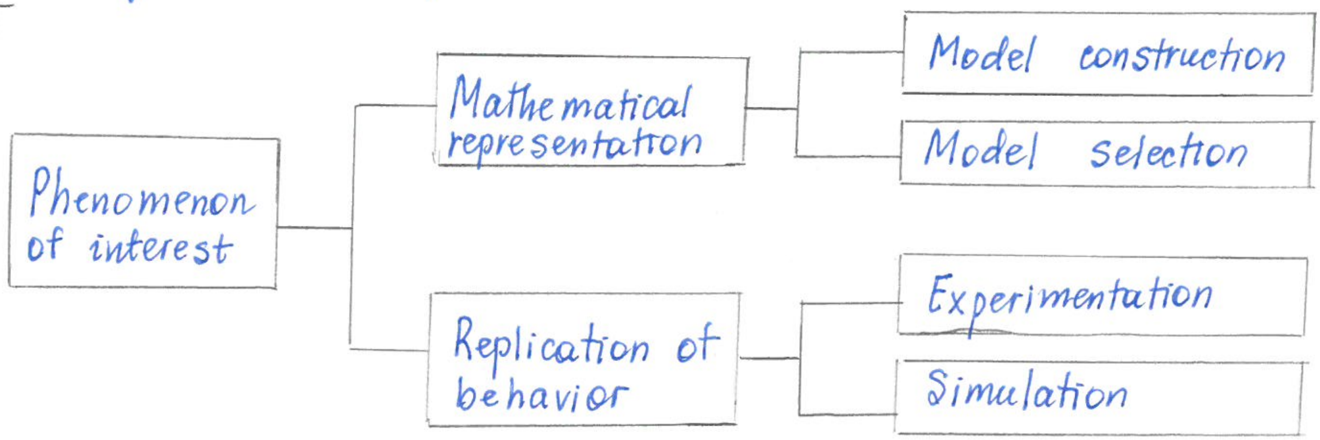
Modeling in this manner we ignore random variation and try to formulate mathematical equations describing the basic fundamental relationships between the variables of the problem.

Statistical models

Statistical models concern the testing (referred to as hypothesis testing) of whether a set of empirical data is from one or another category. These categories are assumed to have particular distributions (with associated means and standard deviations) and the results suggest the data are drawn from one such category. This distribution can then be used to predict the outcome of further trails.

There are mathematical models that already exist that can be identified with some particular real-world phenomenon and selected to study it. Then there are those mathematical models that we construct specifically to study a special phenomenon. The figure below depicts this differentiation between models

FIGURE 9



The nature of the model

Starting from some real-world phenomenon, we can represent it mathematically by constructing a new model or selecting an existing model. On the other hand, we can replicate the phenomenon experimentally or with some kind of simulation.

When it comes to the question of constructing a mathematical model, a variety of conditions can cause us to abandon hope of achieving any success. The mathematics involved may be so complex and intractable that there is a little hope of analyzing or solving the model, thereby defeating its utility. This complexity can occur when attempting to use a system of partial differential equations, for instance. Or the problem may be so large (in terms of the number of factors involved) that it is impossible to capture all the necessary information in a single mathematical model. Predicting the global effects of the interactions of population, use of resources, and pollution is an example of such an impossible situation. In such cases we may attempt to replicate the behavior directly in some manner, or we might observe the behavior directly by conducting various experimental trails. Then we collect data from these trails and analyze the data in some way, possibly using statistical techniques or curve-fitting procedures. From the analysis, certain conclusions could be reached.

On the other hand, we may attempt to replicate the behavior indirectly. We might use an analogue device such as an electrical current to model a mechanical system. (It is possible to design an electrical device whose operation is also described by the same system of equations that the mechanical system under investigation. This simulating electrical device can to some extent help in solving the system of equations, since observing its operation, we thereby observe the behavior of the unknown functions satisfying the system of equations.) We might use a scaled-down model like a scaled model of a jet aircraft in a wind tunnel. Or we might attempt to replicate a behavior on a digital computer – for instance, simulating the global effects of the interactions of population, use of resources, and pollution, or simulating the operation of an elevator system during the morning rush hour.

The distinction between the various model types as depicted in Fig. 9 is made solely for ease of discussion. For example, the distinction between experiments and simulations is based on whether the observations are obtained directly (experiments) or inderectly (simulations). In practical models this distinction is not nearly so sharp; one master model may employ several models as submodels, including selections from existing models, simulations, and experiments. Nevertheless, it is informative to contrast these types of models and compare their various capabilities for portraying the real world.

The Construction of Models

In the preceding discussion we viewed modeling as a process and considered briefly the form of the model itself. Now let's focus attention on the construction of mathematical models. We begin by presenting an outline of a procedure that is helpful in constructing models.

Step 1. Identify the problem: What is it we would like to do or find out? Typically this is a very difficult step because people often have great difficulty in deciding what must be done. In real-life situations no one simply hands us a mathematical problem to solve. Usually we have to sort through large amounts of data and identify some particular aspect of the situation we wish to study. Moreover, we must be sufficiently precise (ultimately) in the formulation of the problem to allow for traslation into mathematical symbology of the verbal statements describing it. This translation is accomplished through the next steps.

First of all one has to recognize that there is a problem and what the problem is.

Step 2 Make assumptions: Generally we cannot hope to capture in a usable mathematical model all of the factors influencing the problem that has been identified. The task is simplified by reducing the number of factors under consideration. Then relationships between the remaining variables must be determined. Again, the complexity of the problem can be reduced by assuming relatively simple relationships. Thus the assumptions fall into two main categories:

a. Classification of the variables: What things influence the behavior we identified in Step 1? We list these things as variables. The variables the model seeks to explain are the dependent variables and there may be several of these. The remaining variables are the independent variables. Each variable is classified as either dependent or independent, or we may choose to neglect it altogether.

We may choose to neglect some of the independent variables for either of two reasons. First, the effect of the variable may be relatively small compared to other factors involved in the behavior. Second, we may also neglect a factor that affects the various alternatives in about the same way, even though it may have a very important influence on the behavior under consideration. For example, consider the problem of determining the optimal shape for a lecture hall where readibility of a chalkboard or overhead projection is a dominant criterion. Lighting is certainly a crucial factor, but it would affect all possible shapes in about the same way. We can simplify the analysis considerably by neglecting such a variable, possibly incorporating it later in a separate, more refined model.

6. Determination of interrelationships among the variables selected for study: Before we can hypothesize a relationship between the variables, we generally must make some additional simplifications. The problem may be sufficiently complex so that we cannot see a relationship among all the variables initially. In such cases it may be possible to study <u>submodels</u>. That is, we study one or more of the independent variables separately. Eventually we will connect the submodels together.

<u>Step 3</u>. Solve or interpret the model: Now we put together all the submodels to see what the model is telling us. In some cases the model may consist of mathematical equations or inequalities that must be solved in order to find out the information we are seeking. Often a problem statement requires a "best" or <u>optimal solution</u> to the model.

Often we will find that we are not quite ready to complete this step. Or we may end up with a model so unwieldy we cannot solve or interpret it. In such situations we might return to Step 2 and make additional simplifying assumptions. Sometimes we will even want to return to Step 1 to redefine the problem.

<u>Step 4</u>. Verify the model: Before we use the model, we must test it out. There are several questions we should ask before designing these tests and collecting - a process that can be expensive and time consuming. First, does the model answer the problem we identified in Step 1, or did we stray from the key issue as we constructed the model? Second, is the model usable in a practical sense; that is, can we really gather the data necessary to operate the model? Third, does the model make common sense?

Once the commonsense tests are passed, we will want to test many models using actual data obtained from empirical observations. We need to be careful to design the test in such a way as to include observations over the <u>same range</u> of values of the various independent variables we expect to encounter when actually using the model.

The assumptions we made in Step 2 may be reasonable over a restricted range of the independent variables, but very poor outside of those values. For instance, a frequently used interpretation of Newton's second law states that the net force acting on a body is equal to the mass of the body times its accelerataion. This "law" is a reasonable "model" until the speed of the object approaches the speed of light.

We must be very careful about the conclusions we draw from any tests. Just as we cannot prove a theorem simply by demonstrating many cases in which the theorem does hold, likewise, we cannot extrapolate broad generalizations from the particular evidence we gather about our model. A model does not become a law just because it is verified repeatedly in some specific instances. Rather we corroborate the reasonableness of our model through the data we collect.

Step 5. Implement the model: Of course our model is of no use just sitting in a filing cabinet. We will want to explain our model in terms that the decision makers and users can understand if it is ever to be of use to anyone. Further, unless the model is placed in a "user-friendly" mode, it will quickly fall into disuse, except possibly for publication in some journal. (Unfortunately, the "publish or perish" sydrome is still with us.) Expensive computer programs sometimes suffer such a demise. Often the inclusion of an additional step to facilitate the collection and input of the data necessary to operate the model determines its success or failure.

Step 6. Maintain the model: We must remember that our model is derived from the specific problem we identified in Step 1 and from the assumptions we made in Step 2. Has the original problem changed in any way, or have some previously neglected factors become important? Does one of the submodels need to be justified?

We summarize the steps for constructing mathematical models in Figure 10. We should not be too enamored with our work. like any model, our procedure is an approximation process and therefore has its limitations.

Step 1. Identify the problem.
Step 2. Make assumptions.
 a. Identify and classify the variables.
 b. Determine interrelationships between the variables and submodels.
Step 3. Solve the model.
Step 4. Verify the model.
 a. Does it adress the problem?
 b. Does it make common sense?
 c. Test it with real-world data.
Step 5. Implement the model.
Step 6. Maintain the model.

Figure 10. The construction of a mathematical model

The process shown in Figure 10 provides a methodology for progressively focusing on those aspects of the problem we wish to study. Furthermore, it demonstrates a rather curious blend of creativity with the scientific method used in the modeling process. The first two steps are more artistic or original in nature. They involve abstracting the essential features of the problem under study, neglecting any factors judged to be unimportant, and postulating relationships that are precise enough to permit completion of the remaining steps.

Let us contrast the modeling process presented in Figure 10 with the scientific method. One version of the scientific method is as follows:

Step 1. Make some general observations of a phenomenon.

Step 2. Formulate a hypothesis about the phenomenon.

Step 3. Develop a method to test that hypothesis.

Step 4. Gather data to use in the test.

Step 5. Test the hypothesis using the data.

Step 6. Confirm or deny the hypothesis.

By design the mathematical modeling process and scientific method have some obvious similarities. For instance, both processes involve making assumptions or hypotheses, gathering real-world data, and testing or verification using that data. These similarities should not be surprising; while recognizing that part of the modeling process is an art, we do attempt to be scientific and objective whenever possible. Nevertheless, there are some subtle differences between the two procedures. One difference lies in the primary goal of the two processes. In the modeling process, assumptions are made in selecting which variables to include or neglect and in postulating the interrelationships among the included variables. The goal in the modeling process is to hypothesize a model, and like the scientific method, evidence is gathered to corroborate that model. However, unlike the scientific method, the objective is not to confirm or deny the model (we already know it is not precisely correct because of the simplifying assumptions we have made), but rather to test its reasonableness. We may decide that the model is quite satisfactory and useful, and elect to accept it. Or we may decide that the model needs to be refined or simplified. In extreme cases we may even need to redefined the problem, in a sense rejecting the model altogether.

Figure 11 amplifies these ideas in viewing the modeling process, and attempts to display graphically its iterative nature.

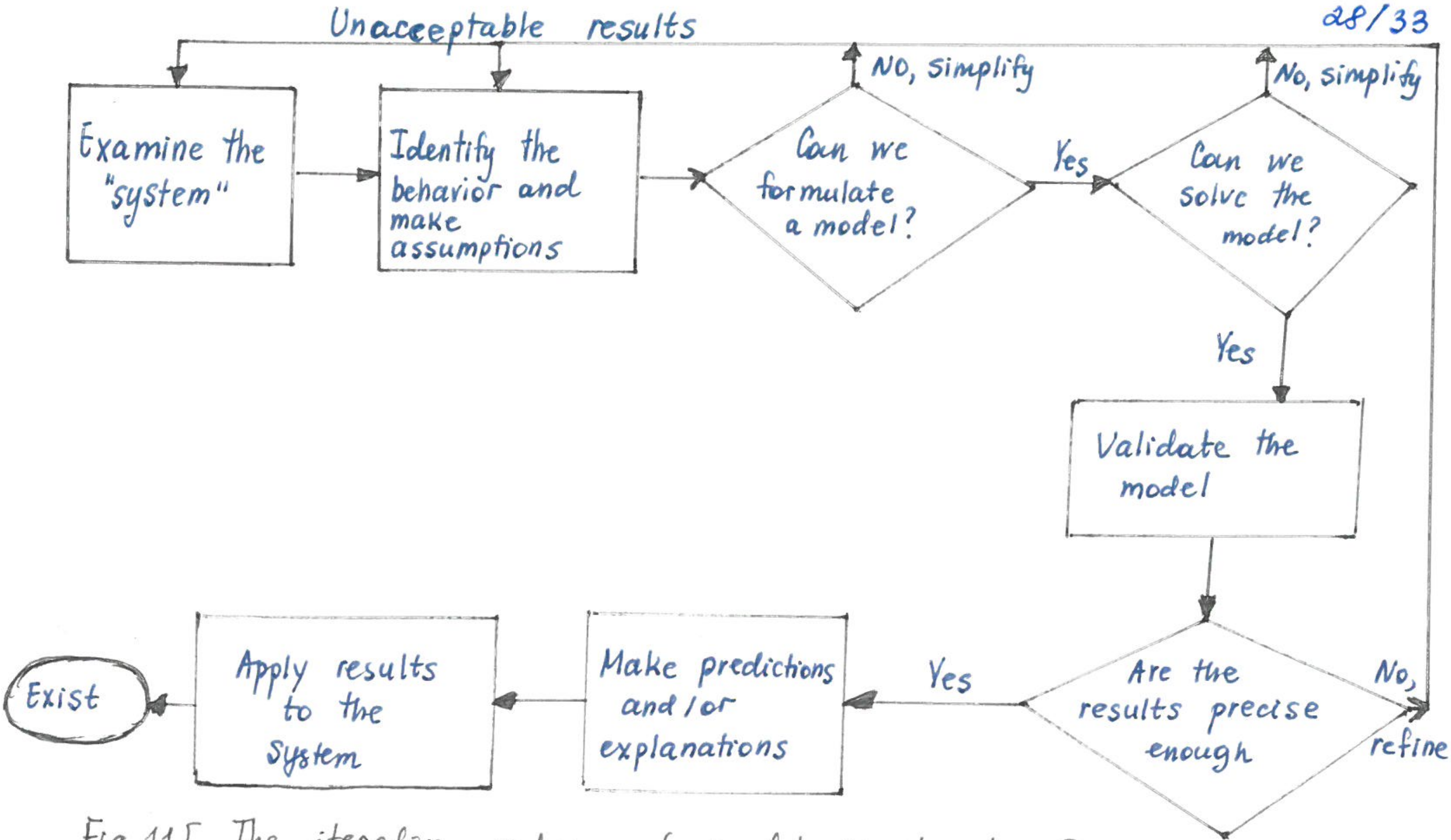


Fig. 11 [The iterative nature of model construction.]

We begin by examining some system and identifying the particular behavior we wish to predict or explain. Next we identify the variables and simplifying assumptions, and then we generate a model. We then attempt to validate the model with appropriate tests. If the results of the tests are satisfactory, we can use the model for its intended purpose. If the results are not satisfactory, there are several possibilities to pursue. We may decide the model needs to be refined by incorporating additional variables, or by restructering a particular submodel. In some cases the test results may be so unsatisfactory that the original problem must be redefined because it turns out to be entirely too ambitious.

The process depicted in Figure 11 not only emphasizes the interative nature of model construction, but also introduces the trade-offs between model simplification and model refinement. We will generally start with a rather simple model, progress through the modeling process, and then refine the model as the results of our validation procedure dictate.

If we cannot come up with a model, or solve the one we have, we must simplify. We simplify a model by treating some variables as constants, by neglecting or aggregating some variables, by assuming simple relationships (such as linearity) in any submodels, or by restricting further the problem under investigation. On the other hand, if our results are not precise enough, we must refine the model. Refinement of a model is generally achieved in the opposite way: We introduce additional variables, assume more sophisticated relationships among the variables, or expand the scope of the problem. By trading off between simplification and refinement, we determine the generality, realism, and precision of our model. This trading-off process cannot be overemphasized and constitutes the "art of modeling".

We complete the description of the construction of models by introducing several terms that are useful in describing models. A model is said to be robust when its conclusions do not depend on the precise satisfaction of the assumptions. On the other hand, a model a fragile if its conclusions do depend on the precise satisfaction of some set of conditions. The term sensitivity refers to the degree of change in a model's conclusions as some condition upon which they depend is varied; the greater the change, the more sensitive is the model to that condition.

Some Illustrative Examples

We now demonstrate the modeling process presented above with a couple of examples. Special emphasis is placed here on identifying the problem and important variables.

Example 1 Vehicular Stopping Distance

Scenario Consider the following rule of thumb often given in driver education classes:

Allow one car length for every ten miles of speed under normal driving conditions, but more distance in adverse weather or road conditions. One way to accomplish this is to use the "Two-Second Rule" for measuring the correct following distance: If you stay two seconds behind the car in front, you have the correct distance no matter what your speed. To obtain that distance, watch the vehicle ahead of you pass some definite point on the highway, like a tar strip or overpass shadow. Then count to yourself "one thousand and one, one thousand and two"; that's two seconds. If you reach the mark before you finish saying those words, then you are following too close behind.

The preceding rule is implemented easily enough, but how good is it?

Problem identification Our ultimate goal is to test this rule of thumb and suggest another rule if it fails. However, the statement of the problem "How good is the rule?" is rather vague. We need to be much more specific and spell out a problem, or ask a question, whose solution or answer will help us accomplish our goal while at the same time permitting a more exact mathematical analysis. Consider the following problem statement: Predict the vehicle's total stopping distance as a function of its speed.

Assumptions We begin our analysis with a rather obvious model for total stopping distance:

total stopping distance = reaction distance + stopping distance

By reaction distance, we mean the distance the vehicle travels from the instant the driver perceives a need to stop to the instant when the brakes are actually applied. Braking distance is the distance required for the brakes to bring the vehicle to a complete stop.

First let's develop a submodel of reaction distance. The reaction distance is a function of many variables, and we start by listing just two of these:

reaction distance = f(response time, speed)

We could continue developing the submodel in as much detail as we like. For instance, response time is influenced by both individual driving factors as well as by the vehicle operating system. System time is the time from which the driver touches the brake pedal until the brakes are mechanically applied. For modern cars we would probably neglect the influence of the system since it is quite small in comparison to the human factors. The portion of the response time determined by the driver depends on many things such as reflexes, alertness, visibility, and so forth. Since we are only developing a rule of thumb, we could just incorporate average values and conditions for these latter variables and so specify. Once all of the variables we deem important to the submodel have been identified, we can begin to determine interrelationships among them. Let us suggest that the submodel is constructed.

Next consider the braking distance term. The weight and speed of the vehicle are certainly important factors to be taken into account. The efficiency of the brakes, the kind and condition of the tires, the road surface, and weather conditions are other legitimate factors. As before, we would most likely assume average values and conditions for these latter factors. Thus our initial submodel gives braking distance as a function of vehicular weight and speed:

$$\text{braking distance} = h(\text{weight}, \text{speed}) \quad [\, d_b \propto v^2 \,]$$

Finally, let's discuss briefly the last three steps in the modeling process for this problem. We would want to test our model against real-world data. Do the predictions afforded by the model agree with the real driving situations? If not, we would want to assess some of our assumptions and perhaps restructure one (or both) of our submodels. If the model does predict real driving situations accurately, then does the rule of thumb stated in the opening discussion agree with the model? The answer to the latter question then gives an objective basis for answering "How good is the rule?". Whatever rule we come up with (in order to implement the model), it must be easily understood and easy to use if it is going to be effective.

In this example, maintenance of the model does not seem to be a particular issue. Nevertheless, we would want to be sensitive to the effects on the model of such changes as power brakes or disk brakes, or a fundamental change in tire design, and so forth.

Example 2 The Assembly Line

Scenario There is increasing concern by both individual laborers and labor unions for both wages and job quality. On the other hand, corporations are feeling the competition from abroad. To increase domestic productivity, more assembly line operations have evolved. A question management faces is how to assign employees to the various jobs on the assembly line, realizing that the jobs are often quite tedious and repetitive, and demand various worker skill levels.

Problem identification The classical approach to making employee job assignments is to ensure maximum company profits. Typically, people are assigned to machines in such a way as to maximize the profit margin on the product being produced. In situations of fixed product demand, this approach amounts to minimizing the costs of production.

The difficulty with the classical approach is that it focuses on only one aspect of the problem — namely, short-term profits. In many situations the results of such assignment procedures include job dissatisfaction, excessive absenteeism, poor workmanship and product quality, and diminished worker efficiency and overall productivity. The net effect often is rising production costs coupled with reduced consumer demand. Thus while mangement attempts to maximize short-term profits, long-term profits actually suffer.

Although the maximization of profit is certainly a primary consideration, other factors, such as product quality and job satisfaction, must be taken into account. The precise reconciliation of the various factors into a problem statement depends on the particular situation. Let's assume that the firm has a contract and has the plant capacity to produce a fixed number of items per month.

We then define the problem as follows:
Minimize production cost while meeting specified levels of demand, quality control, and job satisfaction.

Assumptions Under this problem identification, production cost is a function of individual wages, each individual's productivity on each job, the expected number of defects each individual makes on each job per unit time, and the reduction in both quality and productivity as a function of the time an individual spends at a particular job assignment.

We will not construct a model, for now, let's ask whether or not we can obtain the needed data. Certainly the hourly wage presents no problem. In many instomces we can build a history of an individual's productivity and workmanship by measuring the hourly output on each machine as well as the number of defects that are produced. The reduction in efficiency incurred by assigning an individual the same job for a prolonged period may be difficult to measure. Hopefully, after observing and interviewing several employees, we can establish guidelines to prevent excessive diminution in productivity while simultaneously maintaining job satisfaction. For example, we may determine that an individual should not be assigned to a particular tedious job for more than two hours each day.

The advantage of this approach is that it forces the modeler to consider many more aspects of the problem. The mathematical modeler must guard against the tendecy to ignore factors just because they are difficult to quantify. In the preceding problem there may also be salutary effects from involving management in the data collection process. A lesson to be learned from this example is that a careful modeler considers all pertinent factors and then attempts to determine the sensitivity of the model to the various assumptions made.