General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

CR-171 875
C.1

TECHNICAL INFORMATION RELEASE

management and technical services company

TIR 2114-MED-5005

FROM	**TO**
Joel I. Leonard, Ph.D.	N. Cintron, Ph.D./SD4

DATE	**CONTRACT NO:**	**T.O. OR A.D. REF:**	**MIS OR OTHER NASA REF:**
April 22, 1985	NAS9-17151		

SUBJECT The Modeling and Simulation of Feedback Control Systems

(NASA-CR-171875) THE MODELING AND SIMULATION OF FEEDBACK CONTROL SYSTEMS (Management and Technical Services Co.) 14 p HC A02/MF A01 CSCL 09B G3/63 N85-26225 Unclas 21222

This document contains a brief description of the principles of mathematical models and their development. It should serve as an introduction to those unfamiliar with this topic. It was originally prepared as an Appendix for a NASA Reference Publication and is documented here in preliminary form.

Joel I. Leonard, Ph.D.

/db

Attachment

Unit Manager **Approving** F.A. Kutyna **NASA Manager** F.A. Kutyna, Ph.D. **Concurrence**

DISTRIBUTION

NASA/HQS:
R. J. White

MATSCO/HQS:
L. Griffiths

NASA/JSC:
BE3/B.J. Jefferson
SA/W. H. Shumate
SB/J. Vanderploeg
SD2/J. Logan
SD3/M. Bungo
/J. Charles
/P. Johnson
/V. Schneider

MATSCO/HO:
R. F. Meyer
Mission Sci.Group
TIR Files

GE/HO:
T.D. Gregory(2)
Contracts File

THE MODELING AND SIMULATION OF FEEDBACK CONTROL SYSTEMS

Prepared by

Joel I. Leonard, Ph.D.

Management and Technical Services Company

Houston, Texas

Prepared for

National Aeronautics and Space Administration

Contract NAS9-17151

1985

The Modeling and Simulation of Feedback Control Systems

The purpose of this appendix is to explain the basic vocabulary and principles of model development so that the reader may better grasp the description of the simulation models. Examples which describe the steps leading to a computer algorithm of a model subsystem are provided, and the simulation techniques used to assess model behavior and accuracy are discussed.

DEFINITIONS OF TERMS AND CONCEPTS

Parameters and Variables

Two general types of quantities are used in mathematical models: *parameters* and *variables*. The value of a *parameter* is generally constant with respect to time, whereas the value of a *variable* changes with time (i.e., is *time-varying*). Biological parameters often vary slowly with time but may be assumed to be constant in the mathematical model.

Parameters may be considered *independent* of system actions. Variables may also be considered independent if they influence the system from the outside. Parameters and *independent variables* are also called *input functions*, or *forcing functions*. More often, variables are *dependent* (also called output functions), since they vary according to the relationships within the system. The objective of a systems analysis is to specify the state of the system; that is, to specify the values of all dependent variables at every instant of time.

Classification of Mathematical Model Systems

Mathematical models are classified according to the types of equations they employ (ref. A-1).

Distributed- and lumped-parameter systems—In *distributed-parameter systems*, the values of variables depend on time as well as space coordinates, and the system must be analyzed by solving *partial differential equations*. In a *lumped-parameter system*, each element is treated as if it were concentrated ("lumped") at one particular point, which is called a *node* or a *compartment*. The use of lumped parameters greatly simplifies the analysis because *ordinary differential equations* are used to describe the changes with time. To account for spatial changes that occur in such a system (e.g., the value of blood pressure is different in arteries, capillaries, and veins), compartments are added for each major space location until one reaches the level of compartmentalization (lumping) commensurate with the basic purpose of the model being designed.

Generally, as subdivisions are included in a model, the fidelity and accuracy of the model's response increases. For example, in this project, two circulatory models were employed, one composed of 7 compartments and the other of 28 compartments. The 7-compartment model could simulate mean blood pressures in arteries and veins, whereas the 28-compartment model could simulate pulsatile flow and provide systolic and diastolic pressure for all anatomical portions of the vasculature (i.e., aorta, arteries, arterioles, capillaries, venules, veins, and vena cava).

All models described in this publication are lumped-parameter systems. Whether the lumping of components and the simplifying of input-output relationships are appropriate depends entirely on the intended purpose of the model. Lumped, compartmental modeling is in keeping with a basic facet of the systems approach: to simplify the problem and define the essentials of its solution.

Linear and nonlinear systems—In modeling, systems are generally taken to consist of components with a known (or assumed) relationship. In a broad sense, a component of a system may be thought of as transforming certain inputs into certain outputs. Such relationships are often (loosely) termed *transfer functions* (although transfer function has a rigorous meaning only in terms of Laplace transforms).

Mathematically, the components of a system may be represented by several operators, which transform one function into another. If all the operators for the system are linear, in the loose sense that their operation on two independent functions produces two independent results, then the system itself is said to be *linear*. Otherwise, the

system is said to be *nonlinear*. Mathematically, much more is known about the behavior of linear systems.

Unfortunately, most biological systems—with their threshold and saturation behavior, sigmoidal dose-response relationship, and dead time—are nonlinear; analysis of such systems revolves around *numerical methods* and the use of large *computers*. Obtaining rapid, accurate, numerical solutions of nonlinear system equations is absolutely essential in biological modeling and often represents a separate challenge.

Other classifications—Models can be developed for either *predictive* or *descriptive* purposes. A predictive model must only produce accurate predictions of the output variables in the system. For example, a single equation may be used to predict the fractional saturation of hemoglobin with oxygen at different levels of oxygen partial pressure. In the latter case, the modeler is not concerned with exact replication in the model of the interactions between the variables in the system. The relationships employed in the predictive model to generate the predicted output need not conform to the mechanisms in the real system which lead to the same outputs. In contrast, descriptive models not only must generate predictions in agreement with real system output but also must employ intermediate relationships that are realistic representations of the true processes which generate the observed outputs.

The vast majority of biological systems are *continuous* systems, in which values are always changing with time, and they are best described by *differential equations*. These equations are typically solved with numerical procedures on digital computers by transforming them into *finite-difference equations*. In this process, the variables actually appear to be constant for the duration of the *integration step size* and, therefore, more properly belong to a *discrete system*. This approximation of continuous biological systems by discrete numerical systems can produce inaccuracies and instabilities in the solution, unless appropriate care is taken and the system is carefully verified.

The models described in this section can also be said to be *deterministic*, since the input-output relationships for each component are based on simple physical laws (i.e., flow-pressure, diffusion, mass action) and are therefore fixed, predictable, and reproducible. In reality, most biological systems are *stochastic*, in the sense they are subject to random noise, and their responses can be described by statistical laws and in terms of probabilities and expected values.

Riggs (ref. A-1) makes the observation that "all naturally-occurring systems are, in the final analysis, time-varying distributed-parameter, quantized, stochastic, nonlinear systems. That we can sometimes obtain useful information by treating these wayward creatures as if they were fixed lumped-parameter, continuous, deterministic, linear systems is little short of miraculous."

Biological Control Systems

It is difficult to conceive of any biological quantity which is not controlled or influenced by one or more factors. For example, blood pressure influences and controls the rate of baroreceptor neural firing. In this situation, baroreceptor neural firing is *controlled*, or *regulated*, within narrow limits at the expense of the other system quantities, the limits of which may vary widely. The relationship between blood pressure and baroreceptor neural firing would be construed as an *open-loop* system for controlling baroreceptor neural firing. Further examination of the process shows that the baroreceptor afferent signal is processed by the central nervous system (CNS) and results in an efferent neural signal that controls peripheral resistance. This relationship between baroreceptor firing and peripheral resistance is an open-loop system for controlling peripheral resistance. However, it is known that peripheral resistance directly regulates blood pressure. Once this control is added to the system, the complete loop of physical changes (blood pressure, peripheral resistance) and information transmission (afferent and efferent signals) forms a *closed-loop* control system. The difference between the closed-loop and open-loop systems is obviously the presence of the relationship of peripheral resistance to blood pressure. Blood pressure in the closed-loop system is called the *feedback variable*.

A feedback control system, as described previously, is composed of two major elements, or components: the controlled system and the *controlling system* or *controller*. Figure A-1 is a diagram illustrating the relationships of these elements. The purpose of the controlled system is to transform inputs from the outside (*load inputs*, *disturbances*, or *stress stimuli*) and inputs from the controller (*con-*

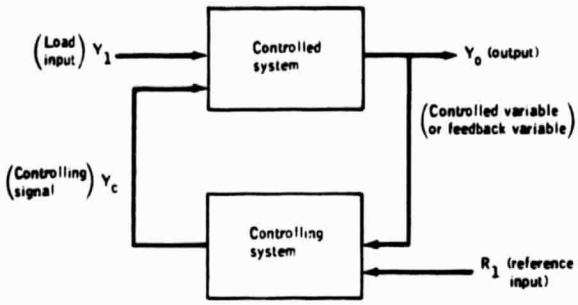


FIGURE A-1.—A generalized feedback control system. Most biological homeostatic systems can be reduced to these basic components.

troliing signals) into *responses* (*outputs*). The reference input of figure A-1 is also referred to as a *set-point*, and it represents a normal or desired val e of a feedback variable. Differences between the set-point and the actual feedback variable result in corrective action by the controller; this correction permits the output to revert toward normal values in the face of the disturbance. Consequently, the presence of *feedback control* in biological systems allows important quantities to fluctuate within limits necessary for maintaining life in the face of many metabolic and environmental disturbances encountered by the organism.

In manmade technological control systems, the set-point is a physical quantity that can be varied at will (e.g., a thermostat setting for a home heating system). Biological systems, however, often have neither a reference input nor an error detector. The establishment of a set-point, or an *operating point*, in mathematical models of biological systems is often included for convenience, only because the real system behaves as if it were controlled by such a reference value. (See refs. A-1 and A-2 for further information.)

In terms of the example described previously, blood pressure is the controlled feedback variable and peripheral resistance can be considered the controlling variable. The controlled system consists of components that transform changes in blood pressure into changes in peripheral resistance (i.e., baroreceptor afferent signals, CNS processing, efferent signals, effect of autonomics on vasculature resistance). A typical blood pressure load disturbance might be an infusion of blood into the circulatory system (which implies that a description of the volume-pressure relationships of the circulation must be included in the controlled system) or the introduction of an upright tilt disturbance (which implies that a description of gravity effects on blood volume distribution must be included in the controlled system).

FORMULATION OF A COMPUTER MODEL

It would be instructive to review the steps that lead from conceptualization of a model to digital computer implementation.

Model systems can be represented by some combination of boxes which represent the elements of the system. Such combinations of boxes are called *block diagrams*, which include logic diagrams, schematics, flow diagrams, and system diagrams. The first step in constructing a model of any system is to draw a block diagram representing these interconnections. This type of diagram, or series of block diagrams, becomes a "roadmap" or a "key" for taking a system apart and putting it back together. Some of the most common basic symbols used in a special form of block diagram, called an analog diagram, are shown in figure A-2. The

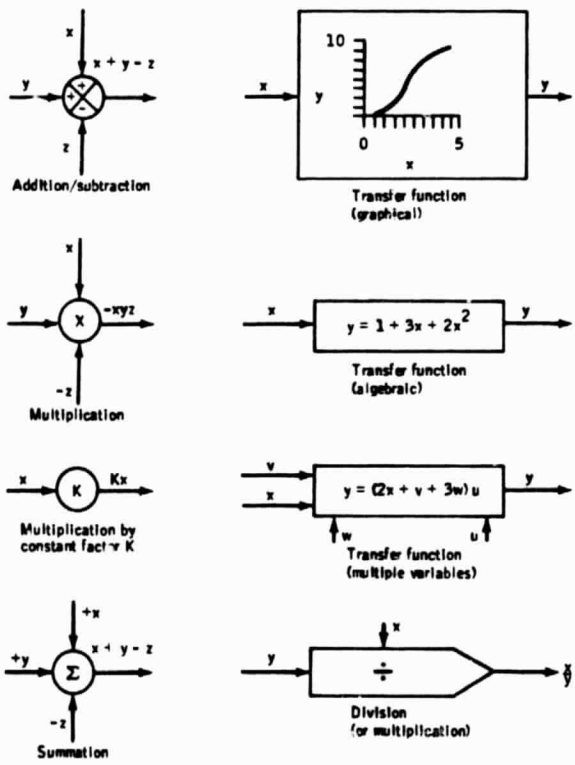
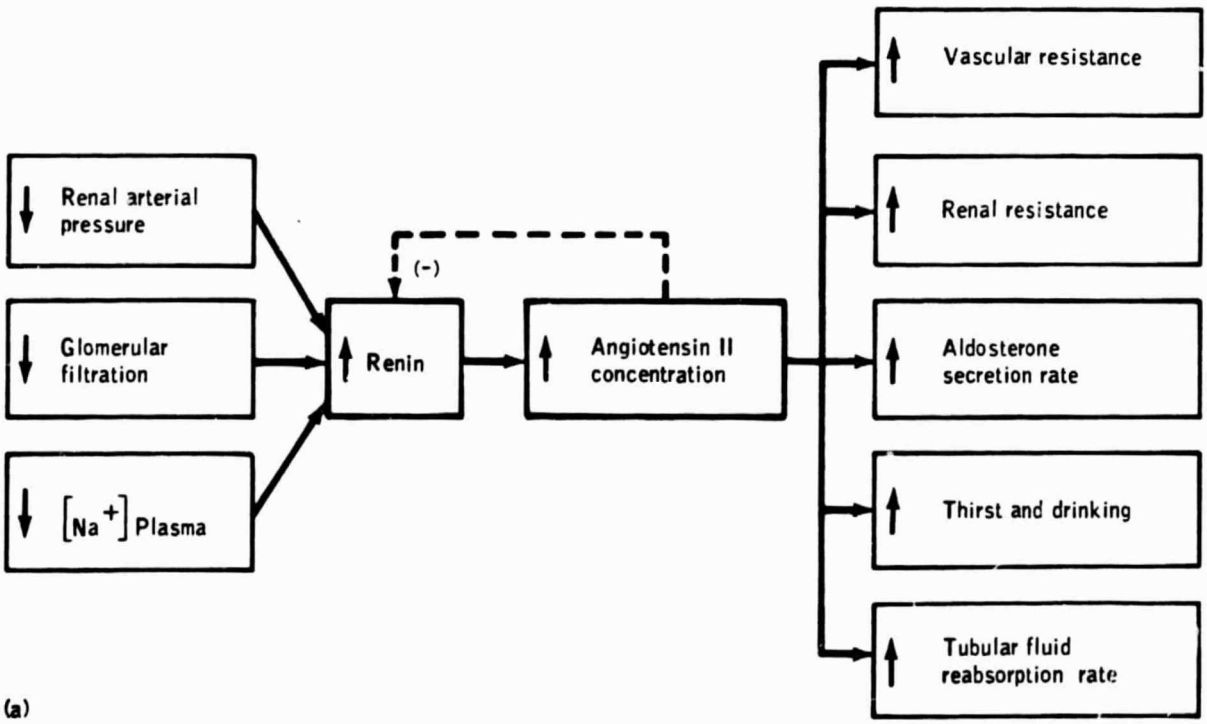


FIGURE A-2.—Symbols used in block diagrams representing mathematical operations.

analog diagram is an adaptation of the diagrammatic method so useful with linear systems. The mathematical operations are self-explanatory. As an example of the loosely termed "transfer function," the graphical transfer function in the figure might represent the relationship between cardiac output x and right atrial pressure y. This transfer function can then represent a measured cardiac function curve.

As a specific example of the programing process, the subsystem of the circulatory, fluid, and electrolyte model that deals with angiotensin formation has been chosen. Figure A-3(a) can be construed as a hypothesis diagram for the open-loop control of renin-angiotensin. This diagram indicates the major factors which control angiotensin formation and the many effects of angiotensin on other subsystems. It is generally agreed that two of the major influences on renin release (a precursor to angiotensin formation) from the juxtaglomerular cells of the kidneys are renal perfusion pressure and the load of sodium filtered through the tubules (i.e., glomerular filtration times plasma sodium concentration). Renin enters the circulation and forms angiotensin, which has a slight negative feedback effect on renin formation. (See dashed line in fig. A-3(a).) The final effect of angiotensin is considered to be widespread, affecting vascular resistance, renal resistance, aldosterone secretion rate, thirst and drinking, and the rate of tubular fluid reabsorption.

To progress to a computer model of this system, each of the pathways connecting any two variables must be described in as much detail as is possible, or as much as is commensurate with the objectives of the model design. Therefore, the next level of detail is shown in the block diagram of figure A-3(b), in which the relationships between variables are qualitatively organized. Functional relationships are identified wherever data describing hormonal secretion rates or hormonal effects are available. This diagram was developed from some



(a)

FIGURE A-3.—Formulation of a control system algorithm. (a) Hypothesis diagram showing factors which influence renin-angiotensin production and the quantities which they affect. These represent the assumptions for one of the hormonal subsystems in the Guyton model. (b) A more detailed block diagram of the hypothesis diagram of figure A-3(a) showing the mathematical operators relating each quantity. (c) An analog computer diagram of the renin-angiotensin system. This diagram contains sufficient information to wire a patch panel for programing an analog computer. (d) A Fortran algorithm for the renin-angiotensin control system as it appears in the Guyton model. The symbols can be interpreted from the preceding diagrams.

very general algebraic relationships, and the mathematical operations of summation, division, multiplication, and integration are indicated.

Figure A-3(c) is the same diagram, reorganized to reflect the computer program names for each variable and the quantitative transfer functions for each element. It was common at one time to simulate models on analog computers, and figure A-3(c) is known as an *analog diagram* because it represents a circuit diagram for programing this segment of the model on an analog computer. However, a digital computer requires a different language, and figure A-3(d) is the Fortran version of the renin-angiotensin subsystem. Note that the first statement represents a function relating renal pressure and renin secretion. In the analog computer, this function would be represented by a function generator; in the digital computer, it is represented by a table of *x-y* values (not shown) from which the desired operating points are interpolated. The exponential functions (e.g., EXP(−I/RNK)) in figure A-3(d) are numerical approximations for integration, in which "I" is the integration step size.

It can be appreciated that the discipline of translating an ordinary physiological hypothesis diagram into a quantitative computer representation can lead to a better understanding of the biological system. Available data from diverse sources are integrated into a common framework; other data are excluded as being nonessential for the given level of detail (i.e., importance of data can be ranked), and missing information, which suggests the need for new experiments, is quickly identified.

SIMULATION TECHNIQUES

Once a model is implemented on a computer and verification procedures ensure that it is operating appropriately, the model is ready to be tested for accuracy. A model that is deemed credible can also be used to describe the behavior of the system in terms familiar to control engineers and can be used to predict system responses in terms that can be tested by biological experimentation. Some of the more important techniques used to produce model credibility and to predict system behavior are discussed in the following paragraphs.

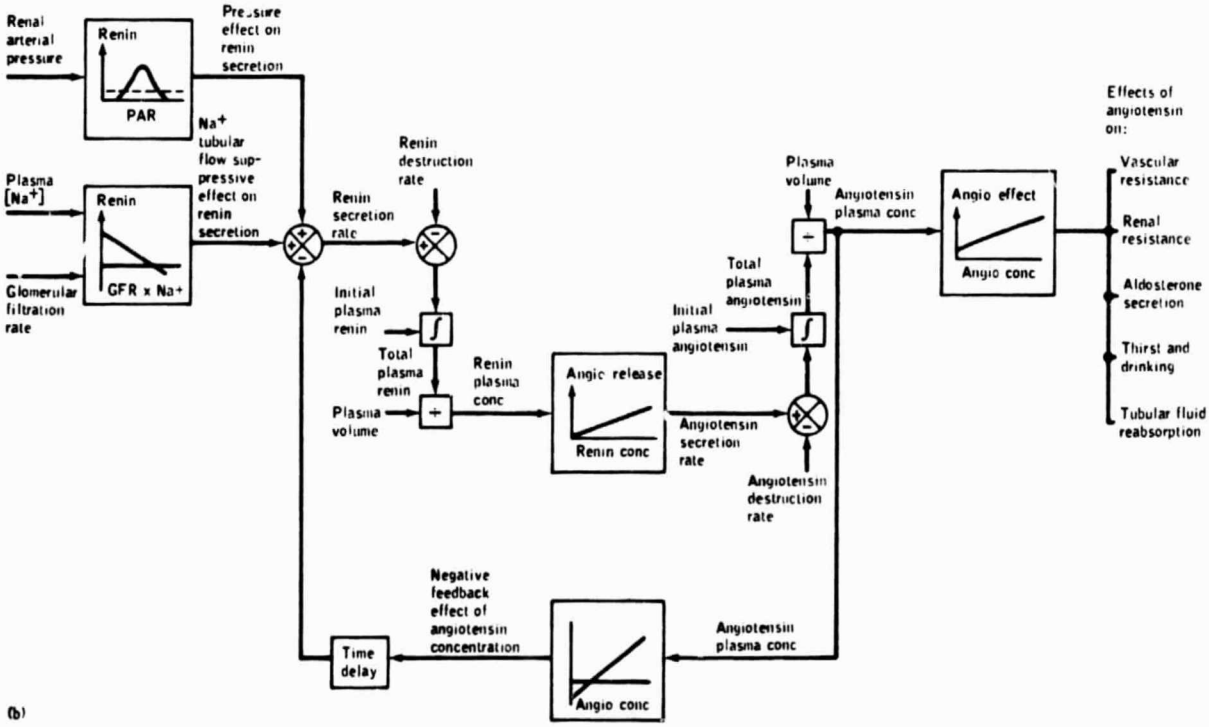


FIGURE A-3.—Continued.

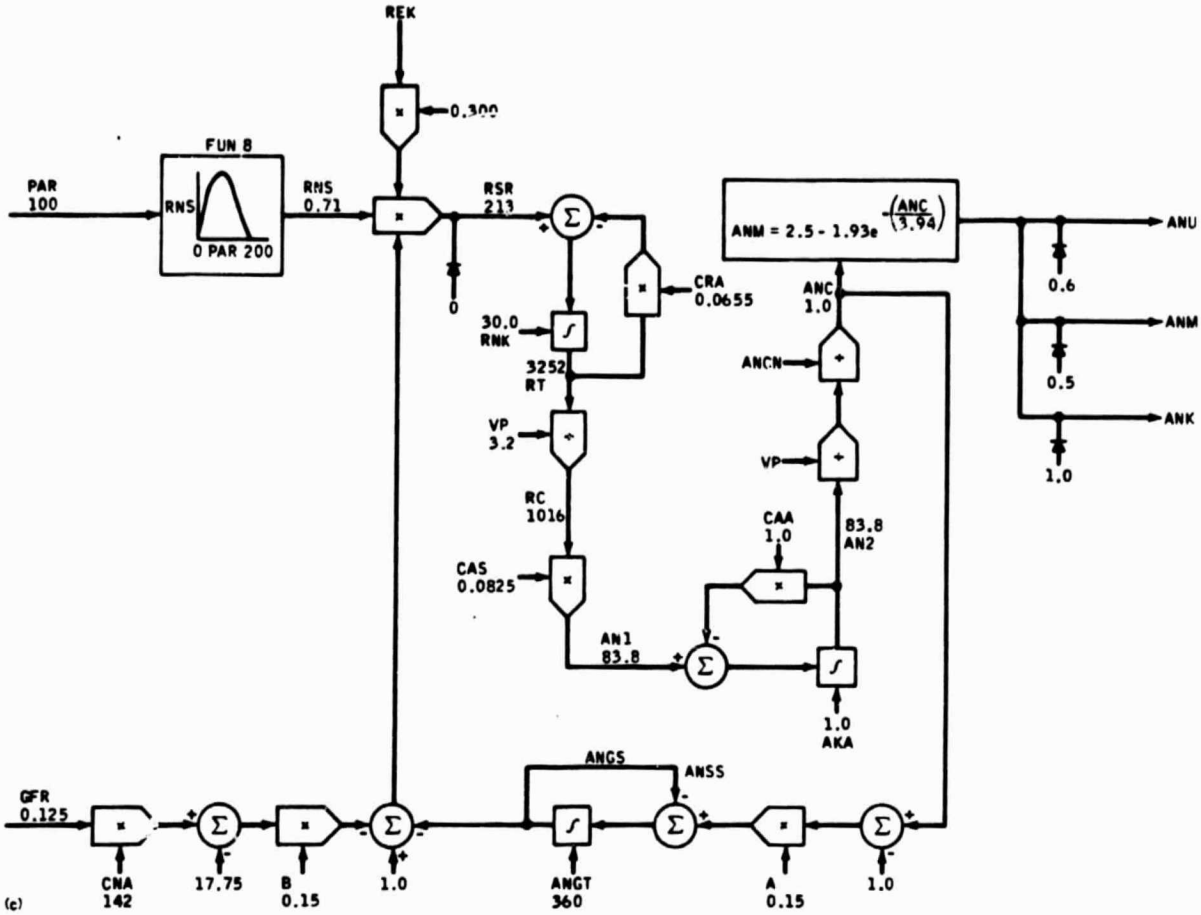


FIGURE A-3.—Continued.

```
CALL FUNCTN(PAR, RNS, FUN8)
RSR = 300.*REK*RNS*(1.-ANGS-B*(GFR*CNA-17.75))
IF (RSR.LT.0.)) RSR = 0.0
RT = RT * (RSR-CRA*RT)*(1.-EXP(-I/RNK))
RC = RT/VP
AN1 = CAN+CAS*RC
AN2 = AN2+(AN1-CAA*AN2)*(1.-EXP(-I/AKA))
ANC = AN2/VP/ANCN
ANM = ANMM-AN3*EXP(-ANC/ANTC)
IF (ANM.LT.0.5) ANM = 0.5
ANSS = A*(ANC-1.)
ANGS = ANGS+(ANSS-ANGS)*I/ANGT
ANK = ANM
IF (ANK.LT.1.)ANK = 1.
ANU = ANM
IF (ANU.LT.0.6) ANU = 0.6
```
(d)

FIGURE A-3.—Concluded.

Dynamic Simulation

In the context of this study, obtaining the solution to a model means introducing some type of load disturbance (i.e., a parameter perturbation) and solving the model's differential equations iteratively, using numerical techniques. This process, known as dynamic simulation, is accomplished using high-speed digital computers and results in time-varying values of the dependent variables. These responses are examined in qualitative terms for their reasonableness by analysts who are familiar with the physiological system or, more often, they are compared quantitatively with experimental data. In this project, simulation responses were available from digital computers in tabular and graphical form and could be compared with previously stored data.

Sensitivity Analysis

Sensitivity analysis is a method for studying system responses due to variation in parameters (ref. A-3). The conceptual basis of sensitivity analysis is simple; small variations are made in the values of the system parameters of a model, and the effects of these changes are observed individually in the solution. (See app. E.) Sensitivity functions which describe the observed effects may be computed, and these are interpreted to extract information about the dynamic system that could not be obtained from simply finding solutions with a particular set of input conditions (refs. A-4 and A-5). Sensitivity analysis provides the following.

1. A quantitative means of comparing the relative importance of individual parameters on any system variable
2. A means of determining interactive effects of two or more parameters on model behavior
3. A tool to help assess the validity of a particular model without the need to collect and use extensive measurements from the real system
4. Information in a form that can be easily interpreted by those familiar with the subject matter of the model but not necessarily knowledgeable of simulation techniques
5. A means of assigning relative importance to all parameters, a process that can be valuable both to the simulater in performing parameter estimation or stability analysis and to the experimenter in allocating resources for data collection
6. A practical method of analyzing and comparing two different models designed to represent the same physical system

A sensitivity analysis is very useful when performed early in model development, before model validation. This analysis is particularly important to the experimenter, who can help evaluate the model based on the relative sensitivities of the parameters without really knowing much about the model. The technique also is useful for involving the experimenter early in the modeling process, another important factor for eventual model acceptance (ref. A-6). The results of the sensitivity analysis must be evaluated in the light of other known information about the real system. The fact that a parameter has a very significant effect on a particular variable of the model is of little importance if it is known that the parameter in the real system is relatively constant or that changes in other parameters are capable of canceling the original effects.

Although sensitivity analysis can be considered to be a special case of dynamic simulation, there are several important differences between these two procedures.

1. Sensitivity analyses are characterized by comparatively small perturbations.
2. Sensitivity analyses often are performed by varying one parameter at a time.
3. Sensitivity analyses often are performed to obtain sensitivity functions rather than solutions of the dependent variable
4. Sensitivity analyses usually entail comparisons between two or more simulation runs rather than between model results and experimental data.

Examples of sensitivity analysis are provided in the descriptions of the erythropoietic model and of the thermoregulatory model. (See app. E.)

Variation of Parameters

Once a parameter has been identified as particularly influential, either by sensitivity analysis or from direct knowledge of the real system, it is often desirable to determine its effect on different parameter values. For example, the volume of blood is known to be important in the blood flow/pressure response to upright tilt from the supine position. It is reasonable to ask, "How does the response change as more and more blood is removed?" Another way to pose this question is, "What is the effect of hemorrhage on standing?" Documenting this effect with a simulation model of circulatory control is relatively straightforward; the parameter representing blood volume is assigned a series of values (i.e., a percentage of the control or normal value) and, at each level, a dynamic simulation is performed. The resulting time-varying responses (of, for example, heart rate, cardiac output, or blood pressure) can be plotted as overlays on the same graph. If steady-state responses are desired, the graph often is constructed with blood volume on the abscissa and the response variable on the ordinate. (See Sec. V, "Cardiovascular Subsystem.")

Error Analysis

Parameter values are never known with 100 percent accuracy. If the standard deviation (SD) around the mean value can be estimated for each parameter, it is possible to place statistical confidence limits on a model's behavior. For the example discussed previously, assume an experiment is performed in which blood volume reduction is measured as −10 percent ± 1.5 percent (SD). Dynamic simulations can be performed for three values of blood volume: −8.5 percent, −10 percent, and −11.5 percent, corresponding to the mean minus SD, the mean, and the mean plus SD, respectively. The response, for example, for heart rate could also be expressed in terms of a mean and a deviation. This expression would represent a prediction of the minimum error interval in the response variable, because of the inherent design of the system and the uncertainty in measuring blood volume, but would not include any experimental errors that could occur in measuring the response variables. Conversely, if a decrease in the confidence interval of a response variable to a given width was desired, it would be possible, using these techniques, to determine the minimum experimental accuracy required in measuring the independent parameter.

Error analysis becomes especially desirable in large-scale systems containing many parameters that promulgate errors through the simulation, and in certain nonlinear systems in which the interactive effects of different parameters lead to amplification of individual errors. Unfortunately, even though the techniques to accomplish this analysis are rather straightforward, there are few examples in the literature of physiological systems.

A related problem that has application to error and sensitivity analysis is the effect of noise on the behavior of the system. Noise can be described as a statistical disturbance of a particular variable, and it is characterized by statistical properties such as mean value, probability distribution, or spectral density. Model output variation which results from noise can be found by including a distribution function for each parameter or variable that exhibits noisy behavior. This problem becomes extremely relevant in parameter estimation analysis when model output is compared to data having a significant noise level (ref. A-7).

Stability Analysis

It is appropriate to mention stability analysis of dynamic systems because of the inverse relationship between sensitivity and stability in negative feedback systems. In general, sensitivity to disturbing factors can be reduced by an increase in feedback gain (in technological systems at least). However, instability occurs as a consequence of this gain increase. Thus, systems with high gain may have low sensitivity to external perturbations but may also be operating on the borderline of instability. Most biological systems are normally stable, and they do exhibit low sensitivity. Whether they are working somewhere near the stability limit by way of high gain factors is not known but should be studied on a case-by-case basis (ref. A-8). An analysis of stability can become an important measure of the competence of a model, in that if both model and actual system can be thrown into instability by the same parametric changes, there is reason for having greater confidence in the mathematical representation.

Little practical work has been done on stability analysis of complex physiological systems. Formal techniques for investigating stability in linear systems and in simple nonlinear systems have been reported (refs. A-7 and A-9); but, for the most part, studying large-scale nonlinear models is a trial-and-error experience. Systems that exhibit oscillatory or periodic behavior in normal operation (e.g., eye movement, respiration) can often be made unstable more easily than systems that behave monotonically. Inherent instability is dependent on the properties of the system and is normally not a function of the specific disturbance. If the system is inherently stable, all transients will ultimately disappear regardless of the disturbance causing them. Conversely, any disturbance to an unstable system will initiate oscillations that increase in amplitude with time.

Stability can arise from either inherent features of the real system or from structural features of the mathematical model (such as long integration interval). The techniques of sensitivity analysis can reveal both types, although it is not always possible to distinguish between the two. Like sensitivity, stability is a function of the operating point; therefore, all possible operating points must be tested for stability. A careful, systematic sensitivity

analysis may often reveal not only points of instability but their causes as well.

Parameter Estimation

The object of parameter estimation (or identification) analysis is to determine the value of one or more parameters in a model. The parameters selected for estimation are usually impossible or difficult to measure directly in the real system. In practice, the technique involves repetitive adjustment of the parameter values until some objective judgment of acceptable correlation between model output and corresponding measurements in the system prototype has been satisfied. (See fig. A-4.) Because the automatic optimization of parameters has been the object of considerable attention (ref. A-10), there is a large body of literature on parameter estimation in physiological systems and algorithms.

The error criterion used in parameter estimation is usually a difference function of the form

$$e(t) = y(t) - y^*(t)$$

where y^* is a dependent variable that has been measured in the real system, y is the corresponding model variable, and t is time. The error criterion E is a function of e, usually $|e|$ or e_2, integrated over a specified time interval; e.g.,

$$E = \int_0^T |e| \, dt$$

Since y is dependent on the system parameters q_i, e can be expressed as $e(t) = e(t, q_1, q_2, ..., q_m)$. The criterion for the best fit between data and model is achieved when E reaches a minimum value.

A more powerful use of sensitivity analysis, but used infrequently, is the determination of parameters that could be estimated most accurately by means of the curve-fitting procedure discussed previously. Parameter estimation is used most effectively on parameters exerting a strong influence on a particular model variable which can be easily measured in the real system. If sensitivity analysis is used before parameter estimation, it is possible to select those parameters with the highest sensitivity coefficients as the best candidates for parameter estimation analysis. When the parameter sensitivity is low, then that parameter value cannot be estimated with certainty using that criterion. The low-sensitivity parameter should be set at a reasonable constant value determined from other sources.

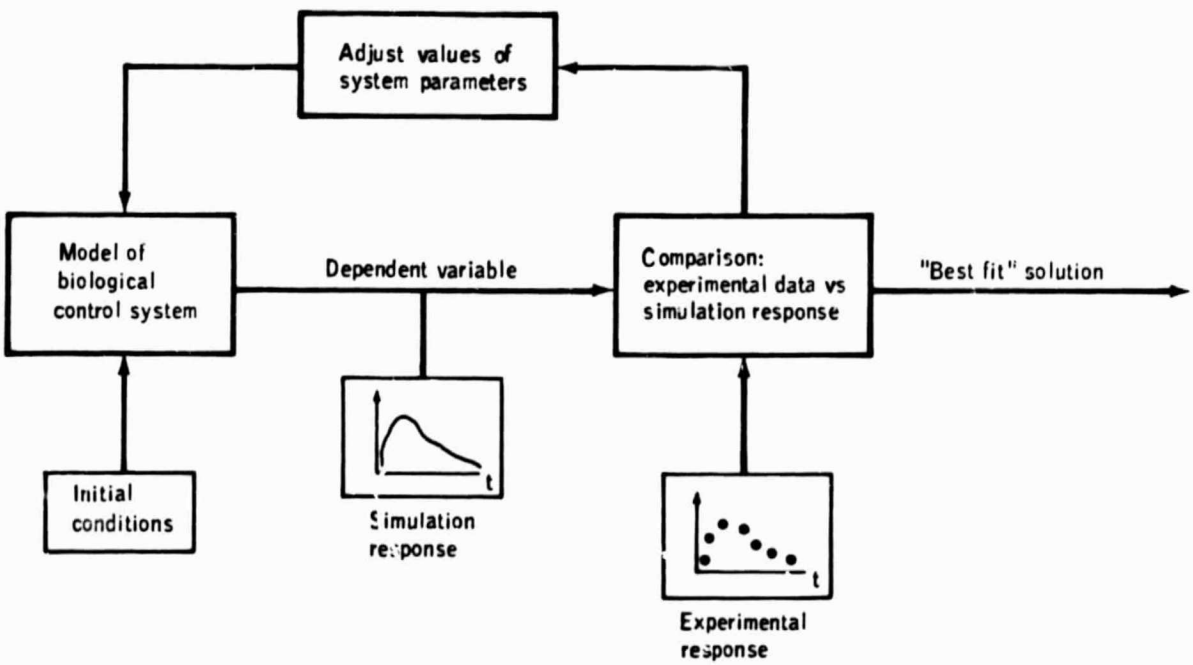


FIGURE A-4.—Simulation procedure for parameter estimation. This method of fitting model output to experimental results can produce values for system parameters that are difficult to measure directly.

If the problem of sensitivity analysis is expressed as determining the behavior of a model given all the parameter variations, then the inverse problem would be to determine (or identify) the parameter variations capable of producing a given behavior of the real system. Unlike sensitivity analysis, variation of parameters, error analysis, and stability analysis, parameter estimation requires data measurements from the real system. This inverse problem may not have a unique solution. Nevertheless, it would be valuable to know the various solutions possible, since this knowledge would be a great aid in hypothesis testing. If several different parameter perturbations could produce similar model results, it might be possible to accept the most reasonable, based on physiological plausibility; alternatively, this information could provide the basis for further experimental testing.

MODEL VALIDATION

The validation process is primarily concerned with demonstrating the accuracy and the capability of simulation models. Two general criteria must be met before model credibility can be established. First, a quantitative variable or parameter validity criterion must be met, and, second, a qualitative "plausibility" criterion must be met. The first condition refers to tests in which model output is compared directly to experimental data, whereas the second condition includes all other tests in which only the general behavior of the model is examined on a more subjective basis. In the first case, a high degree of fidelity in model response is expected, whereas in the second, the model responses need only be "reasonable." No validation procedure is appropriate for all models. Rather, validation depends on the nature of the model and the goals and objectives of the modeling study.

Quantitative Tests

An important aspect of validation involves comparing the behavior of the model's dependent variables with that of their experimental counterparts for the same stress. Differences between model behavior and experimental data can often be corrected or minimized either by introducing new, previously omitted elements into the model's structure or by modifying the existing structure (i.e., adjusting the value of parameters that are not well known).

The extent to which the validation process can be carried is often limited by data availability. Thus, if only steady-state data are available, validation in the dynamic, or transient, mode cannot be performed, even though the model has that capability. Similarly, if only a relatively small number of experimental variables have been measured during a particular stress, then it is possible to validate the model for the responses of only those measured variables; simulated values of all other variables can be obtained but should be considered as predictions requiring experimental verification. The response of the body to a given stress is almost always related to the level or intensity of that stress, and, more often than not, this relationship is nonlinear. Therefore, to validate the model properly, it is desirable to obtain data not merely for a given stress but for a range of intensities of that stress. If it is important to simulate more than one type of stress, the validation process will result in a more accurate model if the experimental response of the same variables is known for each of the desired stresses. Thus, an idealized set of experimental data suitable for complete validation of a complex model should include the following.

1. Steady-state data
2. Transient data
3. Data for a wide range of stress intensities
4. Data for all major dependent variables of interest or importance
5. Data for a variety of stresses

In addition, since experimental protocol, measurement techniques, and number and type of subjects may vary widely from one investigation to another, even when studying the same stress, it is desirable to obtain many of these data from the same experimental study.

The data used to validate the model should meet several conditions.

1. The data used in the verification process (i.e., establishing the correctness of computer coding and ensuring that the model runs as intended) are the same as those used in the development of the model. However, to ascertain model validity, data that were not originally included in the model's formulation must be used. The model must be capable of predicting beyond the data from which the model was generated.

2. The data must be of sufficient precision to make the test meaningful. The data should cover

the range of interest and should have a minimum level of noise. The latter condition is best assured by including data representing a large subject population.

3. The objectives of the modeling studies should be kept in perspective. Not all data are appropriate for use in validation. The assumptions about the biological system being studied are often reflected in the experimental protocol used to gather the data, and these are often not the same assumptions used in the model.

Obtaining good agreement between a simulation response and data is especially important when trying to simulate a diverse number of variables having constantly changing values. Agreement between simulation response and data is often improved by adjusting parameter values. This procedure is the same as that described under parameter estimation techniques. More than one combination of parameter values may lead to a good "fit." In these cases, it is important that all changes be reasonable and consistent with known physiological processes.

Qualitative Tests

The purpose of modeling and simulation, in the context of the current study, is not necessarily to produce an optimal fit between experimental data and model output, although this result would not be undesirable. Rather, the objectives are to help understand the behavior and interactions of the system and its components and to assist in new experimental approaches. Model credibility, in this case, can also be established by performing fewer quantitative tests.

Often, the model analyst is content to verify initially that the "shape" of the data and model output agree. This situation would occur if the modeler were primarily interested in the validation of model dynamics as contrasted to the exact fit of model output and data. In this case, it is not considered critical that the absolute magnitude of the response is in error; this type of discrepancy can often be remedied by the adjustment of a system parameter.

An important criterion for indicating whether a model is good enough to be used for forming conclusions about the real system is that a one-to-one correspondence and similarity of form must exist between model and real system. So-called "black-box models" (i.e., models which represent overall behavior of systems without representing their underlying mechanisms) are not good models for making predictions regarding general system behavior; at best, they may be used as descriptions of data. Thus, there is always some model that can fit a particular set of experimental data (usually by adjusting one or more parameters), but only the biological plausibility of a particular model justifies preferring it to all others. Therefore, the inclusion of a greater number of adjustable parameters in a model, although perhaps providing a better agreement with the data, does not necessarily add insight into the physiological mechanisms.

The techniques described earlier in this appendix—sensitivity analysis, error analysis, stability analysis, and variation of parameters—do not require extensive data sets. They can be very useful in establishing the plausibility of a model without necessitating excessive analysis of the mathematics of the model and all the explicit and implicit assumptions. Sensitivity analysis, in particular, can be important in this regard by quickly and systematically analyzing these component relationships without the need of actual subject data. This capability is particularly useful in comparing two different models.

For a model to contribute to a particular investigative field, it must ultimately be judged by scientists familiar with that area. When these scientists are not the people who developed and validated the model, it is important that lines of communication between them be established as early in the modeling process as possible. Model validation, ideally, should be an interdisciplinary process. During this project, the experience has been that general simulations, sensitivity analyses, or parameter variation studies performed without comparison to good experimental data appear to make less of an impact on experimenters not familiar with systems analysis than the same work supplemented with at least a single simulation showing reasonably good agreement with experimental data.

REFERENCES

A-1. Riggs, D. S.: Control Theory and Physiological Feedback Mechanisms. Williams and Wilkins Co. (Baltimore), 1970.

A-2. Milhorn, H. T., Jr.: The Application of Control Theory to Physiological Systems. W. B. Saunders Co. (Philadelphia), 1966.

A-3. Leonard, J. I.: The Application of Sensitivity Analysis to Models of Large Scale Physiological Systems. (General Electric Co., Houston, Tex.; TIR 741-MED-4028, Contract NAS 9-12932.) NASA CR-160228, 1974.

A-4. Cruz, Jose' B., Jr.: System Sensitivity Analysis. Benchmark Papers in Electrical Engineering and Computer Science. Dowden, Hutchinson, and Ross (Stroudsburg, Pa.), 1973.

A-5. Tomovic, Rajko; and Vukobratovic, M.: General Sensitivity Theory. American Elsevier (New York), 1972.

A-6. Miller, D. R.: Model Validation Through Sensitivity Analysis. Proceedings of the 1974 Summer Computer Simulation Conference, Vol. II, AFIPS Press (Montvale, N.J.), 1974, pp. 911-914.

A-7. Rosen, R.: Dynamical System Theory in Biology, Vol. I: Stability Theory and Its Applications. Biomed Engineering Series, Wiley-Interscience (New York), 1970.

A-8. Jones, R. W.: Principles of Biological Regulation: An Introduction to Feedback Systems. Academic Press (New York), 1973.

A-9. Tomovic, Rajko (David Thornquist, transl.): Sensitivity Analysis of Dynamic Systems. McGraw-Hill (New York), 1963.

A-10. Johnson, L. E.: Computers, Models, and Optimization in Physiological Kinetics. CRC Crit. Rev. Bioeng., vol. 2, Feb. 1974, pp. 1-37.