MATCHACURVE-3:

Multiple-Component and Multidimensional Mathematical Models for Natural Resource Studies

Chester E. Jensen

INTERMOUNTAIN FOREST & RANGE EXPERIMENT STATION
Ogden, Utah 84401

USDA Forest Service Research Paper INT-146, 1973

THE AUTHOR

CHESTER E. JENSEN is Principal Statistician for the Intermountain Forest and Range Experiment Station. Previously, he held the same position at the Northeastern and Central States Forest Experiment Stations. Mr. Jensen, who holds a master of forestry degree from Michigan State University, received his training in graduate statistics at Iowa State University.

USDA Forest Service
Research Paper INT-146
December 1973

MATCHACURVE-3:

Multiple-Component and Multidimensional Mathematical Models for Natural Resource Studies

Chester E. Jensen

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Forest Service
U.S. Department of Agriculture
Ogden, Utah 84401
Robert W. Harris, Director

CONTENTS

ABSTRACT

Mathematical model development procedures for graphed relations between variables are presented. As an extension of single-component two-dimensional model alternatives given in the two previous papers of the Matchacurve series, the author concentrates on multiple-component and multidimensional modeling. These procedures are particularly useful in describing unique main effects and interactions. A detailed application is given for a heavily convolute surface developed from "live" data.

INTRODUCTION

Phenomenal development of computer capability almost everywhere has sparked a surge in the use of mathematical models. Their use varies from the relatively brief statistical evaluation of model performance on new data sets to the automation of repeated model estimates as inputs to extensive compilation processes.

Of primary concern in this paper is the development of such models from a specific source, namely, *graphed relations* between continuous variables. The graphs may have been generated from theory, experience, graphic analysis of data, or from some combination of these. But, whatever the source, it is assumed that the graphs *are at hand* and that mathematical expressions for these relationships are required by the analyst.

We limit our task to finding transformations of the independent variables that, when scaled to the graphed relation, emulate the latter with a degree of accuracy satisfactory to the analyst.

Hoerl (1954) assembled a broad array of mathematical curve-form alternatives and presented numerous examples graphed to a relatively constant scale. Those conforming generally to the data trends (or curves from any other source) to be emulated could be fitted to the data (e.g., by least squares) or simply scaled to any graph. Performance could then be evaluated by the analyst and the best alternative adopted. Bartlett (1947) and Draper and Hunter (1969) recognized the necessity for allowing the analyst to make this judgment based on his own acceptance criteria. Isolated families of mathematical forms were discussed by Dolby (1963), Box and Tidwell (1962), and by Draper and Hunter (1969).

After considering these and other relevant publications on transformation, additional systemization of curve-selection processes seems in order. Moreover, it is felt that we should implement new efforts to bridge the communication gap between the practicing analyst and the more mathematically erudite transformation architect. Tukey and Wilk (1965) may have had such efforts in mind when they wrote: "As in the past, much, perhaps most, of even carefully collected data will not be completely analyzed. In part, this is because ... the technology of data analysis is still unsystematized and many of those who could put its tools to good use are unable to do so effectively."

Also, publications to date dwell heavily on the two-dimensional aspects of curve-form description; they place little emphasis on the details of specific approaches to the isolation and description of interactions in three or more dimensional relations. Such interactions are implicit in the majority of relations for the multitude of disciplines we encounter in Forest Service research.

Researchers have expressed a need for modeling techniques that are sensitive to unique main effects and interactions, techniques that can be applied by the scientist group in general. In response, we have elected to work with two fairly flexible families of mathematical forms (Jensen and Homeyer 1970, 1971). Curve-form selection from these sources has been systematized for two-dimensional relations. In *this* paper, the additional descriptive capability of multiple-component mathematical forms is demonstrated and an approach to descriptor development for three or more dimensional relations detailed.

We are concerned primarily with the mathematical description of established graphic relationships. However, it seems advisable to begin by putting the whole modeling system (graphic-plus mathematical development) in analytical perspective, particularly where the graphed model is developed *from data*.

Here, guided by constraints known to exist in the relationship being modeled, unique main effects and interactions visible in the data can be drawn with interpretive freedom and few mathematical obstacles. Effects that cannot be identified in the data, are negligible in a practical sense, or are inconsistent with expectation, can be deleted. The graphic model thus derived can be described mathematically and fitted as a unit to original or new data sets by least squares. Finally, the fitted model can be evaluated statistically for performance.

At the graphing stage, shapes and scales for the effects of variables are not likely to be developed with mathematically rigorous attention to the fitting process. This is a potential source of bias and/or information loss. Ultimately, of course, a complex form-development problem is fraught with similar difficulties in any other modeling system. The solutions are generally heavily assumptive.

As visualized here, the screening of components in a graphic model is strongly subjective and probably insensitive to lesser effects, but it is still rational. The loss in screening power here, compared to that of rigorous testing systems, is counterbalanced by high sensitivity in the identification of complex curve forms admitted to the graphic screening process.

Given an accurate mathematical descriptor of the graphed model, the model can be fitted as a unit to a pertinent data set; i.e., a gross, least-squares adjustment for elevation and scale of the model in space. So, at least at one stage in model development here, objectivity is achieved in the fitting process.

In summary, the strength of this model-building system seems to lie in its sensitivity to complex forms and simplicity of curve-form development. Its weakness is lack of objectivity and lack of sensitivity to minor effects in any screening effort. However, for relations wherein strong interactions are expected, we judge the advantages to predominate.

But, for the majority of applications we have witnessed, prior information is so weak that it provides only general curve-form guidance; it offers little or no information on specific algebraic transformations of the independent variables likely to emulate response curves in the population involved. As a result, the analyst often submits an elementary array of rather arbitrarily selected components (e.g., X_1, X_2, X_1^2, X_2^2, X_1X_2) to the screening process. Here, the potential is great for curve-form bias and for unexploited information in the data, especially when there is a general expectation for curves and/or interactions in the relation. Speed and economy, but not necessarily sensitivity, characterize such an analysis.

The screening process above can be amended to include the development of progressively more sensitive components based on the analysis of residuals from sequentially fitted models. If so, it more nearly possesses the attributes of the graphic model-development approach; i.e., the components are selected to match data trends *visible* in the data.

By way of summation, there is substantial incentive for developing a model graphically when only general expectations exist for a relation (the usual case) and when those expectations include curvilinear main effects or interactions of any form (again, the usual case).

GRAPHIC MODELING

To begin with, let's consider graphic forms as they are necessary to our mathematical modeling strategy. For example, a graph for a two-dimensional (2-D) relation might appear as in figure 1.

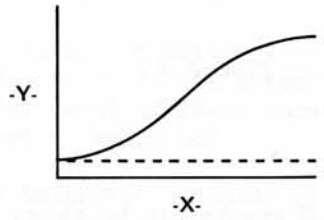


Figure 1

For 3-D relations, a graph might be presented as a complete surface (fig. 2).

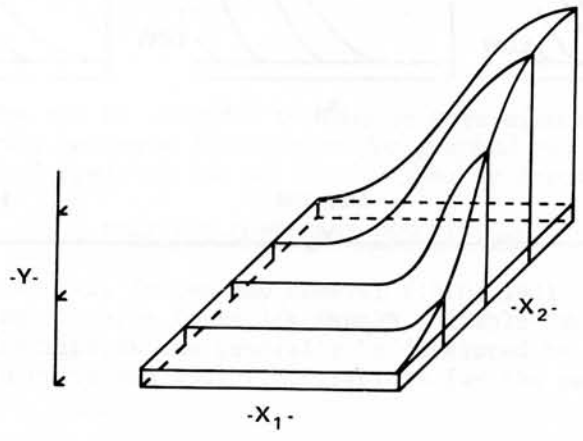


Figure 2

But, for purposes of model development here, it is only necessary to show 2-D graphs (e.g., Y over X_1) at representative points in the third dimension (fig. 3).

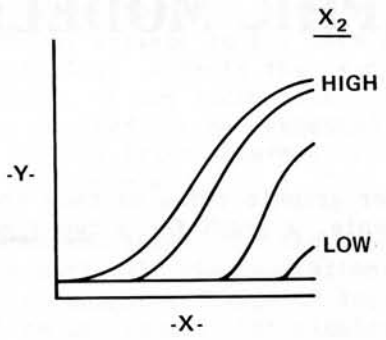


Figure 3

For 4-D relations, the 3-D sets of 2-D graphs can be shown at representative points in the fourth dimension (fig. 4),

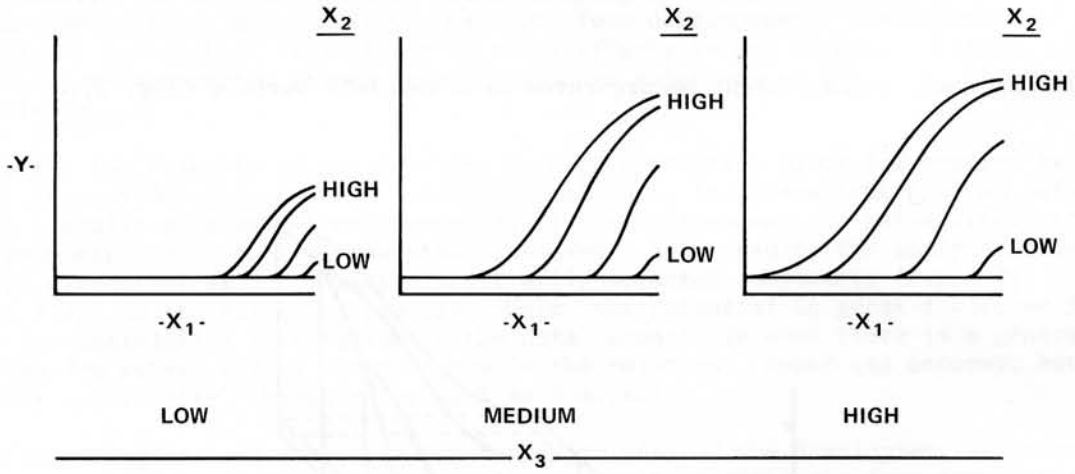


Figure 4

and so on for more dimensions.

MATHEMATICAL DESCRIPTORS

Two Dimensions

GENERAL

This system of graphic display might be described as *ultimately 2-D*, as are the mathematical models developed here. For example, in any *one* of the foregoing 3-D graphs, the specific 2-D sigmoidal shape of Y over X_1 would be identified from Matchacurve-1 (Jensen and Homeyer 1970) for each level of X_2.

Changes in the shape-controlling parameters (P_i) of these sigmoids would then be expressed as 2-D functions of X_2, as would the intercepts and scalars (the differences between maximum heights and intercepts) of these curves. The descriptor for the surface, composed of 2-D components, would be:

$$Y = \text{intercept} + \text{scalar (sigmoid)}$$

where:

$$\text{intercept} = \text{constant in the above examples, but } \textit{could}$$

$$= f_a(X_2)$$

$$\text{scalar} = f_b(X_2)$$

$$\text{sigmoids} = f(P_i), \text{ and}$$

$$P_i = f_i(X_2)$$

The same procedures can be extended to four or more dimensions. *Then, the analyst's capability for developing accurate descriptors for graphed relations involving two or more dimensions depends largely on his ability to find or create accurate* 2-D *descriptors.*

MULTIPLE-COMPONENT DESCRIPTORS

In Matchacurves-1 and -2, Jensen and Homeyer (1970, 1971) provide 2-D descriptors for a fairly broad array of curve forms.[1/] When a suitable "match" for a graphed curve cannot be found, 2-D descriptors can generally be developed by sequentially "matching" *segments* of the graphed curve and adding descriptors for the parts to arrive at the whole.

[1/]The reader should be familiar with the content of both publications to ease assimilation of the procedures herein.

Introductory Example

A commonly encountered kind of curve form that departs from the single-component Standards of Matchacurves-1 and -2 is represented by the graphed curve in figure 5.

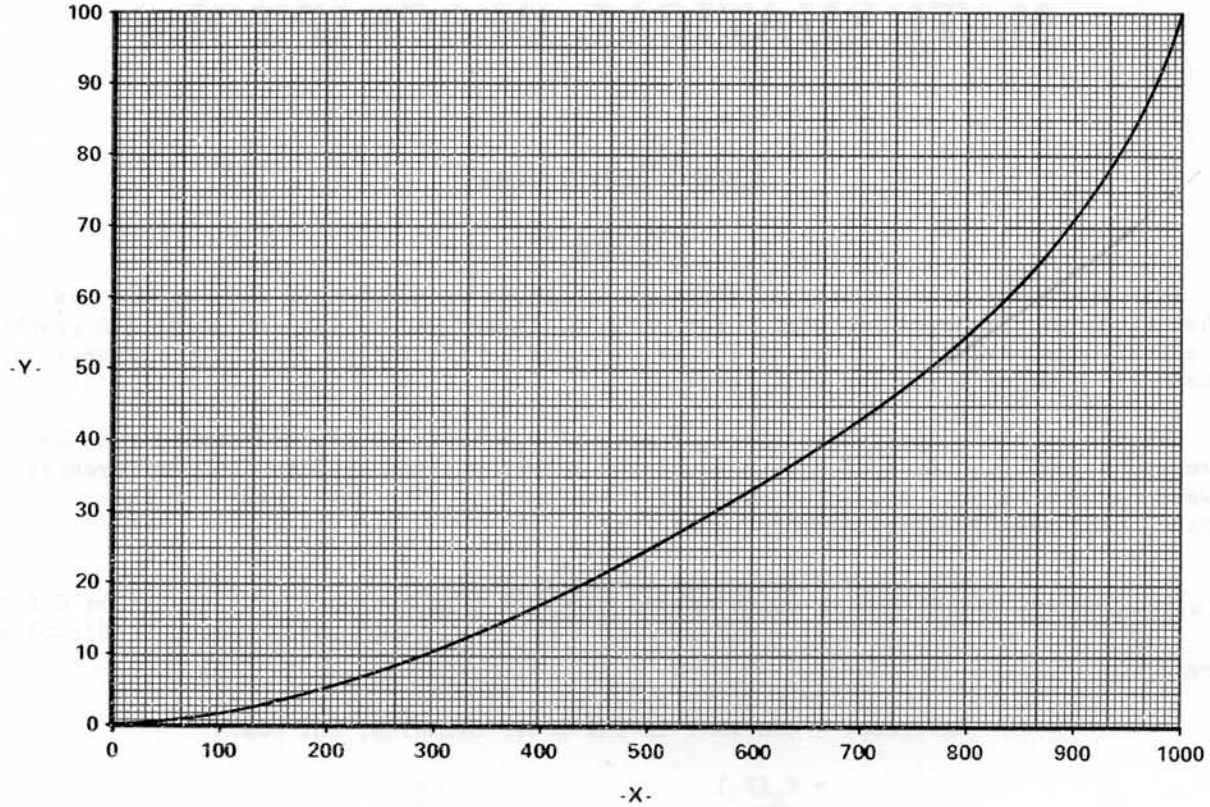


Figure 5

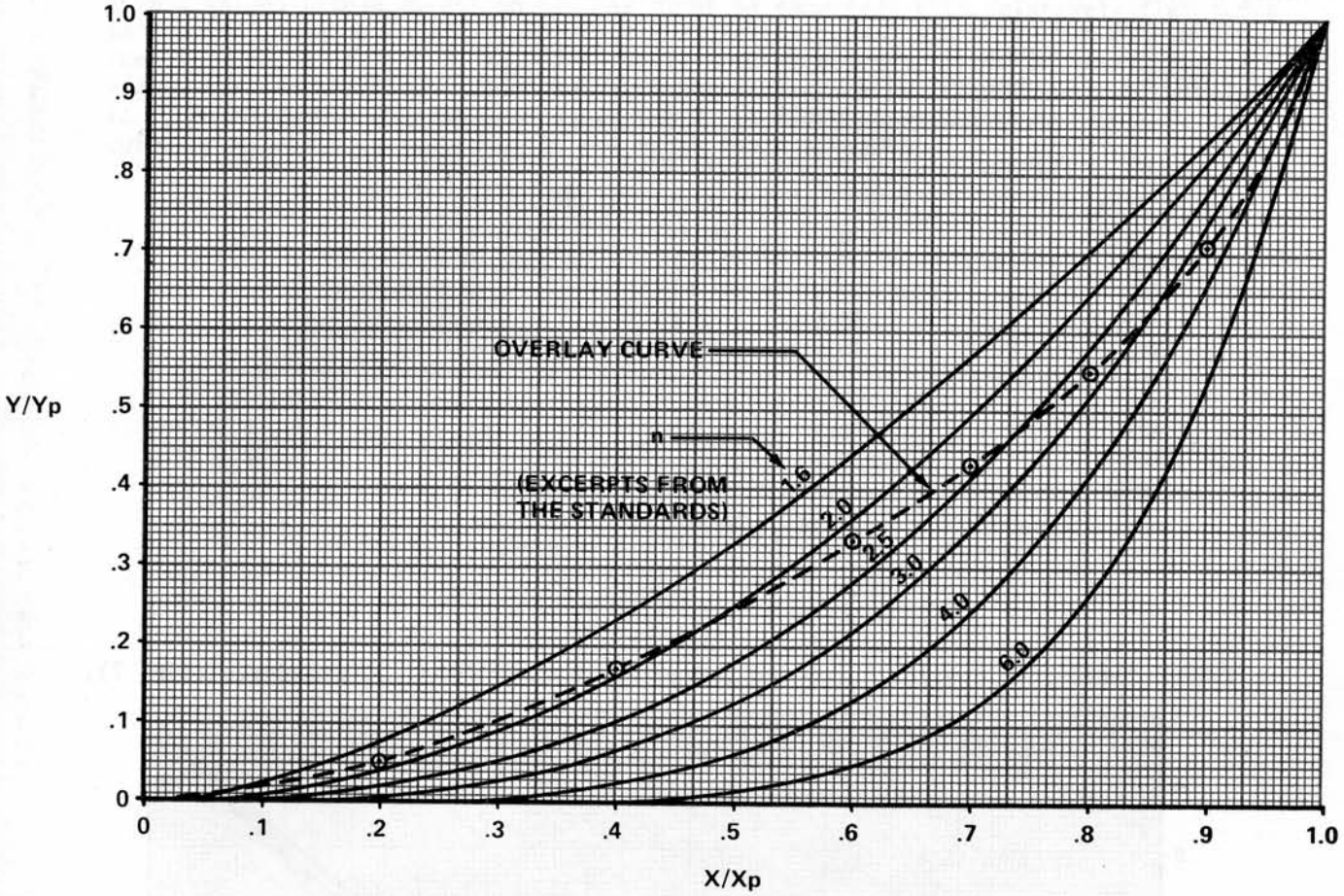
Assume this is the curve to be described. Now, a discerning analyst will immediately *reject* it as a single-component exponential form because of the second pronounced bend in the curve at $700 \leq X \leq 1000$. But, let us run it through Matchacurve-2 procedures to provide a contrast for a two-component form to be developed subsequently.

Representative points of the curve are scaled to a maximum of 1.0 in both X and Y below,

Table 1

X	(X/1000)	Y	(Y/100)
0	0.0	0.0	0.00
200	.2	5.2	.05
400	.4	17.0	.17
600	.6	33.5	.34
700	.7	43.0	.43
800	.8	54.8	.55
900	.9	71.1	.71
1000	1.0	100.0	1.00

and are replotted as a 5- by 7.5-inch (reduced to 4.25 by 6.25 inches for publication) overlay curve, the same size as the Standards (fig. 6):



A-1. – Standards for position A, set 1 – (X-transform to be fitted by least squares = $(X)^n$, $0 \leq X \leq X_p$)

Figure 6

As is evident, there is no suitable match for the overlay curves in the n > 1.0 Standards (set A-1). Although overlay curves are not shown, the same holds true in the 0 < n < 1.0 (set A-2) and the n < 0 Standards (set A-3).

Having thus exhausted the *single-component* Matchacurve-2 alternatives, we now resort to description of the graphed curve in two pieces. The first part selected covers the range $0 \leq X \leq 700$ since that portion of the curve in figure 5 has only one

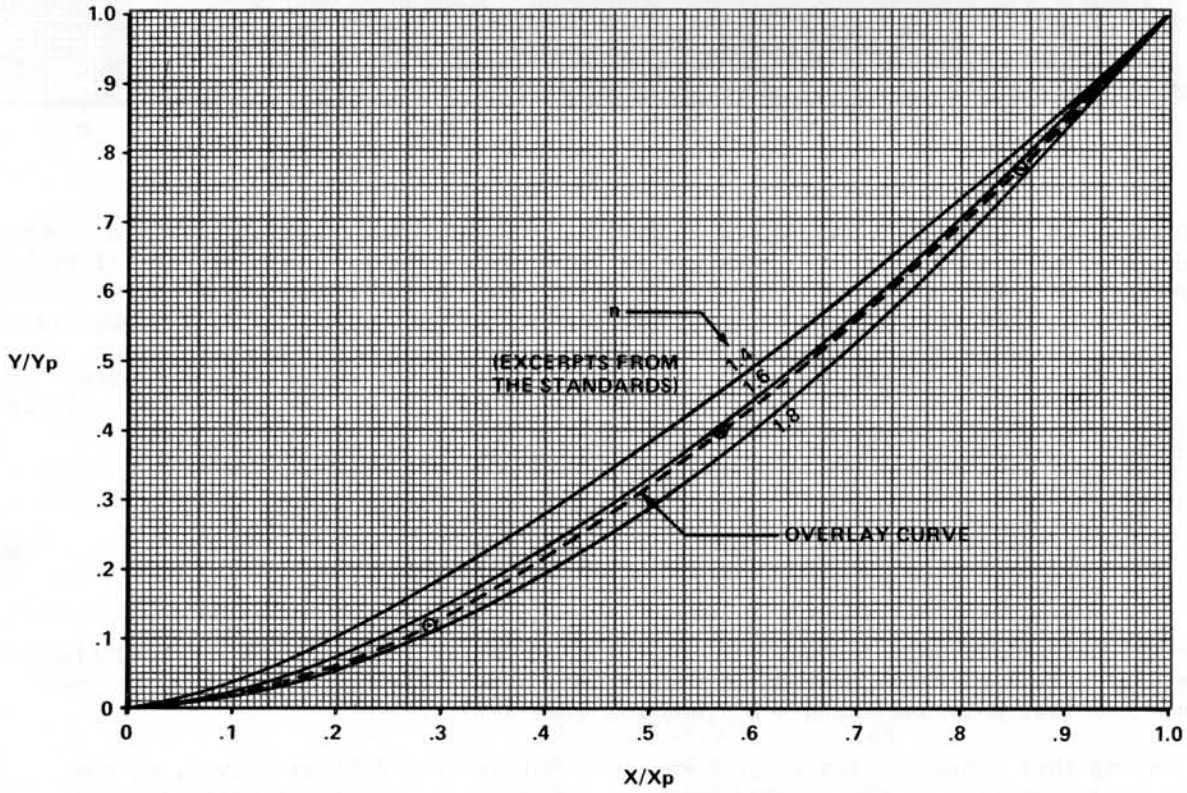
point of bend and seems reasonably similar to the conformation of the n > 1.0 Standards. Next, X and Y are scaled to 1.0 at X = 700 (table 2, columns 1-4).

Table 2

X	(X/700)	Y	(Y/43.0)	⋮	$bX^{1.65}$ [1/]	$Y-bX^{1.65}$
0	0.00	0.0	0.00		0.0	0.0
200	.29	5.2	.12		5.4	- .2
400	.57	17.0	.40		17.1	- .1
600	.86	33.5	.78		33.3	.2
700	1.00	43.0	1.00		43.0	.0
800		54.8			53.6	1.2
900		71.1			65.1	6.0
1000		100.0			77.5	22.5

[1/] $b = 43.0/(700)^{1.65} = 0.0008691$

After plotting an overlay curve and comparing it to the n > 1.0 Standards (fig. 7),



A-1. – Standards for position A, set 1 – (X-transform to be fitted by least squares = $(X)^n$, $0 \leq X \leq X_p$)

Figure 7

we find that $X^{1.65}$ gives a good match for the curve in the $0 \leq X \leq 700$ range.[2] $X^{1.65}$, scaled to 43 at X = 700, gives a fairly close approximation of the representative Y values (table 2, column 5). As might be expected, this relatively flat curve is too flat when extended to represent the desired curve in the $700 < X \leq 1000$ range. The differences, 1.2, 6.0, and 22.5 (table 2, last column), represent the curve of values that must still be added to match the desired curve. Needed in the descriptor is a second component that is zero in the $0 \leq X \leq 700$ range and about equal to the curve of values just indicated in the range $700 \leq X \leq 1000$.

By scaling X and $(Y-bX^{1.65})$ to 1.0 at X = 1000 (table 3, columns 1-4),

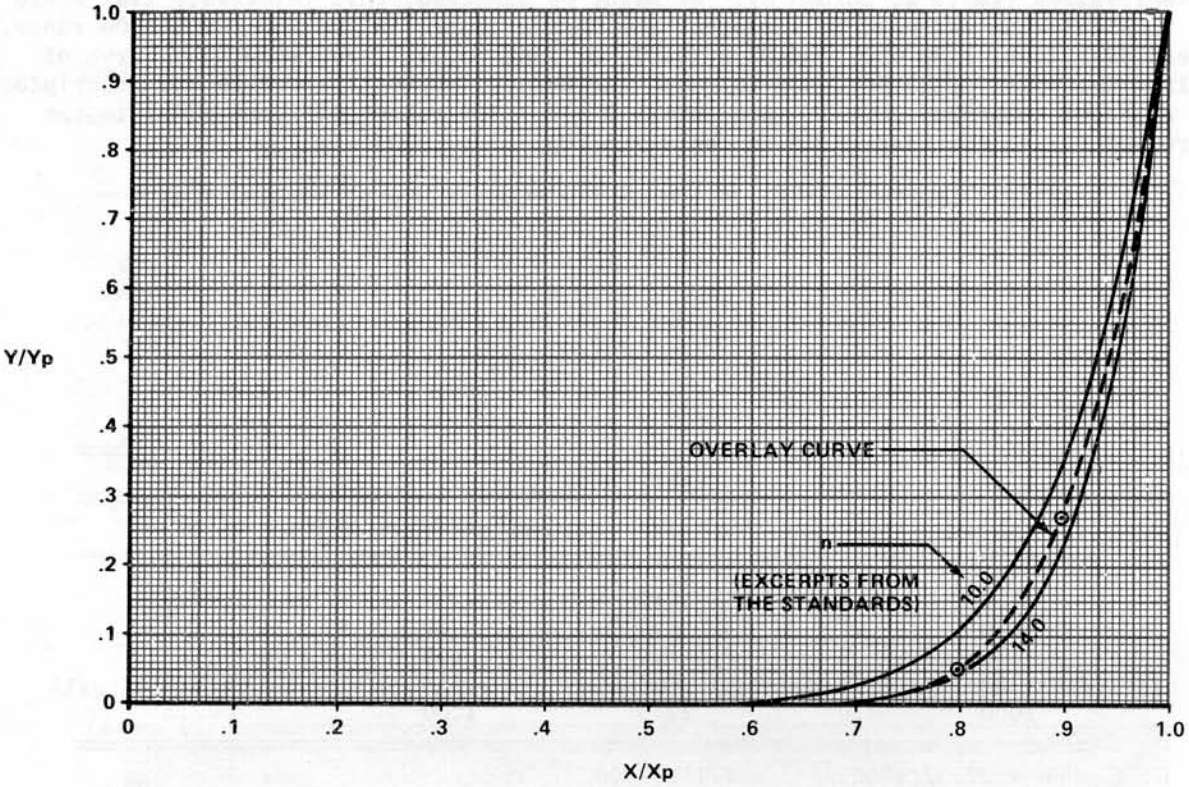
Table 3

X	(X/1000)	$(Y-bX^{1.65})$, Y_1	$(Y_1/22.5)$	$bX^{12.5}$ [1]
0	0.0	0.0	0.00	0.0
700	.7	.0	.00	.3
800	.8	1.2	.05	1.4
900	.9	6.0	.27	6.0
1000	1.0	22.5	1.00	22.5

[1] $b = 22.5/(1000)^{12.5} = 7.115 \times 10^{-37}$

[2] The generating of curves that lie between the Standards of either Matchacurve-1 or -2, along with the compounding and verification of more complex forms, virtually necessitates ready access to a computer. A small, desk-top computer is ideal for the tasks at hand. It should have at least: 250 words of program storage, 10 storage registers, capability for X^n, e, conventional mathematical operations, and recorded input and output potential.

making an overlay curve, and comparing it to the n > 1.0 Standards (fig. 8),



A-1. – Standards for position A, set 1 – (X-transform to be fitted by least squares = $(X)^n$, $0 \leq X \leq X_p$)

Figure 8

we find that X^{13} might match the required curve fairly well, but, on trial, $X^{12.5}$ performs a little better. Scaling $X^{12.5}$ to 22.5 at X = 1000 gives values close to those desired in the $700 \leq X \leq 1000$ range (table 3, column 5). Then the complete descriptor is:

$$\hat{Y} = (8.691 \times 10^{-4})\ X^{1.65} + (7.115 \times 10^{-37})\ X^{12.5}$$

with final values listed in table 4. In this case, $\hat{Y}$ is regarded as being satisfactorily close to Y.

Table 4

X	Y	$\hat{Y}$
0	0.0	0.0
200	5.2	5.4
400	17.0	17.1
600	33.5	33.4
700	43.0	43.3
800	54.8	55.0
900	71.1	71.1
1000	100.0	100.0

The computational ideas just shown are applicable to other curve forms, but a larger array of ideas may be necessary to independent descriptor efforts by the reader. The examples that follow are presented with decreasing explanatory detail as seems appropriate to the stage of discussion. We will start with an abbreviated version of descriptor procedures for the above example (fig. 5).

Section A, Multiple Exponentials Without Inflection

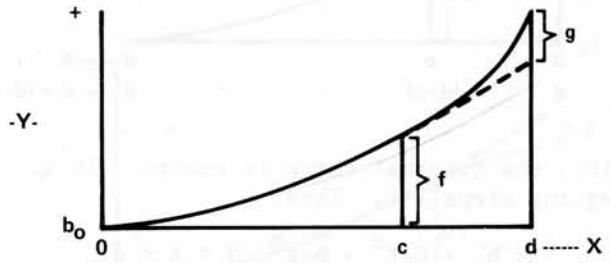


Figure 9

1.--Read a set of representative XY points from the graphed curve (fig. 9). Select one X-value in this set, point c, such that the curve over the range 0 to c appears to have about the same general conformation (one point of bend) as the Standards of Matchacurve-1 (sets A-1, A-2, or A-3 for this spacial orientation).

2.--"Match" this exponential curve in the range 0 to c and determine n in X^n.

3.--Scale X^n to f at c, $b_1 = f/c^n$. Then, $\hat{Y}_1 = b_o + b_1X^n$.

Extend $\hat{Y}_1$ over the range c to d (dotted line).

4.--Match the exponential curve of residuals over the range 0 to d and determine m in X^m.

5.--Scale X^m to g at d, $b_2 = g/d^m$. Then, $\hat{Y}_2 = b_2X^m$.

6.--Add components $\hat{Y}_1$ and $\hat{Y}_2$. Then,

$$\hat{Y} = b_o + b_1X^n + b_2X^m, \quad b_o = 0.$$

If the intercept is some value other than zero, the above procedures apply to the *differences* between the intercept and the graphed curve. And, as before,

$$\hat{Y} = b_o + b_1X^n + b_2X^m, \quad \text{but } b_o \neq 0$$

Alternative orientations of the above curve in space can be handled simply, as follows:

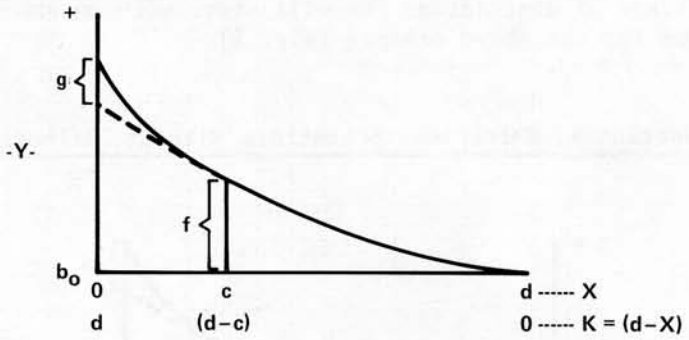


Figure 10

In this case (fig. 10), the original curve is reversed in X. Substitute K = (d-X) for X and follow the foregoing steps 1-6. Then,

$$\hat{Y} = b_o + b_1K^n + b_2K^m, \quad 0 \leq X \leq d$$

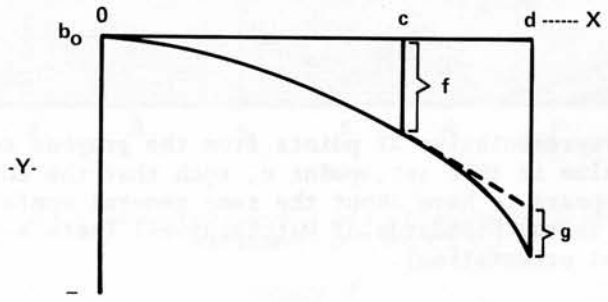


Figure 11

In Figure 11, the original curve is inverted about the intercept b_o. Work with absolute differences between b_o and the curve, for the first component. For the second component, work with absolute differences between the first component curve and the original curve. Follow steps 1-6, but subtract the two components from b_o. Then,

$$\hat{Y} = b_o - b_1X^n - b_2X^m$$

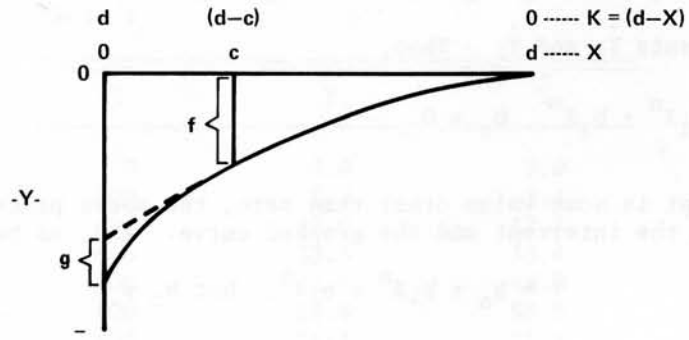


Figure 12

In figure 12, the original curve is both reversed in X and inverted about b_o. Substitute K for X and work with absolute differences as in figure 11. Follow steps 1-6 and *subtract* the two components from b_o. Then,

$$\hat{Y} = b_o - b_1K^n - b_2K^m, \quad 0 \leq X \leq d$$

These mechanics for reorienting a curve have general application and greatly facilitate descriptor-development efforts.

Section B, Multiple Exponentials With Inflection

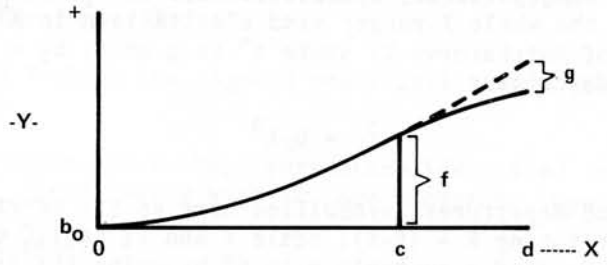


Figure 13

Figure 13 is a variant of the original curve in Section A. Steps 1-6 still apply, except that the last component is *subtracted* from the first. Then,

$$\hat{Y} = b_o + b_1X^n - b_2X^m$$

Alternative orientations of this curve in space can be handled as described under Section A along with appropriate changes in sign for the last component.

Section C, Multiple Exponentials With Flat Central Segment

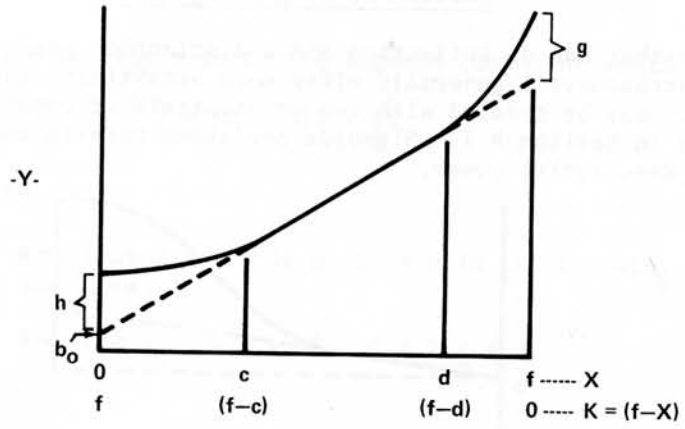


Figure 14

The descriptor for a curve that has an extended flat midsection (fig. 14) can often be treated as follows:

1.--Determine the intercept, b_o, by graphic extension of the flat segment. Calculate the slope of the line. Then,

$$\hat{Y}_1 = b_o + b_1X$$

2.--Determine the residuals of the curve from the straight line for representative points over X. Be sure to include three or four such points at each end of X, where the curve bends away from the straight line.

3.--For the right-end departures, symbolized here as Y_2: scale X and Y_2 to 1.0 over the whole X-range; find a suitable n in X^n by using the Standards of Matchacurve-2; scale X^n to g at f, $b_2 = g/f^n$. Then, the right-end descriptor is:

$$\hat{Y}_2 = b_2X^n$$

4.--For the left-end departures, symbolized here as Y_3: reverse the X-axis by substituting K = (f-X); scale K and Y_3 to 1.0 over the whole range of X; find a suitable m in K^m by using the Standards; and scale K^m to h at f, $b_3 = h/f^m$. Then, the left-end descriptor is:

$$\hat{Y}_3 = b_3K^m$$

5.--Then the entire descriptor is:

$$\hat{Y} = b_o + b_1X + b_2X^n + b_3K^m, \quad 0 \leq X \leq f$$

Variants from this curve include negative departures from either or both ends of the straight line. In such cases, the signs of the corresponding components simply become negative.

Of course, there is no limit to the number of components that can be used in a descriptor, unless the analyst has algebraic simplicity rather than accuracy as an overriding objective. With electronic computer service available almost everywhere, computational complexity should not be a limiting factor.

Section D, Single Sigmoids

For any curve that has an inflection and a distinctive peak (positive or negative), the sigmoids of Matchacurve-1 generally offer more sensitivity than the exponentials. (Note that sigmoids may be created with two exponentials of opposite sign, as is partially detailed in Section B.). Sigmoids are shown here in combination with linear effects for added descriptive power.

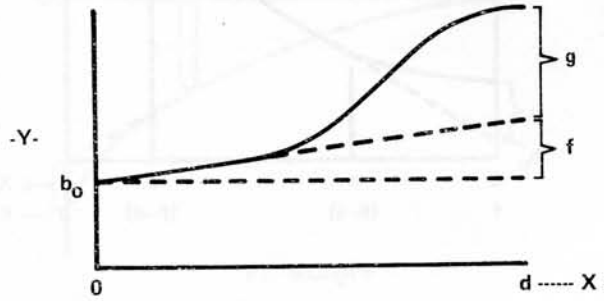


Figure 15

In this case (fig. 15), we can use a straight line with slope $b_1 = f/d$. The intercept plus the straight line is $\hat{Y}_1 = b_o + b_1X$. Then add to $\hat{Y}_1$ a sigmoid[3/] matching the curve differences, Y_d, from the linear function, scaled to g at the point in X (d here) where Y_d peaks. The sigmoid matching process is analogous to steps 1-5 in Section A. Representative sets of X- and corresponding Y_d-values are scaled to 1.0. An overlay curve is made and compared to the sigmoid Standards of Matchacurve-1, and n and I for a matching curve are noted.

The final sigmoid component is $Y_2 = gE$, where:

g is the scalar for E

E, specified as Y/Y_p in Matchacurve-1 (page 4), symbolizes the e-transform, which varies from zero to one in the range XP ± XP.

XP is the point in X where the sigmoid peaks in either a positive or a negative vertical direction.

n and I are appropriate curve-shape parameters identified from the matching curve in the Standards. These, along with XP, are necessary inputs to E.

The final descriptor is then:

$$\hat{Y} = b_o + b_1X + gE$$

Alternative arrangements of the sigmoid about the straight line would require treatment as detailed under Section A, but a brief explanation follows:

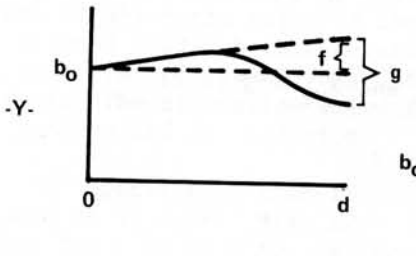


Figure 16

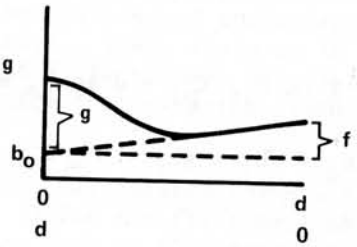


Figure 17

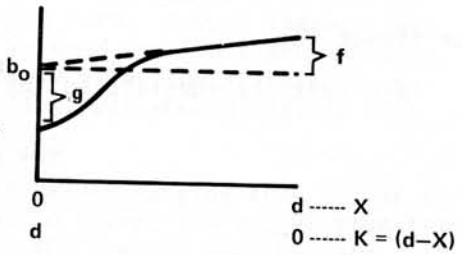


Figure 18

For figure 16

$X_p = d$ in the X-scale and the sign of the sigmoid component would become negative:

$$\hat{Y} = b_o + b_1X - gE$$

For figure 17

K = (d-X) would be substituted for X in E, $X_p = d$ in the K-scale, and the sigmoid would be positive:

$$\hat{Y} = b_o + b_1X + gE, \quad 0 \leq X \leq d.$$

[3/]Appropriate sigmoids can be selected from the Standards of Matchacurve-1.

For figure 18

Again, K = (d-X) is substituted for X in E, X_p = d in the K-scale, and the sigmoid is *subtracted*:

$$\hat{Y} = b_o + b_1X - gE, \quad 0 \leq X \leq d.$$

The same sigmoid alternatives exist when the slope of the straight line is negative, the only change in the descriptors would be that the sign of the first component would be *negative*.

Section E, Multiple Sigmoids

Sometimes the combination of several sigmoidal (upright or inverted) components can be used effectively to describe asymmetrical curves. Each sigmoid is matched and scaled independently and both are summed in the final descriptor.

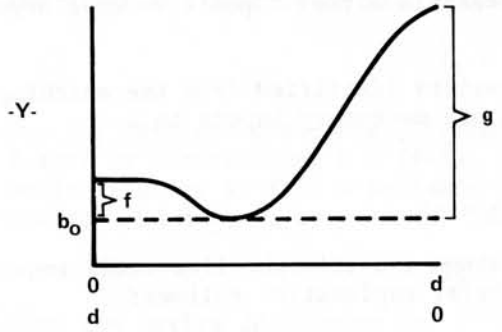


Figure 19

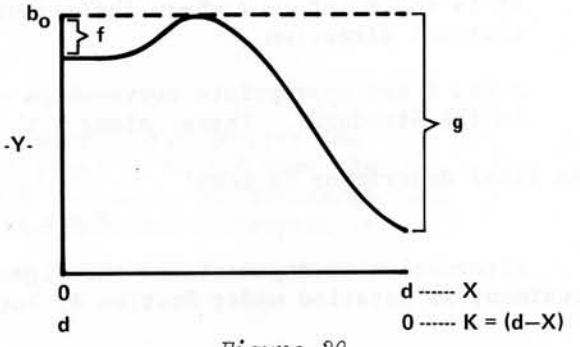


Figure 20

For figure 19

K = (d-X) is substituted for X in E_1, X_{p1} = d in K, and X_{p2} = d in X:

$$\hat{Y} = b_o + fE_1 + gE_2, \quad 0 \leq X \leq d$$

For figure 20

K = (d-X) is substituted for X in E_1, X_{p1} = d in K, and X_{p2} = d in X:

$$\hat{Y} = b_o - fE_1 - gE_2, \quad 0 \leq X \leq d$$

Section F, Exponentials and Sigmoids

Of course, exponentials and sigmoids can be used in combination (fig. 21).

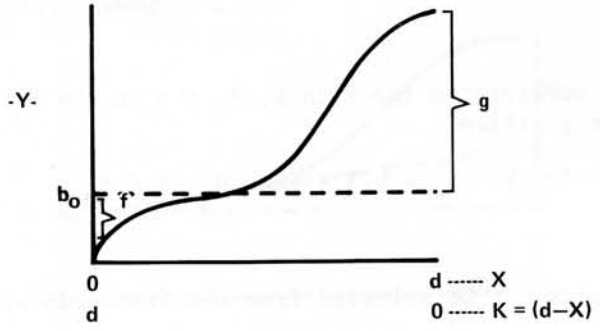


Figure 21

Where:

$$X_P = d \text{ in } X \text{ for } E, \text{ and } \hat{Y} = b_o + gE - (f/d^n)K^n, \quad 0 \leq X \leq d$$

Section G, Bell-Shaped Curves

Bell-shaped curves, in whole or in part, can be summed in a descriptor. And, in this particular example (fig. 22),

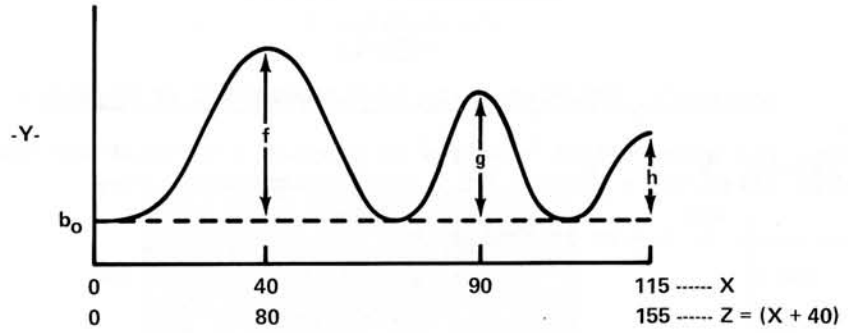


Figure 22

the peaks of the two leftmost bell-shaped curves occur at points other than the extremes of the X-range, which is not true in previous examples involving sigmoids. A peculiarity of E is that its controlled use is within XP ± XP, where XP is the point in the X-range at which the Y-peak (either positive or negative) occurs and XP ± XP are the points at which the tails of the bell-shaped curve drop to zero. Beyond these points, the sigmoid values become negative. Then, where an XP is used such that XP ± XP does not cover the pertinent range of X, it becomes necessary to alter the X-scale so that it does. The discussion below for the foregoing graph (fig. 22) should clarify both the problem and its solution.

"E" specifies a symmetrical, bell-shaped curve with values ranging from 1.0 at XP to zero at XP ± XP. With XP = 40 for the first bell-shaped curve, the range of application for E is 40 ± 40, or from zero to 80. Beyond X = 80, E would actually dip below zero by some unspecified amount. The problem is simply remedied by adding a large enough constant to X such that XP + XP would cover the original maximum of 115. Z = (X + 40) accomplishes this nicely; so the descriptor would be:

For the first curve,

$\hat{Y}_1 = fE_1$, where Z replaces X in E and $X_P = 80$ in Z;

for the second curve,

$\hat{Y}_2 = gE_2$, where X is used without transformation in E, since XP = 90 gives an applicable range for X of zero to 180, covering the X-extreme of 115 as required;

for the third curve, the left half of a bell-shaped curve,

$\hat{Y}_3 = hE_3$, where X is also used without transformation in E since XP ± XP (115 ± 115) covers the pertinent range of X, 0-115.

Then, the entire descriptor is:

$$\hat{Y} = b_o + fE_1 + fE_1 + gE_2 + hE_3, \quad 0 \leq X \leq 115$$

where:

E_1 involves Z

E_2 involves X

E_3 involves X

Section H, Matching Parts of Exponentials or Sigmoids

On occasion, the added effort required to compile a multiple-component descriptor can be avoided by utilizing a *portion* of a single-component curve.

Assume the curve in figure 23 (A) below

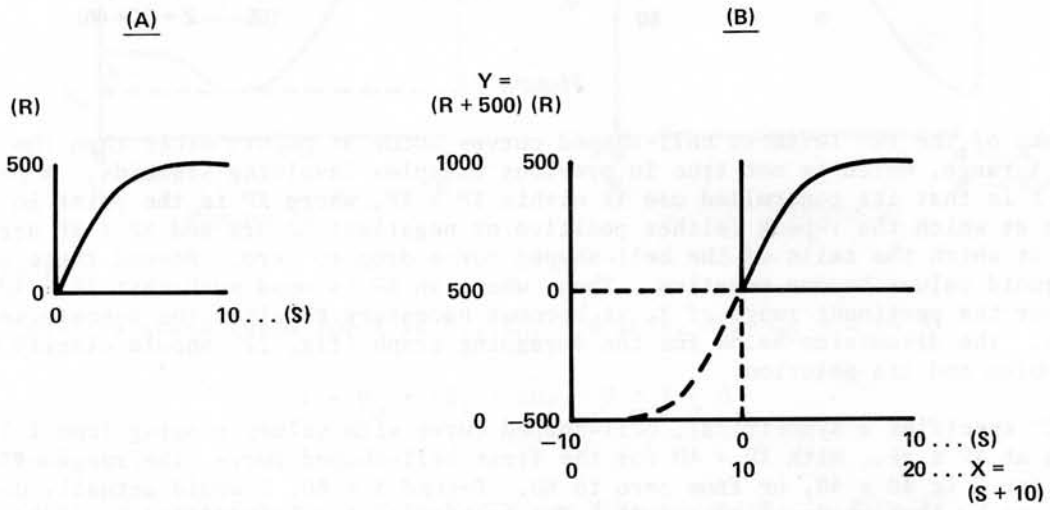


Figure 23

is to be described mathematically and that no satisfactory matching curve has been found in Matchacurves-1 or -2. As an alternative, the analyst *could* undertake a multiple-component description, but another single-component possibility exists. Extend the desired curve in (a) such that the resulting form approximates that of any sigmoid shown in the Standards. Add a constant to the abscissa such that XP ± XP covers the pertinent range of X. (S + 10) turns the trick here. Also, add a suitable constant, 500 in this case, to R for overlay scaling purposes only. For representative points from the right half of this curve, scale the X-Y pairs to a maximum of 1.0, construct an overlay curve, and find a matching Standard for the upper half of the overlay. The adopted single-component standard fitted to the X- and Y-values then represents the original curve in the region $10 \leq X \leq 20$. The resulting descriptor is:

$$\hat{R} = -500 + 1000\ E$$

where:

$$X = (S + 10) \text{ in } E, \ XP = 20, \text{ and } 0 \leq S \leq 10.$$

Procedures for any fraction of any form are analogous.

Summary Statement

In concluding the discussion for *2-D multiple-component descriptors*, we reiterate that the analyst's capability for developing sensitive multidimensional models depends largely on his 2-D talents. *The examples shown here, if understood, give enough procedural background to permit intuitive extension of the system to include more flexible curve-form alternatives.* The "live" example, which will be presented later, merely serves to reinforce a few of these ideas through actual application.

Three Dimensions

GENERAL

Given 2-D curves over X_1 for representative levels of X_2 (fig. 24)

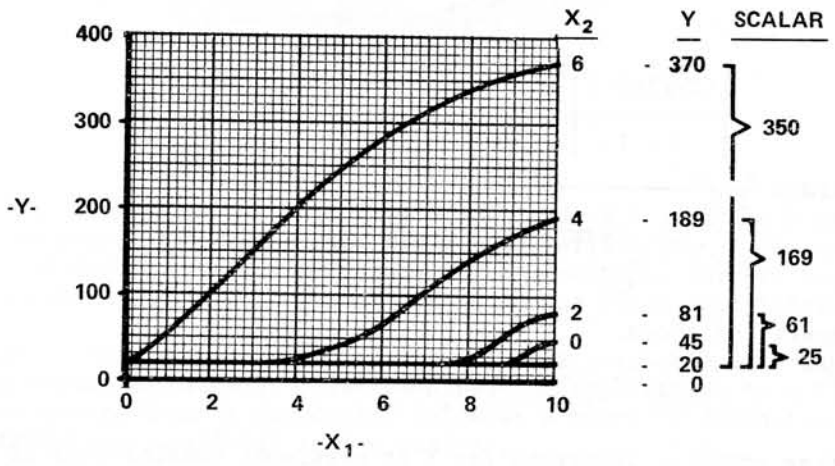


Figure 24

and given a 2-D descriptor,

$$\hat{Y} = \text{intercept} + \text{scalar}\ (X_1\text{-transform})$$

. . . for each such curve, the analyst is in a position to formulate a descriptor for the implied surface. He simply expresses the changing intercepts, scalars, and parameter(s) of the X_1-transform in terms of X_2. He then substitutes appropriately in the original 2-D descriptor.

Assume pertinent information for the four sigmoid curves above has been assembled by using Matchacurve-1 Standards (table 5, columns 1-6).

Table 5

Sigmoid parameters							
X_2	intercept	XP	n	I*	scalar	$\hat{\text{scalar}}$	$\hat{I}$
0	20	10	3.0	0.90	25	25.0	0.900
2	20	10	3.0	.87	61	61.1	.870
4	20	10	3.0	.66	169	169.4	.663
6	20	10	3.0	.10	350	350.0	.100

*Let I = sigmoid inflection point in X_1 as a proportion of the range in X_1.

Here, the intercept and the sigmoid parameters, XP and n, are constant over X_2; the scalar *increases* and I *decreases* exponentially with increasing X_2-values. Using Matchacurve-2,

$$\text{scalar} = 25 + 9.028(X_2)^2$$

$$\hat{I} = 0.9 - 0.003704(X_2)^3$$

By substituting in the original equation for $\hat{Y}$,

$$\hat{Y} = \text{intercept} + \text{scalar}\ (X_1\text{-transform})$$

$$\hat{Y} = 20 + (25 + 9.028(X_2)^2)\ (\text{sigmoid})$$

where:

$$\text{sigmoid} = \frac{e^{-\left|\frac{(X_1/10)-1}{1-I}\right|^3} - e^{-\left(\frac{1}{(1-I)}\right)^3}}{1 - e^{-\left(\frac{1}{(1-I)}\right)^3}}$$

and I is as specified above.

By plotting computer solutions for $\hat{Y}$ at pertinent combinations of X_1 and X_2, we can see that the descriptor lies quite close to the original curves (fig. 25).

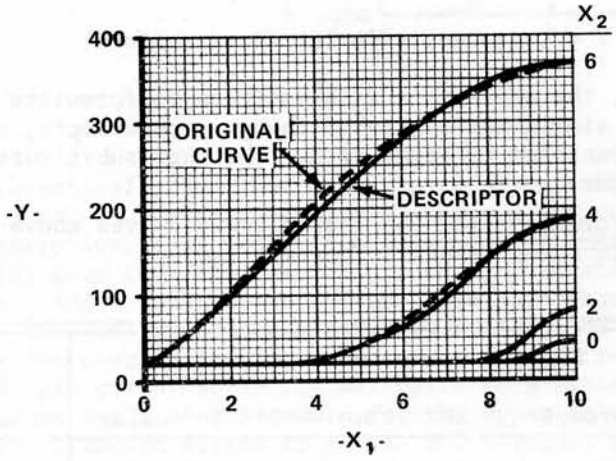


Figure 25

Then, the entire predicted surface appears as follows (fig. 26):

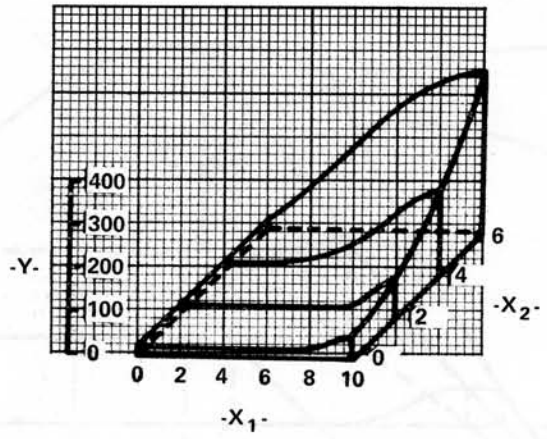


Figure 26

Note that the original curves may be given over X_2 instead of X_1, in which case, the roles of X_1 and X_2 in descriptor development would be interchanged. Also, where multiple-component 2-D descriptors are necessary for the original curves, each component can be treated as a separate X-transform to be added in the final descriptor.

This descriptor system is applicable to virtually any surface and is especially useful in emulating strong interactive relationships. An application to a "live" data set[4] involving such interaction is documented below as a means of demonstrating actual use of some 2-D alternatives available to the analyst.

"LIVE" DATA EXAMPLE

On each of 16 passes over a level target area, 600 gal of fire retardant were dropped from an aircraft. Square-foot coverage and volume of material reaching the ground per 100 ft^2 were measured for each drop along with drop height, windspeed and direction, aircraft speed, temperature, humidity, and so forth. Interest centered on the change in coverage of $\geq$ 2 gal/100 ft^2 over a controlled range of drop heights within a partially controlled range of windspeeds.[5]

Suppose the plane were rolling along the ground (drop height $\doteq$ 3 ft) at normal flying speed and in the absence of wind, it would be expected that a 600-gal drop would be distributed over a relatively narrow strip of ground and that the area covered by $\geq$ 2 gal/100 ft^2 would be held to some nominal value. Increased aircraft (and drop) height should result in greater dispersal; maximum coverage should be reached at some optimal height. As drop height is increased beyond the optimum, dispersal of the retardant from air friction and evaporation should become more complete; coverage should finally reach zero at some relatively great drop height.

So expectation at zero wind is for a bell-shaped curve truncated to the left of the peak. With increasing wind, optimal drop height and peak coverage should diminish,

---

[4] The author is indebted to the Northern Forest Fire Laboratory, Fire Control Technology Project, Missoula, Mont., for the "live" data used here (George and Blakely 1973).

[5] The effects of all other variables were either negligible or unidentifiable in this data set.

since dispersal of the retardant would be accentuated by the horizontal shearing force of the wind. Thus, the expected surface was as shown in figure 27 below.

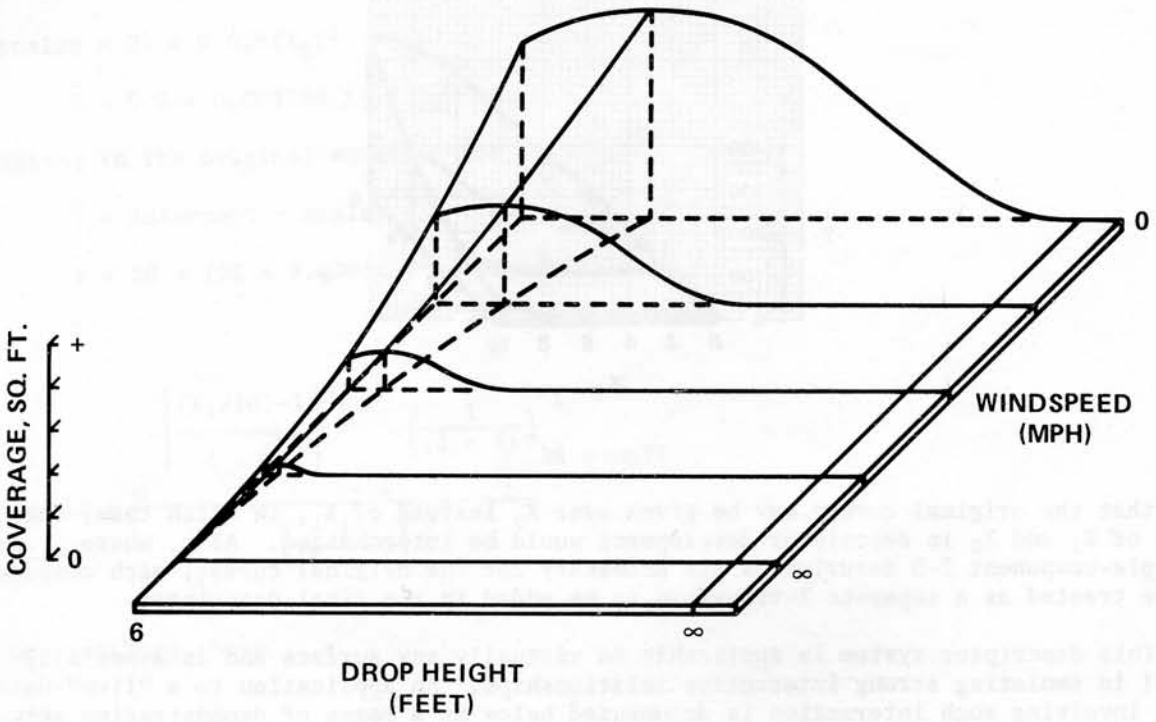


Figure 27

Using these expectations as guides, coverage, C, was curved over height, H, fitting the curves to actual data points by approximate least deviations (Karst 1958) for each of three data groups in wind, W (fig. 28):

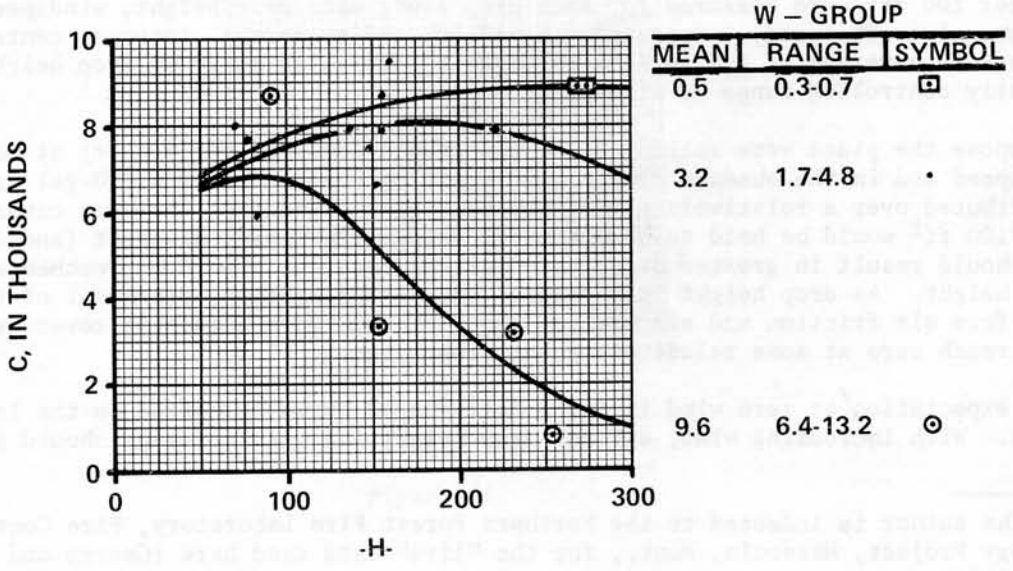


Figure 28

These curves, plotted at their respective W means, were connected over W to form the implied surface, figure 29. And, it can be seen that major features of the expectation are reflected in the data--*even within the limited range of H and W.*

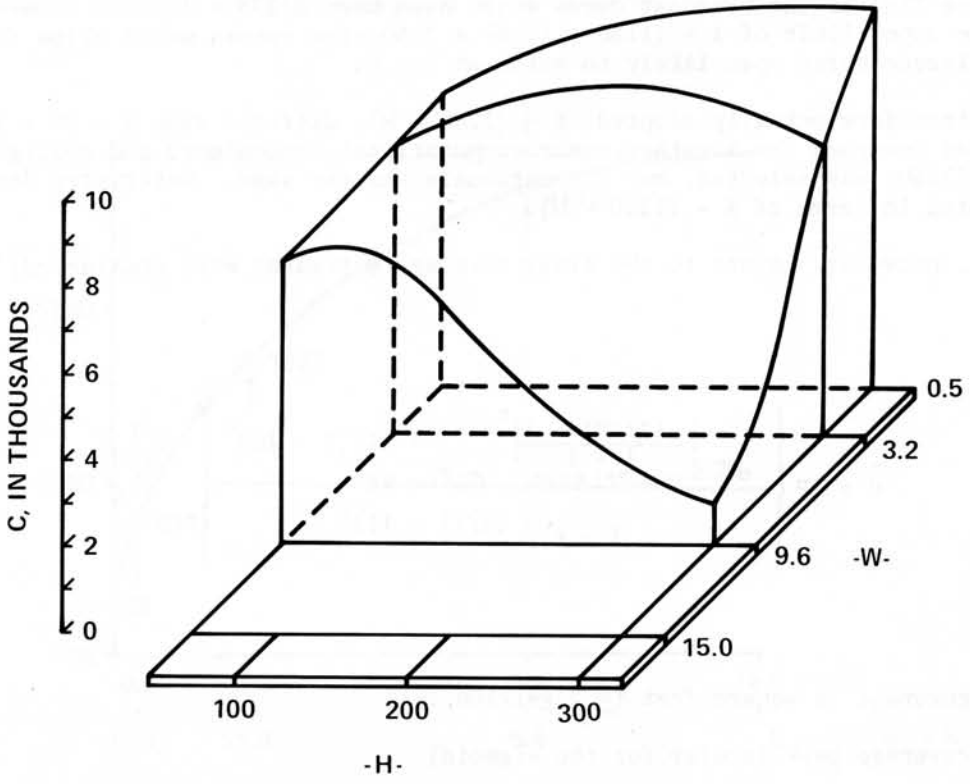


Figure 29

Figure 29, then, is the basic surface for which a descriptor was developed. Assuming that the curves over H (figs. 27, 28) could be suitably represented by segments of symmetric bell-shaped forms, the e-transform of Matchacurve-1 was deemed applicable and the problem of identifying matching e-transforms approached.

The e-transform is limited in that it will yield values of from zero to one in the full bell-shaped form only *within* the range HP ± HP, where HP is the point in H at which the C-values peak. And, it was a unique feature of this example that the maximum H-values for all three curves of the set exceeded 2HP (fig. 30).

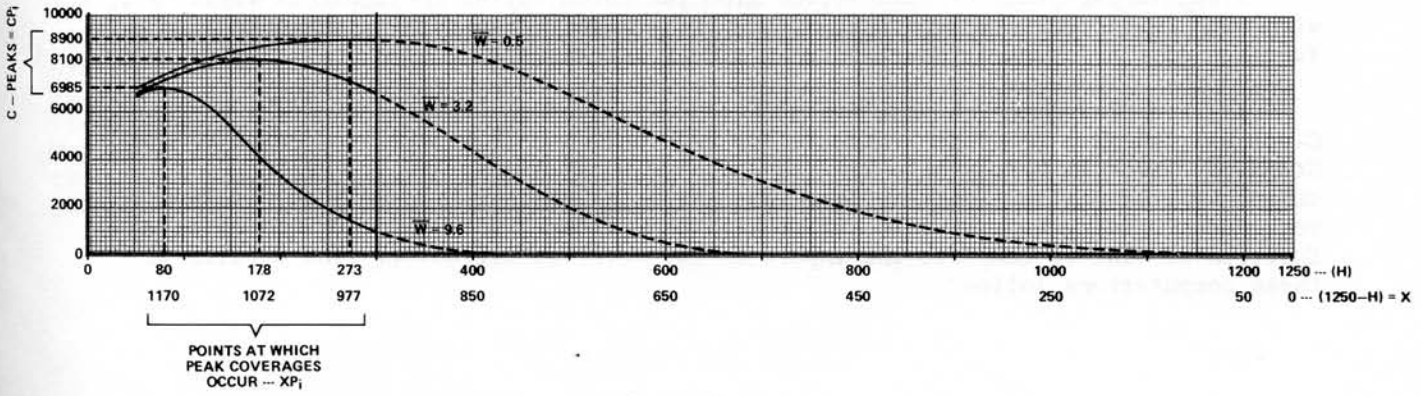


Figure 30

The broadest curve (for W = 0.5) was estimated to peak at H = 273. In this case, the e-transform would have dropped below zero where H exceeded 2HP, or 2(273) = 546, long before the approximate graphed limit of H = 1150 was reached.

One solution would have been to supplant H with a transform such as X = (H + 1000). Then, since 2XP for the broadest curve would have been 2(273 + 1000) = 2546--safely beyond the upper limit of X = (1150 + 1000) = 2150--the excess would allow for the slightly larger curve span likely to exist at W = 0.

The transform actually adopted, X = (1250 - H), differed from X = (H + 1000); the H-scale was reversed for a rather minor computational convenience and a slightly larger constant (1250) was selected, but the rationale was the same. Descriptor development is presented in terms of X = (1250 - H).

Next, necessary inputs to the final coverage estimator were considered:

$$\hat{C} = CP \left\{ \frac{e^{-\left|\frac{(X/XP) - 1}{1 - I}\right|^{n}} - e^{-(1/(1 - I))^{n}}}{1 - e^{-(1/(1 - I))^{n}}} \right\}$$

where,

C = coverage in square feet ($\geq$ 2 gal/100 ft^2)

CP = coverage peak (scalar for the sigmoid)

X = (1250 - H)

XP = the point in X at which C peaks

I = the proportional point in X at which the inflection point of the bell-shaped curve occurs (a) in the range X = 0 to XP for the left half of the curve or (b) in the range (2XP - X) for the right half (fig. 30)

n = the power of the transform that dictates the degree of curvature above and below any inflection point.

Since the XP- and CP- values read from the three curves in figure 27 varied with wind in accord with expectation, they were each expressed as a suitable function of wind by using Matchacurve-1 and -2 techniques.

XP- and CP-values from the foregoing equations along with representative X- and C-values read from one side of the XP for each bell-shaped curve were scaled to the Standard curves in Matchacurve-1. Standards suitably similar to the scaled data curves were identified, and the corresponding n (held constant) and I-values were recorded. Varying over wind in accord with expectation, "I" was described as a function of wind to complete the inputs to the coverage estimator. The details of these computations follow:

For XP = f(W)

Paired W- and XP-values (0.5, 977; 3.2, 1072; and 9.6, 1170) were read from figure 30. These points were plotted and a smooth curve was drawn through them (fig. 31) that extended over the relevant range of W (0-15 mi/h).

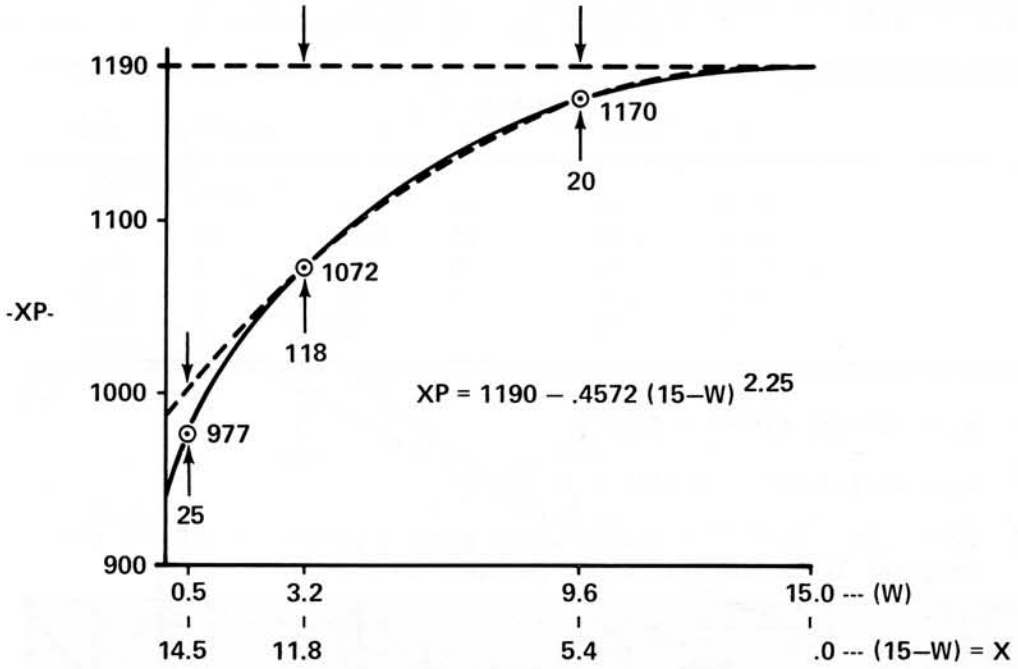


Figure 31

The initial descriptor approach here was to subtract a single-component exponential curve on the reversed W-scale from the intercept, 1190. W was reversed in X = (15-W) to aline large values of the independent variable with those of XP. Next, the three data point X-values and associated coverage differences (absolute) from 1190 were scaled to 1.0 at X = 14.5. An overlay curve was plotted for Matchacurve-2, but was not well matched by any curve in sets A-1, A-2, or A-3 of the Standards; so a two-component approach was adopted. This time, the differences were fitted at X = 0, 5.4, and 11.8 to start with (see the dotted line below the intercept, 1190, in figure 31). Then, the difference (0.25) between the first curve and the desired curve (for $11.8 \leq X \leq 14.5$) was added as a second component. Computations are summarized in table 6 and overlay curves are shown in figure 32.

Table 6

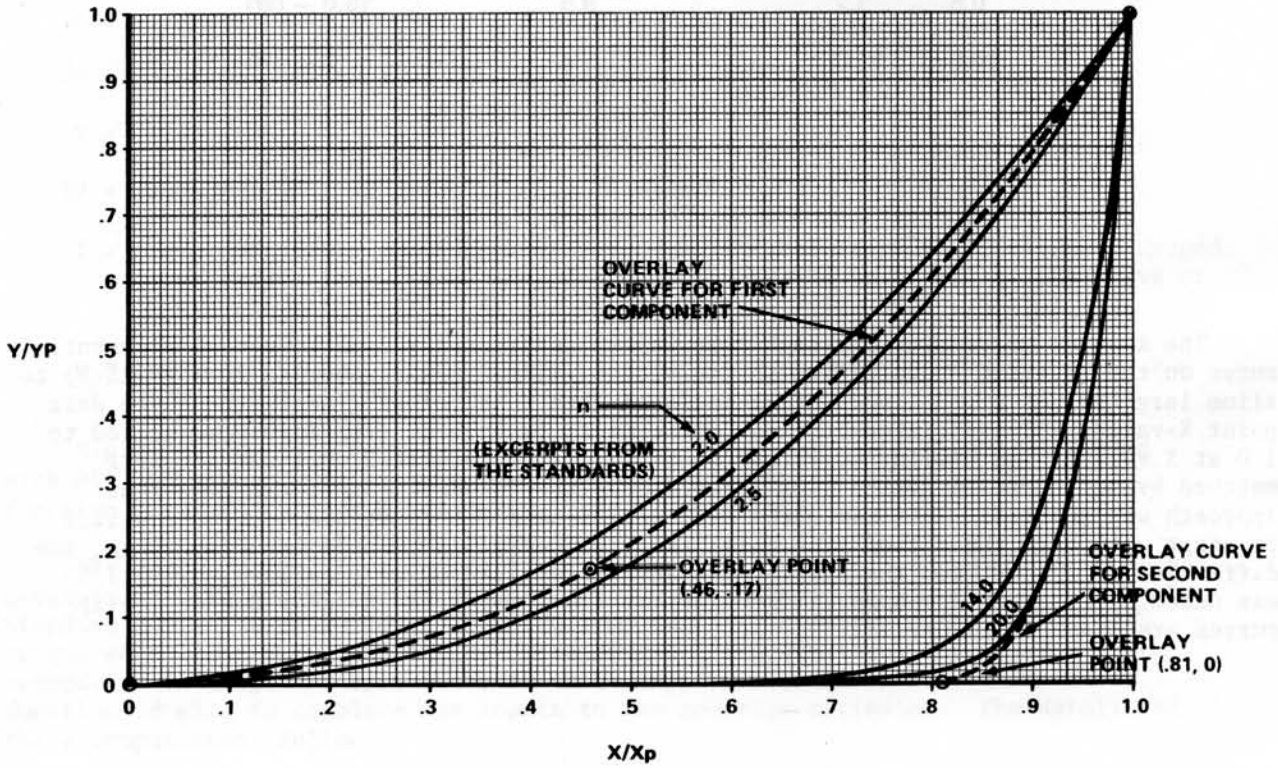
W	XP	(15-W), X	(X/11.8)	(1190-XP), Y	Y/118	$(b_1X^{2.25})$ 1/
0.0	--	15.0	--	--	--	--
.5	977	14.5	--	213	--	188
3.2	1072	11.8	1.00	118	1.00	118
9.6	1170	5.4	.46	20	.17	20
15.0	1190	.0	.00	0	.00	0

	X	X/14.5	$(Y-b_1X^{2.25})$, d	d/25	(b_2X^{20}) 2/	$\hat{XP}$ 3/
	15.0	--	--	--	--	936
	14.5	1.00	25	1.00	25	977
	11.8	.81	0	.00	0	1072
	5.4	.37	0	.00	0	1170
	.0	.00	0	.00	0	1190

1/ $b_1 = 118/(11.8)^{2.25} = 0.4572$

2/ $b_2 = 25/(14.5)^{20} = 1.4811 \times 10^{-22}$

3/ $\hat{XP} = 1190 - b_1X^{2.25} - b_2X^{20}$--which gives a perfect match for the original XP-values at control points 0-14.5 in wind.



A-1. – Standards for position A, set 1 – (X-transform to be fitted by least squares = $(X)^n$, $0 \leq X \leq X_p$)

Figure 32

The completed function was then:

$$\hat{X}P = 1190 - 0.4572\ (15\text{-}W)^{2.25} - 1.4811 \times 10^{-22}\ (15\text{-}W)^{20},\quad 0 \leq W \leq 15$$

For CP = f(W)

Paired W- and CP-values (0.5, 8900; 3.2, 8100; and 9.6, 6970) were read from figure 30. These points were plotted and a smooth curve drawn through them (fig. 33) that extended over the relevant range of wind, 0-15 mi/h.

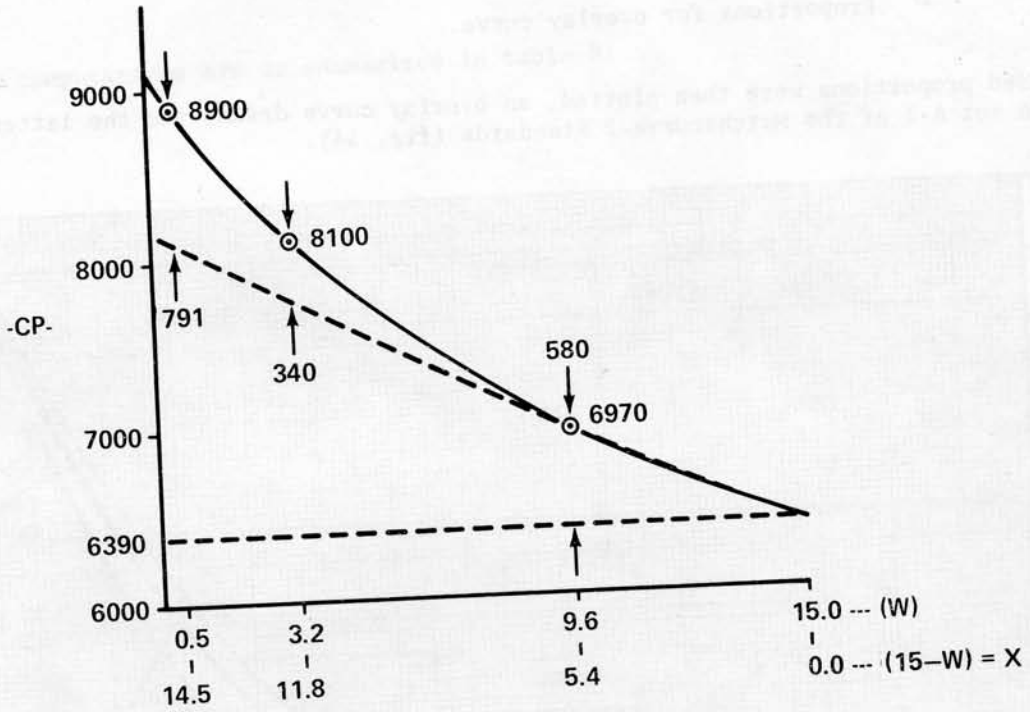


Figure 33

Again, a single-component exponential was first tried for the descriptor here. Reversing the W-axis to X = (15-W) to aline large values of the independent variable with those of CP, the X- and CP-values were scaled to 1.0 at X = 14.5. An overlay curve was constructed and compared to the Standards of Matchacurve-2. But, none of the Standards were acceptably close to the overlay; so a two-component model was tried next.

To start with, the lower end of the curve ($0.0 \leq X \leq 5.4$) was fitted with an intuitively selected flat form ($X^{1.1}$); it was obvious without scaling and overlays that the curve was extremely flat in that range of X (fig. 33). The $X^{1.1}$ was then scaled to the difference (580) between the intercept (6390) and the desired curve height (6970) at X = 5.4; so the first component, a partial descriptor for CP, was:

$$CP = 6390 + (580/(5.4)^{1.1})\ (15\text{-}W)^{1.1}$$

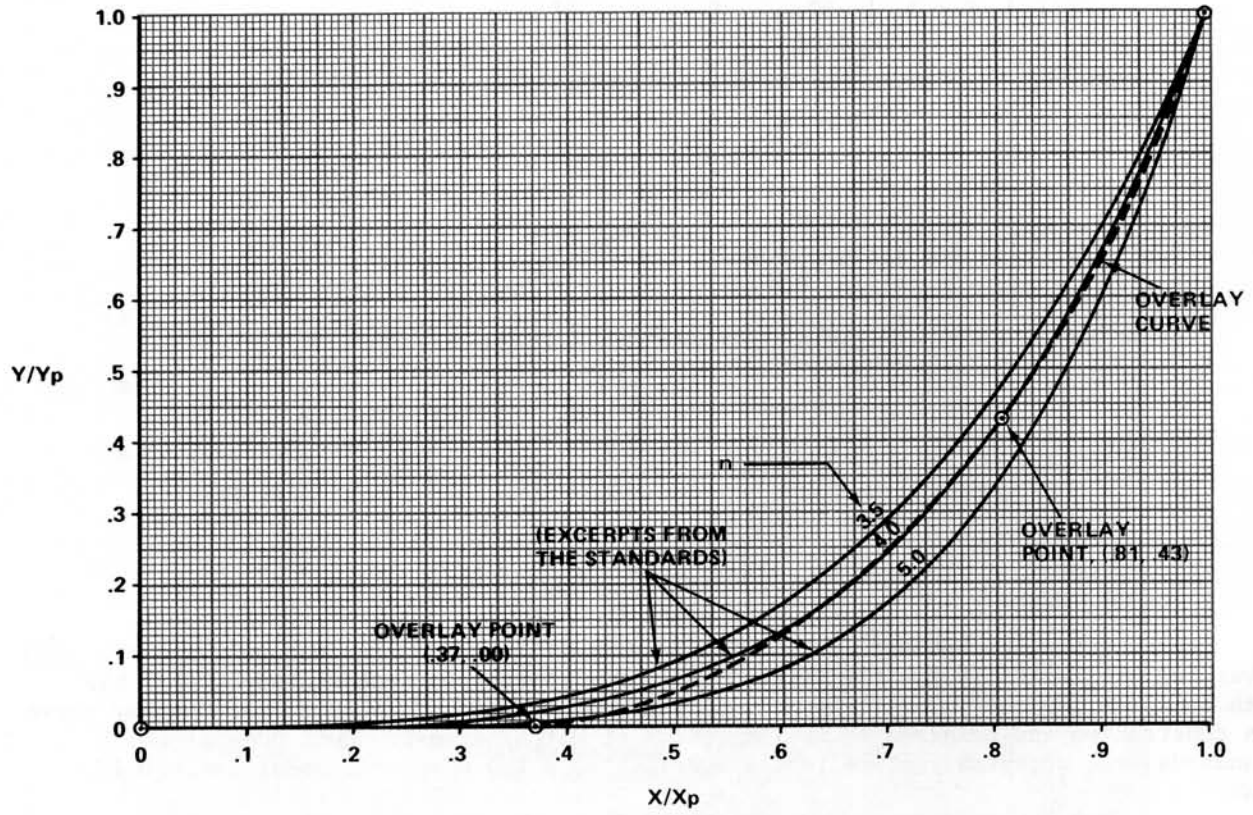
This is the dashed line above the intercept (6390) in figure 33. The remaining differences, 340 and 791, and the associated X-values were then scaled to 1.0 at X = 14.5 (table 7).

Table 7

W	X = (15-W)	(X/14.5)[1/]	difference, d	d/791[1/]
0.0	15.0	--	--	--
.5	14.5	1.00	791	1.00
3.2	11.8	.81	340	.43
9.6	5.4	.37	0	.00
15.0	.0	.00	0	.00

[1/] Proportions for overlay curve.

The paired proportions were then plotted, an overlay curve drawn, and the latter compared to set A-1 of the Matchacurve-2 Standards (fig. 34).



A-1. – Standards for position A, set 1 – (X-transform to be fitted by least squares = $(X)^n$, $0 \leq X \leq X_p$)

Figure 34

The Standard curve with n = 4.0 matched the overlay fairly well. So the fourth power of X was scaled to the difference, 791, at X = 14.5 and added to the partial function already in hand. Then, the complete function was:

$$\hat{CP} = 6390 + (580/(5.4)^{1.1})(15-W)^{1.1} + (791/(14.5)^4)(15-W)^4$$

$$= 6390 + 90.739\ (15-W)^{1.1} + 0.017894\ (15-W)^4,$$

$$0 \leq X \leq 15$$

where the computations are as summarized in table 8:

Table 8

W	CP	(15-W), X	(X/5.4)	(CP-6390), Y	(Y/580)	$(b_1X^{1.1})$ [1/]
0.0	--	15.0	--	--	--	--
.5	8900	14.5	--	2510	--	1719
3.2	8100	11.8	--	1710	--	1370
9.6	6970	5.4	1.00	580	1.00	580
15.0	6390	.0	.00	0	.00	0

	X	(X/14.5)	$(Y-b_1X^{1.1})$, d	(d/791)	(b_2X^4) [2/]	$\hat{CP}$ [3/]
	15.0	--	--	--	--	9080
	14.5	1.00	791	1.00	791	8900
	11.8	.81	340	.43	347	8107
	5.4	.37	0	.00	15	6985
	.0	.00	0	.00	0	6390

1/ $b_1 = 580/(5.4)^{1.1} = 90.739$

2/ $b_2 = 791/(14.5)^4 = 0.017894$

3/ $\hat{CP} = 6390 + b_1X^{1.1} + b_2X^4$. These are sufficiently close to the original CP-values for our purposes.

Matching Bell-Shaped Curves

Tables 9-11 list $\hat{XP}$- and $\hat{CP}$-values from the equations developed in the foregoing sections along with some representative X- and C-values from either side of the XP-values for the three curves in figure 30. Also listed are C-value estimates obtained by *using matching e-transforms* from Matchacurve-1 Standards (see the overlay curves in figure 35).

Table 9

W	H	(1250-H), X	(2XP-X),[1] K	(K/977)	C	(C/8900)	$\hat{C}$[2]
0.5	50	1200	754	0.772	6860	0.771	7137
	100	1150	804	.823	7800	.876	7793
	150	1100	854	.874	8450	.949	8322
	200	1050	904	.925	8770	.985	8692
	273	[977] [3]	[977]	1.000	[8900] [3]	1.000	8900

[1] Temporary transform for plotting the overlay curve only. The X-values need to be in ascending order toward XP = 977.

[2] Estimated coverage using the e-transform with CP = 8900, XP = 977, I = 0.510, n = 2.0, and a zero intercept.

[3] Smoothed values from the XP- and CP-estimators, each a function of wind (shown earlier in the text).

Table 10

W	H	(1250-H), X	(2XP-X), K	(K/1072)	C	(C/8107)	$\hat{C}$[1]
3.2	50	1200	944	0.881	6730	0.830	6843
	100	1150	994	.927	7600	.937	7612
	150	1100	1044	.974	8040	.992	8042
	178	[1072]	[1072]	1.000	[8107]	1.000	8107

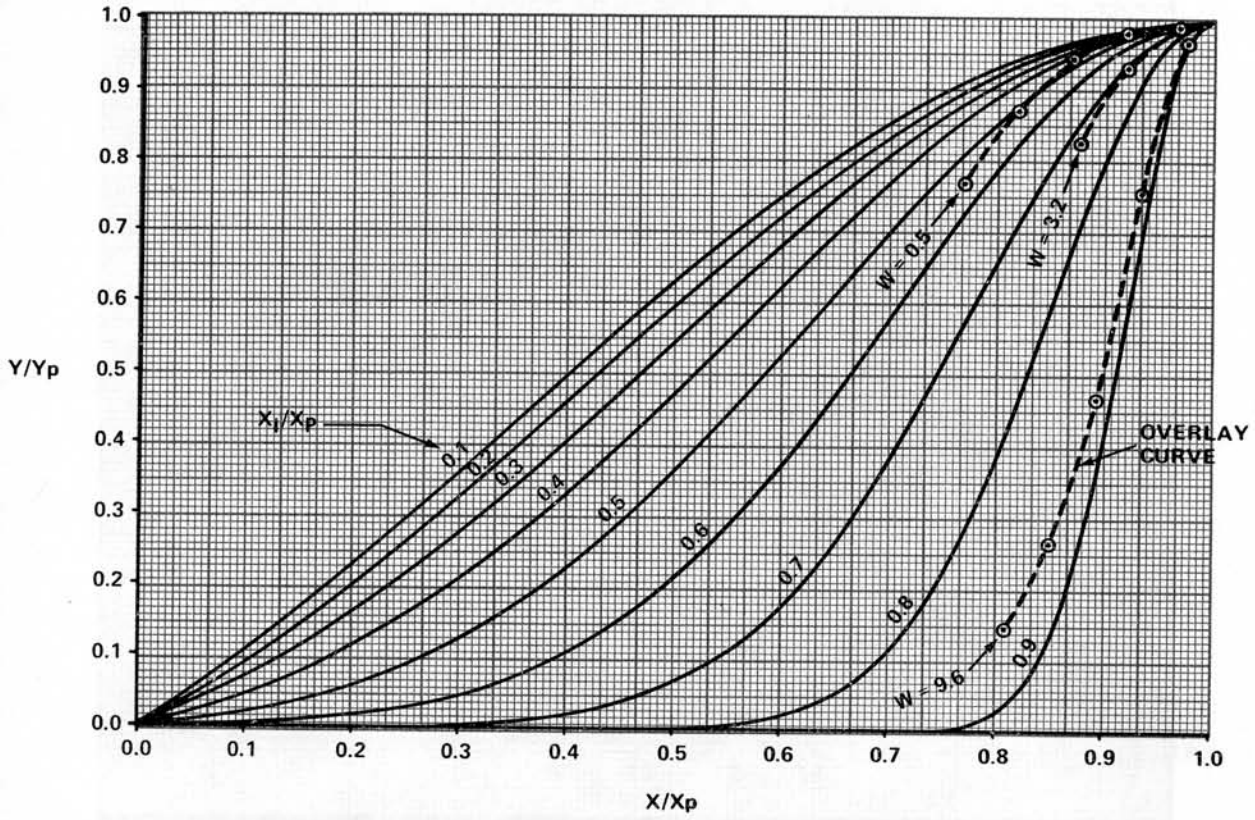
[1] Estimated coverage using the e-transform, CP = 8107, XP = 1072, I = 0.710, n = 2.0, and a zero intercept.

Table 11

W	H	(1250-H),[1] X	(X/1170)	C	C/6985	$\hat{C}$[2]
9.6	80	[1170]	1.000	[6985]	1.000	6985
	100	1150	.983	6820	.976	6845
	150	1100	.940	5300	.759	5448
	200	1050	.897	3270	.468	3364
	250	1000	.855	1830	.262	1612
	300	950	.812	997	.143	600

[1] Since the right half of this curve (fig. 30) was read, X-values are already ascending toward XP = 1170 and no transform of X is necessary for plotting the overlay curve.

[2] Estimated coverage using the e-transform, CP = 6985, XP = 1170, I = 0.880, n = 2.0, and a zero intercept.



Set 2. — Standards for n = 2.0 and scaled inflection points X_I/X_P, ranging from 0.1 to 0.9.

Figure 35

As may be seen in tables 9, 10, and 11, the coverage values of each curve were scaled to 1.0 by using the appropriate CP as the maximum. The values of X or of (2XP-X) were similarly scaled by using the appropriate XP as the maximum. After the paired proportions for H-levels within W-levels were plotted and smoothed as three overlay curves, they were compared to the Matchacurve-1 Standards for n = 2 (fig. 35).

Although better matches for some of these overlays can be found in Standards that have n other than 2.0, *it was decided to hold n constant over wind to avoid surface irregularities that may develop when both n and I are allowed to vary.* Then, the I-values (XI/XP in the Standards), corresponding to the best matching curve alternatives in this set, were estimated to be 0.50, 0.70, and 0.88 for W = 0.5, 3.2, and 9.6, respectively. On scaling and checking these and alternative curves (by means of a small computer) at the tabled H-values by W-group, some improvement was achieved with the final I-array, 0.51, 0.71, and 0.88. The last column in each of tables 9-11 shows the final scaled coverage estimates. There was close proximity of these values to actual coverage, C, so I = f(W) was next developed.

For I = f(W)

Using the paired wind and I-values just determined (0.5, 0.510; 3.2, 0.710; and 9.6, 0.880), "I" was plotted over W and a smooth "initial" curve drawn through the three data points (fig. 36).

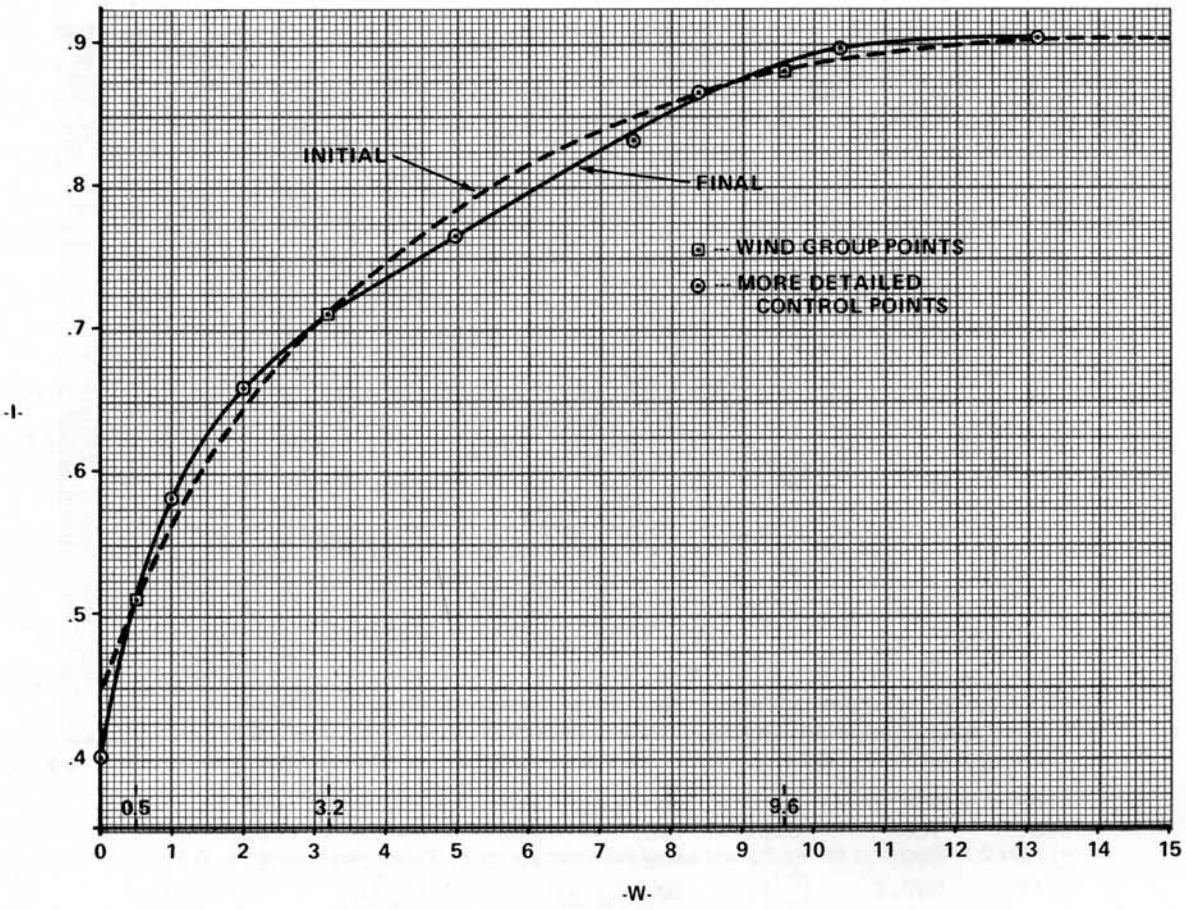


Figure 36

On checking the performance of the e-transform over the range of W, with smoothed CP-, XP-, and initial I-values, it was found that the left edge of the surface undulated unacceptably. Further, performance in the high-wind group was not well balanced for the relatively divergent wind levels there (6.4, 8.4, 10.4, 13.2). These problems were largely overcome by iteratively adjusting to the "final" curve in figure 36. This was the one for which a descriptor was developed.

This curve was judged to be a multicomponent model because neither the sigmoids nor the exponentials have the flattened central segment found there. Since a suitable single-component exponential was not found, the flat segment was represented by a straight line; the negative differences at the left end by an exponential; and those at the right by a portion of a sigmoid (fig. 37).

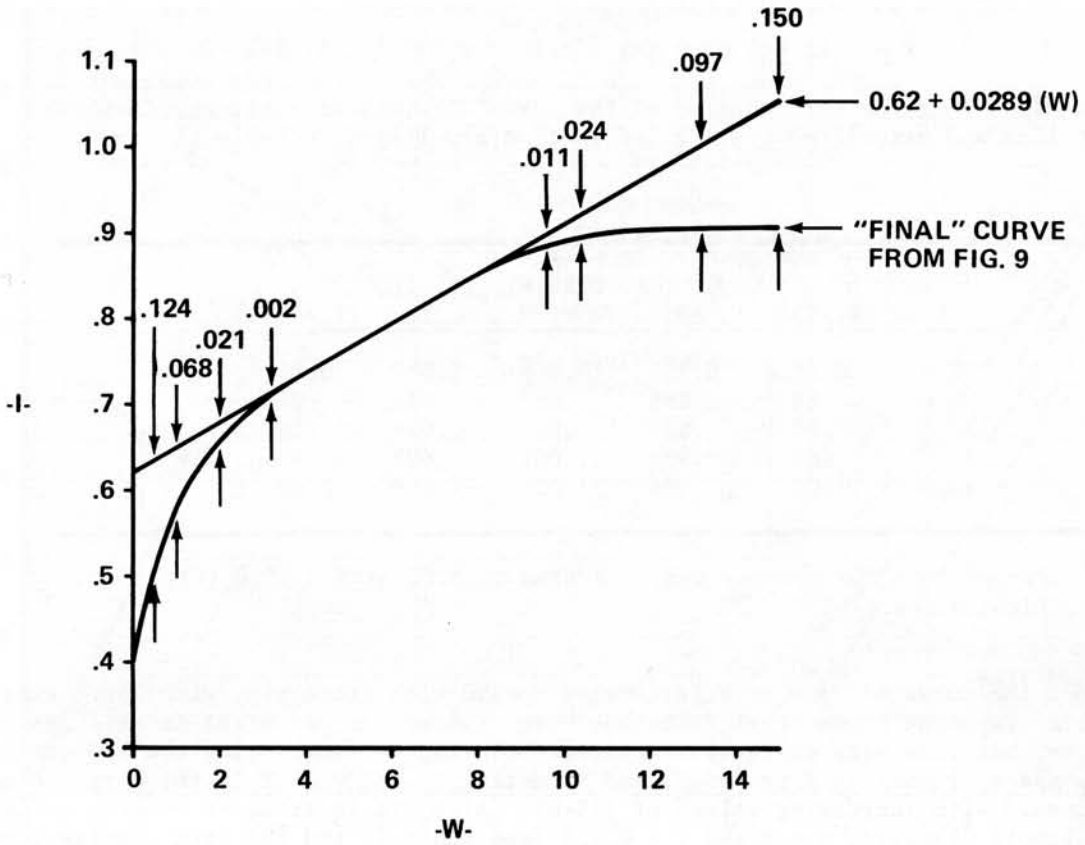


Figure 37

The Flat Segment: A straight line was drawn along the "final" curve. The intercept, 0.62, and values necessary to the coefficient for W, (1.054-0.620)/15 = 0.0289, were read directly from the graph.

The Left End: This portion of the curve was satisfactorily matched by a single exponential of the reversed W-axis, scaled at W = 14.5 (table 12).

Table 12

W	(15-W), X	(X/14.5)[1/]	I	(0.62 + 0.0289 W), Y_1	(Y_1-I), d	(d/0.124)[1/]	($b_2 X^{16.75}$), [2/] $\hat{d}$
0.0	15.0	--	0.400	0.620	0.220	--	0.219
.5	14.5	1.00	.510	.634	.124	1.00	.124
1.0	14.0	.97	.581	.649	.068	.55	.069
2.0	13.0	.90	.657	.678	.021	.17	.020
3.2	11.8	.81	.710	.712	.002	.02	.004

[1/] Proportions for an overlay curve that is not presented here in the text.

[2/] $b_2 = 0.124/(14.5)^{16.75} = 4.3703 \times 10^{-21}$. The $\hat{d}$-values are sufficiently close to the original d-values for our purposes.

So the partial form at this point:

$$\hat{Y} = 0.62 + 0.0289\ (W) - 4.3703\ X\ 10^{-21}\ (15\text{-}W)^{16.75}$$

The Right End: For this portion of the curve, the differences, Y_2, from the straight line and associated scaling information are listed in table 13.

Table 13

W	(W/15)[1/]	I	0.62 + 0.0289(W), Y_1	(Y_1-I), Y_2	(Y_2/0.15)[1/]
8.4	0.56	0.863	0.863	0.000	0.00
9.6	.64	.886	.897	.011	.07
10.4	.69	.897	.921	.024	.16
13.2	.88	.904	1.001	.097	.65
15.0	1.00	.904	1.054	.150	1.00

[1/] Proportions for overlay curve, scaled to 1.00 at W = 15.0 (fig. 38, right side).

Since the curve of Y_2 over W is concave upward with increasing wind, both single- and double-component exponential functions were examined as potential descriptors for this curve, but none were suitably accurate. However, by reorienting the exponential curve in space, reversing W to (15-W) and inverting Y_2 to (0.15-Y_2), the curve becomes convex upward with increasing values of (15-W). Also, it is at least roughly matched by the sigmoid Standard, n = 8 and I = 0.108 (see table 14 and the left overlay curve in fig. 38). This sigmoid was the result of an interpolation between the two Standards I = 0.1 and 0.2 at n = 8.

Table 14

W	(15-W), X	(X/15)[1/]	Y_2	(0.150-Y_2), Y_3	(Y_3/0.150)[1/]	(0.150 e^{-T}), $\hat{Y}_3$[2/]
0.0	15.0	1.00	0.000	0.150	1.00	0.150
8.4	6.6	.44	.000	.150	1.00	.146
9.6	5.4	.36	.011	.139	.93	.139
10.4	4.6	.31	.024	.126	.84	.130
13.2	1.8	.12	.097	.053	.35	.053
15.0	.0	.00	.150	.000	.00	.000

[1/] Proportions for overlay curve, scaled to 1.00 at X = 15 (fig. 38, left side).

[2/] The scaled e-transform with n = 8, I = 0.108, XP = 15, and intercept = zero.

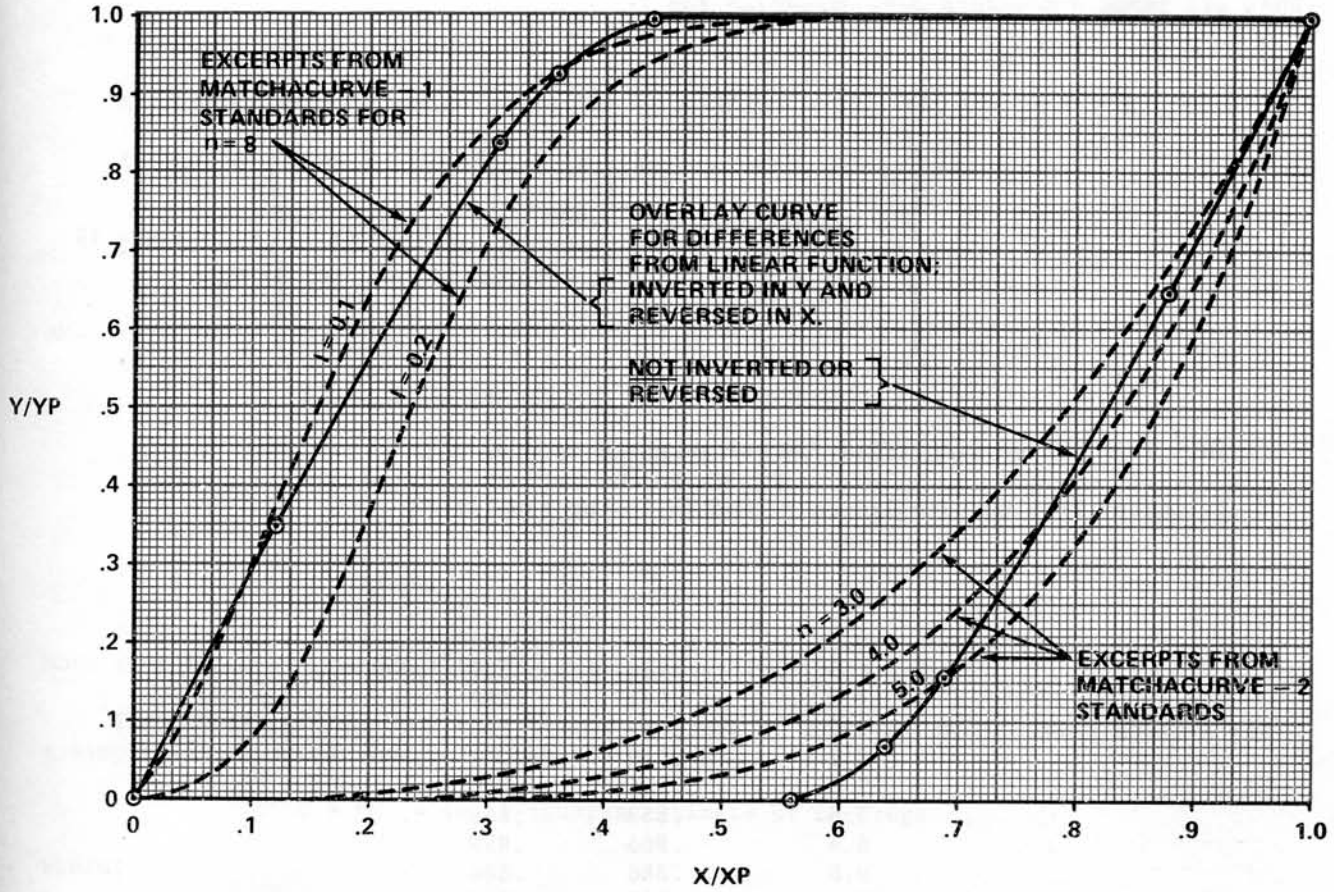


Figure 38

Then, the right end was specified as:

$$0.15 - Y_2 = 0.150 \text{ e-transform, based on } X = (15-W)$$

or

$$Y_2 = 0.150 - 0.150 \text{ e-transform,}$$

and

$$Y_2 = 0.150 - 0.150 \left\{ \frac{e^{-\left|\frac{(15-W)}{15} - 1\right|^8} - e^{-\left\{\frac{1}{1 - 0.108}\right\}^8}}{1 - e^{-\left\{\frac{1}{1 - 0.108}\right\}^8}} \right\}$$

When all three components were assembled for I:

$$\hat{I} = 0.62 + 0.0289\,(W) - 4.3703 \times 10^{-21}(15-W)^{16.75} - \left\{ 0.15 - 0.15 \left\{ \frac{e^{-\left|\frac{\frac{(15-W)}{15} - 1}{.892}\right|^{8}} - 0.08249}{0.91751} \right\} \right\}, \quad 0 \leq W \leq 15$$

This form can be simplified further as shown on page 37.

Estimated I-values are listed in the following tabulation:

W	*I*	$\hat{I}$
0.0	0.400	0.401
.5	.510	.510
1.0	.581	.580
2.0	.657	.658
3.2	.710	.709
5.0	.766	.764
7.5	.838	.835
8.4	.863	.859
9.6	.886	.886
10.4	.897	.900
13.2	.904	.905
15.0	.904	.904

where I and $\hat{I}$ were suitably close over the range of W.

At this point, all necessary inputs had been derived and the complete descriptor was:

$$C' = CP \left\{ \frac{e^{-\left|\frac{(X/XP) - 1}{1 - I}\right|^{2.0}} - e^{-(1/(1-I))^{2.0}}}{1 - e^{-(1/(1-I))^{2.0}}} \right\}$$

where:

C' = coverage in square feet, estimated from the pre-least squares fit model

CP = coverage peak for any specified windspeed, W, or

$CP = 6390 + 90.739\,(15 - W)^{1.1} + 0.017894\,(15 - W)^{4}, \quad 0 \leq W \leq 15$

$X = (1250 - H)$

H = drop height, $50 \leq H \leq 300$

XP = point in X at which CP occurs, or

$XP = 1190 - 0.4572\ (15 - W)^{2.25} - 1.4811 \times 10^{-22}\ (15 - W)^{20}$

I = inflection point in X expressed as a proportion of XP (XI/XP in Matchacurve-1),

and

$$\hat{I} = 0.4565 + 0.0289(W) - 4.3703 \times 10^{-21}(15-W)^{16.75} + 0.16349\ e^{-\left|\frac{\frac{(15-W)}{15} - 1}{0.892}\right|^{8}}$$

Note that the last component has been simplified (see page 36).

Then, fitting this form back to the original 16 observations by least squares in a simple linear model forced through the origin

$\hat{C} = bC'$ = least-squares estimate of coverage

where:

$b = \frac{CC'}{(C')^2}$ over the 16 observations

The final model is:

$\hat{C} = 1.000569(C')$

with $\overline{C} = 6962$ = mean coverage

$s_{y \cdot x} = 850$ = standard error of estimate

$R^2 = 0.89$ = coefficient of determination

The coefficient, b, is very near 1.0000; so it is evident that even before least-squares adjustment, the original descriptor was well alined with the data spacially. But, in checking the form of the descriptor by comparing predicted and actual coverages (fig. 39), at least one anomaly appears. The three high-wind points probably lie *too* close to the surface since greater variation occurs about the surface at lower wind levels; i.e., reasonable variance about the surface would be expected at high-wind levels also. The accuracy of the standard error of the estimate shown, 850, cannot be assessed since an unknown number of degrees of freedom have been sacrificed by developing the detailed curve-form hypothesis on the data set from which the estimate was made.

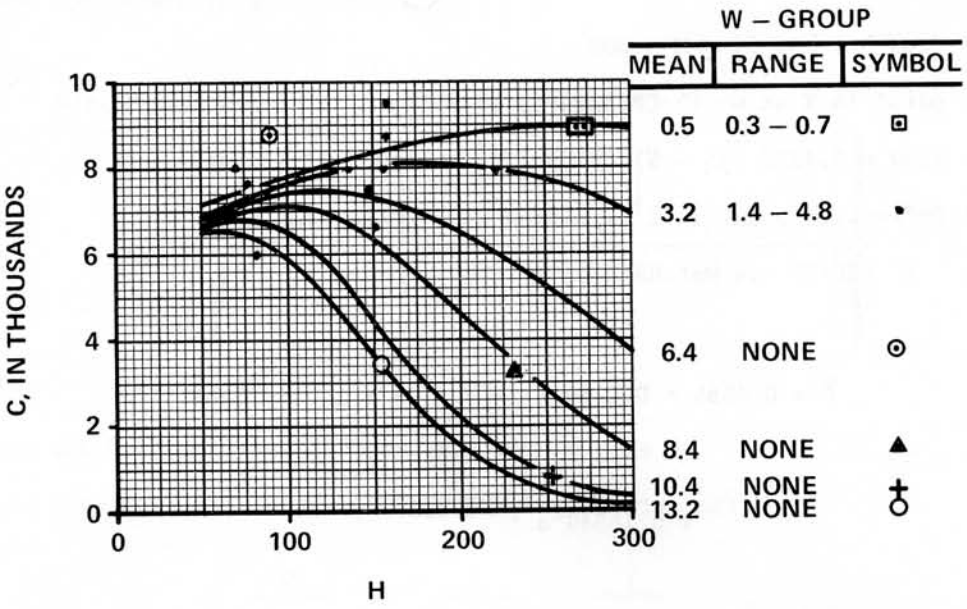


Figure 39

Recall that through the use of controlled I-values (fig. 36) for curves passing close to the three high-wind points, low variation of these points from the surface was assured. Alternative I-values could be used that would allow for variation from the surface, in accord with some assumed variance criteria. This refinement was not undertaken and the model was adopted as developed. Predicted values (table 15) and the associated surface (fig. 40) for the model are as follows:

Table 15

Wind (mi/h)	Drop height (feet)					
	50	100	150	200	250	300
0	7198	7804	8316	8710	8969	9081
5	6830	7525	7707	7337	6491	5338
10	6617	6707	4925	2619	1009	282
15	6345	5663	3460	1447	414	81

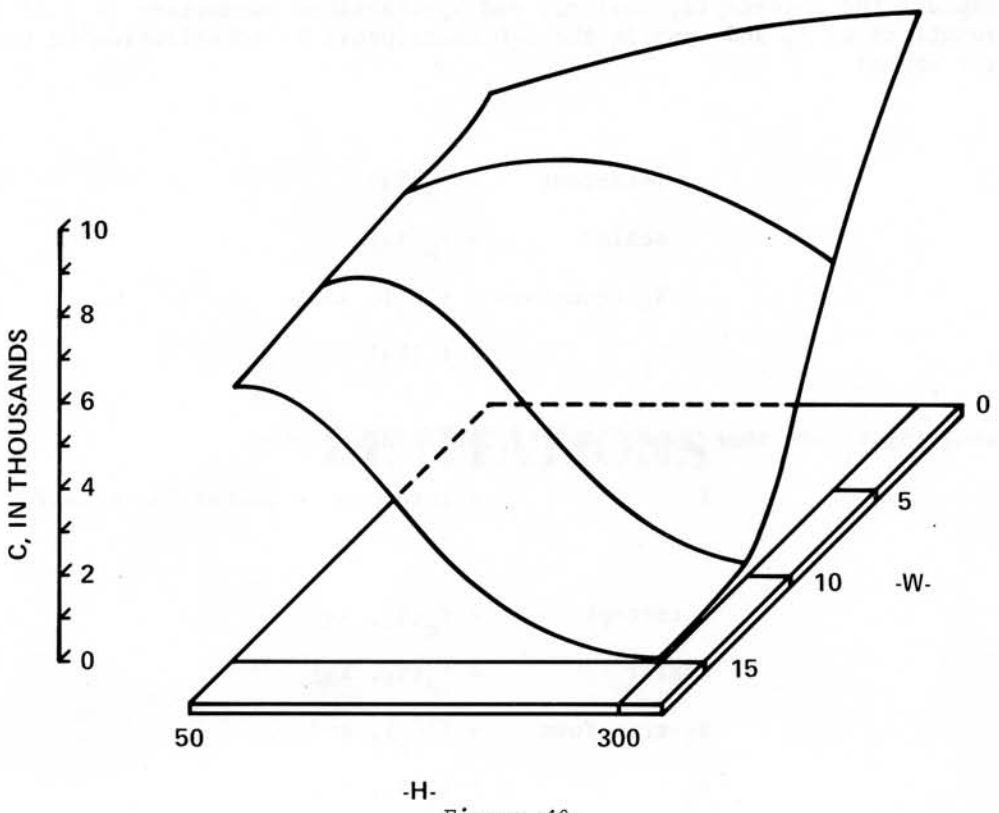


Figure 40

Since the objective of this paper is to show the analyst a system that can be used to exercise a great deal of curve-form control in model development, it is appropriate to review the imposing list of constraints that can be met in models like this one. Controllable items include: the intercept for the whole surface; the elevation, breadth, and positioning of the peaks over the interacting independent variables; the trend in Y at the truncated left edge of the surface; the magnitude of curvature above and below the inflection points in the bell-shaped curves; and the positioning of the inflection points themselves within the ranges of the independent variables.

Four or More Dimensions

Descriptor procedures outlined for three dimensions can be extended to more dimensions. For example, given that we have a set of 4-D graphs, Y over X_1, at four points in X_2 and that each such set occurs at three points in X_3 (fig. 4, page 4). Assume also that we have developed 2-D descriptors for these 12 curves.

The corresponding 12 intercepts, scalars, and X_1-transform parameters (call this group of variables K_i) may vary systematically over X_2 *and* X_3. So treat each K as a 3-D relation, ($K = f(X_2, X_3)$). Plot it over X_2 and fit a smooth curve through the resulting points at each of the specified levels of X_3. Identify a suitably accurate 2-D, X_2-descriptor for each curve:

$$\hat{K} = \text{intercept} + \text{scalar } (X_2\text{-transform})$$

Now, express the intercepts, scalars, and X_2-transform parameters (K_i) of *these* curves as functions of X_3 and compile the 3-D descriptors by substitution in the basic X_2-descriptor above:

where:

$$\text{intercept} = f_a(X_3)$$

$$\text{scalar} = f_b(X_3)$$

$$X_2\text{-transform} = f(P_j), \text{ and}$$

$$P_j = f_j(X_3)$$

The resulting $\hat{K}_i$ are then inputs to the 2-D, X_1-descriptor:

$$\hat{Y} = \text{intercept} + \text{scalar } (X_1\text{-transform})$$

where:

$$\text{intercept} = f_c(X_2, X_3)$$

$$\text{scalar} = f_d(X_2, X_3)$$

$$X_1\text{-transform} = f(P_j), \text{ and}$$

$$P_j = f_j(X_2, X_3)$$

The expansion of the system to five or more dimensions is analogous and the amount of descriptor development increases exponentially with the dimensions. This is the cost of more complete curve-form control and is a factor to consider in selecting one descriptor process from the array of alternatives.

LIMITATIONS

The descriptor developed for a specific relationship by using the procedures outlined here is likely to be more complex algebraically than one developed from limited, standardized transformations,--fairly popular at present. However, to the analyst who has access to at least a small desk-top computer, the added computational complexity is literally of no consequence. Such a computer is a minimum requirement for the application of Matchacurve procedures.

Some new, but certainly not insurmountable, problems of communicating through reports and publications *are* created. Complex descriptors are not easily translated into spacial characteristics, often vital for reader comprehension. The solution to this problem is to *devote suitable discussion* to the shapes and magnitudes of trends and to the limits of application. To sidestep a potentially impossible task of hand calculation by the reader, tables of computer-calculated values may be included for pertinent levels of the independent variables. Rough interpolation, if required, can be left to the reader.

In reality, when graphic forms are developed from data, the effects of variables incorporated in the model are being scaled visually by the analyst. Then, as is shown, he makes further visual scale selections in quantifying this model mathematically. A simple linear, least-squares adjustment of the entire model to some pertinent data set (where available) helps to identify scales more objectively. By way of contrast, a mechanized fitting process, such as least squares, could be substituted for the *visual scaling* portion of the system and some benefits could accrue in the overall fitting effort.

However, the practicality of this alternative is lessened in the presence of the complex, nonlinear (in the parameters) mathematical models that often result from application of the Matchacurve procedures. Estimating the parameters for such models requires the use of rather specialized fitting techniques (Hartley 1961; Spang 1962; Marquardt 1963) and, quite possibly, mathematics beyond the average practicing analyst. Marquardt's iterative system seems to be the most acceptable, but no system appears to be universally applicable (Draper and Smith 1966).

LITERATURE CITED

Bartlett, M. S.
1947. The use of transformations. Biometrics 3:39-52.

Box, G. E. P., and P. W. Tidwell
1962. Transformation of the independent variables. Technometrics 4:531-550.

Dolby, J. L.
1963. A quick method for choosing a transformation. Technometrics 5:317-325.

Draper, N. R., and W. C. Hunter
1969. Transformations: some examples revisited. Technometrics 11(1):23-40.

Draper, N. R., and H. Smith
1966. Applied regression analysis, p. 267-273. Wiley and Sons, New York.

George, C. W., and A. D. Blakely
1973. An evaluation of the drop characteristics and ground distribution patterns of forest fire retardants. USDA For. Serv. Res. Pap. INT-134, 60 p.

Hartley, H. O.
1961. The modified Gauss-Newton method for fitting of nonlinear regression functions by least squares. Technometrics 3:269-280.

Hoerl, A. E.
1954. Fitting curves to data. Chem. Bus. Handb. 20:55,77.

Jensen, C. E., and J. W. Homeyer
1970. Matchacurve-1 for algebraic transforms to describe sigmoid- or bell-shaped curves. USDA For. Serv., Intermt. For. & Range Exp. Stn., 22 p.

Jensen, C. E., and J. W. Homeyer
1971. Matchacurve-2 for algebraic transforms to describe curves of the class X^n. USDA For. Serv. Res. Pap. INT-106, 39 p.

Karst, O. J.
1958. Linear curve fitting using least deviations. J. Am. Stat. Assoc. 53:118-132.

Marquardt, D. W.
1963. An algorithm for least squares estimation of nonlinear parameters. J. Soc. Ind. & Appl. Math. 2:431-441.

Spang, H. A.
1962. A review of minimization techniques for nonlinear functions. J. Soc. Ind. & Appl. Math. 4:343-365.

Tukey, J. W., and M. B. Wilk
1965. Data analysis and statistics: techniques and approaches. I.E.E.E. Symposium on Information Processing in Sight and Sensory Systems Proc. 1965:7-27. Cal. Inst. Tech. Univ., Pasadena.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974 — 784-488 / 40 REGION NO. 8