NASA TECHNICAL MEMORANDUM

NASA TM X-1797

NASA TM X-1797

A SYSTEMATIC APPROACH TO MODEL DEVELOPMENT BY COMPARISON OF EXPERIMENTAL AND ANALYTICAL REGRESSION COEFFICIENTS

by John H. Lynch
Lewis Research Center
Cleveland, Ohio

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MAY 1969

NASA TM X-1797

A SYSTEMATIC APPROACH TO MODEL DEVELOPMENT BY COMPARISON OF EXPERIMENTAL AND ANALYTICAL REGRESSION COEFFICIENTS

By John H. Lynch

Lewis Research Center
Cleveland, Ohio

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

ABSTRACT

An approach is suggested for the systematic investigation of mathematical models used to describe physical processes. The substance of the approach consists of reducing specific results obtained from a model and from experiments to orthogonal response surfaces and comparing the regression coefficients. This polynomial representation displays the behavior of the model in a form that is much more easily analyzed than simultaneous differential or integral equations, especially when the model is numerical. A sample application to a reactor physics problem is included.

A SYSTEMATIC APPROACH TO MODEL DEVELOPMENT BY COMPARISON OF EXPERIMENTAL AND ANALYTICAL REGRESSION COEFFICIENTS

by John H. Lynch

Lewis Research Center

SUMMARY

A systematic procedure is suggested for locating and examining the sources of disagreement between calculations and measurements. The procedure consists essentially of comparing regression coefficients of orthogonal response polynomials developed for calculations and measurements.

In cases where the coefficients disagree significantly, the comparison may suggest evolutionary improvements in the model. The comparison may also be used to force agreement between the mathematical model and the physical process if all of the salient variables are treated explicitly. When the same forcing corrections are found for several different response characteristics of the same process, the source of the discrepancies between experiments and calculations can be considered to have been located and evaluated.

This approach to model development may in some cases uncover information that would otherwise go unnoticed. A regression analysis of the type suggested also has a distinct economic motivation when developing numerical models. The regression polynomials display the behavior of the model in a form which is much more readily analyzed than an intuitively selected set of computer calculations. This report describes this approach to model development and includes a detailed sample analysis.

The terms calculations, model, calculational model, and calculated results, and similar terms used throughout the text, are synonymous insofar as they relate to the theoretical or analytical aspects of any scientific investigation. Similarly, the terms experimental data, measurements, observations, observed data, and the like, all refer to physical data obtained by observation. Experiment design refers to the planning of the investigation and applies in this report to both calculations and measurements.

INTRODUCTION

In the course of investigating some physical process, recourse is frequently made to models which simulate the process. The model is usually chosen so that certain of the physical variables interact in the model according to physical theories which describe the process. The primary advantage of using a model is that it provides information that is usually much less expensive to generate than when full-scale experiments are performed. Also, models frequently provide insight into the interrelations between variables that is not readily obtained with full-scale tests.

Models may take several forms. They may differ only in size from the full-scale process. Reaction kinetics, for example, is frequently studied using scaled-down chemical reactors. Compressible flow over immersed bodies, on the other hand, is usually studied using wind tunnels in which a physical analogy is retained by preserving certain similarity parameters such as Mach and Reynolds numbers. One class of models which has found more widespread use than any other is the class called mathematical models. This model is the familiar representation of the physical process by sets of equations which embody the physical theory. Such models are used in virtually every phase of scientific investigation. In some areas they are indispensable. The design of atomic devices or the prediction of nuclear power reactor lifetime, for example, would not be experimentally feasible.

Mathematical models may range from simple equations to more complex sets of simultaneous relations which incorporate experimentally derived constants and which must be used with computing machinery. The behavior of simpler mathematical models, such as closed-form solutions, is readily analyzed by inspection in most cases. More complex mathematical models involving numerical methods, however, are not so easily understood because a computer solution is nothing more than a set of values that satisfies the governing finite difference equations. Thus, the exact dependence of the calculated results on the input is not generally known and, in fact, must be computed for each different configuration.

This lack of insight, especially with numerical models, leaves the investigator in a position of ignorance as to the source of error in his model when his calculations do not agree with experimental data. To obtain the desired agreement, he must adjust (develop) his model in some way. When the system under study is complex, development of the model frequently becomes, in practice, an intuitive procedure. A common intuitive approach, for example, is to examine several cases in which single parameters are varied one at a time. Based on the sequential results, some judgment is exercised after each investigation as to what constitutes an improvement in the model. This judgment is then used as a guide in the selection of subsequent investigations.

Intuitive methods such as this usually arise from an inherent complexity found in many mathematical models. For complex systems, several approximations are frequently employed simultaneously, and an economically feasible systematic plan for evaluating each approximation and its interplay with the other approximations does not normally suggest itself.

Intuitive developments, however, can be expensive, are unsystematic, and can easily lead to wrong conclusions. Further, they may be inadequate for the process being studied. Strong interactions between variables could go unnoticed. What is needed for model development is a more systematic, logical, and relatively inexpensive method which locates the cause of the disagreement between calculations and measurements and enables physically realistic corrections to be made. The description and use of such a method is the subject of this report. Although the presentation assumes the development of mathematical models, similar procedures could be used with nonmathematical models.

METHOD

The method which is discussed herein makes use of a statistical experimental design tool known as regression analysis. This technique is frequently used with processes which are too complex to treat analytically. Examples of such processes are agricultural and chemical yield or material flow in manufacturing operations. Regression analysis is also frequently used to obtain analytical representations of experimental data. The product of such an analysis is a response polynomial in which the effects of identified independent variables and their interactions appear as coefficients. An additional motivation behind such an experiment design is to make efficient use of the experimental data. Effects of intermediate values of the variables are obtained by inference from the response polynomial. The main intent, then, of regression analysis of measured data is to provide a mathematical french-curve to represent the data without particular regard for the physical mechanisms underlying the results.

Regression analysis has also been used to some extent for calculations. The motivation here is to replace several simultaneous equations by a single equation. The calculations, or "synthetic experiments" as they are sometimes called, are performed using specific values of input parameters. The results of the calculations are represented by one or more response polynomials. Thus, calculations involving intermediate values of the input parameters need not be performed. This can result in a considerable savings in computing costs. Efficiency is the underlying purpose in this case.

In essence, the method to be discussed suggests that by simultaneous application of regression analysis to both the calculational model and the experiments, considerable

insight into the capability and effectiveness of the model can be obtained. A term-by-term comparison of the regression coefficients for the model and the experiments can identify the shortcomings of the model. This identification can then be used either to improve basic model parameters or to provide effective corrections to the existing parameters. This transformation of model and observed physical behavior into regression polynomial form displays the behavior of both model and experiment in a form which is readily analyzed. First-order, second-order, and possibly higher-order effects, as well as interaction effects of all independent variables, are completely separated if the experiment design is orthogonal. The extent to which a particular set of coefficients agrees or disagrees can be taken as a measure of model adequacy for the particular variable whose effect is indicated by these coefficients. This direct indication of the source of discrepancy between model and experiment could be much more revealing than an arbitrary comparison between calculations and experiments.

The qualitative information obtained provides a basis for systematic examination of the equations and physical parameters that comprise the calculational model. In the interest of providing effective forcing corrections to the model, the fact that the polynomial demonstrates the model behavior is useful. If the model predicts the correct average behavior (pure constant term in the polynomial), the response may be corrected by adjusting an independent variable using the correction f_i derived in appendix A. If the model does not predict the correct average response (at the zero coded level of all independent variables), a further modification can be made to the independent variables to account for their individual contribution to the error in the constant term. This is also discussed in appendix A.

If response polynomials for several dependent parameters are constructed from the same set of measurements and calculations, forcing corrections can be obtained for each set. If these corrections are the same or nearly the same, the investigator can feel confident that the discrepancies between model and experiment have been located and evaluated. If the indicated corrections are not similar (they probably will not be exact because of the behavior or inherent error in the polynomial fits), then using any particular set can be viewed only as a correction for the associated parameter. In this case, it may cause arbitrary behavior of the model with respect to the other dependent parameters.

REGRESSION TRANSFORMATION

The regression transformation is essentially an orthogonal fit of data taken from specific calculations or measurements. Detailed discussions of regression methods can be found in reference 1. The discussion that follows is intended to abstract the salient properties of regression analysis as it applies to the ideas suggested in this report. It

thus provides all information necessary to generating regression polynomials but does not include any derivations. Only the second-order fit in n independent variables is discussed. A second-order fit would be adequate for a wide range of applications. For more complex fits, reference 1 should be consulted.

The regression polynomial has the form

$$Q = A_0 + A_1x_1 + A_2x_2 + \ldots + A_nx_n + A_{11}x_1^2 + A_{22}x_2^2 + \ldots + A_{nn}x_n^2 + A_{12}x_1x_2$$

$$+ A_{13}x_1x_3 + \ldots + A_{1n}x_1x_n + A_{23}x_2x_3 + A_{24}x_2x_4 + \ldots + A_{n-1,n}x_{n-1}x_n \quad (1)$$

where

Q	dependent variable of interest
$x_1, x_2, \ldots, x_n$	coded values of selected independent variables
A_0, A_n, A_{nn}	regression coefficients

The coefficients in equation (1) are obtained by using

$$A_0 = \sum_{l=1}^{m} \frac{Q'_l}{m} - K \sum_{j=1}^{n} A_{jj} \quad (2)$$

$$A_n = \frac{\sum_{l=1}^{m} Q'_l x_{n,l}}{\sum_{l=1}^{m} \left(x_n\right)_l^2} \quad (3)$$

$$A_{nn} = \frac{\sum_{l=1}^{m} Q'_l \left(x_n^2 - K\right)_l}{\sum_{l=1}^{m} \left(x_n^2 - K\right)_l^2} \quad (4)$$

$$A_{n-1,n} = \frac{\sum_{l=1}^{m} Q'_l \left(x_{n-1}\right)_l \left(x_n\right)_l}{\sum_{l=1}^{m} \left(x_{n-1}\right)_l^2 \left(x_n\right)_l^2} \tag{5}$$

$$K = \frac{\sum_{l=1}^{m} \left(x'_n\right)_l^2}{m} \tag{6}$$

In equations (1) to (6)

Q'_l observed value of Q from l^{th} experiment (or calculation)

l experiment (or calculation) number

m total number of experiments (or calculations)

$\left(x'_n\right)_l$ fixed coded value of n^{th} independent variable for l^{th} experiment (or calculation)

n total number of independent variables

The coefficients given by equations (2) to (5) are associated with a particular type of fractional factorial experiment design called an orthogonal composite design (ref. 2). Only this design is discussed herein because several of its features make it one of the most attractive (see ref. 1 for other possible designs). The orthogonal composite design is particularly attractive because the coefficients may be evaluated without computing machinery. And the required number of experiments with this fractional factorial design is considerably less than with a full factorial design. Also, the orthogonality of the design guarantees separation of effects so that analysis by coefficient comparison is meaningful.

When performing the required experiments (or calculations), the coded variables x_n are assigned particular values for each experiment (or calculation). These values, or levels as they are frequently called, are chosen from a diagram to obtain specific combinations and depend on the range of the variables to be studied. Diagrams and detailed discussion for several variable designs can be found in references 2 and 3. A few diagrams are repeated in appendix B for reference.

The procedure for generating an orthogonal response polynomial is as follows. First, the variables that are believed to significantly affect the parameter of interest

are listed. Generally, these can be either qualitative or quantitative. A feasible size (number of experiments) is next assigned to the experimental program to determine whether regression techniques are practical. Then, in view of the limited number of experiments that are feasible, the number of variables is reduced until an experiment design is obtained that is within the stated number of experiments available. The number of experiments required m as a function of the number of independent variables treated explicitly n is given by

$$m = 2^n + 2n + 1 \tag{7}$$

The number of independent variables in the preliminary list can be reduced in the following ways:

(1) By assuming that certain variables do not have a strong enough effect to merit their explicit treatment (These variables would then be ignored at the risk of having made an incorrect assumption.)

(2) By combining certain of the variables that can be expected, on physical grounds, to act according to known mathematical relations (Current and resistance might logically be multiplied together, thus making voltage the independent variable. Dimensionless parameters are other examples of this technique.)

(3) By randomizing certain variables (When a variable is qualitative rather than quantitative, there are ways to treat it explicitly (ref. 1); also, however, in the interest of reducing the number of experiments, it can be randomized so that its effect is confounded with the residual or natural random error of the experiments. This prevents these variables from introducing a systematic bias in the results.)

Once the significant variables have been selected, the range of variation over which the study is to be performed is chosen. This range establishes the delta δ limits of the design (ref. 3). Intermediate levels are then established using linear interpolation, as will be shown in the example that follows. The experiments and calculations are then performed according to the factorial design diagram, and the polynomial coefficients are generated using equations (2) to (6).

ERROR IN POLYNOMIALS

Error in the fits can affect conclusions drawn from the coefficients. Some quantities that give an indication of the ability of the regression equation to represent the experimental (or calculational) data are as follows:

$$R^2 = \text{Goodness-of-fit parameter} = 1 - \frac{\sum_{l=1}^{m} (Q - Q')_l^2}{\sum_{l=1}^{m} (Q')_l^2} \tag{8}$$

A value of nearly 1.0 for this parameter indicates that the regression equation represents the observed data well.

$$\sigma = \text{Standard deviation} = \left[\frac{\sum_{l=1}^{m} (Q - Q')_l^2}{m - \text{Number of coefficients in regression equation}} \right]^{1/2} \tag{9}$$

This can be used when computing the confidence interval for the fit (ref. 1). Also, it provides some measure of the dispersion of error in the fit.

SAMPLE ANALYSIS

Problem

To demonstrate the coefficient comparison technique, an example problem has been chosen from the field of nuclear reactor physics. The problem is to determine the perturbing effect on the thermal neutron flux in a slab of material having arbitrary thickness and nuclear properties. The quantity of interest (i.e., the flux perturbation factor) is given by

$$F \equiv \frac{\overline{\varphi}_p}{\overline{\varphi}_o} = \frac{\text{Spatially averaged thermal neutron flux in the slab}}{\text{Spatially averaged thermal neutron flux in absence of slab (averaged over same space)}} \tag{10}$$

A common condition is for the slab to be immersed in water. Thus, an infinite water

medium is assumed to surround the slab. The slab is assumed to be infinite in two dimensions.

The "model" selected to calculate perturbation factors is the one-dimensional neutron diffusion equation

$$-\nabla \cdot D(\hat{x})\nabla\varphi(\hat{x}) + \Sigma_a(\hat{x})\varphi(x) = \hat{S}(\hat{x}) \tag{11}$$

which governs the neutron flux at each point in the system shown in figure 1. Symmetry boundary conditions apply at the right and left boundaries, and the system is assumed to be infinite in the dimensions normal to x. Thirty centimeters of water is an arbitrarily

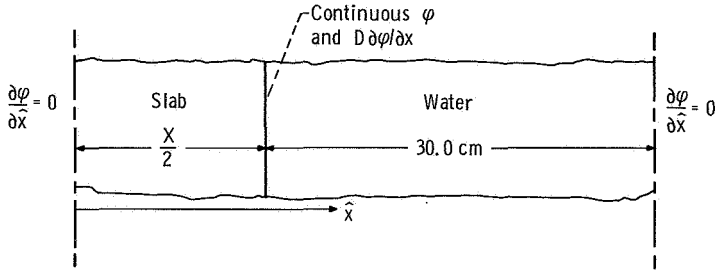

Figure 1. - Sample problem geometry.

chosen distance required for the flux to recover from the effects of the slab. In equation (11), D and Σ_a are the diffusion and absorption coefficients (ref. 4). Although the analytical solution of equation (11) is straightforward, it is supposed for purposes of this example to be obtainable only by computer calculation.

The questions that we wish to answer are

(1) How well does this model represent the physical situation?

(2) If it does not work, why not?

(3) How can it be made to work?

Since this sample problem is intended only as an illustration, the "experiments" will be replaced by a more elegant model. This model will be the solution of the Boltzmann transport equation in one dimension over the same space (ref. 5):

$$\hat{\Omega} \cdot \nabla\psi(\hat{x}, \hat{\Omega}) + \Sigma^t(\hat{x})\psi(\hat{x}, \hat{\Omega}) = S(\hat{x}, \hat{\Omega}) \tag{12}$$

where

$$\int_\Omega \psi(\hat{x}, \hat{\Omega})d\hat{\Omega} = \varphi(\hat{x})$$

$\hat{\Omega}$ direction of neutron flow

Σ^t thermal neutron total cross section

and

$$\int_{\Omega} S(\vec{x}, \hat{\Omega}) d\hat{\Omega} = \hat{S}(\hat{x})$$

The transport approximation is usually a much better representation of the physical problem than the diffusion model of equation (11). Thus, it can realistically be used as the "experiment" for illustrative purposes.

The perturbation factor for this example problem is defined by

$$F \equiv \frac{1}{\varphi(x/2 + 30)} \left[\frac{\int_0^{X/2} \varphi(x) dx}{\int_0^{X/2} dx} \right] \tag{13}$$

Regression Model

Inspection of equation (11) and figure 1 shows that the perturbation factor is a function of D, Σ_a, and X of the slab and also of D and Σ_a of the water through the continuity of φ and $D\varphi'$ at the slab-water interface. It is assumed that the dependence on the water properties is small in order to reduce the number of variables to three.

Next, the range of these variables is selected. To simplify the example, the range from 0.5 to 1.5 was arbitrarily chosen. Intermediate values were determined by using $\delta = 1.2154$ (ref. 3). These values and their corresponding coded levels are shown in table I. The fact that all chosen values are the same at each level has no significance.

TABLE I. - INDEPENDENT VARIABLE VALUES FOR EXAMPLE

Level	Coded level	Thermal neutron absorption cross section, Σ_a, cm^{-1}	Transport corrected scattering cross section, $(1 - \overline{\mu}_o)\Sigma_s$, cm^{-1}	Slab thickness, X, cm
1	-1.2154	0.5	0.5	0.5
2	-1.0	.5886	.5886	.5886
3	0	1.0	1.0	1.0
4	1.0	1.4114	1.4114	1.4114
5	1.2154	1.5	1.5	1.5

The choice at coded levels 1 and 5 was arbitrary. The absorbing slab was assumed to be heavy so that $\overline{\mu}_o$ was zero (ref. 4).

Next, using the orthogonal composite design shown in figure 2 and D of $\left[3(\Sigma_a + \Sigma_s)\right]^{-1}$ for the absorber, we obtain the combinations shown in table II. The constants for water were selected and are the same for all experiments.

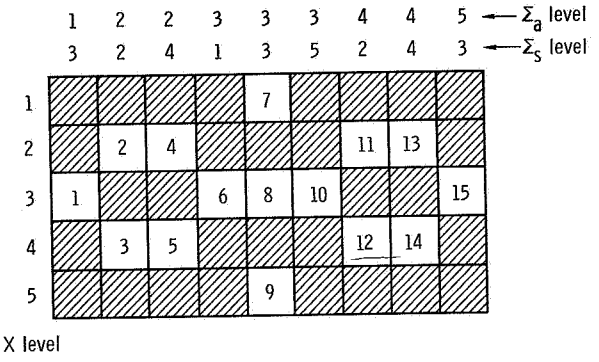

Figure 2. - Experiment design. Numbers in blocks are experiment numbers.

TABLE II. - VALUES OF CONSTANTS USED IN DIFFUSION CALCULATIONS

Material	Experiment	Slab thickness, X, cm	Thermal neutron cross sections		Transport corrected scattering cross section, $(1 - \overline{\mu}_o)\Sigma_s$	Thermal neutron diffusion coefficient, D	Source function in diffusional model, $\hat{S}$
			Absorption, Σ_a	Transport, Σ_{tr}			
Slab	1	1.0	0.5	1.5	1.0	0.2222	0
	2	.5886	.5886	1.1772	.5886	.2832	
	3	1.4114		1.1772	.5886	.2832	
	4	.5886		2.0	1.4114	.1667	
	5	1.4114		2.0	1.4114	.1667	
	6	1.0	1.0	1.5	.5	.2222	
	7	.5		2.0	1.0	.1667	
	8	1.0		2.0	1.0	.1667	
	9	1.5		2.0	1.0	.1667	
	10	1.0		2.5	1.5	.1333	
	11	.5886	1.4114	2.0	.5886	.1667	
	12	1.4114		2.0	.5886	.1667	
	13	.5886		2.8228	1.4114	.1181	
	14	1.4114		2.8228	1.4114	.1181	
	15	1.0	1.5	2.5	1.0	.1333	
Water	1 to 15	(a)	.0195	2.086	2.066	.1598	1.0

[a]Thickness of water is 30 cm.

Results

Solutions to equation (11) were obtained analytically, and equation (12) was solved by using the numerical method of reference 5. A P_3S_4 approximation was used. Values of F were constructed according to equation (13). Results of these calculations and coefficient comparisons are shown in table III.

Analysis

Inspection of the polynomial coefficients shown in table III provides both quantitative and qualitative information that is not obvious from the computed and "measured (transport)" values of F. These coefficients are also more amenable to systematic appraisal than intuitively or randomly chosen comparisons between measurements and calculations.

The qualitative aspect of the analysis is considered first. Generally speaking, the coefficients have (1) sign, (2) magnitude, (3) relative size compared to other coefficients in the same polynomial, and (4) relative size compared to the analogous coefficient in the opposite polynomial. When these four characteristics are considered, the coefficients are found to yield the following qualitative information:

(1) A_0: The model (diffusion) calculated high F values on the average, over the ranges of parameters studied. This suggests that the model is systematically in error and by how much (A_0 is the value of F at the origin of n-dimensional space).

(2) A_1: The model gives too large a decrease in F for a given change in Σ_a. This tendency of diffusion theory can be shown from known physics of this particular problem (the diffusion neutron flux is higher). The point is that this behavior was discovered without physical arguments.

(3) $A_2, A_{22}, A_{12}, A_{23}$: The model and experiment (transport model) both suggest that the x_2 variable need be explored no further. This is because these coefficients are within the standard error of the fits. Had the experimental coefficients been much larger, some development in the model in the treatment of Σ_s would have been indicated.

(4) A_3: The comments for A_1 apply here also. This too was expected on physical grounds because Σ_a and thickness are closely related in terms of neutron flux.

(5) A_{11}, A_{33}: These coefficients show the model overestimating the second-order effects of Σ_a and of thickness. This is consistent with the effects seen in A_1 and A_3.

(6) A_{13}: This comparison shows that Σ_a and thickness have a somewhat weaker interdependence than the model indicates.

TABLE III. - SUMMARY OF CALCULATED RESULTS FOR EXAMPLE

Experiment	Thermal neutron flux perturbation factors, F			
	Diffusion model		Transport (experiment) model	
	Exact	Polynomial	Exact	Polynomial
1	0.185596	0.186161	0.171724	0.172466
2	.251756	.251286	.231429	.231328
3	.119975	.120266	.110970	.110656
4	.249458	.249026	.230832	.229884
5	.116985	.117066	.109311	.109832
6	.102208	.101780	.094086	.093827
7	.187954	.188714	.172417	.173200
8	.101009	.101750	.093598	.094349
9	.068006	.066668	.063797	.062516
10	.094872	.099726	.093027	.092789
11	.121624	.121758	.111365	.111646
12	.052238	.052886	.049052	.049562
13	.121654	.121578	.110887	.110762
14	.051082	.051766	.048396	.049298
15	.068774	.067763	.064186	.062948

Model	Calculated regression coefficients, A(K = 0.730)						
	A_0	A_1	A_2	A_3	A_{11}	A_{22}	A_{33}
Diffusion	0.101750	-0.048707	-0.000845	-0.050208	0.017068	-0.000675	0.017561
Transport	.094349	-.045054	-.000427	-.045534	.015812	-.000705	.015915

	Calculated regression coefficients, A(K = 0.730)			Standard deviation, σ	Goodness-of-fit parameter, R^2
	A_{12}	A_{13}	A_{23}		
Diffusion	0.000520	0.015537	-0.000235	0.00108	0.999988
Transport	.000140	.014647	-.000155	.00118	.999973

(7) A_1/A_3: The model and the experiment show Σ_a and thickness as having the same importance. This confirms the equivalence of Σ_a and thickness as treated by the model.

In summary, the qualitative information primarily indicates (1) Σ_a and thickness are weighted too heavily, (2) Σ_s is unimportant, and (3) the model calculates too high. From this it may be inferred that forcing corrections such as those shown in appendix A would tend to correct the model response.

Use of quantitative information depends on whether the investigator wishes to explore the coefficient disagreement by using more fundamental investigations (a remeasurement of Σ_a for this example), or instead wishes only to adjust the independent

parameters in his model to obtain the correct response. If an adjustment of parameters is desired, the actual calculated coefficients for measurements and calculations can be used, as shown in appendix A. The point here is that the source of the disagreement (Σ_a and thickness) has been located by the coefficient comparison method. And effective corrections are possible. The corrections are not arbitrary, as they might be were an intuitive procedure used. They are, instead, made in a logical manner and to the variables that need correcting (and only to these).

CONCLUDING REMARKS

The philosophy underlying this approach to model development springs from the fact that a few causal factors can produce a relatively enormous number of effects. Computer analysis typifies this phenomenon. It is virtually impossible to analyze the effectiveness of a complex model by examination of these effects (computer output) without some clearcut path connecting causes with effects. The polynomials provide this path. Corrective measures may thus be administered to the source of discrepancies rather than resorting to normalization of experimental and calculated results as is frequently done.

The proposed method of model evaluation and development provides a procedure which separates the various effects of each variable on selected responses. By comparing regression coefficients from measurements and calculations, something may be learned that is not intuitively obvious. This procedure can possibly be applied in pure research as well as in design and development. It requires only an elementary understanding of factorial experiment designs, and all computations required for the fits may be performed by hand using the equations shown in reference 3. Skill comes into play only when selecting the variables and analyzing the results. This method will not replace sound analytical or physical arguments, but, in some cases, it may provide information that cannot be obtained in any other way.

Lewis Research Center,
National Aeronautics and Space Administration,
Cleveland, Ohio, February 27, 1969,
120-27-04-54-22.

APPENDIX A

DERIVATION OF EFFECTIVE x_i VALUES

The following correction may be used to correct an independent variable so that it will produce the correct (experiment) response. The derivation is shown for a second-order polynomial (with three variables), but it is adaptable to higher-order designs.

The problem may be stated in two parts. First, given the polynomials

$$F = A_0 + A_1x_1 + A_2x_2 + A_3x_3 + A_{11}x_1^2 + A_{22}x_2^2 + A_{33}x_3^2 + A_{12}x_1x_2 + A_{13}x_1x_3 + A_{23}x_2x_3 \tag{A1}$$

and

$$F' = A'_0 + A'_1x_1 + A'_2x_2 + A'_3x_3 + A'_{11}x_1^2 + A'_{22}x_2^2 + A'_{33}x_3^2 + A'_{12}x_1x_2 + A'_{13}x_1x_3 + A'_{23}x_2x_3 \tag{A2}$$

(where the prime denotes the model), if the A_0 and A'_0 terms are alike, we want to select a function f_i such that, for any x_i, x_j, and x_k,

$$A_ix_i + A_{ii}x_i^2 + A_{ij}x_ix_j + A_{ik}x_ix_k = A'_i(f_ix_i) + A'_{ii}(f_ix_i)^2 + A'_{ij}x_j(f_ix_i) + A'_{ik}x_k(f_ix_i) \tag{A3}$$

where x_i, x_j, and x_k are the three variables involved and $i \neq j \neq k$. Solving equation (A3) for f_i gives

$$f_i = -\frac{b}{2} \pm \frac{1}{2}\left(b^2 - 4c\right)^{1/2} \tag{A4}$$

where

$$b \equiv (A'_{ii}x_i)^{-1}(A'_i + A'_{ij}x_j + A'_{ik}x_k) \tag{A5}$$

and

$$c \equiv (A'_{ii}x_i)^{-1}(A_{ii}x_i + A_{ij}x_j + A_i + A_{ik}x_k) \tag{A6}$$

Second, if the A_0 and A'_0 coefficients do not match, the model is not calculating the correct nominal value of the response because A_0 and A'_0 are the adjusted averages of all responses. Also, they are the responses at the center point of the design (at $x_i = x_j = x_k = 0$). No multiplicative correction is possible because at the zero level of the independent variable the correction will have no effect.

If the coefficient inspection reveals that only one variable, say x_i, is important, an additive correction Δ can be determined by equating the polynomials at the point in question. This would give

$$x_i^{eff} = x_i + \Delta \tag{A7}$$

where

$$\Delta = -\frac{d}{2} \pm \frac{1}{2}\left(d^2 - 4e\right)^{1/2} \tag{A8}$$

with

$$d = \left(2x_i + \frac{A'_i}{A'_{ii}}\right) \tag{A9}$$

and

$$e = (A_0 - A'_0) + (A_i - A'_i)x_i + \left(A_{ii} - A'_{ii}\right)x_i^2 \tag{A10}$$

If more than one independent variable is important and also if the constant terms (A_0 and A'_0) do not match, two or more equations are required to determine the additive corrections that will force agreement. In this case, polynomials can be constructed for other responses, and these polynomials can also be equated.

Other corrective measures may suggest themselves, depending on the physical problem. The point is that corrective measures become much more logical when the polynomials are compared on a coefficient basis.

APPENDIX B

ORTHOGONAL COMPOSITE FACTORIAL DESIGNS

Factorial designs are given for four, five, and six variables.

TABLE IV. - FOUR-VARIABLE DESIGN[a]

CD \ A	1	1	2	3	3	3	4	4	5
B	3	2	4	1	3	5	2	4	3
13					11				
22		2	6				17	21	
24		3	7				18	22	
31					12				
33	1			10	13	16			25
35					14				
42		4	8				19	23	
44		5	9				20	24	
53					15				

[a]

Coded level	Level
$-\delta = -1.4142$	1
-1.0	2
0	3
1.0	4
1.4142	5

TABLE V. - FIVE-VARIABLE DESIGN[a]

ABC	1 3	2 2	2 4	3 1	3 3	3 5	4 2	4 4	5 3	D E
133										
222		2	3				4	5		
224		6	7				8	9		
242		10	11				12	13		
244		14	15				16	17		
313					18					
331					19					
333	20			21	22	23			24	
335					25					
353					26					
422		27	28				29	30		
424		31	32				33	34		
442		35	36				37	38		
444		39	40				41	42		
533					43					

[a] Coded level Level

Coded level	Level
$-\delta = -1.5960$	1
-1.0	2
0	3
1.0	4
1.5960	5

TABLE VI. - SIX-VARIABLE DESIGN[a]

	1	2	2	2	2	3	3	3	3	3	4	4	4	4	5	A
	3	2	2	4	4	1	3	3	3	5	2	2	4	4	3	B
	3	2	4	2	4	3	1	3	5	3	2	4	2	4	3	C
133								36								
222		2	10	18	26						45	53	61	69		
224		3	11	19	27						46	54	62	70		
242		4	12	20	28						47	55	63	71		
244		5	13	21	29						48	56	64	72		
313								37								
331								38								
333	1	1				34	35	39	43	44					77	
335								40								
353								41								
422		6	14	22	30						49	57	65	73		
424		7	15	23	31						50	58	66	74		
442		8	16	24	32						51	59	67	75		
444		9	17	25	33						52	60	68	76		
533								42								
DEF																

[a] Coded level Level

$-\delta = -1.7606$	1
-1.0	2
0	3
1.0	4
1.7606	5

These designs will give a response equation the same as equation (1). Coefficients are given by equations (2) to (6).

APPENDIX C

SYMBOLS

A_n	first-order regression coefficient for experiment
A'_n	first-order regression coefficient for model
A_{nn}	second-order regression coefficient for experiment
A'_{nn}	second-order regression coefficient for model
D	thermal neutron diffusion coefficient
F	thermal neutron flux perturbation factor
K	constant in response equation
m	total number of experiments or calculations
n	number of independent variables
Q'	dependent variable in response equation
Q_1	observed value of Q from an experiment or calculation
R^2	goodness-of-fit parameter
S	source function in transport model
$\hat{S}$	source function in diffusion model
X	slab thickness in example calculation
$\hat{x}$	space variable in example calculation
x_n	independent variable in response equation
x'_n	fixed value of independent variable in an experiment or calculation
Δ	additive correction
δ	variable level required for orthogonality
$\overline{\mu}_o$	average cosine of thermal neutron scattering angle
Σ^t	thermal neutron total cross section
Σ_a	thermal neutron absorption cross section
Σ_s	thermal neutron scattering cross section
Σ_{tr}	thermal neutron transport cross section
σ	standard deviation
φ	nondirectional thermal neutron flux
ψ	directional neutron flux in transport model
$\hat{\Omega}$	direction of neutron flow in transport model

Subscripts:

i, j, k	independent variable indices
l	number of experiment

REFERENCES

1. Draper, N. R.; and Smith, H.: Applied Regression Analysis. John Wiley & Sons, Inc., 1966.
2. Davies, Owen L.: The Design and Analysis of Industrial Experiments. Second ed., Hafner Publishing Co., 1956.
3. Krase, J. M.; and Cyl-Champlin, C.: Synthetic Experiment Design Techniques in Reactor Analysis. Rep. APEX-303, General Electric Co., Nov. 30, 1956.
4. Glasstone, Samuel; and Edlund, M. C.: The Elements of Nuclear Reactor Theory. D. Van Nostraud Co., Inc., 1952.
5. Barber, Clayton E.: A FORTRAN IV Two-Dimensional Discrete Angular Segmentation Transport Program. NASA TN D-3573, 1966.

NASA-Langley, 1969 —— 19 E-4916

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

POSTMASTER: If Undeliverable (Section 158
Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546