PART 1

BIOMEDICAL SYSTEMS ANALYSIS
CHAPTER 1

MODELING AND SIMULATION OF BIOMEDICAL SYSTEMS

Narender P. Reddy
University of Akron, Akron, Ohio

1.1 MATHEMATICAL MODELING

Models are conceptual constructions that allow formulation and testing of hypotheses. A mathematical model attempts to duplicate the quantitative behavior of the system. Mathematical models are used in today’s scientific and technological world because of the ease with which they can be used to analyze real systems. The most prominent value of a model is its ability to predict as yet unknown properties of the system. The major advantage of a mathematical or computer model is that the model parameters can be easily altered and the system performance can be simulated. Mathematical models allow the study of subsystems in isolation from the parent system. Model studies are often inexpensive and less time consuming than corresponding experimental studies. A model can also be used as a powerful educational tool, since it permits idealization of processes. Models of physiological systems often aid in the specification of design criteria for the design of procedures aimed at alleviating pathological conditions. Mathematical models are useful in the design of medical devices. Mathematical model simulations are first conducted in the evaluation of the medical devices before expensive animal testing and clinical trials. Models are often useful in the prescription of patient protocols for the use of medical devices. Pharmacokinetic models have been extensively used in the design of drugs and drug therapies.

There are two types of modeling approach: the black box approach and the building block approach. In the black box approach, a mathematical model is formulated on the basis of the input-output characteristic of the system without consideration of the internal functioning of the system. Neural network models and autoregressive models are some examples of the black box approach. In the building block approach, models are derived by applying the fundamental laws (governing physical laws) and constitutive relations to the subsystems. These laws together with physical constraints are used to integrate the models of subsystems into an overall mathematical model of the system. The building block approach is used when the processes of the system are understood. However, if the system processes are unknown or too complex, then the black box approach is used. With the building block approach, models can be derived at the microscopic or at the macroscopic
levels. Microscopic models are spatially distributed and macroscopic models are spatially lumped and are rather global. The microscopic modeling often leads to partial differential equations, whereas the macroscopic or global modeling leads to a set of ordinary differential equations. For example, the microscopic approach can be used to derive the velocity profile for blood flow in an artery, but the global or macroscopic approach is needed to study the overall behavior of the circulatory system including the flow through arteries, capillaries, and the heart.

Models can also be classified into continuous time models and models lumped in the time domain. While the continuous time modeling leads to a set of differential equations, the models lumped in time are based on the analysis of discrete events in time and may lead to difference equations or sometimes into difference-differential equations. Random walk models and queuing theory models are some examples of discrete time models. Nerve firing in the central nervous system can be modeled by using such discrete time event theories. Models can be classified into deterministic and stochastic models. For example, in deterministic modeling, we could describe the rate of change of volume of an arterial compartment to be equal to the rate of flow in minus the rate of flow out of the compartment. However, in the stochastic approach, we regard the probability of increase in the volume of the compartment in an interval to be dependent on the probability of transition of a volume of fluid from the previous compartment and the probability of transition of a volume of fluid from the compartment to the next compartment. While the deterministic approach gives the mean or average values, the stochastic approach yields means, variances, and covariances. The stochastic approach may be useful in describing the cellular dynamics, cell proliferations, etc. However, in this chapter, we will consider only the deterministic modeling at the macroscopic level.

The real world is complex, nonlinear, nonhomogeneous, often discontinuous, anisotropic, multilayered, multidimensional, etc. The system of interest is isolated from the rest of the world by using a boundary. The system is then conceptually reduced to that of a mathematical model by using a set of simplifying assumptions. Therefore, the model results have significant limitations and are valid only in the regimes where the assumptions are valid.

### 1.2 COMPARTMENTAL MODELS

Compartmental models are lumped models. The concept of a compartmental model assumes that the system can be divided into a number of homogeneous well-mixed components called compartments. Various characteristics of the system are determined by the movement of material from one compartment to the other. Compartmental models have been used to describe blood flow distribution to various organs, population dynamics, cellular dynamics, distribution of chemical species (hormones and metabolites) in various organs, temperature distribution, etc.

As an example, let us consider a simple one-compartment model for the prescription of treatment protocols for dialysis by an artificial kidney device (Fig. 1.1). While the blood urea concentration (BUN) in the normal individual is usually 15 mg\% (mg\% = milligrams of the substance per 100 mL of blood), the BUN in uremic patients could reach 50 mg\%. The purpose of the dialysis is to bring the BUN level closer to the normal. In the artificial kidney, blood flows on one side of the dialyzer membrane and dialysate fluid flows on the other side. Mass transfer across the dialyzer membrane occurs by diffusion due to concentration difference across the membrane. Dilysate fluid is a makeup solution consisting of saline, ions, and the essential nutrients that maintains zero concentration difference for these essential materials across the membrane. However, during the dialysis, some hormones also diffuse out of the dialyzer membrane along with the urea molecule. Too-rapid dialysis often leads to depression in the individual because of the rapid loss of hormones. On the other hand, too-slow dialysis may lead to unreasonable time required at the hospital.

Simple modeling can be used to calculate the treatment protocols. Let us consider a one-compartment model of the tissue where we assume that the blood and tissue are well mixed and that the concentration of urea is uniform throughout the body. Let \( C_o \) be the concentration of urea at the outlet of the body, i.e., at the inlet of the dialyzer in the arterial line that takes blood into the dialyzer, and let \( C_i \) be the concentration of urea at the inlet of the compartment, i.e., at the exit of the dialyzer.
FIGURE 1.1 A one-compartment model of the human body to analyze the patient-dialyzer interactions.

in the venous line that brings the blood back to the body. Mass balance demands that the rate of change of mass in the body be equal to the net rate of mass coming into the body from the dialyzer, plus the metabolic production rate.

\[ V \frac{dC}{dt} = Q(C_i - C_o) + m = -QC_o \left(1 - \frac{C_i}{C_o}\right) + m \]  

(1.1)

where 

- \( V \) = tissue volume plus the blood volume
- \( Q \) = blood flow rate to the kidney
- \( m \) = metabolic production rate of urea in the body

It should be noted that mass is equal to volume \( V \) times the concentration \( C \).

Usually, the dialysate flow rate in the artificial kidney is much larger than that of the blood flow rate. Regardless of the type of the dialyzer (cocurrent, countercurrent, or mixed flow), the extraction ratio \( E \) can be expressed as (Cooney, 1980):

\[ E = 1 - \exp\left(-\frac{kA}{Q}\right) \]  

(1.2)

where \( A \) is the interfacial membrane surface area for mass transfer and \( k \) is the permeability of the membrane for that particular solute (urea in the present context). Since \( Q \) does not change during dialysis, and since \( k \) and \( A \) are design parameters, extraction ratio \( E \) remains a constant. Extraction can be further expressed in terms of the concentrations as follows:

\[ E = 1 - \frac{C_i}{C_o} \]  

(1.3)

It should be pointed out that \( C_o \) is the concentration at the outlet of the body and therefore at the inlet.
of the dialyzer, and \( C_i \) is the concentration in the blood coming into the body and therefore going out of the dialyzer. Also, it should be noted that the concentration \( C_o \) in the blood going out of the body is the same as the concentration in the body, since we assumed that the entire body (tissue and blood) constitutes a homogeneous, well-mixed compartment. Therefore, Eq. (1) can be rewritten as follows on substitution of Eq. (3)

\[
V \left( \frac{dC}{dt} \right) = -QCE + m
\]  

(1.4)

When the dialyzer is turned on, metabolic production rate \( (m) \) can be neglected when compared to the other term in the equation, and upon integration will result in

\[
C = C^0 \exp \left[ -\left( \frac{QE}{V} \right) t \right]
\]  

(1.5)

where \( C^0 \) is the initial concentration of urea in the tissue.

When the patient is not on dialysis, then the blood flow to the dialyzer \( Q \) is zero, and therefore

\[
V \left( \frac{dC}{dt} \right) = m
\]  

(1.6)

When the patient is not on dialysis, the concentration of urea will increase linearly if the metabolic production rate is constant or will increase exponentially if the metabolic production rate is a linear function of the concentration (first-order reaction). When the patient is on dialysis, the concentration would decrease exponentially. This way, the treatment protocol can be prescribed after simulating different on and off times (e.g., turn on the dialyzer for 4 hours every 3 days) to bring the BUN under control.

Now, let us examine the limitations of the one-compartment model. First, the entire blood and tissue are assumed to be in equilibrium. However, it is well known that intracellular urea concentration may be significantly different from the extracellular compartment. Moreover, urea may be preferentially produced in certain organs like the brain, heart, muscle, etc. An accurate treatment requires a multicompartment model.

Let us consider a two-compartment model (Fig. 1.2) consisting of an intracellular pool (compartment 1) and an extracellular pool (compartment 2). Urea is produced by the intracellular pool and is transported across the cell membrane into the interstitial fluids and then into the bloodstream. Mass balance for these two compartments can be expressed as

\[
V_1 \left( \frac{dC_1}{dt} \right) = m - B(C_1 - C_2)
\]  

(1.7)

where \( B \) (\( C_1 - C_2 \)) = the interfacial transfer from compartment 1 to compartment 2 (from intracellular pool to extracellular pool)

\( C_1 \) and \( C_2 \) = concentrations of urea in compartments 1 and 2

\( B = \) a constant

The constant \( B \) is a product of permeability of the cellular membrane for urea and the interfacial surface area.

\[
V_2 \left( \frac{dC_2}{dt} \right) = B(C_1 - C_2) - QC_2E
\]  

(1.8)

Blood flow to the dialyzer \( (Q) \) is zero when the patient is not on the dialysis machine. Babb et al. (1967) simulated the two-compartmental model and the model results were in agreement with the experimental data (Fig. 1.3). However, the two compartmental model may not be sufficient if one wants to find the concentration of urea in the brain tissue. A multicompartment model involving...
FIGURE 1.2 A two-compartmental model of the body used to simulate the patient-dialyzer interactions.

separate compartments for brain, heart, kidney, lean tissue, etc., may be needed to accurately determine the concentration of urea in critical organs. Similar compartmental models have been used in the development of metabolic modeling.

Now, let us consider an example of cellular dynamics in bone. Cellular dynamics plays an important role in natural and stress-induced bone remodeling. There are four kinds of functionally distinct cells: mesenchymal cells, osteoclasts, osteoblasts, and osteocytes. Mesenchymal cells have an outstanding capacity for proliferation and are capable of further differentiating into an osteoclast or an osteoblast. While the mesenchymal cell undergoes cell division, osteoblasts, osteoclasts, or the osteocytes do not undergo cell division. The osteoclast is responsible for bone resorption. The osteoblast is responsible for bone formation. The osteoclast can transform into an osteoblast and vice versa. The osteocyte is a resident cell of the bone. When the osteoblast has surrounded itself with a matrix (after new bone formation), it becomes an osteocyte and it loses the characteristics of an actively secreting cell. Let us assume that these four types of cells can be divided into four different, well-mixed homogeneous compartments (Fig. 1.4). Let \( W \) be the number of mesenchymal cells, \( X \) be the number of osteoclasts, \( Y \) be the number of osteoblasts, and \( Z \) be the number of osteocytes at any given time. Let compartments 1 through 4, respectively, represent these four types of cells. The birth rate and death rate of cells in a compartment depend on the number of cells in that compartment. Let \( B \) be the birth rate of mesenchymal cells per individual cell per unit time. Let \( D_1 \) and \( D_2 \) represent the death rates of osteoclasts and osteocytes per cell per unit time. Equations representing the cell population dynamics can be expressed as follows by integrating the notions of bone cell physiology with the concepts of compartmental modeling.

The rate of change of the number of mesenchymal cells \((W)\) can be expressed as

\[
\frac{dW}{dt} = BW - C_1W - C_2W
\]  

(1.9)
**FIGURE 1.3** Model simulation results of the patient BUN (blood urea nitrogen levels) in mg% (mg% = mg of the substance per 100 mL of blood) plotted as a function of time. [From Babb et al. (1967).] Closed circles are experimental observations.

**FIGURE 1.4** A four-compartment model of bone cells: Compartment 1 represents mesenchymal cells, compartment 2 represents osteoclasts, compartment 3 represents osteoblasts, and compartment 4 represents osteocytes. Mesenchymal cells reproduce and transform either into osteoclasts or into osteoblasts. The osteoclast can transform into an osteoblast and an osteoblast can transform into an osteoclast. The osteoblast can transform into an osteocyte.
The first term on the right-hand side represents the birth rate of mesenchymal cells, the second term represents the rate of number of mesenchymal cells transforming into an osteoclast, and the last term represents the rate of transformation of mesenchymal cells into osteoblasts.

The rate of change of the number of osteoclasts \((X)\) as a function of time can be expressed as

\[
\frac{dX}{dt} = C_1W - C_3X + C_4Y - D_1X \tag{1.10}
\]

The first term on the right-hand side represents the rate of mesenchymal cells transforming into an osteoclast, the second term represents the rate of transformation of osteoclasts into osteoblasts, the third term represents the rate of transformation of osteoblast into osteoclasts, and the last term represents the rate of death of osteoclasts.

The rate of change of the number of osteoblasts \((Y)\) as a function of time can be expressed as

\[
\frac{dY}{dt} = C_2W + C_3X - C_4Y - C_5Y \tag{1.11}
\]

The first term on the right-hand side represents the rate of transformation of mesenchymal cells into osteoblasts, the second term represents the rate of transformation of osteoclasts into osteoblasts, the third term represents the rate of transformation of osteoblasts into osteoclasts, and the last term represents the rate of transformation of osteoblasts into osteocytes.

The rate of change of the number of osteocytes \((Z)\) as a function of time can be expressed as

\[
\frac{dZ}{dt} = C_5Y - C_6Z \tag{1.12}
\]

The first term in the above equation represents the rate of osteoblast transformation into osteocytes, and the last term represents the rate of death of osteocytes. Reddy and Joshi (1987) simulated the stochastic compartmental model of bone cells in which the equation for the population means are the same as the above equations. In addition, the stochastic analysis provides information about the variations and covariances of cellular populations. Figure 1.5 shows the normalized number of osteoblasts plotted as a function of age of the individual when \(C_1\) and \(C_2\) are assumed to be sinusoidal functions in time. These simulation results of Reddy and Joshi (1987) are consistent with experimental observations of Frost (1963).

Compartmental models are used in the analysis of thermal interactions. Simon and Reddy (1992) formulated a mathematical model of the infant-incubator dynamics. Neonates who are born preterm often do not have the maturity for thermal regulation and do not have enough metabolic heat production. Moreover, these infants have a large surface area to volume ratio. Since these preterm babies can not regulate heat, they are often kept in an incubator until they reach thermal maturity. The incubator is usually a forced-convection heating system with hot air flowing over the infant. Incubators are usually designed to provide a choice of air control or skin control. In air control, the temperature probe is placed in the incubator air space and the incubator air temperature is controlled. In the skin control operation, the temperature sensor is placed on the skin and the infant’s skin temperature is controlled. Simon and Reddy (1992) used a four-compartment model (Fig. 1.6) to compare the adequacy of air control and skin control on the core temperature of the infant. They considered the infant, air, mattress, and the wall to be four separate, well-mixed compartments.

### 1.3 Electrical Analog Models

Electric analog models are a class of lumped models and are often used to simulate flow through the network of blood vessels. These models are useful in assessing the overall performance of a system or a subsystem. Integration of the fluid momentum equation (longitudinal direction, in cylindrical
coordinates) across the cross section results in the following expression (Reddy, 1986; Reddy and Kesavan, 1989):

\[
\rho \frac{dQ}{dt} = \frac{\pi a^2 \Delta P}{l} - 2a\tau_w. \tag{1.13}
\]

where \(\rho\) is the fluid density, \(Q\) is the flow rate, \(a\) is the wall radius, \(P\) is the pressure, \(l\) is the length, and \(\tau_w\) is the fluid shear stress at the wall. If we assume that the wall shear stress can be expressed by using quasi-steady analysis, then the wall shear stress can be estimated by \(\tau_w = 4 \mu Q/a^3\). Substituting for the wall stress and rearranging results in

\[
\left(\frac{\rho l}{\pi a^2}\right) \frac{dQ}{dt} = \Delta P - \left(\frac{8\mu l}{\pi a^4}\right)Q \tag{1.14}
\]

which can be rewritten as

\[
L \frac{dQ}{dt} = \Delta P - RQ \tag{1.15}
\]

where \(L = \rho l/(\pi a^2)\) and \(R = 8\mu l/(\pi a^4)\).
It can be easily observed that flow rate $Q$ is analogous to electrical current $i$, and $\Delta P$ is analogous to the electrical potential drop (voltage) $\Delta E$. In Eq. (1.15), $L$ is the inductance (inertance) and $R$ is the resistance to flow. Therefore, Eq. (1.15) can be rewritten as

$$\frac{di}{dt} = \Delta E - Ri \tag{1.16}$$

The fluid continuity equation when integrated across the cross section can be expressed as

$$\frac{dV}{dt} = \Delta Q = Q_{in} - Q_{out} \tag{1.17}$$

where $V$ is the volume. However, volume is a function of pressure (from the momentum balance for the vessel wall):

$$P = P_{ext} + (h/R_0)\sigma \tag{1.18}$$

where $P_{ext}$ is the external pressure on the outside of the vessel wall, $h$ is the wall thickness, and $\sigma$ is...
The hoop stress in the wall. The hoop stress is a function of wall radius $R$ and modulus of elasticity $\Omega$ of the wall, and can be expressed as

$$\sigma = \Omega \left( \frac{R}{R_0} - 1 \right)$$  \hspace{1cm} (1.19)

Equations (1.17), (1.18), and (1.19) can be combined as

$$\frac{dP}{dt} = b \frac{dR}{dt}$$  \hspace{1cm} (1.20)

where $b$ is a constant equal to $\Omega h(R_0^3)$.

Equation (1.20) can be expressed in terms of volume instead of the radius:

$$\frac{dV}{dt} = C \frac{dP}{dt}$$  \hspace{1cm} (1.21)

$C$ is often referred to as the compliance or capacitance.

Substituting Eq. (1.21) in Eq. (1.17) results in

$$C \frac{dP}{dt} = Q_{in} - Q_{out}$$  \hspace{1cm} (1.22)

Equation 22 can be expressed in terms of an electrical equivalent as follows:

$$E = \left( \frac{1}{C} \right) \int i \, dt$$  \hspace{1cm} (1.23)

Equations 16 and 23 [or (1.15) and (1.22)] can be used to simulate either a segment of a blood vessel or the entire blood vessel itself. In small blood vessels, the inductance $L$ is very low compared to the resistance term $R$, and therefore the inductance term can be neglected in small arteries, arterioles, and capillaries. Since there is no oscillation of pressure in the capillaries, the inductance term can be neglected in vessels downstream of the capillary (i.e., venules, veins, and vena cava, etc.).

An electrical analog model of the circulation in the leg is illustrated in Fig. 1.7. Let us consider the flow from the femoral artery into the small leg arteries. There is no inductance in small leg arteries; there is only resistance. Since the small arteries are distensible, they have capacitance (compliance). The muscular pressure ($P_{\text{m}}$) acts as the external pressure on the majority of small leg arteries. Consequently, $P_{\text{m}}$ is used as the reference pressure across the capacitor. The arterioles do not have inductance, but have a variable resistance, which is controlled by neurogenic and metabolic factors. In this model, the precapillary sphincters and the capillaries are lumped together. Since the capillaries are rather rigid, they do not have any capacitance (compliance), but the combined resistance of the sphincters and capillaries is variable subject to metabolic control. For instance, precapillary sphincters dilate in the presence of lactic acid and other end products of metabolism. Venules have resistance and a variable capacitance. This capacitance is subject to neurogenic control, since the diameter of the venule is under neurogenic control. From the venules, the flow goes into leg small veins which have a resistance and a variable capacitance subject to neurogenic control. In addition, the venules have valves that permit only unidirectional flow. These valves can be modeled as diodes. Again, the reference pressure for the capacitor is the muscle pressure $P_{\text{m}}$. It is well known that the blood flow in the legs is aided by the muscle pump, which is essentially the external pressure oscillations on the blood vessel wall due to periodic skeletal muscle contractions during walking, etc. The muscle pump is absent in bedridden patients. Extremity pumps are used on such patients to enhance blood flow to the legs. These extremity pumps provide a periodic, graded sequential
external compression of the leg. The electrical analog model in Fig. 1.7 can be easily modified to simulate the effect of these extremity pumps.

Electrical analog models have been used in the study of cardiovascular, pulmonary, intestinal, and urinary system dynamics. Recently, Barnea and Gillon (2001) have used an electrical analog model to simulate flow through the urethra. Their model consisted of a simple \( LRC \) circuit with a variable capacitor. The time-varying capacitor simulated the time-dependent relaxation of the urethra. They used two types of resistance: a constant resistance to simulate Poiseuille-type viscous pressure drop and a flow dependant resistance to simulate Bernoulli-type pressure loss. With real-time pressure-flow data sets, Barnea and Gillon (2001) have used the model to estimate urethral resistance and changes in urethral compliance during voiding, and have suggested that the urethral elastance (inverse of compliance) estimated by the model provides a new diagnostic tool. Ventricular and atrial pumping can be modeled by similar techniques. The actual pump (pressure source) can be modeled as a variable capacitor. Figure 1.8 shows a model of the left heart with a multisegment representation of the ventricle (Rideout, 1991).
1.14 BIOMEDICAL SYSTEMS ANALYSIS

Time delay and memory processes occur in several biomedical disciplines. An example of such an application occurs in modeling of the immune system (Reddy and Krouskop, 1978). In cellular immune response, lymphocytes are sensitized to a foreign material and have memory. The immune response is significantly enhanced if the similar material is reintroduced after a certain lag time. Another example is stress-induced bone remodeling. Modeling of the nervous system would involve time delays and memory. Similar hereditary functions are used to describe the material responses of viscoelastic materials. The effect of environmental pollutants can be modeled by using such hereditary functions. Stress-induced bone remodeling involves time lags between the actual application of stress and actual new bone formation, and also involves stress/strain histories. To illustrate the modeling of the effects of memory and time delay, let us consider a model of cellular immunity.

1.4 MODELS WITH MEMORY AND MODELS WITH TIME DELAY

Time delay and memory processes occur in several biomedical disciplines. An example of such an application occurs in modeling of the immune system (Reddy and Krouskop, 1978). In cellular immune response, lymphocytes are sensitized to a foreign material and have memory. The immune response is significantly enhanced if the similar material is reintroduced after a certain lag time. Another example is stress-induced bone remodeling. Modeling of the nervous system would involve time delays and memory. Similar hereditary functions are used to describe the material responses of viscoelastic materials. The effect of environmental pollutants can be modeled by using such hereditary functions. Stress-induced bone remodeling involves time lags between the actual application of stress and actual new bone formation, and also involves stress/strain histories. To illustrate the modeling of the effects of memory and time delay, let us consider a model of cellular immunity.

1.4.1 Modeling Cell-Mediated Immunity in Homograft Rejection

In cell-mediated immunity, lymphocytes in the tissue become sensitized to the target (graft) cells and travel to the regional lymph nodes where they initiate an immunological response by increasing the production of immunocompetent lymphocytes. The newly produced lymphocytes are then transported into the blood stream via the thoracic duct. Lymphocytes recirculate from the blood stream through the tissue and return to the blood stream via the lymphatic system. When foreign cells are introduced into the tissue, blood lymphocytes migrate into the tissue at an increased rate and bring about the destruction of the target cells. Lymphocytes have memory and they exhibit an increased
secondary response; e.g., if, after the rejection of the first graft, a second graft is introduced into the host, the second graft is rejected much faster. A similar situation occurs in delayed hypersensitivity, which is another cell-mediated reaction. In this analysis, let us assume that blood and tissue are well-stirred compartments and that the newly produced lymphocytes are introduced into the blood compartment (Reddy and Krouskop, 1978).

For sensitization to occur, a lymphocyte has to come in contact with a target cell. The number of lymphocytes becoming sensitized at any given time \(L_s(t)\) is a function of the number of lymphocytes in the tissue \(L_T(t)\) and the number of target (foreign) cells \(g(t)\):

\[
L_s(t) = C_1 L_T(t)g(t) \quad (1.24)
\]

Certain lymphocytes, on encountering target cells, are transformed into memory cells. The memory cell formation depends on the number of lymphocytes in the tissue and the number of target cells. The number of memory cells formed at any time \(t\) may thus be expressed as

\[
L_{m_s}(t) = C_1 L_T(t)g(t) \quad (1.25)
\]

Sensitized lymphocytes stimulate the production of immunocompetent lymphocytes, and the effect of each sensitized cell lasts for a given period of time. For the purpose of the present analysis, it is assumed that the effect of each sensitized lymphocyte decays exponentially over a period of time. The production rate of blood lymphocytes at any time \(t\) due to the primary response \(\left(\frac{dL_B}{dt}\right)_{\text{prim}}\) would then be equal to the sum of the residual effect of all the lymphocytes sensitized between time 0 and time \(t - \Phi_1\), where \(\Phi_1\) is the time lag between sensitization and production of the lymphocytes.

The number of lymphocytes produced because of primary response between time \(t\) and time \((t - \Phi_1)\) would be

\[
L_B(t) - L_B(t - \Delta t) = C_3 \left[ L_s(t - \Phi_1) \Delta t + L_s(t - \Phi_1 - \Delta t) \Delta t + L_s(t - \Phi_1 - 2 \Delta t)e^{-\kappa_1 \Delta t} \Delta t + \cdots \right]
\]

Due to lymphocytes sensitized at \(t - \Phi_1\)

Due to lymphocytes sensitized at \(t - \Phi_1 - \Delta t\)

Due to lymphocytes sensitized at \(t - 2\Phi_1 - \Delta t\)

\[
+ L_s(t - \Phi_1 - r \Delta t)e^{-\kappa_1 r \Delta t} \Delta t + \cdots \right] \quad (1.26)
\]

\[
= C_3 \sum L_s(t - \Phi_1 - r \Delta t)e^{-\kappa_1 r \Delta t} \Delta t \quad (1.27)
\]

Dividing by \(\Delta t\) and taking the limits as \(\Delta t \rightarrow 0\) changes the left-hand side to a derivative and the right-hand side to an integral in terms of the hereditary function:

\[
\left(\frac{dL_B(t)}{dt}\right)_{\text{prim}} = C_3 \int_0^{t-\Phi_1} L_s(\tau)e^{-\kappa_1 (t-\Phi_1-\tau)} d\tau \quad (1.28)
\]

Substituting for \(L_s\) in terms of \(L_T\) gives

\[
\left(\frac{dL_B(t)}{dt}\right)_{\text{prim}} = k_2 \int_0^{t-\Phi_1} L_T(\tau)e^{-\kappa_1 (t-\Phi_1-\tau)} d\tau \quad (1.29)
\]

For the secondary response to appear, a memory cell must encounter a target cell, and therefore the secondary response depends on the number of memory cells and the number of target cells. As in the
above equation, Reddy and Krouskop (1978) expressed the secondary response in terms of a hereditary function

\[
\left( \frac{dL_B(t)}{dt} \right)_{\text{secondary}} = k_3 \int_0^{t-\phi_2} L_1(\tau) g(\tau) e^{-k(t-\phi_2-\tau)} g(t-\Phi_3) \, d\tau
\]  

(1.30)

In developing the above equation, it is assumed that the effect of a memory cell also decays exponentially over a period of time. Thus the production rate of blood lymphocytes at time \( t \) due to secondary response \( \frac{dL_B}{dt} \) is due to the sum of the residual effects of all the memory cells formed between 0 and time \( t - \phi_2 \), where \( \phi_2 \) is the time lag between memory cell formation and the appearance of the secondary response.

The net rate in change of blood lymphocytes may then be described as

\[
\frac{dL_B}{dt} = k_2 \int_0^{t-\phi_1} L_1(\tau) e^{-k(t-\phi_1-\tau)} \, d\tau + k_3 \int_0^{t-\phi_2} L_1(\tau) g(\tau) e^{-k(t-\phi_2-\tau)} g(t-\Phi_3) \, d\tau
\]

Due to primary response

Due to secondary response

\[
\begin{align*}
\text{Recirculation} & \quad K_d L_T - K_d L_B \\
\text{Death in the blood} & \quad -K_d L_B \\
\text{Migration into tissue due to target cell presence} & \quad -K_d L_B g
\end{align*}
\]  

(1.31)

The rates of change of tissue lymphocytes and the number of target cells can be described by a similar mass balance:

\[
\frac{dL_T}{dt} = K_d L_B g - K_d L_T + K_d L_B - K_d L_T g
\]

Increased migration \quad Recirculation \quad Loss due to target cell destruction  

(1.32)

\[
\frac{dg}{dt} = \left( \frac{dg}{dt} \right)_{\text{input}} - k_{10} L_T g
\]

(1.33)

These equations were simulated by Reddy and Krouskop. Figure 1.9 shows the production rate of lymphocytes and the number of target cells when the target cells were introduced on day 0 and again on day 4. Figure 1.10 shows the production rate of lymphocytes and the number of target cells present in the tissue when the target cells were introduced continuously.

### 1.4.2 A Model to Predict the Number of Engineers in the United States

An easy to understand example of a deterministic model with time delay and memory is a model to predict the number of engineers in the United States at any given time. Let us restrict our analysis to a single discipline such as biomedical engineering. Let \( E \) be the number of engineers (biomedical) at any given time. The time rate of change of the number of engineers at any given time in the United States can be expressed as:

\[
\frac{dE}{dt} = G + I - R - L - M
\]

(1.34)

where \( G = \) number of graduates entering the profession (graduating from an engineering program) per unit time  
\( I = \) number of engineers immigrating into the United States per unit time  
\( R = \) number of engineers retiring per unit time  
\( L = \) number of engineers leaving the profession per unit time (to become doctors, lawyers, managers, etc.)  
\( M = \) number of engineers dying (before retirement) per unit time
FIGURE 1.9 The simulation results of the production rate of lymphocytes (a), and the number of target cells or foreign cells (b) plotted as a function of time in days. In the simulation, the target cells were introduced on day 0 and again on day 4. [From Reddy and Krouskop (1978).] Note the increased secondary response.
FIGURE 1.10 The simulation results of the production rate of lymphocytes (a), and the number of target cells or foreign cells (b) when the target cells were continuously introduced. [From Reddy and Krouskop (1978).]
In this equation, we have lumped the entire United States as a single region (a well-stirred compartment) with homogeneous distribution. In addition, we have not made any distinction with regard to age, sex, or professional level. We have considered the entire pool as a well-stirred homogeneous compartment. In reality, there is a continuous distribution of ages. Even with this global analysis with a lumped model, we could consider the age distribution with a series of compartments with each compartment representing engineers within a particular age group. Moreover, we have assumed that all engineering graduates enter the workforce. A percentage of them go to graduate school and enter the workforce at a later time.

The number of graduates entering the profession is a function of the number of students entering the engineering school 4 years before:

\[ G(t) = k_1 S(t - 4) \]  

where \( S(t) \) is the number of students entering the engineering school per unit time. The number of students entering the engineering school depends on the demand for the engineering profession over a period of years, i.e., on the demand history.

The number of engineers immigrating into the United States per unit time depends on two factors: demand history in the United States for engineers and the number of visas that can be issued per unit time. Assuming that the immigration visa policy is also dependent on the demand history, we can assume that \( I \) is dependent on demand history. Here we have assumed that immigrants from all foreign countries are lumped into a single compartment. In reality, each country should be placed in a separate compartment and intercompartmental diffusion should be studied.

The number of engineers retiring per unit time is proportional to the number of engineers in the profession at the time:

\[ R(t) = k_2 E(t) \]  

The number of engineers leaving the profession depends on various factors: the demand for the engineering profession at that time, demand for various other professions at that time, and several personal factors. For the purpose of this analysis, let us assume that the number of engineers leaving the profession in a time interval is proportional to the number of individuals in the profession at that time:

\[ L(t) = k_3 E(t) \]  

The number of engineers dying (before retirement) per unit time is proportional to the number of engineers at that time:

\[ M(t) = k_4 E(t) \]  

The demand for engineers at any given time is proportional to the number of jobs available at that time \( (J(t)) \) and is inversely proportional to the number of engineers available at that time:

\[ D(t) = kJ(t)/E(t) \]  

The number of jobs available depends on various factors, such as government spending for R&D projects, economic growth, sales of medical products, and number of hospitals. Let us assume in this case (biomedical engineering), the number of jobs is directly proportional to the sales of medical products \( (p) \), directly proportional to government spending for health care R&D \( (e) \), and directly proportional to the number of new medical product company start-ups \( (i) \):

\[ J(t) = (k_6 e + k_7 c + k_8 I + k_9 + kp) \]  

Although we assumed that the number of jobs at the present time is dependent on \( e(t), c(t), h(t), i(t), \) and \( p(t) \), in reality the number of jobs at present may depend on previous values of these parameters, or may depend on the history of these parameters.
Let us now analyze the demand history. This history depends on the memory function. Let us assume that the effect of demand existing at a time decays exponentially (exponentially decaying memory). The net effect of demands from time 0 to \( t \) can be expressed as

\[
H_1(t) = \int_{\tau=0}^{\tau=t} \{D(\tau) \exp[-k_{10}(t - \tau)]\} \, d\tau
\] (1.41)

The number of students entering engineering school per unit time is

\[
S(t) = k_{11}H_1(t)
\] (1.42)

Immigration rate can similarly be expressed as:

\[
I(t) = k_{12}H_2(t)
\] (1.43)

where

\[
H_2(t) = \int_{\tau=0}^{\tau=t} \{D(\tau) \exp[-k_{13}(t - \tau)]\} \, d\tau
\] (1.44)

\( H_1 \) and \( H_2 \) are called hereditary functions. Instead of an exponential decay of memory, we could have a sinusoidal or some other functional form of memory decay, depending on the physical situation.

\[
\frac{dE}{dt} = k_1k_{10}H_1(t-4) + k_{11}H_2(t) - (k_2 + k_3 + k_4)E(t)
\] (1.45)

In this analysis, making various assumptions, we have formulated a lumped-parameter deterministic model to predict the number of engineers (biomedical) present in the United States at any given time. If we want to know the geographical distribution, we can take two approaches. We can divide the entire United States into a number of compartments (e.g., northeast, east, west, etc.) and study the intercompartmental diffusion. Alternatively, we can make \( E \) a continuous variable in space and time, \( I(x,y,t) \), and account for spatial diffusion.

### 1.5 Artificial Neural Network Models

Neural network models represent the black box type of model. These models are used where the precise functioning of the system is not understood but the sample input-output data are known. Neural networks represent a new generation of information processing systems that are artificially (virtually) constructed to make use of the organizational structure of the neuronal information processing in the brain. A neural network consists of several interconnecting neurons, also called nodes. These nodes are organized into an input layer, one or more hidden layers, and an output layer. The number of input layer nodes in the input layer depends on the number of input variables. The number of nodes in the output layer is determined by the number of output parameters. Each input parameter is represented by a node in the input layer, and each output parameter is represented by a node in the output layer. The number of nodes in the hidden layer could be variable. Each node in a feed-forward neural network is connected to every node in the next level of nodes. That is, each input node is connected to all the nodes in the hidden-layer neurons. Let us for simplicity consider only one hidden layer. Now, each node in the hidden layer is connected to all the nodes in the output layer. Figure 1.11 shows a network with four output nodes and five input nodes with four hidden-layer nodes. The connection strengths are determined by the weights.

Let us assume that \( W_{ij} \) represents the weight of the connection from \( j \)th node in the hidden layer to the \( i \)th node in the output layer, and let us assume that \( w_{ij,k} \) represents the weight of connection from \( k \)th input node to \( j \)th node in the hidden layer. Let \( X_i \) represent the value of \( k \)th input node. The
The sum of weighted inputs to the $j$th node in the hidden layer is
\[ I_j = \sum w_{ij} X_k \] (1.46)

In other words,
\[ I_1 = w_{1,1} X_1 + w_{1,2} X_2 + w_{1,3} X_3 + w_{1,4} X_4 \] (1.47)

where $X_1$, $X_2$, $X_3$, and $X_4$ are the values of the four input parameters.

The output of a hidden-layer neuron is a function of its input:
\[ H_j = f(I_j) \] (1.48)

This function $f$ is called the activation function. An example of this function is the sigmoid function
\[ H_j = k \left[ \frac{2}{1 + \exp(-a I_j + B)} - 1 \right] \] (1.49)

where $k$, $a$, and $b$ are constants. $B$ is called the bias, and can be zero. In general, any monotone, nondecreasing differentiable signal function can be used as the activation function.

The input $G_i$ to the $i$th node in the output layer is the sum of its weighted inputs:
The output of the node in the output layer is some function of the input node.

\[ Y_i = F(G_i) \]  

(1.51)

The activation function \( F \) of the output neurons can be any monotone, nondecreasing differentiable function. Sigmoid or logistic functions are usually used.

If the weights \( W_{jk} \) and \( W_{ij} \) are all known, then given the inputs \( X_k \), the output \( Y_i \) of the system can be calculated. The weights are determined through a training algorithm by using the sample input-output data.

There are several training techniques, and the most popular technique is the back propagation technique. Let us assume that, for a set of sample inputs \( X_k \), we know the actual outputs \( d_i \). Initially, we do not know the weights, but we could have a random initial guess of the weights \( W_{jk} \) and \( W_{ij} \). As an example, we could define all weights initially to be \( W_{jk} = W_{ij} = 0.2 \) or \( 0.5 \). Using the above equations along with the sample input vector \( X_k \), we can calculate the output of the system \( Y_i \). Of course, this calculated value is going to be different from the actual output (vector, if there is more than one output node) value \( d_i \), corresponding to the input vector \( X_i \). The error is the difference between the calculated output value and the actual value. There are various algorithms to iteratively calculate the weights, each time changing the weights as a function of the error. The most popular of these is the gradient descent technique.

The sum of the error in the \( m \)th iteration is defined as

\[ e_i^m = d_i - y_i^m = d_i - F(\Sigma W_{ij}^m H_j) = d_i - F(\Sigma W_{ij}^m f(\Sigma W_{jk}^n X_k)) \]  

(1.52)

The instantaneous summed squared error at an iteration \( m \) corresponding to the sample data set \( n \) can be calculated as

\[ E_n^m = \frac{1}{2} \sum (e_i^m)^2 \]  

(1.53)

The total error \( E \) at each iteration, for all the sample data pairs (input-output), can be calculated as the sum of the errors \( E_n \) for the individual sample data.

Adjusting the weights for each iteration for connections between the hidden layer neurons and the output layer neurons \( W_{ij} \) can be calculated as

\[ W_{ij}^{m+1} = W_{ij}^m - Z \left( \frac{\delta E^m}{\delta W_{ij}} \right) \]  

(1.54)

where \( Z \) is the learning rate.

The error gradient can be expressed as

\[ \frac{\delta E^m}{\delta W_{ij}} = \left( \frac{\delta E^m}{\delta F} \right) \left( \frac{dF}{dW_{ij}} \right) \]  

(1.55)

For a sigmoid activation function, it turns out that the differential is a simple function of the sigmoid:

\[ dF = b(1 - F)F \]  

(1.56)

where \( b \) is a constant. Thus,

\[ \frac{dF}{dW_{ij}} = b(1 - F(W_{ij}))F(W_{ij}) \]  

(1.57)

For adjusting the weights for connections between the input and the hidden-layer neurons, the error
is back-propagated by calculating the partial derivative of the error $E$ with respect to the weights $w_{j,k}$ similarly (Haykin, 1999).

The whole process of calculating the weights by using the sample data sets is called the training process. There is a Neural Network package in MATLAB that can be easily used in the training process. There are several algorithms in the package including back-propagation, modified back-propagation, etc., that the user can choose in the MATLAB software. Once the weights are calculated by MATLAB or any other software, it becomes a matter of obtaining the output vector for a given input vector by using matrix multiplications. The most important aspect of a neural network is that it should be tested with data not used in the training process.

Neural networks have been used for classification and control. For instance, Reddy et al. (1995) used neural networks to classify the degree of the disease in dysphagic patients by using noninvasive measurements (of throat acceleration, swallow suction, pressure, etc.) obtained from dysphagic patients during swallowing. These measurements were the inputs to the network and the outputs were normal, mild, moderate, and severe. Neural network performance depends on a black-box approach, and since the performance depends on the sample data, initial weights, etc., Reddy et al. (1995) trained several networks with various initial conditions and activation functions. On the basis of some initial testing with known data, they recruited the best five networks into a committee. A majority opinion of the committee was used as the final decision. Reddy and Buch (2000) and Das et al. (2001) obtained better results with a committee of neural networks (Fig. 1.12) when compared to a single network, and the majority opinion of the committee was in agreement with clinical or actual classification.

1.6 MECHANICAL MODELS

Mechanical models consisting of combinations of springs and dashpots are very popular in numerous disciplines. Spring-dashpot models have been used to model the mechanical behavior of viscoelastic
materials and can be used to represent the one-dimensional behavior of tissue and other biological materials. In a linear spring, the force is proportional to the change in length or the strain. On the other hand, the force in a dashpot is proportional to the rate of change in strain. Consider a mass supported by a spring and a dashpot in parallel. Let a force \( F \) be acting on the mass. Application of Newton’s law results in:

\[
m\left(\frac{d^2X}{dt^2}\right) + b\left(\frac{dX}{dt}\right) + kX = F
\]

where \( X \) = elongation or change in length with respect to the steady-state value
\( b \) = constant of the dashpot
\( k \) = spring constant

It should be pointed out that the mechanical equation (1.58) is similar to the following electrical equation:

\[
L\left(\frac{di}{dt}\right) + Ri + \frac{1}{C} \int i \, dt = E
\]

where \( L \) = inductance
\( R \) = resistance
\( i \) = current
\( E \) = voltage

This equation can be expressed in terms of the charge \( q \) instead of the current as

\[
L\left(\frac{dq}{dt}\right) + R\left(\frac{dq}{dt}\right) + \frac{1}{C} q = E
\]

Equations (1.58), (1.59), and (1.60) are similar. Therefore, mass is analogous to the inductor, the dashpot is analogous to the resistor, and the spring is analogous to the capacitor. The spring and the capacitor are storage units, whereas the dashpot and the resistor are dissipators of energy. The charge is analogous to the deformation or elongation, the current is similar to the velocity, and force is analogous to the voltage. Therefore, any electrical system can be modeled by mechanical analogs and any mechanical system can be modeled by electrical analogs.

Lumped mechanical models have been used to analyze impact dynamics. Generally, muscle is represented by a combination of a spring and a dashpot, whereas a ligament is modeled by a spring. Human body vibrations can be analyzed by similar models (Fritton et al., 1997).

### 1.7 MODEL VALIDATION

This chapter discussed the art of modeling, with a few examples. Regardless of the type of model developed, a mathematical model should be validated with experimental results. Validation becomes very important in the black-box type of models such as the neural network models. Moreover, the model results are valid only within certain regimes where the model assumptions are valid. Sometimes, any model can be fit to a particular set of data by adjusting the parameter values. The techniques of parameter estimation were not presented in this chapter; the presentation was limited to lumped-parameter analysis, or macroscopic modeling.

## REFERENCES


