Sensitivity Analyses of Continuous and Discrete Systems in the Time and Frequency Domains

William J. Karnavas, Paul J. Sanchez, and A. Terry Bahill, *Fellow, IEEE*

Abstract—**The technique of sensitivity analysis is old and well known, but few modern papers include them. Perhaps this is because of the subtle tricks and customizations that have to be done to reap their benefits. The paper shows how to overcome some of the difficulties of performing sensitivity analyses. It draws examples from a broad range of fields: physics, systems theory, physiology, expert systems, bioengineering, control theory, simulation, queueing theory, and system design. The paper generalizes the important points that can be extracted from literature covering diverse fields and long time spans.**

I. Introduction

A SENSITIVITY analysis shows how a model changes with variations in its parameters. The results of a sensitivity analysis can be used to 1) validate a model, 2) warn of strange or unrealistic model behavior, 3) suggest new experiments or guide future data collection efforts, 4) point out important assumptions of the model, 5) suggest the accuracy to which the parameters must be calculated, 6) guide the formulation of the structure of the model, 7) adjust numerical values for the parameters, and 8) allocate resources. The sensitivity analysis tells which parameters are the most important and most likely to affect predictions of the model. Following a sensitivity analysis, values of critical parameters can be refined while parameters that have little effect can be simplified or ignored. If the sensitivity coefficients are calculated as functions of time, it can be seen *when* each parameter has the greatest effect on the output function of interest. This can be used to adjust numerical values for the parameters. The values of the parameters should be chosen to match the physical data at the times when they have the most effect on the output.

Sensitivity functions can be used to set system design specifications. In the manufacturing environment they can be used to allocate resources to critical parts allowing casual treatment of less sensitive parts. In one modeling study the author expended great time computing an optimal control input to minimize a complex performance criterion. A subsequent sensitivity analysis showed that the computed control signal was meaningless, because normal environmental variations in one of the model parameters swamped the effects of variations in the control signal. Systems engineers selecting the best design among several feasible designs do sensitivity analyses to ensure that the decision about the best design is not extremely sensitive to any particular performance requirement.

The earliest sensitivity analyses that we have found are the genetics studies on the pea reported by Gregor Mendel in 1865 (see [1]) and the statistics studies on the Irish hops crops by Gosset writing under the pseudonym Student (see [2]). Since then sensitivity analyses have been used to validate social models [3], engineering models [4], physiological models [5]–[7], numerical computations [8], expert systems [9], [10], and discrete event simulations, where the techniques are called response surface methodology [11], frequency domain experiments [12]–[14], and perturbation analysis [15], [16]. When changes in the parameters cause discontinuous changes in system properties the sensitivity analysis is called that of singular perturbations: Kokotovic and Kahalil [17] include over 60 papers on singular perturbations spanning the 1960's, 70's, and 80's.

Three levels of sensitivity analysis have evolved in the literature. At the first level a sensitivity analysis was preformed and the results were shown. For example the authors took a data-set or a model and carried out a regression analysis or another sensitivity analysis and reported the results [18], [19]. But this type of paper is becoming unusual in engineering journals. The development of models is central to engineering and so the sensitivity analysis of models is appropriate when accompanying model development. For example the authors developed a mathematical model of system and did a sensitivity analysis by plotting pole-zero trajectories produced by parameter changes and tuned the system performance based on the plots [7], [20], [21]. In another example of the first level, the authors developed a mathematical model, reported the sensitivity analysis and used the sensitivity analysis to determine the values of parameters that had little experimental basis [6], [22].

By the second level, a sensitivity analysis of a system or model was no longer of interest, but sophisticated techniques for calculating sensitivities were regularly published. A typical example from electrical engineering shows an improved sensitivity calculation leading to faster solution for switching-mode circuit simulations [23]. In the field of control engineering the efficient calculation of parameter sensitivities of high order linear dynamic systems has been a continuing area of publication with each paper extending the techniques to broader classes of linear dynamic systems. In a good example of this progression, the first paper [24] lays the groundwork

Manuscript received October 21, 1990; revised March 28, 1992 and August 5, 1992.

W. J. Karnavas was with the Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721 and is now with the Department of Neurosurgery, University of Pittsburgh Medcial School, Pittsburgh, PA.

P. J. Sanchez and A. T. Bahill are with the Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721.

IEEE Log Number 9206208.

for determining the sensitivity function of all parameters of a system of linear dynamic equations by running the original system and a perturbed second system of equal size. The technique is then extended to systems with multiple inputs and nonzero initial conditions [25]. Following papers carry out further simplifications in the system requirements and the transformations necessary [26], [27]. Of general interest is a method for determining the sensitivity of the variance of a general mathematical model to the variance of its parameters [28]. Finally, very detailed and sophisticated methods of calculating sensitivities in models with real and imaginary parameters have been presented [29], [30].

At the third level of sensitivity analysis the techniques that use sensitivities or minimize sensitivities to reach some other goal are emphasized. The authors expect the readers to be familiar with sensitivity analyses. For example sensitivity functions have been used in the design of adaptive controllers [31], [32]. And sensitivity analyses have been used as tools in robust quality engineering as championed by Taguchi: his methods have been well explained in English and both the positive and negative aspects of the methods have been reviewed [33], [34]. Taguchi's method follows a fairly simple progression: select a quality measure, choose an objective value for this quality measure and then try to minimize the variability of the objective. This is accomplished by running factorial experiments on the system to determine the sensitivity of the product design and manufacturing design. The dollar value of the variability reduction is a quadratic function of variation from the objective: this is the primary difference with traditional quality control practices that only worry about variability beyond a threshold. A value is given to the variability reduction, or increase in quality, and is used to decide which improvements are worthwhile.

There are many common ways to do sensitivity analyses. A traditional root-locus plot graphically displays the results of a sensitivity analysis: it shows the movement of the system's closed-loop poles as a function of the system's gain. Spread sheets, like Lotus 1-2-3, are convenient for doing sensitivity analyses of systems that are not described by equations. A partial derivative can be a sensitivity function. In a system described by analytic functions, for example, calculating partial derivatives constitutes a sensitivity analysis. This paper explores many techniques for doing sensitivity analyses. The examples in this paper come from a wide variety of fields. The reader may skip unfamiliar examples without loss of continuity.

There are two classes of sensitivity functions: analytic and empirical. Analytic sensitivity functions are used when the system under study is relatively simple, well defined, and mathematically well behaved (e.g., continuous derivatives). These generally take the form of partial derivatives. They are convenient because once derived, they can often be used in a broad range of similar systems and are easily adjusted for changes in other parameters. They also have an advantage in that the sensitivity of a system to a given parameter is given as a function of all the other parameters, including time or frequency (depending on the model of the system), and can be plotted as functions of these variables.

Empirical sensitivity functions are often just point evaluations of a system's sensitivity to a given parameter(s) when other parameters are at known, fixed values. Empirical sensitivity functions can be functions estimated from the point sensitivity evaluations over a range of a parameter(s). They are generally determined by observing the changes in output of a computer simulation as model parameters are varied from run to run. Their advantage is that they are often simpler (or feasible) than their analytic counterparts or can be determined for an unmodeled, physical system. If the physical system is all that is available, the system output is monitored as the parameters are changed from their nominal values.

II. Analytic Sensitivity Functions

This section will explain three different analytic sensitivity functions. They are all based on finding the partial derivative of a mathematical system model with respect to some parameter. A short example is then given for each functional form of sensitivity function.

A. The Absolute Sensitivity Function

The *absolute-sensitivity* of the function F to variations in the parameter α is given by

$$S_{\alpha}^{F} = \left.\frac{\partial F}{\partial \alpha}\right|_{\text{NOP}}$$

where NOP means the partial derivative is evaluated at the normal operating point (NOP) where all the parameters have their nominal values. In this paper the function F may also be a function of other parameters such as time, frequency, or temperature. Absolute-sensitivity functions are useful for calculating output errors due to parameter variations and for assessing the times at which a parameter has its greatest or least effect. They are also an important part of adaptive control systems (this use of sensitivity functions is illustrated in [31], [35] but is not addressed in the present paper because of its complexity). The absolute-sensitivities of the optimal solution of a linear programming (LP) problem to its constraints (some of its parameters) are given by the values of the dual variables associated with each constraint. These values are referred to as the marginal value or shadow price of each constraint and are interpreted as the increase in the optimal solution obtainable by each additional unit added to the constraint [36].

The following two examples show the use of absolute-sensitivity functions. The first shows how to use an absolute-sensitivity function to calculate output errors due to parameter variations. And the second shows how to use an absolute-sensitivity function to see when a parameter has its greatest effect.

Example 1: A pendulum clock (based on [4]).

The period of oscillation of a pendulum clock is given by $P = 2\pi\sqrt{l/g}$ where l is the length of the pendulum and g is the gravitational acceleration, 9.8 m/s^2. From this equation it can be seen that a one meter pendulum will have the typical 2-s period. Assume that a clock keeps good time at the nominal operating point, i.e., at temperature T_0 where the length is l_0. However, if the temperature increases so that

$T = T_0 + \Delta T$ then the length becomes $l = l_0(1 + k\Delta T)$ where $k = 2 \times 10^{-5}/°C$ for a brass rod. Let us use the absolute-sensitivity function of P with respect to T, S_T^P, to calculate how many seconds per day the clock will lose if the temperature rises by 10°C. First calculate the sensitivity function, then evaluate it at the nominal operating point, and finally multiply it by the change in the parameter value to get the change in the output value:

$$S_T^P = \left.\frac{\partial 2\pi\sqrt{l/g}}{\partial T}\right|_{\mathrm{NOP}} = \left.\frac{\pi}{\sqrt{g}} \frac{l_o k}{\sqrt{l_o(1 + k\Delta T)}}\right|_{\mathrm{NOP}}$$

at the nominal operating point $\Delta T = 0$ so

$$S_T^P = k\pi\sqrt{l_0/g} = 2 \times 10^{-5}\,\mathrm{s}/°\mathrm{C}.$$

Therefore, the change in the period

$$\Delta P = S_T^P \Delta T = k\pi\sqrt{l_o/g}\Delta T = 2 \times 10^{-4}\,\mathrm{s}.$$

So the number of seconds lost per day will be 2×10^{-4} s/period (0.5 period/s) (60 s/min) (60 min/hr) (24 hr/day) = 8.64 seconds per day.

Example 2: A single-pole system with a time-delay. Next we will show an example of using an absolute-sensitivity function to determine when a parameter has its greatest effect. For this example we will use the transfer function

$$M(s) = \frac{Y(s)}{R(s)} = \frac{Ke^{-\theta s}}{\tau s + 1}$$

and find when the parameter K has its greatest effect on the step response of the system. The step response of the system is

$$Y_{sr}(s) = \frac{Ke^{-\theta_s}}{s(\tau s + 1)}$$

and the absolute-sensitivity function of the step response with respect to K is

$$S_K^{Y_{sr}}(s) = \frac{e^{-\theta_0 s}}{s(\tau_0 s + 1)},$$

which transforms into

$$S_K^{y_{sr}}(t) = 1 - e^{-(t-\theta_0)/\tau_0}, \qquad \text{for } t > \theta_0.$$

This tells us that the parameter K has its greatest effect when the response reaches steady state, which is what our intuition also tells us.

B. The Relative Sensitivity Function

If we want to compare the effects of different parameters we should use relative-sensitivity functions. The *relative-sensitivity* of the function F to the parameter α evaluated at the nominal operating point is given by

$$\overline{S}_\alpha^F = \frac{\%\,\text{change in } F}{\%\,\text{change in } \alpha} = \frac{\partial F/F}{\partial \alpha/\alpha} = \left.\frac{\partial F}{\partial \alpha}\right|_{\mathrm{NOP}} \frac{\alpha_0}{F_0}$$

where NOP and the subscripts 0 mean that all functions and parameters assume their nominal operating point values [37]. Relative-sensitivity functions are formed by multiplying the partial derivative (the absolute-sensitivity function) by the nominal value of the parameter and dividing by the nominal value of the function. Relative-sensitivity functions are ideal for comparing parameters, because they are dimensionless, normalized functions. In the field of economics the lack of dimensions of the relative-sensitivity function is exploited to allow easy comparison of parameters' changes on model outputs even though the parameters may describe widely varying aspects of the model and have different units. Economists refer to the relative sensitivity function $\overline{S}_A^B$ as the elasticity of B with respect to A, and denote it as E_B, A [38].

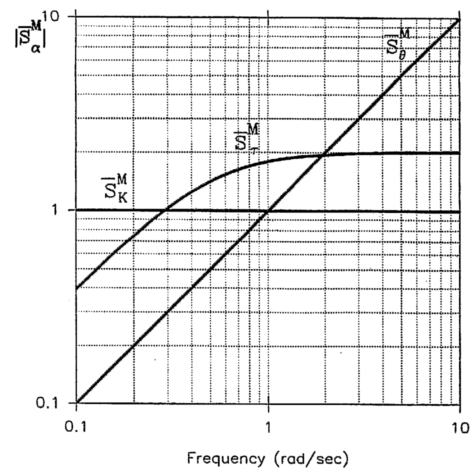


Fig. 1. Relative-sensitivity functions as a function of frequency for the transfer function $M(s) = 32e^{-1.0s}/(2s + 1)^2$.

Example 3: A double-pole system with a time-delay (based on [39]). Let us now show how to use relative-sensitivity functions to compare parameters. Consider the transfer function

$$M(s) = \frac{Ke^{-\theta s}}{(\tau s + 1)^2}.$$

Which of the parameters is most important? To answer this question let us compute the relative-sensitivity functions:

$$\overline{S}_K^M = \left. S_K^M \right|_{\mathrm{NOP}} \frac{K_0}{M_0} = 1$$

$$\overline{S}_\theta^M = \left. S_\theta^M \right|_{\mathrm{NOP}} \frac{\theta_0}{M_0} = -s\theta_0$$

$$\overline{S}_\tau^M = \left. S_\tau^M \right|_{\mathrm{NOP}} \frac{\tau_0}{M_0} = \frac{-2s\tau_0}{\tau_0 s + 1}.$$

The magnitudes of these three relative-sensitivity functions are plotted in Fig. 1 using their nominal values of 32, 1, and 2 for K_0, θ_0, and τ_0, respectively (from a crayfish model in [39]). By looking at these plots we can see that

- for low frequencies, e.g., $\omega < 0.3, K$ has largest magnitude,
- for midrange frequencies, e.g., $\omega \approx 1, \tau$ has largest magnitude,
- and for high frequencies, e.g., $\omega > 2, \theta$ has largest magnitude.

Now, of course, this was a simple example. And for this example, most people would intuitively say that the gain is the most important parameter for low frequencies (i.e., steady state) and the time-delay is the most important parameter for high frequencies. But it was still nice that this sensitivity analysis gave us a quantitative justification for our intuition.

Example 4: An expert system (from [10]). Most sensitivity functions are functions of time or frequency. But this is not mandatory. For example, let us compute the relative-sensitivity functions for an expert system that uses Mycin style certainty factors [9]. Assume we have a rule of the following form.
If

$$\text{premise1} = \text{true}\,(CF_1)$$

and

$$\text{premise2} = \text{true}\,(CF_2)$$

or

$$\text{or premise3} = \text{true}\,(CF_3)$$

then

$$\text{conclusion} = \text{true},\ CF_4.$$

In a Mycin style system the certainty of the rule involves the minimum of the certainties of the AND clauses. Therefore, only one of CF_1 and CF_2 can have an effect on the final certainty CF_F. That is, if $CF_1 < CF_2$ then $\overline{S}_{CF_2}^{CF_F} = 0$. Whereas if $CF_1 > CF_2$ then $\overline{S}_{CF_1}^{CF_F} = 0$. And if $CF_1 = CF_2$ then $\overline{S}_{CF_1}^{CF_F} = S_{CF_2}^{-CF_F}$. Therefore, for simplicity, we will assume that $CF_1 < CF_2$ and that changes are small enough so that this inequality is not violated. The final certainty factor becomes

$$CF_F = \frac{CF_1 CF_4}{100} + \left(1 - \frac{CF_1 CF_4}{100}\right)\frac{CF_3 CF_4}{10^4}.$$

The relative-sensitivity functions are

$$\begin{aligned}
\overline{S}_{CF_1}^{CF_F} &= \left[1 - \frac{CF_3 CF_4}{10^4}\right]\frac{CF_4}{100}\Bigg]_{\text{NOP}}\frac{CF_{1_0}}{CF_{F_0}}\\
\overline{S}_{CF_2}^{CF_F} &= 0\\
\overline{S}_{CF_3}^{CF_F} &= \left[1 - \frac{CF_1 CF_4}{10^4}\right]\frac{CF_4}{100}\Bigg]_{\text{NOP}}\frac{CF_{3_0}}{CF_{F_0}}\\
\overline{S}_{CF_4}^{CF_F} &= \left[\frac{CF_1}{100} + \frac{CF_3}{100} - \frac{2CF_1 CF_3 CF_4}{10^4}\right]_{\text{NOP}}\frac{CF_{4_0}}{CF_{F_0}}.
\end{aligned}$$

It is now time to plug in some numbers and see what we can learn. Assume $CF_2 = CF_3 = CF_4 = 80$ and $CF_1 = 79.9$. Then from the previous equation the nominal value of the final certainty becomes

$$CF_{F_0} = 87$$

and

$$\begin{aligned}
\overline{S}_{CF_1}^{CF_F} &= 0.26\\
\overline{S}_{CF_2}^{CF_F} &= 0\\
\overline{S}_{CF_3}^{CF_F} &= 0.26\\
\overline{S}_{CF_4}^{CF_F} &= 0.53.
\end{aligned}$$

This means that changes in CF_4 are twice as important as changes in either CF_1 or CF_3. And that increases in CF_2 will have no affect on the final certainty. In general this is strange: changing the value of a particular certainty factor may produce no effect on the output for reasons that are unrelated to that particular certainty factor.

Certainty factors are usually assigned by the domain expert based on gut level feelings. Often experts say this is difficult and frustrating. After the certainty factors are all assigned, the knowledge engineer often tweaks the certainty factors looking for changes in output certainty. Note that this technique, in general, will not work. For example, if there is a rule with many conjunctive premises (AND clauses), the certainty factor of only one of them will affect the certainty of the output.

What if the certainty factors are not all equal? What is the effect of their magnitude on their sensitivities? Let $CF_1 = 50, CF_2 = 60, CF_3 = 70$, $CF_4 = 80$, and $CF_{F_0} = 74$. The sensitivities become

$$\begin{aligned}
\overline{S}_{CF_1}^{CF_F} &= 0.24\\
\overline{S}_{CF_2}^{CF_F} &= 0\\
\overline{S}_{CF_3}^{CF_F} &= 0.46\\
\overline{S}_{CF_4}^{CF_F} &= 0.70.
\end{aligned}$$

This means that the premises with bigger certainty factors are more important, e.g., changing the certainty of premises 1 would have about half the effect of changing the certainty of premise 3.

In summary, the certainty factor assigned by the knowledge engineer to the conclusion of a rule, is more important than the certainty factors derived for the premises during the consultation. Clauses with larger certainty factors are more important than clauses with smaller certainty factors. The final certainty, the output of the system, is often completely insensitive to many premises.

C. Limitations of the Relative Sensitivity Function

The relative-sensitivity function is limited in usefulness in analytic studies since it has different meanings in the time and frequency domains. This is the result of the relative-sensitivity function being the product of two functions (the partial derivative and the original function), and the Laplace transform of a product is not the product of the Laplace transforms, e.g., $L[x(t)\times y(t)] \neq Lx(t)\times Ly(t)$. Therefore, the frequency domain relative-sensitivity function is often difficult to compute in the time domain (requiring convolutions) and a similar problem arises in computing the time domain relative-sensitivity function in the frequency domain. One way to overcome this difficulty is to define two different relative-sensitivity functions; one for the time domain and one for the frequency domain.

The time-domain relative-sensitivity function of $f(t)$ for the parameter α is defined to be

$$\overline{S}_{\alpha}^{f(t)} = \frac{\partial f(t)/f(t)}{\partial\alpha/\alpha} = \frac{\partial f(t)}{\partial\alpha}\Bigg]_{\text{NOP}}\frac{\alpha_0}{f(t)_0}$$

where NOP and the subscripts 0 mean that all functions and parameters assume their nominal operating point values. The frequency-domain relative-sensitivity function (sometimes called the Bode Sensitivity Function [4]) of $F(s)$ for the parameter α is defined to be

$$\hat{S}_{\alpha}^{F(s)} = \frac{\partial F(s)/F(s)}{\partial \alpha/\alpha} = \left.\frac{\partial F(s)}{\partial \alpha}\right|_{\mathrm{NOP}} \frac{\alpha_0}{F(s)_0}$$

where NOP and the subscripts 0 mean that all functions and parameters assume their nominal operating point values.

These two relative-sensitivity functions produce different results. For example, consider a simple closed-loop system with unity feedback and $G = K/s$. The closed-loop transfer function is

$$M(s) = \frac{Y(s)}{R(s)} = \frac{K}{s+K}.$$

The time-domain step response of the system is

$$y_{sr}(t) = 1 - e^{-Kt}$$

where the subscript sr stands for step response. The time-domain relative-sensitivity function of the step response with respect to the gain K is

$$\overline{S}_{K}^{y_{sr}} = te^{-Kt}\rfloor_{\mathrm{NOP}} \frac{K_0}{1-e^{-K_0 t}} = \frac{K_0 t e^{-K_0 t}}{1-e^{-K_0 t}}.$$

For most systems the time-domain behavior is the most important. Usually we only use the frequency domain to help understand the time domain behavior. However, let us now calculate the frequency-domain relative-sensitivity function of the step response $Y_{sr}(s)$ with respect to the gain K. First,

$$Y_{sr}(s) = \frac{K}{s(s+K)}.$$

Then

$$\hat{S}_{K}^{Y_{sr}}(s) = \frac{s}{s+K_0}.$$

To make it obvious that this is different from the time-domain relative-sensitivity function of the step response of the system, let us take the inverse Laplace transform of this function to get

$$\hat{S}_{K}^{Y_{sr}}(t) = \delta - K_0 e^{-K_0 t}$$

where δ represents the unit impulse. As can be seen from Fig. 2 this is not the same at the time-domain relative-sensitivity function of the step response of the system. This difference between the time and frequency domain functions is one of the reasons for using the semirelative-sensitivity function.

D. The Semirelative Sensitivity Function

We have used the absolute-sensitivity function to see when a parameter had its greatest effect on the step response of a system, and we have used the relative-sensitivity function to see which parameter had the greatest effect on the transfer function. Now suppose we wish to compare parameters, but we want to look at the step response and not the transfer function. What happens if we try to use the relative-sensitivity function to compare parameters effects on the step response? We get into trouble. For a step response the nominal output value y_0 varies from 0 to 1 and division by zero is frowned upon. Furthermore, the relative-sensitivity function

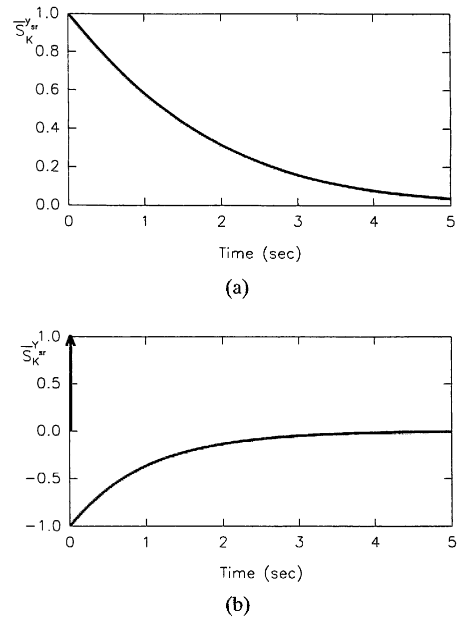


Fig. 2. Time-domain relative-sensitivity function with respect to the gain K for the step response of a closed-loop negative-feedback control system with $K = H = 1$ and $G = 1/s$ (a), and the inverse Laplace transform of the frequency-domain relative-sensitivity function with respect to the gain K for the step response of the same system (b).

gives undue weight to the beginning of the response when y_0 is small. Therefore, let us investigate the use of the *semirelative-sensitivity function,* which is defined as

$$\tilde{S}_{\alpha}^{F} = \left.\frac{\partial F}{\partial \alpha}\right|_{\mathrm{NOP}} \alpha_0$$

where NOP and the subscript 0 mean that all functions and parameters assume their nominal operating point values.

As can be seen by the definition, semirelative-sensitivity functions will have the same shape as absolute-sensitivity functions. They are just multiplied by the constant parameter values. But this scaling allows comparisons to be made of the effects of the various parameters.

Example 2 (revisited): A single-pole system with a time-delay. Let us use the same example we used for the absolute-sensitivity function. Namely

$$M(s) = \frac{Y(s)}{R(s)} = \frac{Ke^{-\theta s}}{(\tau s+1)}.$$

The step response becomes

$$Y_{sr}(s) = \frac{Ke^{-\theta s}}{s(\tau s+1)}.$$

Now calculating the absolute-sensitivity function we get

$$S_{K}^{Y_{sr}} = \frac{e^{-\theta_0 s}}{s(\tau_0 s+1)}.$$

Therefore the semirelative-sensitivity function becomes

$$\tilde{S}_{K}^{Y_{sr}} = \left.S_{K}^{Y_{sr}}\right|_{\mathrm{NOP}} K_0 = \frac{K_0 e^{-\theta_0 s}}{s(\tau_0 s+1)}$$

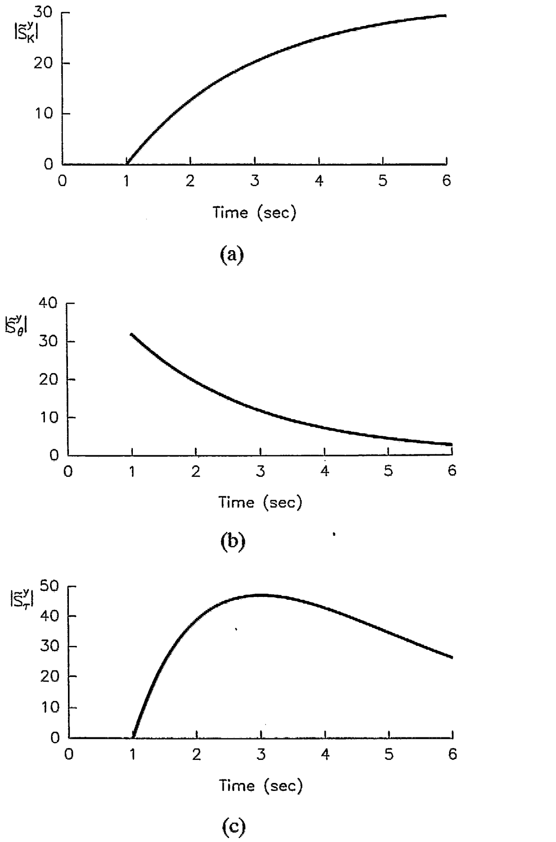


Fig. 3. Semirelative-sensitivity functions with respect to K, θ, and τ for $M(s) = 32e^{-s}/(2s+1)$.

which is the same as the step response given previously. This transforms to

$$\tilde{S}_K^{y_{sr}}(t) = K_0(1 - e^{-(t-\theta_0)/\tau_0}) \qquad \text{for } t > \theta_0.$$

Second:

$$\tilde{S}_\theta^{Y_{sr}} = \frac{-K_0\theta_0 e^{-\theta_0 s}}{\tau_0 s + 1)}$$

which transforms to

$$\tilde{S}_\theta^{y_{sr}}(t) = -K_0\theta_0 e^{-(t-\theta_0)/\tau_0} \qquad \text{for } t > \theta_0.$$

Finally,

$$\tilde{S}_\tau^{Y_{sr}} = \frac{-K_0\tau_0 e^{-\theta_0 s}}{(\tau_0 s + 1)^2}$$

which transforms to

$$\tilde{S}_\tau^{y_{sr}}(t) = -K_0\tau_0(t - \theta_0)e^{-(t-\theta_0)/\tau_0} \qquad \text{for } t > \theta_0.$$

These three semirelative-sensitivity functions are plotted in Fig. 3 using the nominal values previous used, namely 32, 1, and 2 for K_0, θ_0, and τ_0, respectively. This sensitivity analysis gives us a wealth of information. It says that if our model does not match the physical response early in the rise, then we should adjust the time-delay of the model; because in the beginning the time-delay has its greatest effect, but the sensitivity functions of the gain and the time constant are still zero. If we wish to affect the steady state behavior of the model, then we should change the gain, because the effects of the time-delay and the time constant will have decayed to zero. It also tells us that the time constant will have its greatest effect right in the middle of the movement. Sometimes a sensitivity analysis like this will show parameters that have their peaks at the same time, these parameters can be treated as a group with tradeoffs between their individual values.

Once again the results of the sensitivity analysis agree with our intuitions: the time-delay has its greatest effect in the beginning of the movement, the time constant has its greatest effect in the middle of the movement, and the gain has its greatest effect at the end of the movement.

III. Empirical Sensitivity Functions

For many models, calculating analytical sensitivity functions is difficult or impossible, so the sensitivity functions are derived empirically from models. This procedure is called simulation, and usually takes the form of "running" a mathematical model of a system on a computer. Various methods are deployed to vary parameters across or during runs of the simulation, so that the output(s) of the system will yield useful information about how the parameters affect the system output or performance. In the Direct Observation approach (employed in factorial type experiments), in different runs of the model each parameter is given a different value. A more sophisticated method is Sinusoidally Varied Parameters, which attempts to determine (or at least identify) the effects of many parameters in a single run.

Direct Observation can be used to estimate any of the previously defined sensitivity functions by properly choosing the simulation and parameters varied. The absolute sensitivity of the period of a pendulum with respect to temperature (an extension of the first example) is calculated in the first example of the next section while the semirelative sensitivity of an eye movement with respect to one specific parameter is shown in the second example of the next section. Both simulations were performed by running the model at 1) the nominal operating point, and 2) with slightly higher parameter values. In this case, the sensitivity determined is valid only near the nominal operating point. By using many parameter values from the range of interest for the runs, an equation for the sensitivity over a range of the parameter may be estimated. This defines a surface for the response of the system with respect to its parameter(s) and is consequently referred to as response surface methodology (RSM) [11]. One problem with using Direct Observation in practice is that the number of runs increases geometrically with the number of factors being studied.

A second empirical sensitivity technique has grown out of the complexity of determining a multidimensional response surface for nondeterministic systems. This method uses one or few simulations of a stochastic system where the parameters of interest have been sinusoidally modulated during the simulation. By the proper choice of parameters and modulation frequencies, a qualitative measure of the effects of parameter variations can be seen by analysis of the simulation output(s). A simple transfer function and a well studied nondeterministic system will be presented.

A. Direct Observation of Simulations

The Direct Observation method is easy to use and evaluate. For many systems it is the only alternative if the analytic sensitivity functions cannot be calculated. In nondeterministic applications where the model is small and parameter interactions are not a concern, it is probably the best empirical method of evaluating sensitivity of a system to parameter changes.

The primary objective of Direct Observation is to eliminate the analytic evaluation of the (partial) derivatives of the three previously defined sensitivity functions. This is done by referring back to the derivative. By definition,

$$\frac{\partial F}{\partial \alpha} = \lim_{h \to 0} \frac{F(\alpha + h) - F(\alpha)}{h}.$$

This derivative can be estimated by

$$\frac{\partial F}{\partial \alpha} \approx \frac{F(\alpha + h) - F(\alpha)}{h}.$$

If $\partial F / \partial \alpha$ is fairly constant around the nominal operating point, then this approximation is not only good, but valid over a wide range. However, if the function F has a nonconstant slope around α, then h must be small. And if F has a discontinuity, then the approximation may be terrible.

Example 1 (continued): The pendulum clock. In this case we wish to determine the sensitivity of a pendulum clock to temperature by Direct Observation (a simulation). As before, we will use the sensitivity of the period with respect to temperature S_T^P, but this time we will approximate the partial derivative with respect to temperature, with the difference in period due to a small temperature change divided by the small temperature change, so

$$S_T^P = \frac{P\rfloor_{\mathrm{NOP}+\Delta T} - P\rfloor_{\mathrm{NOP}}}{\Delta T}.$$

This is a good approximation for small changes in temperature. If the values for all parameters are substituted into the equation for the period at the nominal temperature, the period is found to be 2.00709 seconds. Increasing the length of the pendulum by 2×10^{-5} for a 1°C increase for a 1 meter rod, the period becomes 2.00711 seconds. Substituting these numbers into the previous equations yields

$$S_T^P = \frac{2.00711 - 2.00709\,\mathrm{s}}{1°\mathrm{C}} = 2 \times 10^{-5}\,\mathrm{s}/°\mathrm{C}$$

which is the same as the analytic answer. This same answer could also have been found by experimenting with the actual pendulum, but it would require measuring time with five decimal places of precision. This answer does not lead the user to know over what temperature ranges or changes it is valid, or if it is transferable to different lengths pendulum, etc.

Example 5: A bioengineering model (from [6]). Fig. 4 shows the results of such a sensitivity study of the linear homeomorphic model for human eye movements [6]. This model was a linear sixth-order biomechanical model, with 18 parameters specifying force generators, springs, masses, and dashpots. So many parameters were needed because the model was homeomorphic, that is, each parameter corresponded to a physical component, or the results of a particular physiological experiment. The high order and large number of parameters, made it quite difficult to compute the sensitivity functions. So they were derived by running the model. For small parameter changes in this linear system the semirelative-sensitivity function became

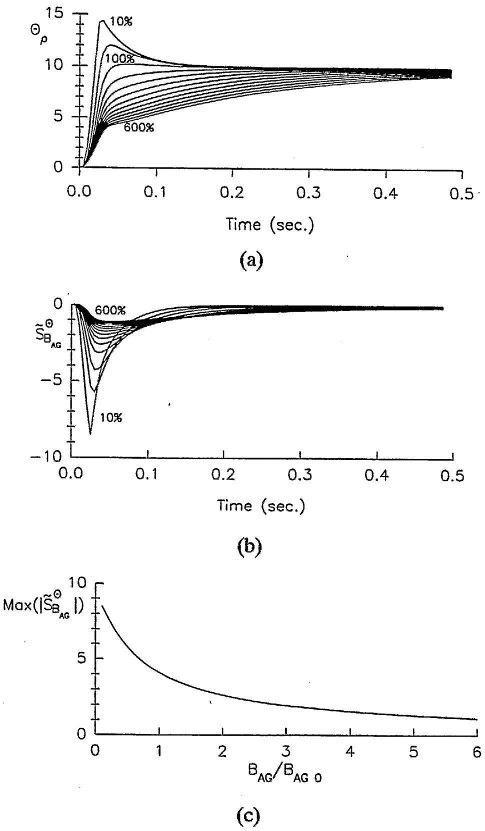


Fig. 4. Output of the linear homeomorphic model for human eye movements for a 10 degree saccade as a function of time (a), the semirelative-sensitivity function of the output with respect to the parameter B_{AG} (b), and the maximum value of this sensitivity function as a function of the value of B_{AG} (c).

$$\tilde{S} = \frac{\Delta y}{\Delta \alpha}\alpha_0.$$

To perform the sensitivity analysis, a ten-degree movement was simulated (line labeled 100% in the top graph of Fig. 4). Then, one parameter was changed, in this case a dashpot, B_{AG}, and the model was run again, producing a perturbed movement (one of the other lines in Fig. 4, each for a different value of B_{AG}). To compute the sensitivity function, the difference between the nominal and perturbed movements (Δy) was calculated for each millisecond and this difference was divided by the change in the parameter value ($\Delta\alpha$). This ratio was then multiplied by the nominal parameter value (α_0). The center graph of Fig. 4 shows the sensitivity functions corresponding to the lines in the top graph. This process was repeated for each of the 18 parameters in the model. The bottom graph of Fig. 4 shows the maximum value of the semirelative sensitivity of θ with respect to B_{AG} when this parameter took on values of 10%, 50%, 100%, 150%, 200%, 250%, 300%, 350%, 400%, 450%, 500%, 550%, and 600% of its nominal value.

Although the model was linear this does not mean that the sensitivity functions were linear. In fact for real world models, the peak value of any sensitivity function is seldom a linear function of the size of the parameter perturbation. Only one of the 18 parameters in this model showed this type of linearity.

Bahill *et al.* [6] used only one perturbation size, +5%, justifying this with the fact that the model was linear. However a linear system is defined as one where superposition holds for the inputs. Superposition states if input X_1 gives output Y_1, and input X_2 gives output Y_2, then combining the inputs yields a similar combination of the outputs, or $aX_1 + bX_2$ gives output $aY_1 + bY_2$. It does not say that doubling a system parameter doubles its effect on the output. Fig. 4 shows that it does not. The sensitivity of the output θ to the parameter B_{AG} decreases with increasing B_{AG} as can be seen in the top graph of Fig. 4 where the output curves get closer together and in the bottom graph of Fig. 4.

The peak value of the sensitivity function of the output with respect to B_{AG}, a dashpot, decreases in magnitude and moves to the right in time as the perturbation gets bigger. This makes intuitive sense. A dashpot has its greatest effect when the velocity is the greatest, which is just about in the middle of the movement. However, as the value of the dashpot gets bigger the peak velocity of the movement decreases and moves to the right. Therefore, the time when this parameter has its greatest effect also shifts to the right. Also, with a smaller peak velocity the effect of the dashpot is smaller and the magnitude of the sensitivity function decreases.

The results of this sensitivity analysis allowed the authors to discover which parameters were important and which were not. Consequently they devoted their modeling efforts to the important parameters, and treated the nonimportant parameters simply. This sensitivity analysis also showed that several elements had their peak sensitivities at the same point in time, so tradeoffs could be made between these parameters without forcing changes to be made in other parameters.

Example 4 (continued): An expert system.

The parameters of an expert system knowledge base can be varied and the output behavior observed. When this was done in [10] it was found that changing some attributes changed the output: however, changing other attributes did not change the output. When this analysis was presented to the expert, she noticed that one attribute had no effect on the output. Therefore, she recommended that we delete this attribute from the knowledge base. This simplified the expert system proving that this sensitivity analysis was valuable. The problem with this technique is that there is no good way to evaluate the results of such an sensitivity analysis. The analysis is ad hoc, hit or miss.

B. Sinusoidal Variation of Parameters

Direct Observation works well with small systems, but when the number of parameters, test values, or sets of possible parameter interactions becomes large, the number of necessary runs of the model becomes impractical. The number of necessary runs increases combinatorially as the number of parameters and interactions increases. For example, in a five parameter system, trying five values per parameter would require 25 runs for the simple case. If pairwise interactions are also considered, the required number of runs jumps to

$$25 + 5\binom{5}{2} = 75.$$

Compounding the problem for nondeterministic models is the need to make multiple runs (or very long runs) to get reasonable estimates of the output for each (set of) parameter setting(s). The computational difficulty can become prohibitive.

To ameliorate this problem, the parameters can be modulated sinusoidally during a simulation so that the sensitivity to many parameters and their interactions can be evaluated simultaneously. This technique is usually called Sinusoidal Variation of Parameters; in the simulation literature it is called Frequency Domain Experiments [12]–[14].

In this approach two runs of the simulation are performed. The first run has all parameters set at their nominal values. A second run is then performed with each parameter modulated sinusoidally throughout the simulation. The output sequences of the simulations are then transformed into a frequency domain representation (e.g., a power spectrum). The two spectra are compared by taking their ratios at each frequency. Differences that are attributable to factor modulation will manifest as ratios that are greater than one. Differences will be observed at the modulation frequencies and at frequencies related to nonlinear and product effects. This analysis yields an estimate of the importance of each parameter.

So far Sinusoidal Variation of Parameters has been used primarily with nondeterministic systems [12]–[14]. However, it can be used with deterministic systems to evaluate the effects of parameters on the steady state response to noisy (random) inputs and periodic inputs when the periodic input is considered in the output analysis. The output of such studies is often used to guide the design of further simulation experiments using Direct Observation by specifying the interactions and maximum order of parameter effects requiring full analysis.

Example 6: A negative feedback loop around a single-pole. We created a simulation of the simple first order element $G(s) = K/s + A$ in a negative feedback gain H. The simulation was carried out by numerical solution of the differential equation (in the time domain) of the total system

$$M(s) = \frac{Y(s)}{R(s)} = \frac{K}{s + A + HK}$$

using Runge–Kutta integration techniques. For all simulations we used an input signal of a one-hertz (Hz), unit-amplitude sinusoid. The simulation was one second long and consisted of 2048 evenly spaced samples. This can be analyzed for a spectrum with 1 Hz resolution from 0 to 1024 Hz. The parameters were modulated at 5, 30, and 170 Hz. These frequencies were chosen 1) to be away from zero, 2) to go through several cycles during the simulation, and 3) to be spread far enough apart so that the product terms and higher order terms would not conflict. In this type of sensitivity study if a parameter is modulated at ω_a we expect to see power at ω_a. If there is a parabolic nonlinearity we also expect to see power at $2\omega_a$. (Because of the trigonometric identity $2\sin^2 x = 1 - \cos 2x$.) Finally, if the function of interest

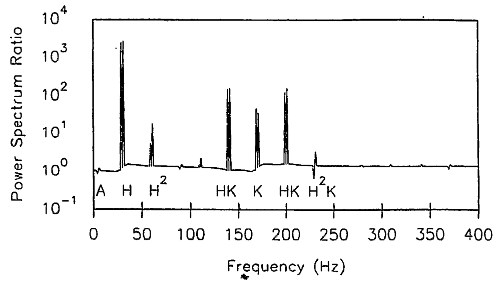


Fig. 5. Ratio of power spectra from $M(s) = K/(s + A + HK)$ when the parameters A, H, and K were modulated sinusoidally.

is sensitive to the product of two parameters modulated at ω_a and ω_b, then we also expect power at $\omega_a \pm \omega_b$. (Because $2 \sin x \sin y = \cos(x-y) - \cos(x+y)$.) These product terms are often called interactions. One final consideration was that we wanted integer frequencies so that an FFT would yield sharp frequency peaks without the use of windowing functions, which tend to blur the spectrum.

The simulation was run twice. The first run was with three parameters, A, K, and H, set to their nominal values of 0.1, 1, and 50, respectively, without modulation: this was the baseline run. In the second run, the signal run, the parameters were modulated according to the following equations.:

$$\begin{aligned} A &= 0.1(1 + 0.5 \sin(5 \times 2\pi t)) \\ K &= 1(1 + 0.5 \sin(170 \times 2\pi t)) \\ H &= 50(1 + 0.5 \sin(30 \times 2\pi t)). \end{aligned}$$

The power spectrum of the output sequence of each run was estimated by the squared magnitude of an FFT. A plot was made by taking the ratio at each frequency of the signal spectrum to the baseline spectrum. A plot of this spectrum ratio is shown in Fig. 5.

Looking at this plot of the spectrum ratio, we see peaks corresponding to the modulation frequencies plus and minus 1 Hz. The dual peaks are due to the modulation effect of the 1 Hz input sine wave on the modulation frequency of each parameter. The peaks around 5 Hz, due to parameter A, are small. The highest peaks appear around 30 Hz, the modulation frequency of the parameter H. In addition to the three primary effects around 5, 30, and 170 Hz, there are significant product effects, because the transfer function contains the product HK. These are the peaks around the frequencies that are the sum and difference of the modulation frequencies of H and K, e.g., 140 and 200 Hz. These product effects are even more important than the direct effect of K. The plot also shows that the parameter H has substantial second order effects. These are the peaks at 60 Hz, which is twice the modulation frequency of H. The second order effect of H also interacts with K, these are the peaks at twice the modulation frequency of H plus and minus the modulation frequency of K, i.e., 110 and 230 Hz. The first order and product effects in the simulations show what we expected from looking at the transfer function and the parameter values. The higher order effects of the parameters are a little surprising at first. However, they can be explained by remembering that the inverse Laplace transform of the simulated transfer function is

$$\begin{aligned} m(t) &= Ke^{-(A+KH)t} \\ &= K\left[1 - (A + KH)t + \frac{(A+KH)^2 t^2}{2} - \cdots\right]. \end{aligned}$$

In this simulation the higher order terms did show significant output.

Note that we can obtain similar information about the sensitivities of the transfer function with respect to H, K, and A from the semirelative-sensitivity functions given here.

$$\begin{aligned} \tilde{S}_K^M &= \frac{K_0(s + A_0)}{(s + A_0 + K_0 H_0)^2} = \frac{(s + 0.1)}{(s + 50.1)^2} \\ \tilde{S}_H^M &= \frac{-K_0^2 H_0}{(s + A_0 + K_0 H_0)^2} = \frac{-50}{(s + 50.1)^2} \\ \tilde{S}_A^M &= \frac{-K_0 A_0}{(s + A_0 + K_0 H_0)^2} = \frac{-0.1}{(s + 50.1)^2}. \end{aligned}$$

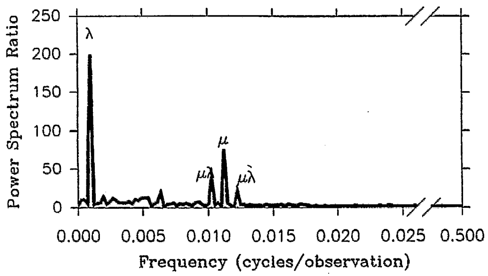
First of all, just as in Fig. 5, it shows that the transfer function is not very sensitive to A. Next comparing the sensitivities of the transfer function to H and K shows the well known fact that negative feedback transfers sensitivity from the plant (K), which may be big, dirty, and inflexible, to the feedback elements (H), which are precise computer components. Looking at these sensitivity functions also shows why the product effects are called interactions in the sinusoidal variation of parameters literature: $\tilde{S}_K^M$ depends on the value of H and $\tilde{S}_H^M$ depends on the value of K. We note once again that superposition does not hold for a sensitivity analysis. A 10% change in both K and H will not be the same as a 10% change in H plus a 10% change in K.

Example 7: An M/M/1 queue. The waiting time in an M/M/1 queue was analyzed with respect to the arrival and service rates using Sinusoidal Variation of Parameters. This is a simple stochastic simulation and has been examined thoroughly in analytic and simulation studies and in fact it is often used as a benchmark problem. The analytic result is well known, so this sensitivity analysis provides us with a chance to look at an old problem in a new light.

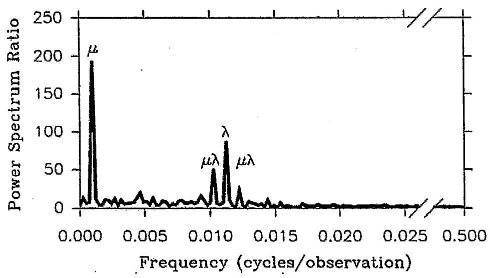
We chose the waiting time in the service queue as the output of interest. The waiting time was calculated at the end of service for each simulated user. The end of service also served as the indexing event for the frequency modulation of the arrival and service rates. By indexing event it is meant that on every departure the modulation "time" was incremented and new values were calculated for the service and arrival rates according to this new modulation time.

The frequencies of modulation for the arrival and service rates were chosen to be very low so that the effects of each parameter change would be more observable. By using low frequencies each parameter also passed through more values from its range than it would have at higher frequencies. Note that a time-delay between the change of a parameter and its observable effect is not detectable or important when carrying out this analysis. Therefore it is not a factor in deciding frequencies.

The frequency analysis was carried out by simulating 33 000 waiting times with the service and arrival rates at their nominal



(a)



(b)

Fig. 6. Ratio of power spectra for two experiments on an M/M/1 queue when its parameters were modulated sinusoidally: (a) λ was modulated at the low frequency and μ at the high frequency and (b) μ was modulated at the low frequency and λ at the high frequency.

values of 0.8 and 0.4, respectively. A second run of 33 000 waiting times was made with

$$\text{service rate} = \mu = 0.8 + 0.2\sin(46/4096 \times 2\pi t)$$
$$\text{arrival rate} = \lambda = 0.4 + 0.2\sin(4/4096 \times 2\pi t)$$

where t is the number of customers served. Since it is a stochastic simulation, longer output sequences were needed than for the previous transfer function example. The autocorrelations of the sequences were calculated, truncated, windowed with a Tukey window, fit with a cosine transform, and the ratio of the nominal and perturbed runs were calculated [40]. The interesting part of this spectrum ratio is plotted in Fig. 6(a). There are distinct peaks (labeled λ and μ) at the modulation frequencies of the arrival rate and service rate. Smaller but still distinct peaks (labeled $\mu\lambda$) are visible at the sum and difference of these frequencies, corresponding to interaction (product) terms. No other peaks appear to be significant. The fact that the peak labeled λ is larger than the peak labeled μ does not mean that λ is more important than μ in this system. Instead the relative peak sizes are the result of the queue acting as a low pass filter. Parameters modulated at lower frequencies have larger peaks, all else being equal. This effect is clearly shown in Fig. 6 (bottom) where the modulation frequencies for λ and μ were interchanged and as a result the relative sizes of their spectral peaks were also interchanged. Comparison of the top and bottom spectra of Fig. 6 enables us to state that λ and μ are equally important in this queue.

The expected waiting time for an M/M/1 queue is $1/(\mu - \lambda)$, where μ is the service rate and λ is the arrival rate. Note that in our experiments μ can equal λ, which would create infinite expected waiting time. This was not a problem in our experiments since it happened rarely and then only for short periods.[1]

In other trials applying Sinusoidal Variation of Parameters to this problem we have found that the size of the spectral peaks are dominated by frequency dependency and extremely low frequencies are necessary to get distinct peaks as shown in Fig. 6. This result suggests caution in interpreting results derived from the current method for implementing this technique. Often frequencies several orders of magnitude higher have been used for Sinusoidal Variation of Parameters on this and similar problems in order to maximize the separation of response peaks in the spectrum [12]. Our experiments with such modulation frequencies gave very poor results. In fact the frequencies and simulation length chosen were a tradeoff between the computation required and the visible responses. By lowering the modulation frequencies further and lengthening the simulation, higher order responses can be made significant. This makes intuitive sense. The waiting time in a queue for a customer is governed by the number of people in the queue ahead and the time necessary to service the present customer. If the modulation frequencies are high, then the conditions necessary to build or empty the queue are present for a short time only and do not get a chance to affect the waiting time of following customers. The queue can be thought of as acting as a low pass filter, and in this case perhaps a leaky integrator [41].

The Sinusoidal Variation of Parameters technique seems to be model dependent. It works best for models that act as bandpass filters. Therefore, before using this type of sensitivity analysis, the input–output behavior of the system should be plotted. The modulation frequencies are then chosen to be in the flat region of the frequency response. If the frequency response of a model is known to be flat in the region of the parameter modulation frequencies, the size of the response peaks may be useful for estimating the importance of parameters. Models that are this well characterized generally do not require analysis by Sinusoidal Variation of Parameters thus limiting the use of the size of the frequency response peaks.

IV. The Pinewood Derby

Since the 1950's over 80 million Cub Scouts have built five-ounce, wooden cars and raced them in Pinewood Derbies. Generally, in derbies that we have observed 80 scouts and parents raced cars. Eight adults were required to run the races. And the events lasted about four hours. Pinewood Derby events could clearly be improved and system design methodology was the obvious tool for improvement.

Originally this system design problem was used as a class project for a graduate course in systems engineering. Then it was redesigned and included as a chapter in a systems design textbook [42]. We used these designs to run actual Pinewood Derbies in four consecutive years. Our group has

[1] In our simulation this is not an actual division: it is a theoretical limit. To understand, think what happens if the server goes to get coffee. The service rate goes to zero, for a little while, but the line only gets longer. The system does not explode. We do not actually see negative waiting times. This is not a problem in our simulations.

invested about three thousand man-hours in this project. We used a CPU-year on an AT&T 3B2/400 computer just finding appropriate round robin schedules. The point of this discussion is that this is a big, real-world problem. The system was designed, built, tested, used, and retired; then this process was repeated in subsequent years. As a part of this system design we did a sensitivity analysis of the tradeoff study that was used to select the best alternatives. This tradeoff study was to choose the best racing format. There were 89 parameters in the tradeoff study. It was originally believed that all of these parameters could effect the recommended alternative: quite surprisingly our sensitivity analysis showed that only three could.

These are some of the figures of merit that we used to compare different design alternatives: Average Races Per Car, Percent Happy Scouts, Number of Irate Parents, Number of Lane Repeats, Acquisition Time and Cost, Total Event Time, and Number of Adults needed. Alternative designs that we considered include: single elimination tournaments, double elimination tournaments, and round robins with three different scoring techniques.

The recommended alternative of the tradeoff study was determined by the *Overall Performance* figure of merit for each alternative: these scores were numbers in the range 0 to 1. The Overall Performance figure of merit is composed of two parts, the *Input/Output Performance* figure of merit and the *Utilization of Resources* figure of merit [42]. Each of these is also a number in the range 0 to 1. Each of these is given a weight (derived from an expert's *Importance Value)* so that their weighted sum is again in the range 0 and 1. The Input/Output Performance and Utilization of Resources figures of merit are each further broken down into subcategories by similar means. The Input/Output Performance figure of merit consists of items like Percent Happy Scouts, Average Races Per Car, and Number of Lane Repeats. The Utilization of Resources figure of merit consists of items like the Acquisition Cost and Total Event Time. Each of these figures of merit was given an importance value by the domain experts.

Each of the figures of merit had a Standard Scoring Function [42], [43] associated with it. In most cases these scoring functions look like sigmoidal curves as shown in Fig. 7. They allowed us to translate input values of the figure of merit, into output scores that have a range of 0 to 1. These functions have parameters that set the valid input range, the input value (the baseline) that gives an output of 0.5, and the slope at this point. The domain experts choose the values of these parameters. These scoring function parameters and the importance values of the figures of merit of the tradeoff study were examined by the sensitivity analysis.

We wanted the sensitivity analysis to answer two questions. First we wanted to know which parameters and figures of merit were the most important and deserved further attention and verification. Clearly a relative-sensitivity measure for each parameter for each alternative would answer this question. The second question was which parameters could be adjusted to change the recommended alternative. This question could be answered by a search algorithm using the sensitivity of each parameter. Since we had five alternatives, 89 parameters, and

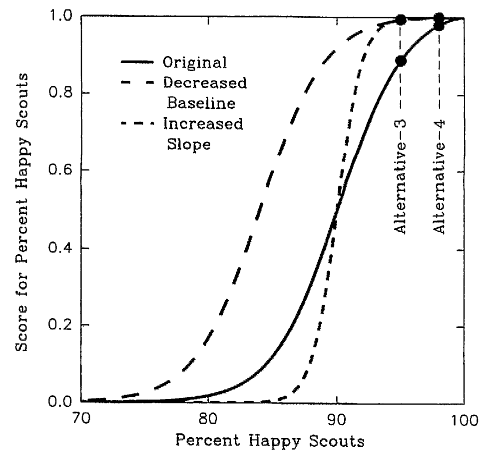


Fig. 7. Scoring function for Percent Happy Scouts figure of merit for a Pinewood Derby system design.

16 figures of merit, we decided against any manual solution and wrote programs that carried out both analyses.

What did we find? First, of the 105 parameters and figures of merit, only 50 to 60 had nonzero sensitivities for any given alternative. The set of parameters with high sensitivities differed for each alternative: this justified our extra work of carrying out the sensitivity analysis of each alternative. If we had taken one alternative as representative, we would have missed parameters that were important in other alternatives. By comparing the lists, we can also see that some parameters are important to all alternatives and, therefore, deserve special attention.

Our sensitivity analysis showed that the most important parameters for the five racing format alternatives were 1) the baseline for the Overall Happiness figure of merit, 2) the baseline for the Percent Happy Scouts figure of merit, 3) the importance value (weight) of the Overall Happiness figure of merit, 4) the baseline for the Number of Repeat Races figure of merit, 5) the input value given for the Percent Happy Scouts figure of merit (note this is an input to the system: it is good that a system is sensitive to its inputs), and 6) the input value given for the Number of Repeat Races figure of merit. The most important parameters concern the happiness of the scouts. Interviews with the designers revealed that this was indeed their primary goal. Our point is that if some obscure parameter, which was given little thought, turned up in the most important list, then the system would have to be redesigned.

Our second important finding was that, of the 89 parameters, only three could be manipulated to change the recommended alternative. To the best of our search program's ability, no change in the other 86 parameters would change the recommended alternative. This gives a strong clue to the strengths and weaknesses of the recommended alternative.

Let us now look at the only three parameters that could change the recommended alternative. In the original tradeoff study the Input/Output Performance figure of merit was given a weight of 0.9 and the Utilization of Resources figure of merit was given a weight of 0.1. The recommended alternative was alternative-4, a round robin event with best time scoring. The sensitivity analysis showed that if the weights were changed to 0.57 for Performance and 0.43 for Resources then alternative-2, a double elimination tournament would win. This is reasonable, because a double elimination tournament requires fewer workers and less money, but it does not perform as well. So if money is more important than the happiness of the scouts, then the double elimination tournament is best. Because these two weights must add up to 1.0, we counted this as only one parameter. The other parameters that could change the recommended alternative were related to the Percent Happy Scouts figure of merit. The scoring function for this figure of merit is shown in Fig. 7. With the original baseline value of 90%, alternative-4 has a higher score than alternative-3. However, if this baseline value is reduced to 84% (as shown in Fig. 7), then the scores of two alternatives become the same for this figure of merit, and the recommended alternative changes from alternative-4, the round robin with best time scoring, to alternative-3, a round robin with mean time scoring. Changing the slope from 0.1 to 0.26 (also shown in Fig. 7) caused the same change.

Bahill has used the Pinewood Derby System Design in lectures since 1989. He told audiences that changing the weights of the figures of merit could change the recommended alternative, but he was never able to demonstrate this. Using trial and error he never found a weight that could change recommendations. This sensitivity analysis showed that his claim was true, but the instances are sparse and are not likely to be found by trial and error. Overall, the sensitivity analysis has enabled us to understand our Pinewood Derby System Design much better.

Finally we would like to mention that two versions of the program were writtten to carry out our sensitivity analyses. The first used empirical methods to approximate derivatives and the second used analytic derivatives. The results of both programs were essentially the same. Small numerical differences in results were noted but they were not significant. The empirical version required the user to define appropriate deltas for estimating derivatives. The analytic version required the programmer to express the derivatives of the scoring functions. For limited use the empirical program would be more economical while the ease of use of the analytic version would pay off for a heavily used system.

V. Discussion

Sensitivity analyses are important tools for validating models in many diverse fields. The types and methods of sensitivity analysis presented in this paper have ranged from precise mathematical definitions yielding exact sensitivities, to empirical methods for determining qualitative sensitivities. Each has its purpose, advantages, and drawbacks.

The precise mathematically defined sensitivity functions yield the maximum information about the sensitivity of a system to its inputs and parameters at the expense of requiring a differentiable set of system equations. The equations could be written in terms of time or frequency. The resulting sensitivities would be functions of all other inputs and parameters (interactions) as well as time or frequency. Relative-sensitivity functions are difficult to compute correctly, because multiplication in the time domain requires convolution in the frequency domain. Consequently, of the analytical sensitivity functions, the semirelative-sensitivity functions are the most useful.

When the system is not defined with a set of differentiable equations or is perhaps not even modeled, the Direct Observation approach yields the sensitivity of the system to a particular parameter about the nominal operating point. If experiments are carried out over a range of nominal operating points, the system can be approximated as a simple function of inputs and parameters. The result is the Response Surface of the system and the gradients with respect to inputs and parameters are the sensitivities. The designs of the experimental procedures necessary to preform such analyses can be found in Montgomery [44]. These estimated sensitivities may suffer from lack of accuracy and limited range of validity.

Determining the Response Surface grows in complexity geometrically for nondeterministic systems and for systems with many inputs, parameters, and parameter interactions. This leads to the use of Sinusoidally Varied Parameters for obtaining qualitative sensitivities of a system. The results of Sinusoidally Varied Parameters experiments are only qualitative, but can be much simpler to obtain than the equivalent Direct Observation results. The results of the Sinusoidally Varied Parameters sensitivities can be used to guide the design of Direct Observation experiments for determining more accurate sensitivities to important parameters.

Once the sensitivities of a system to its inputs and parameters are determined they can be used to guide future research and design. If the sensitivities of a model are quantitatively similar to the sensitivities of the physical system the validity of the model is enhanced. Discrepancies can be used to direct improvements of the model or further testing of the physical system. For example, the field of chaotic systems arose from observations that some complex, nonlinear systems, which were well defined, exhibited extreme sensitivity to initial conditions.

The examples in this paper were chosen to be simple in nature but have one of two properties of interest. They either allowed us to perform multiple types of sensitivity analysis on the same problem or else they were well known problems that could be examined for new information by using sensitivity analysis. One interesting point that showed up was that the sensitivity (a linear function) of a linear system parameter is not necessarily linear. This can take the form of interaction with other parameters or nonlinear direct effects. In examples 2 and 3 the time-delay, θ, had no effect on the amplitude of the step response. So doubling or quadrupling its value had no effect. However, the settling time of the step response did depend on the value of the time-delay, but not in a linear manner. The principle of superposition applies to inputs and not parameters. In most cases the effects of changing two parameters independently are different than changing those same two parameters simultaneously. Some of the examples in this paper were not textbook polished. The Sinusoidally Varied Parameters examples demonstrate a new technique that seems to have promise in the field of simulation, but it has some problems that still have to be resolved.

In general, we do not say that a model is most sensitive to a certain parameter. Rather we must say that a particular facet of a model is most sensitive to a particular parameter at a particular frequency or point in time. For example, the transfer functions of examples 2 and 3 were most sensitive to the gain K at low frequencies, while their step responses were most sensitive to the time-delay θ at the beginning of the movement.

Acknowledgment

We thank F. Szidarovszky and E. Fernandez for helpful suggestions.

References

[1] C. Stern and E. R. Sherwoods, *The Origins of Genetics. A Mendel Source Book.* San Francisco, CA: Freeman, 1966.

[2] J. F. Box, "Gosset, Fisher, and the t distribution," *Amer. Statistician,* vol. 35, no. 2, pp. 61–66, 1981.

[3] A. Ford and P. C. Gardiner, "A new measure of sensitivity for social system simulation models," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-9, pp. 105–114, 1979.

[4] P. M. Frank, *Introduction to System Sensitivity Theory.* New York: Academic, 1978.

[5] F. K. Hsu, A. T. Bahill, and L. Stark, "Parametric sensitivity of a homeomorphic model for saccadic and vergence eye movements," *Comput. Prog. Biomed.,* vol. 6, pp. 108–116, 1976.

[6] A. T. Bahill, J. R. Latimer, and B. T. Troost, "Sensitivity analysis of linear homeomorphic model for human movement," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-10, pp. 924–929, 1980.

[7] S. Lehman and L. Stark, "Simulation of linear and nonlinear eye movement models: Sensitivity analyses and enumeration studies of time optimal control," *J. Cybern. Inform. Sci.,* vol. 4, pp. 21–43, 1979.

[8] S. J. Yakowitz and F. Szidarovszky, *An Introduction to Numerical Computations.* New York: Macmillan, 1989.

[9] B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems.* Reading, MA: Addison-Wesley, 1984, pp. 218–219.

[10] A. T. Bahill, *Verifying and Validating Personal Computer-Based Expert Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 104–112.

[11] J. P. C. Kleijnen, *Statistical Techniques in Simulation.* New York: Dekker, part I, 1974, part II, 1975.

[12] P. J. Sanchez and L. W. Schruben, "Simulation factor screening using frequency domain methods: An illustrative example."

[13] L. W. Schruben and V. J. Cogliano, "An experimental procedure for simulation response surface model identification," *Commun. ACM,* vol. 30, pp. 716–730, 1987.

[14] D. J. Morrice and L. W. Schruben, "Simulation sensitivity analysis using frequency domain experiments," in *Proc. 1989 Winter Simulation Conf.,* 1989, pp. 367–373.

[15] R. Y. Rubinstein and F. Szidarovszky, "Convergence of perturbation analysis estimates for discontinuous sample functions: A general approach," *Adv. Appl. Prob.,* vol. 20, pp. 59–78, 1988.

[16] Y. C. Ho and S. Li, "Extensions of infinitesimal perturbation analysis," *IEEE Trans. Automat. Contr.,* vol. 33, pp. 427–438, 1988.

[17] P. V. Kokotovic and H. K. Khalil, *Singular Perturbations in Systems and Control.* New York: IEEE Press, 1986.

[18] J. T. Heinen and J. G. Lyon, "The effects of changing weighting factors on wildlife habitat index values: A sensitivity analysis," *Photogram. Eng. Remote Sensing,* vol. 55, pp. 1445–1447, 1989.

[19] E. C. Stegmean, B. G. Schatz, and J. C. Gardner, "Yield sensitivities of short season soybeans to irrigation management," *Irrigation Sci.,* vol. 11, pp. 111–119, 1990.

[20] J. R. Latimer, B. T. Troost, and A. T. Bahill, "Linearization and sensitivity analysis of model for human eye movements," in *Modeling and Simulation, Proc. Ninth Annu. Pittsburgh Conf.,* W. Vogt and M. Mickle, Eds. Pittsburgh, PA: Instrum. Soc. Amer., 1978, pp. 365–371.

[21] M. G. Mehrabi and A. Hermami, "Modeling and sensitivity studies of the feeddrive system of a CNC lathe driven by a DC motor," in *Machinery Dynamics—Applications and Vibration Control Problems,* T. S. Sankar, Ed. New York: ASME, 1989, vol. 18-2, pp. 37–44.

[22] A. T. Bahill, *Bioengineering: Biomedical, Medical and Clinical Engineering.* Englewood Cliffs, NJ: Prentice-Hall, 1981.

[23] R. C. Wong, "Accelerated convergence to the steady-state solution of closed-loop regulated switching-mode systems as obtained through simulation," in *Proc. 18th IEEE Power Electronics Specialists Conf.,* 1987, pp. 682–692.

[24] D. F. Wilkie and W. R. Perkins, "Essential parameters in sensitivity analysis," *Automatica,* vol. 5, pp. 191–197, 1969.

[25] D. S. Denery, "Simplification in the computation of the sensitivity functions for constant coefficient linear systems," *IEEE Trans. Automat. Contr.,* vol. AC-16, pp. 348–350, 1971.

[26] S. Bingulac, J. H. Chow, and J. R. Winkelman, "Simultaneous generation of sensitivity functions-transfer function matrix approach," *Automatica,* vol. 24, pp. 239–242, 1988.

[27] W. Hafez, K. Loparo, E. Rasmy, and M. Hashish, "Generation of trajectory sensitivities: Minimal-order realization," *Int. J. Contr.,* vol. 49, pp. 809–825, 1989.

[28] J. F. Genrich and F. S. Sommer, "Novel approach to sensitivity analysis," *J. Petroleum Technol.,* vol. 41, pp. 930–932, 983–985, 1989.

[29] J. M. Wojciechowski, "A general approach to sensitivity analysis in power systems," in *Proc. IEEE Int. Symp. Circuits and Systems,* 1988, pp. 417–420.

[30] J. W. Bandler and M. A. El-Kady, "The method of complex Lagrange multipliers with applications," in *Proc. IEEE Int. Symp. Circuits and Systems,* 1981, pp. 773–777.

[31] A. T. Bahill, "A simple adaptive Smith-predictor for controlling time-delay systems," *IEEE Contr. Syst. Mag.,* pp. 16–22, May 1983.

[32] P. A. Ioannou and P. V. Kokotovic, *Adaptive Systems with Reduced Models.* New York: Springer-Verlag, 1983.

[33] M. S. Phadke, *Quality Engineering Using Robust Design.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[34] J. J. Pignatiello and J. S. Ramberg, "Top ten triumphs and tragedies of Genichi Taguchi," *Quality Eng.,* vol. 4, pp. 211–225, 1991.

[35] I. D. Landau, *Adaptive Control: The Model Reference Approach.* New York: Dekker, 1979.

[36] V. Chátal, *Linear Programming.* New York: Freeman, 1983.

[37] H. W. Bode, *Network Analysis and Feedback Amplifier Design.* New York: Van Nostrand, 1945.

[38] W. Nicholson, *Microeconomic Theory: Basic Principles and Extensions.* Hinsdale, IL: Dryden, 1978.

[39] L. Stark, *Neurological Control Systems, Studies in Bioengineering.* New York: Plenum, 1968, pp. 308–312.

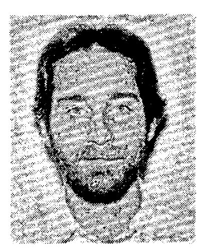
[40] L. W. Schruben, "Simulation optimization using frequency domain methods," in *Proc. 1986 Winter Simulation Conf.,* J. R. Wilson, J. O. Henriksen, and S. D. Roberts, Eds. Piscataway, NJ: IEEE, 1986, pp. 366–369.

[41] D. J. Morrice and L. W. Schruben, "Simulation sensitivity analysis using frequency domain experiments," in *Proc. 1989 Winter Simulation Conf.,* 1989, pp. 367–373.

[42] W. L. Chapman, A. T. Bahill, and W. Wymore, *Engineering Modeling and Design.* Boca Raton, FL: CRC Press, 1992.

[43] W. L. Chapman and A. T. Bahill, "A hypertext software system to help with system design," *IEEE Trans. Syst., Man, Cybern.,* vol. 23, 1993, to be published.

[44] D. C. Montgomery, *Design and Analysis of Experiments.* New York: Wiley, 1984.



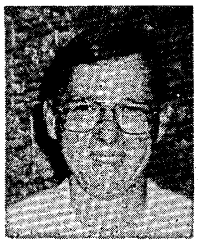
William J. Karnavas was born in Pittsburgh, PA, on May 30, 1963. He received the B.S. degree in 1984 from Carnegie Mellon University, Pittsburgh, and the M.S. degree in 1986 from the University of Pittsburgh, both in electrical engineering, and the Ph.D. degree in systems and industrial engineering in 1992 from the University of Arizona, Tucson.

His research interests are systems engineering theory, modeling physiological systems, and computer-based bioinstrumentation.

Paul J. Sanchez received the B.S. degree in economics in 1977 from Massachusetts Institute of Technology, Cambridge, and the M.S. and Ph.D. degrees in operations research, in 1985 and 1987, respectively, from Cornell University, Ithaca, NY.

He is the President of Indalo Software, which specializes in simulation and statistical software, training and consulting, as well as custom programming on

NeXTSTEP, Unix, and DOS. His research interests include the relationships between the fields of spectral analysis, systems theory, and experimental design; graphical representations of simulation models; statistical issues in simulation such as output analysis, random number testing, and selection of alternative systems; and data structures.



A. Terry Bahill (S'66–M'68–SM'81–F'92) was born in Washington, PA, on January 31, 1946. He received the B.S. degree in electrical engineering in 1967 from the University of Arizona, the M.S. degree in 1970 in electrical engineering from San Jose State University, and the Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, in 1975.

He served as a Lieutenant in the U.S. Navy, and as an Assistant and Associate Professor in the Departments of Electrical and Biomedical Engineering at Carnegie Mellon University, and Neurology at the University of Pittsburgh. Since 1984, he has been a Professor of Systems and Industrial Engineering at the University of Arizona in Tucson. His research interests include systems engineering theory, modeling physiological systems, eye–head–arm coordination, the science of baseball, the system design process, validating expert systems, concurrent engineering, quality function deployment, and total quality management. He has published more 100 papers and has lectured in a dozen countries. His research has appeared in *Scientific American, The American Scientist, The Sporting News, The New York Times, Sports Illustrated, Newsweek, Popular Mechanics, Science85, Highlights for Children, Rolling Stone,* and on the NBC Nightly News. He is the author of *Bioengineering: Biomedical, Medical, and Clinical Engineering* (Prentice-Hall, 1981), *Keep Your Eye on the Ball: The Science and Folklore of Baseball* (with Bob Watts), (W. H. Freeman, 1990), *Verifying and Validating Personal Computer-Based Expert Systems* (Prentice-Hall, 1991), *Linear Systems Theory* (with F. Szidarovszky), (CRC Press, 1992), and *Engineering Modeling and Design* (with Bill Chapman and Wayne Wymore), (CRC Press), 1992.

Dr. Bahill is a member of the following IEEE societies: Systems, Man, and Cybernetics, Engineering in Medicine and Biology, Computer, Automatic Controls, and Professional Communications. For the Systems, Man, and Cybernetics Society he has served three terms as Vice President, six years as Associate Editor, as Program Chairman for the 1985 conference in Tucson, and as Co-chairman for the 1988 conference in Beijing and Shenyang, China. He served four years as an associate editor for the Computer Society's magazine, *IEEE Expert.* He is the Editor of the CRC Press Series on Systems Engineering. He is a Registered Professional Engineer and a member of Tau Beta Pi, Sigma Xi, and Psi Chi.