PATH ANALYSIS: A STATISTICAL METHOD SUITED TO ECOLOGICAL DATA

D. SCOTT

Grasslands Division Substation, Department of Scientific and Industrial Research, Lincoln

SUMMARY: Path analysis is a statistical method akin to multiple regression in fitting a quantitative linear relationship between variables. It has the further advantages of handling simultaneously complex multistage interactions.

Examples are given of its use in determining relationship where there are few variables, many variables, curvilinear relations, feedback interactions, and where *a priori* quantitative information is available from other sources.

The technique is also discussed in terms of the general problem of synthesis of ecological data. It is concluded that the basic concepts of cause and effect among groups of variables and their combination in a multistage causal scheme are applicable in all situations, while the curve fitting properties of the technique and the ability to combine *a priori* quantitative data from several sources, provide a valuable means of synthesis at the empirical stages of a problem.

INTRODUCTION

Ecologists are concerned with the complex interactions between plants, animals and their environment. Whether these studies be at the conceptual level, or with qualitative or quantitative data, the concern is generally with the simultaneous interaction between many variables.

Where the ecologist has quantitative data he will often find that there are limitations in the methods of statistical analysis available to him in considering these simultaneous interactions and in the uncertainties associated with the measurement of any variable.

In an earlier paper (Scott 1966) I outlined some of the basic ideas of path analysis, which is a statistical technique akin to multiple regression in determining the quantitative relationship between variables. But unlike multiple regression it is not confined to the estimation of one variable in terms of a group of independent variables. It can simultaneously consider relationships where variables which are "independent" variables in one relationship may be "dependent" in another. Also in determining the relationships, the method can use already known quantitative relationships. Finally, and possibly the most valuable contribution of the method, is that it is built on a conceptual framework which may provide the basis for the general analysis and synthesis of ecological data.

The purpose of this paper is to describe the role of path analysis in these more complex situations and to discuss the place that path analysis may play in the more general problem of analysis and synthesis of ecological data.

The method was first developed by Wright (1918, 1934 and 1954) for correlation problems in genetics but reached its present development in economics under the title of structural analyses or simultaneous linear equations (e.g. Theil 1958). Further details are given in the following references—Turkey (1954), Li (1955), Theil (1958), Turner and Stevens (1959), Campbell, Turner and Wright (1960), Theil and Goldberger (1961), Ferrari (1963, 1964), Zellner and Theil (1962) and Hamilton (1968).

PATH ANALYSIS AS A STATISTICAL METHOD: SIMPLE EXAMPLE

The information given by Scott (1966) will be briefly recapitulated using as an example the relationship between the rate of leaf elongation of *Notodanthonia setifolia* on several sites above timberline in the Tongariro National Park and the environmental factors of incoming solar radiation, air and soil temperatures (Table 1 and following diagram):

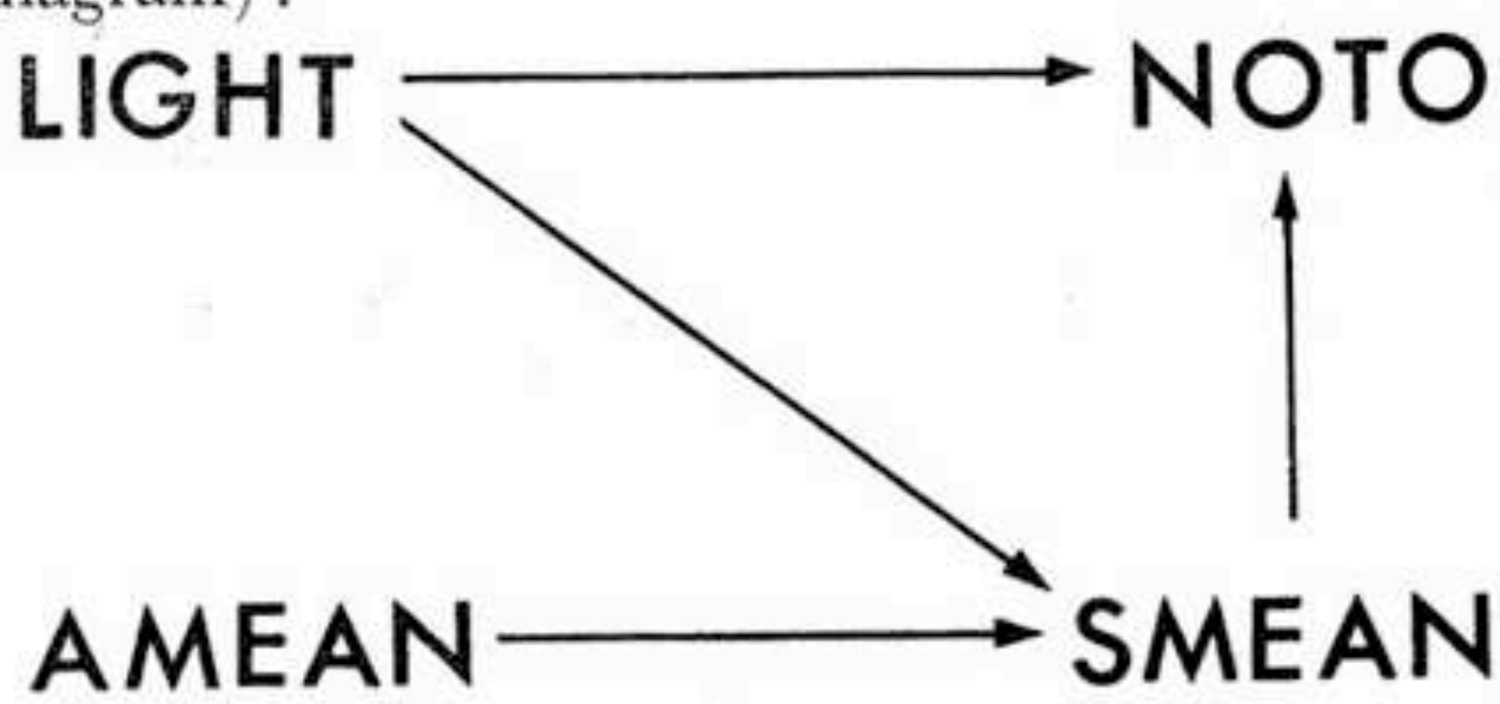


TABLE 1. *Variables and Range in Values in Simple Example of Leaf Elongation in* Notodanthonia setifolia.

Description	Code Name	Range
Incoming radiation (as proportion of mean for all periods)	LIGHT	0.47 to 2.65
Mean daily air temperature at 1m above ground (°C)	AMEAN	2.6 to 14.9
Mean daily soil temperature at 10cm depth (°C)	SMEAN	2.2 to 14.4
Rate of leaf elongation of *Notodanthonia setifolia* (mm week^{-1})	NOTO	0.4 to 6.1

The reader is asked to look at this and subsequent examples primarily as demonstrations of the characteristics of the path analysis method and only secondarily as particular ecological problems.

The basic concept of path analysis is that, in any particular problem involving the relationships among a group of variables, some variables can be recognised as primary causes (LIGHT, AMEAN) and others as effects (NOTO) including the cases where a variable may be a cause in relationship to some variables, and an intermediate effect in relationship to others (SMEAN). These relationships can be shown diagrammatically with the variables listed and the arrows indicating the direction of the interaction.

In the particular example, the grass leaf growth (NOTO) was probably related to temperature and incoming solar radiation (LIGHT)—the latter either positively through photosynthesis or negatively through water stress. Preliminary analysis suggested that it was soil temperature (SMEAN) rather than air temperature which was important to growth. Soil temperature (SMEAN) was expected to be related to both solar radiation (LIGHT) and mean air temperature (AMEAN).

Thus solar radiation and mean air temperature were regarded as "causes" with soil temperature as an "intermediate effect", and leaf growth as a final "effect". For the purposes of the example these were assumed to be the only variables and interactions present.

Structural equations can be written down for each effect in terms of its immediately preceding cause. Where measurements are available of all the variables then the coefficients (path coefficients) of these equations can be determined by statistical means:

$$\text{SMEAN} = 0.605\ \text{LIGHT} + 0.916\ \text{AMEAN}$$
$$\text{NOTO} = 0.813 - 0.287\ \text{LIGHT} + 0.320\ \text{SMEAN}.$$

These latter equations are determined simultaneously and take cognizance of some variables appearing as both dependent and independent variables within the one set of equations.

At this stage it is as well to remember that any statistical method is only concerned with the characteristics of the numerical data and that the ultimate validity of the relationships obtained is more dependent on the investigator's understanding of the expected relationships and the appropriateness of the parameters measured.

With this proviso the completed solution enables several things to be done. Firstly, through using these equations singularly, the values of effect variables can be estimated from values of causes. Secondly, through using the equations as a group they can be used to estimate how changes in one variable will influence others, through its effect along various pathways. It is this ability of path analysis to handle the simultaneous interaction between many variables which seems to have potential in ecology.

The third aspect of the results which was no discussed in the first paper is that, if certain as sumptions are met, tests of the statistical signifi cance of the coefficients can be made. This allow hypothesis testing and determinations of confi dence intervals.

The distinction between path analysis and other methods is most evident in the determination of error variances required for making tests of significance. In most other statistical methods the equation being fitted to describe the relationship is of the form:

$$(y + e) = px_1 + px_2$$

where the dependent variable (y) is related to a linear function of the independent variable x_1, x_2, where the latter is assumed to be measured without error, and where any random effects or errors (e) only appear as differences between the actual and predicted value of the dependent variable. The same applies in path analysis in equations in which the only independent variables are primary causes.

The differences are where intermediate effects (y_1) are the independent variables in other equations with final effects (y_2), e.g.:

$$(y + e)_2 = p(y + e)_1 + px_1.$$

In this case there are random error effects in both some of the independent variables (these being intermediate effects) due to their relationship with other variables in the set, and random effects between the actual and predicted value of the dependent variable (final effect y_2). The statistics of path analysis estimate both classes of random effects which contrast with the single class of errors of the dependent variable determined by other methods.

It will be noted that the term "independent" variable refers to primary causes while "dependent" variables refer to both intermediate and final effects.

There are two consequences of being able to estimate error variance of both dependent and some of the independent variables.

First, it is this which allows the method to determine simultaneously the relationship between different groups of variables within the one set of data — not only the path coefficients but also a consistent set of variances to allow statistical tests to be made on these coefficients. In other methods each relationship tends to be looked at in mathematical and statistical isolation with little regard for the interaction with other relationships containing the same variables. It will be noted that interaction is used not in the sense that the effect of two or more variables on a third may be confounded, but in the sense that there may be relationships between the first group of variables other than their effect on a third.

Also, consideration of random errors in effect variables, whether they be dependent or independent variables in particular relationships, comes closer to reality as there are likely to be errors or uncertainties in the measurement of any variable.

ASSUMPTIONS

The following assumptions are made with regard to the data when they are used in the derivation and testing of path analysis equations.

(a) That the values of all the variables are measured as departures from their mean value. In practice this can be simply circumvented by introducing a constant term into all the structural equations.

(b) That the independent variables (primary causes, x's) are measured without error. No other restriction is placed on the distribution of their values, e.g. they could be selected by the investigator. Measurement without error is unlikely in practice but is the assumption made in almost all statistical techniques. However, environmental variables can often be measured with considerably greater accuracy than plant or animal responses.

(c) That there is a linear and additive response of each dependent variable to changes in each of the other variables which directly influence it. In many cases this could be regarded only as a first approximation. As no assumption is made about the distribution of the values of different variables it is possible to transform the raw data so they better approximate a linear relationship, or to introduce new variables which are functions of other variables.

(d) That the errors in each of the dependent variables are random with zero mean and a variance independent of the values of any other variables or any of the path coefficients. Apart from this there are no other restrictions on the distribution of the values of the dependent variables when they are used to calculate path coefficients and their variances. Errors in dependent variables are often related to the magnitude of the value, e.g. errors in measuring growth are often proportional to the mean size of the individual or population. This may be the second reason why the variables in the raw data may have to be transformed prior to analysis.

(e) To make tests of significance, or to determine confidence intervals for the path coefficients, it is necessary to make an additional assumption —not only that the errors in the dependent variables are independent with zero mean and constant variance, but also that they are normally distributed (or whose distributions have, if other than normal, known characteristics). This is difficult to establish in particular cases. However, many classes of biological data have been shown to have approximately normal distributions.

(f) If it is desired to test the significance of alternative forms of particular structural equations using the multiple correlation coefficient, then (b) must be replaced by the assumption that not only are independent variables measured without error, but also that each variable is derived as random samples from a normal distribution of values.

While these assumptions seem restrictive when given in a list, they are assumptions made in many other methods of statistical analysis (and more frequently ignored). The success of path analysis depends both on the applicability of the causal scheme used, and the degree to which the assumptions are met in collection of the data.

COMPLEX EXAMPLE

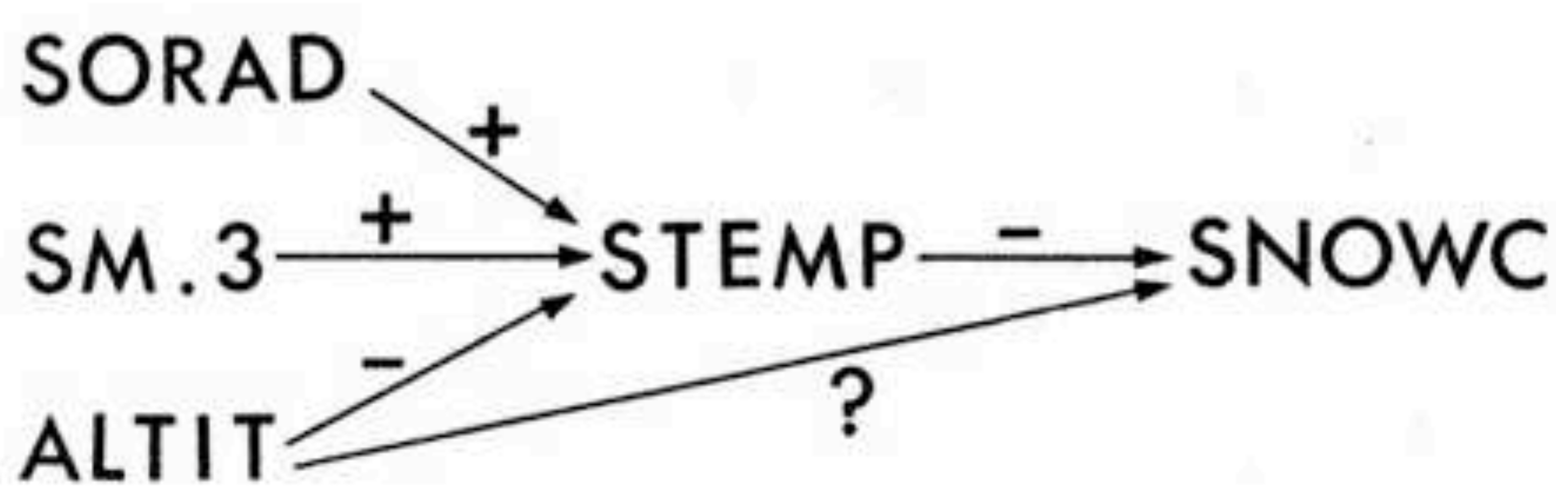
There is no theoretical limit to the number of variables and the number of relationships that can be dealt with by path analysis, apart from the requirement that the number of variables (dependent and independent) in any one structural equation does not exceed the total number of independent variables in the whole scheme. The limitations in practice are measurement of variables and computational facilities.

An example involving many variables is taken from previous work (Scott and Billings 1964) on the relationship between above-ground standing crop of 44 species on 50 sites and 39 environmental factors. Table 2 gives the 26 variables used in the example.

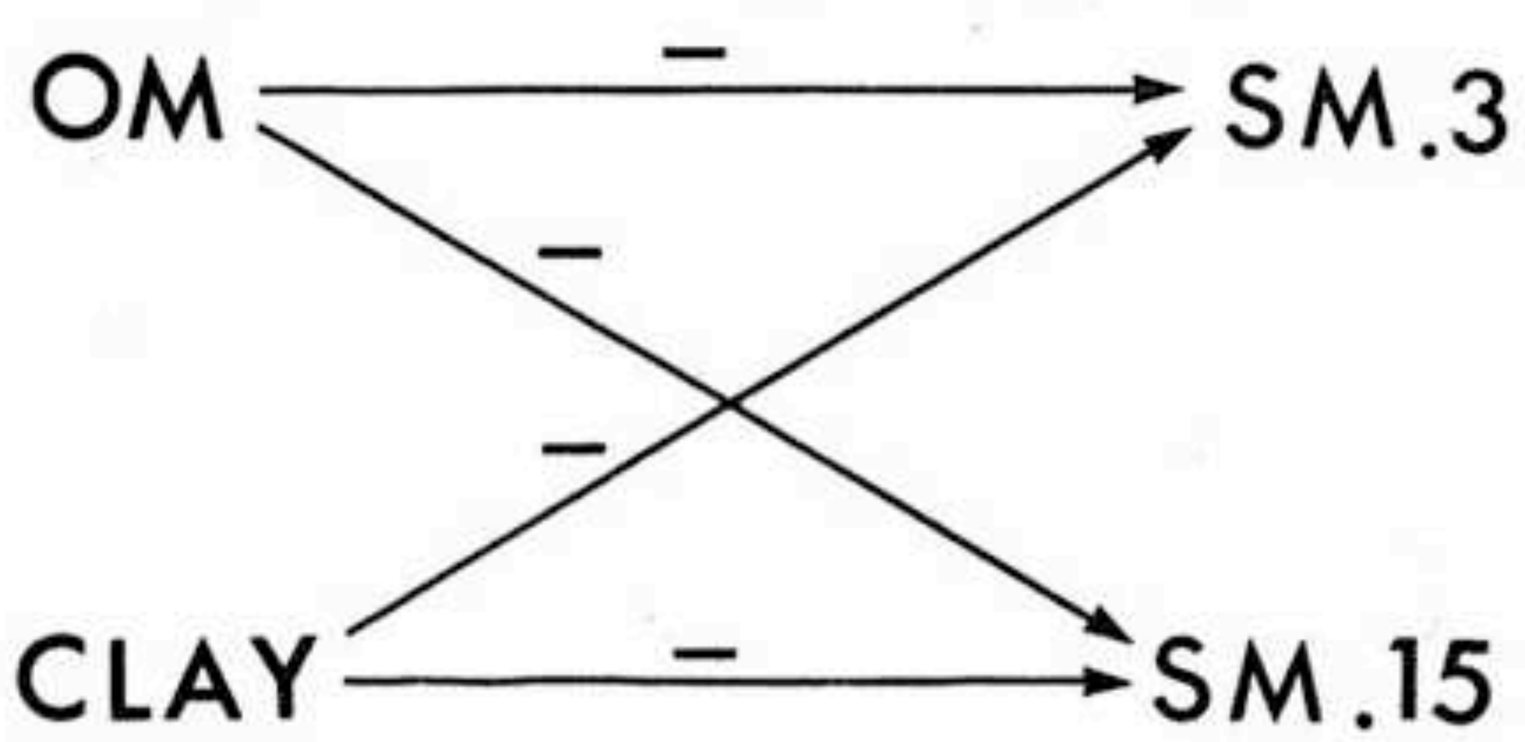
In most problems the relationships between certain variables is probably well understood and well documented. If attention is first concentrated on drawing up the expected relationships between such smaller groups of variables the problem is not as formidable as if initially faced with a long list of variables. As before, the relationships need to be drawn up on the basis of previous knowledge of the particular problem or from consideration of the biological, biochemical or physical principles involved. Ideally this should be done prior to the selection and measurement of the variables.

An example of the relationship between a group of variables from the large list (Table 2) is that between solar radiation, altitude, snow cover, soil temperature and soil moisture. Soil temperature would be expected to decrease with altitude because of adiabatic cooling and changed radiation balance with increased outgoing radiation. Temperature would be dependent on solar radiation since it is the energy source.

Precipitation in the form of snow cover may or may not increase with altitude at the elevations considered (Daubenmire 1943). Also, the lower the temperature the less would be the snow melt and the greater the snow cover. Soil temperature would be influenced by soil moisture through increasing the thermal capacity of the soil resulting in a lower temperature rise for a given input of solar radiation. Thus a partial scheme for the relationship between these variables, and the probable signs of the path coefficients, would be as follows:



Similarly, the moisture tension characteristics of the soil would be expected to be related to the amount and type of soil colloids as measured by soil organic matter and percentage clay:



In this manner small partial schemes have been worked out and combined to show the expected relationships between all the variables:

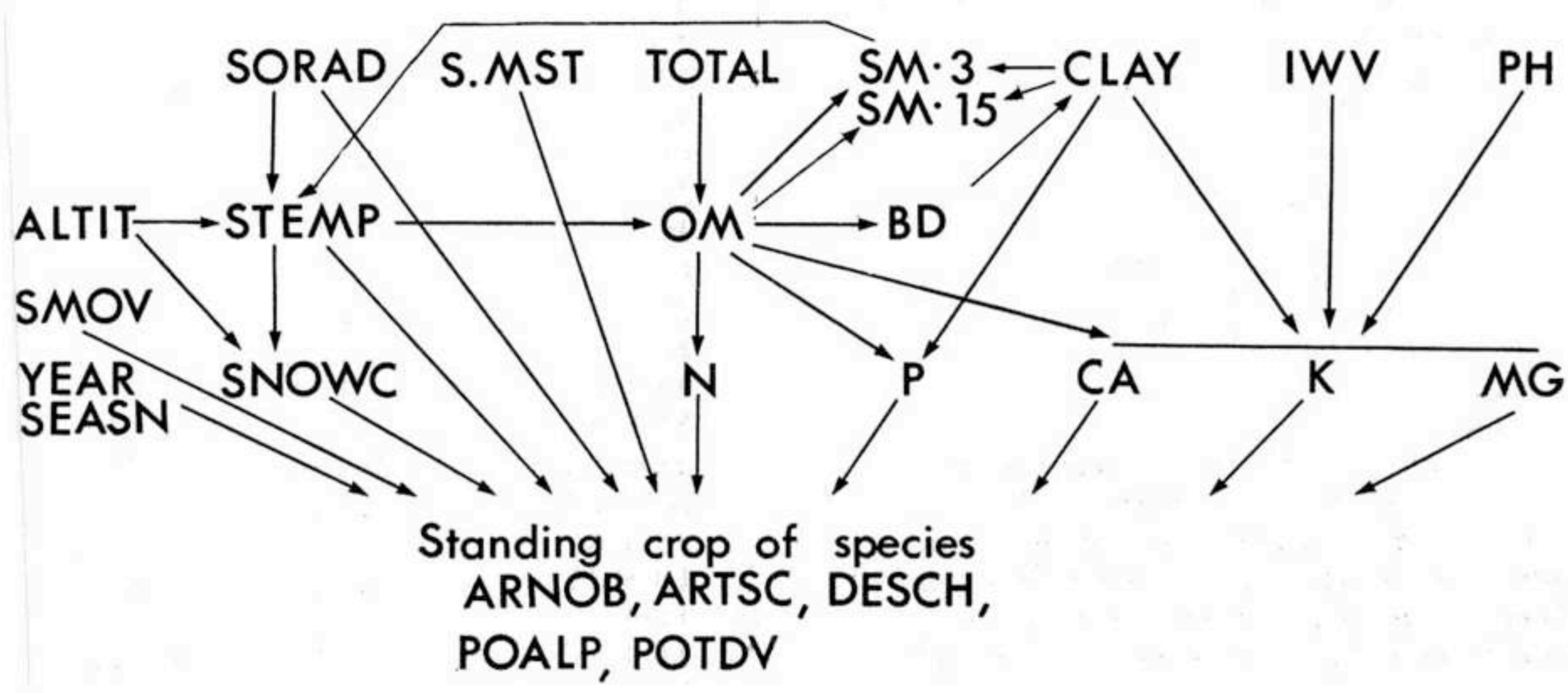
The data are then used to compute the structural equations, path coefficients and standard deviations from the sample data using the expected relationships between the variables. The computation of the standard deviations is given in Hamilton (1968).

The structural equations with path coefficients are listed below. Following each equation, in this and subsequent examples, two sets of numbers in brackets are given. The numbers in the first are the standard error of the equation, and the percentage of variation accounted for by the equation obtained by squaring the multiple correlation coefficient. The second bracket contains the significance of each path coefficient, including the constant term, determined from its standard deviation ('t' test, *=sig. at 5 percent level, **=sig. at 1 percent level), and also the standard deviation as a percentage of the value of the path coefficient. These are listed in the order in which the variables appear in the structural equation.

The structural equations and path coefficients for the relationship between the environmental factors in the example were as follows:

STEMP = 14.8—0.0472 SORAD—0.077 SM.3 —0.470 ALTIT
(2.0, 29) (*40, **25, $^{\text{ns}}$533, $^{\text{ns}}$132)

SNOWC = 8.76—0.194 STEMP + 1.11 ALTIT
(2.5, 2) ($^{\text{ns}}$93, $^{\text{ns}}$383, $^{\text{ns}}$293)



CA = —51.0 + 1.11 OM—0.21 CLAY + 11.7 PH—0.35 IWV
(8.1, 38) (**27, ns98, ns201, **21, ns203)

MG = —8.99 + 0.542 OM + 0.038 CLAY + 2.11 PH—0.254 IWV
(1.7, 17) (**31, *42, ns223, **23, *57)

K = —0.25 + 0.0369 OM + 0.0129 CLAY + 0.105 PH—0.013 IWV
(0.21, 17) (ns145, ns78, ns86, ns60, ns141)

OM = 18.5—0.445 SMST + 0.0377 TOTAL—1.32 STEMP
(7.3, 32) (**31, ns96, *43, *37)

BD = 1.17—0.0213 OM + 0.0117 CLAY
(0.17, 55) (**7, **15, ns61)

SM.3 = 10.2 + 1.84 OM + 0.67 CLAY
(7.6, 79) (**36, **7, *47)

SM.15 = 3.76 + 1.75 OM + 0.21 CLAY
(7.3, 80) (ns92, **7, ns145)

N = 0.0899 + 0.0514 OM
(0.14, 91) **36, **4)

P = 0.79 + 0.440 OM + 0.766 CLAY
(7.46, 12) (ns450, **31, *41).

In these most of the path coefficients reached statistical significance.

The relationship between the logarithm of the standing crop of the species and the group of 12 environmental factors to which they were

TABLE 2. *Description, Code Names, Mean Values and Standard Deviation of Some Environmental Variables Together with Above-Ground Standing Crop of 5 Species, from 50 Alpine Tundra Sites in Medicine Bow Range, Wyoming. (From Scott and Billings 1964).*

Description	Code Name	Mean	S.D.
Altitude of site (m above 3,000)	ALTIT	336	115
Winter snow cover (1-10 scale)	SNOWC	5.8	2.4
Soil moisture regime (1-10 scale)	S.MST	5.7	2.6
Soil movement (1-10 scale)	S.MOV	7.6	2.1
Solar radiation (percent of level site)	SORAD	97	5
30 cm soil temperature (°C)	STEMP	7.8	1.9
Clay (2μ) in top soil (percent)	CLAY	8.0	3.7
Bulk density of top soil (g/cc)	BD	1.0	0.3
Imbibitional water value of top soil (percent)	IWV	19.5	12.7
1/3 atmosphere moisture tension (percent)	SM.3	36.5	16.6
15 atmosphere moisture tension (percent)	SM.15	26.4	12.9
pH of top soil	PH	5.0	0.6
Organic matter in top soil (percent)	OM	11.4	8.9
Total nitrogen in top soil (percent)	N	0.68	0.46
Phosphorus in top soil (ppm)	P	11.9	8.0
Calcium in top soil (meq$100g^{-1}$)	CA	11.4	10.4
Magnesium in top soil (meq$100g^{-1}$)	MG	2.9	1.8
Potassium in top soil (meq$100g^{-1}$)	K	0.54	0.24
Year	YEAR	1.72	0.45
Season (days)	SEASN	0.8	8.0
Total above-ground standing crop (g.m.$^{-2}$)	TOTAL	160.5	70.8
Arenaria obtusiloba standing crop (g.m^{-2})	ARNOB	12.5	23.4
Artemisia scopulorum standing crop (g.m^{-2})	ARTSC	6.5	10.8
Deschampsia caespitosa standing crop (g.m.$^{-2}$)	DESCH	17.1	43.3
Poa alpina standing crop (g.m.$^{-2}$)	POALP	2.7	7.2
Potentilla diversifolia standing crop (g.m.$^{-2}$)	POTDV	5.1	6.8

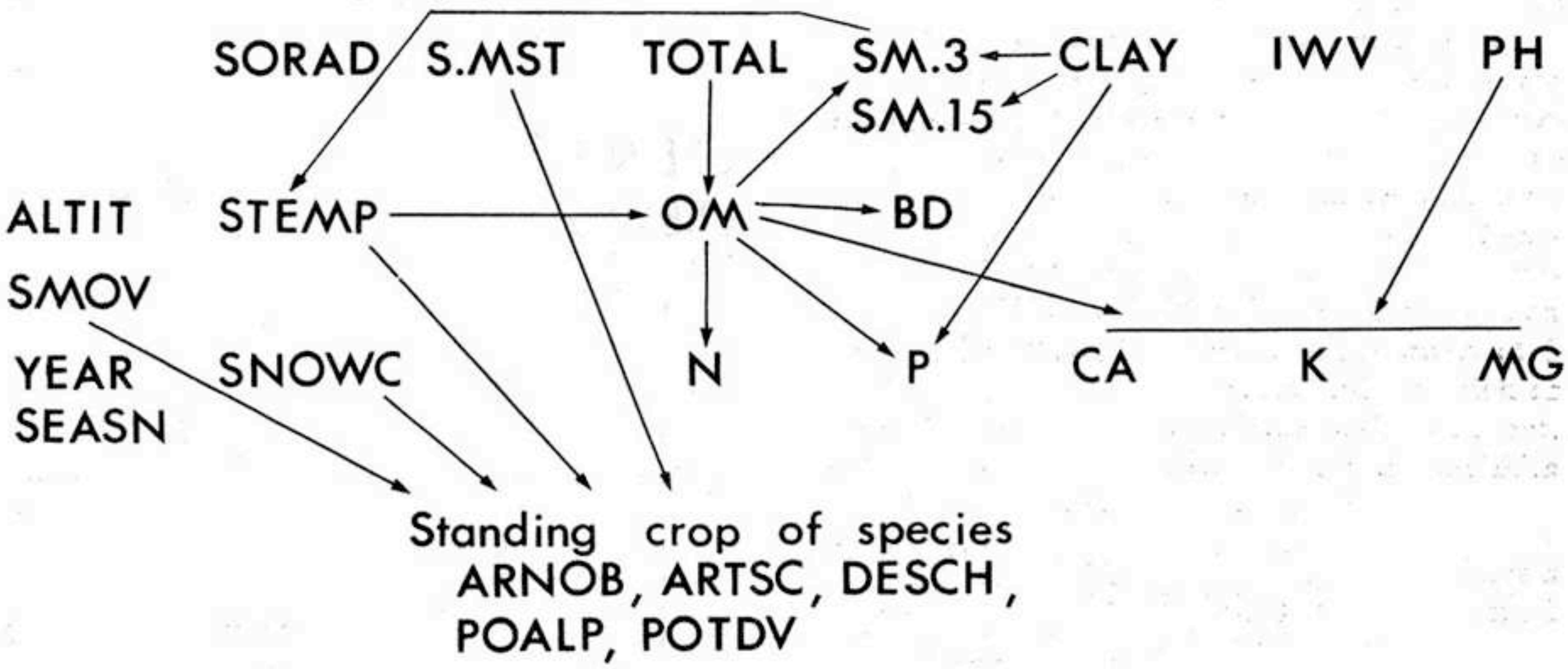
expected to be related were mostly not significant as a group and are not given.

The interpretation of such results depends in part on statistical inference. In this, path analysis, in common with other methods, depends on the test of a null hypothesis derived from the expected relationship against sample data. A prime requirement, which is frequently ignored, is that the hypothesis and the sample data must be derived independently of each other. It is inadmissable to construct a hypothesis from a preliminary analysis of the data and then to use the same data to test that hypothesis statistically (Sanderson 1954, Theil and Goldberger 1961).

Thus, in path analysis, the expected relationships which form the hypothesis should be drawn up without direct reference to the sample data. If, after fitting the data, some of the expected path coefficients are not significant then this may be regarded as evidence for rejecting parts of the original hypothesis. Should the insignificant factors be deleted a new hypothesis is formed and should be tested on a new independent set of data.

The above comments refer only to the statistical tests. The path coefficients of structural equations of any alternative scheme are correctly estimated by the method, but the variances can no longer be used for making statistical tests or estimating the confidence intervals of the path coefficients. Thus, with these limitations in mind, the relationships between the variables can be recalculated to remove all nonsignificant coefficients and the best (minimum variance) values of the remaining coefficients recalculated. While they are the best estimates no statement of their reliability can be made.

In the particular example this reduces the scheme to that shown:



and the best estimates of the modified structural equations containing only the variables which would probably be significant if tested on a further sample were as follows:

$$\mathrm{STEMP} = 10.9 - 0.079\ \mathrm{SM.3}$$
$$(0.31,\ 20)\quad (15,\ 44)$$
$$\mathrm{CA} = -53.0 + 0.614\ \mathrm{OM} + 11.5\ \mathrm{PH}$$
$$(7.6,\ 46)\quad (19,\ 21,\ 17)$$
$$\mathrm{MG} = -7.47 + 0.140\ \mathrm{OM} + 1.75\ \mathrm{PH}$$
$$(1.24,\ 53)\quad (23,\ 15,\ 18)$$
$$\mathrm{K} = 0.0152\ \mathrm{OM} + 0.0736\ \mathrm{PH}$$
$$(1.20,\ 26)\quad (21,\ 12)$$
$$\mathrm{OM} = 16.4 + 0.0406\ \mathrm{TOTAL} - 1.45\ \mathrm{STEMP}$$
$$(7.3,\ 31)\quad (33,\ 39,\ 33)$$
$$\mathrm{BD} = 1.28 - 0.0233\ \mathrm{OM}$$
$$(0.18,\ 53)\quad (3,\ 13)$$
$$\mathrm{SM.3} = 10.2 + 1.84\ \mathrm{OM} + 0.67\ \mathrm{CLAY}$$
$$(7.6,\ 79)\quad (36,\ 7,\ 47)$$
$$\mathrm{SM.15} = 5.78 + 1.72\ \mathrm{OM})$$
$$(7.2,\ 80)\quad (29,\ 7)$$
$$\mathrm{N} = 0.0899 + 0.0514\ \mathrm{OM}$$
$$(0.14,\ 91)\quad (36,\ 4)$$
$$\mathrm{P} = 0.461\ \mathrm{OM} + 0.826\ \mathrm{CLAY}$$
$$(7.36,\ 13)\quad (20,\ 18)$$

and for the relationship between species (logarithm standing crop) and environmental factors:

ARNOB = —63.8 + 10.6 SMST
(29.4, 46) (16, 15)
ARTSC = 4.84 SMOV—5.42 STEMP
(32.6, 18) (30, 25)
DESCH = 0.98 SMOV—1.59 SNOWC
(3.15, 56) (14, 11)
POALP = —52.3—6.02 SNOWC + 6.36 SMOV
(36.0, 10) (52, 48, 39)
POTDV = 4.06—0.805 SNOWC
(3.83, 0) (45, 37).

The particular results will not be discussed apart from the general comment that the variation accounted for by each of the structural equations is generally low, particularly for those between the standing crop and immediately preceding environmental factors.

CURVILINEAR RELATIONSHIPS

No assumptions are made in path analysis regarding the distribution of values of the variables. Accordingly it is possible to introduce additional variables which are squares, polynomial, product (interaction) or other transformations of other variables. This allows curvilinear relationships to be fitted. The effect of such transformations is to introduce further independent variables (primary causes) into the scheme. That one variable may be the transform of another is not taken into account in the fitting procedures and the variables are treated as if unrelated. Path analysis is similar to other methods in this (e.g. analysis of variance and multiple regression).

An example is from an experiment in which the carbon dioxide (CO_2) exchange of shoots of *Trifolium repens* was measured at a number of light intensities and temperatures (Scott and Menalda 1970). The relationship between CO_2 exchange and both these factors is known to be strongly curvilinear. In addition, because of the size of the plants and their interaction with light, there was a difference between leaf temperatures and those of the adjacent walls of the sampling chamber. It was expected that CO_2 exchange would be dependent on leaf temperature and light intensity, while leaf temperature in turn would be dependent on light intensity and chamber wall temperature. The expected curvilinear relationship was dealt with by introducing quadratic terms. The description of the variables, range of values and coding are given in Table 3 and the expected relationships shown diagrammatically as follows:

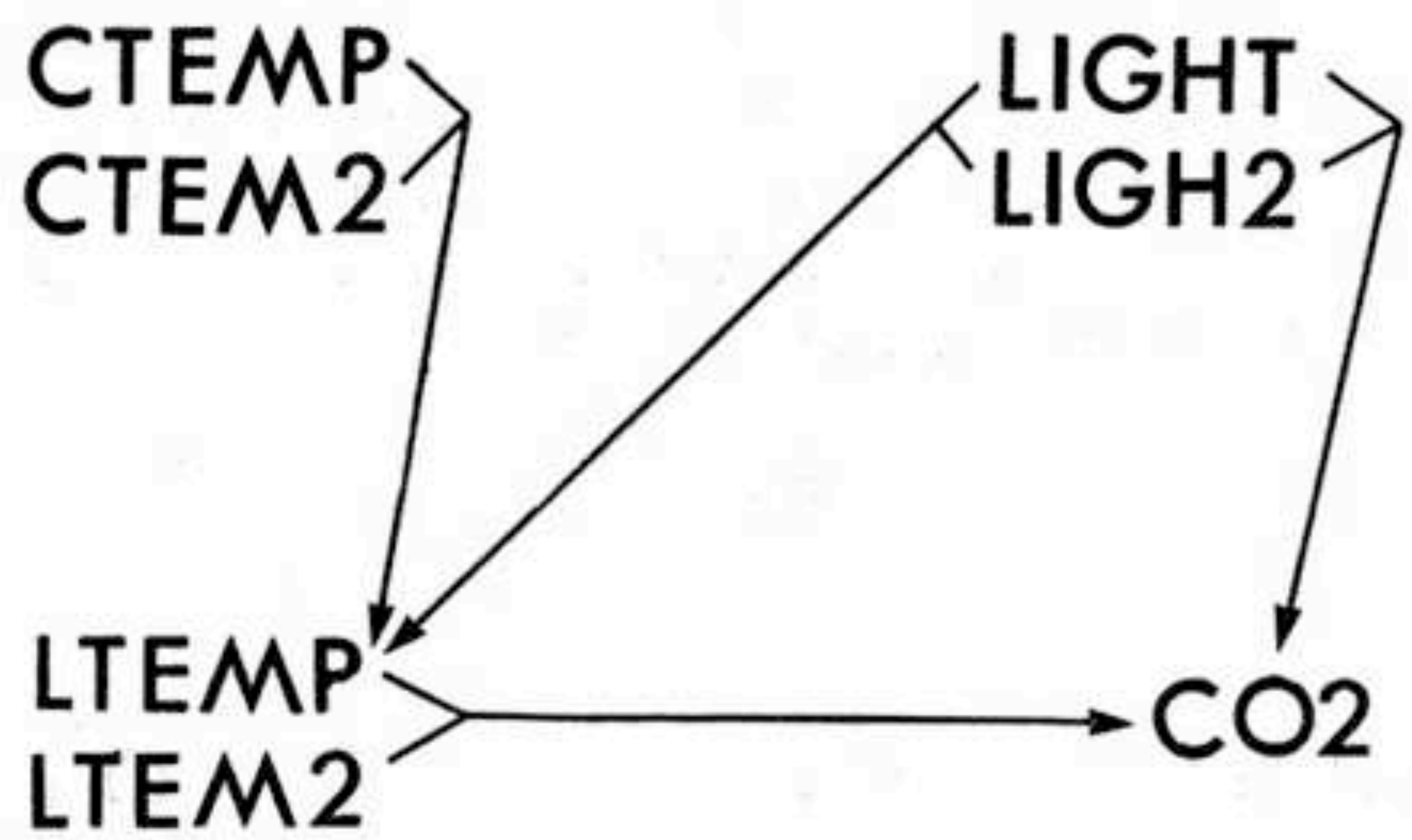


TABLE 3. *Description and Range of Variables in Curvilinear Example of CO_2 Exchange of Shoots of* Trifoloum repens.

Description of Variables	Code Name	Range
CO_2 exchange of shoots (mgm $CO_2.g^{-1}.hr^{-1}$)	CO2	—22 to 70
Temperature of fully illuminated leaves (°C)	LTEMP	7.0 to 50.3
$LTEMP^2$	LTEM2	
Temperature of sampling chamber wall (°C)		
$CTEMP^2$	CTEMP	2.6 to 50.3
Light intensity in sampling chamber (watt.m^{-2} total)	LIGHT	0 to 1,000
$LIGHT^2$	LIGH2	

From these the following structural equations and path coefficients were determined:

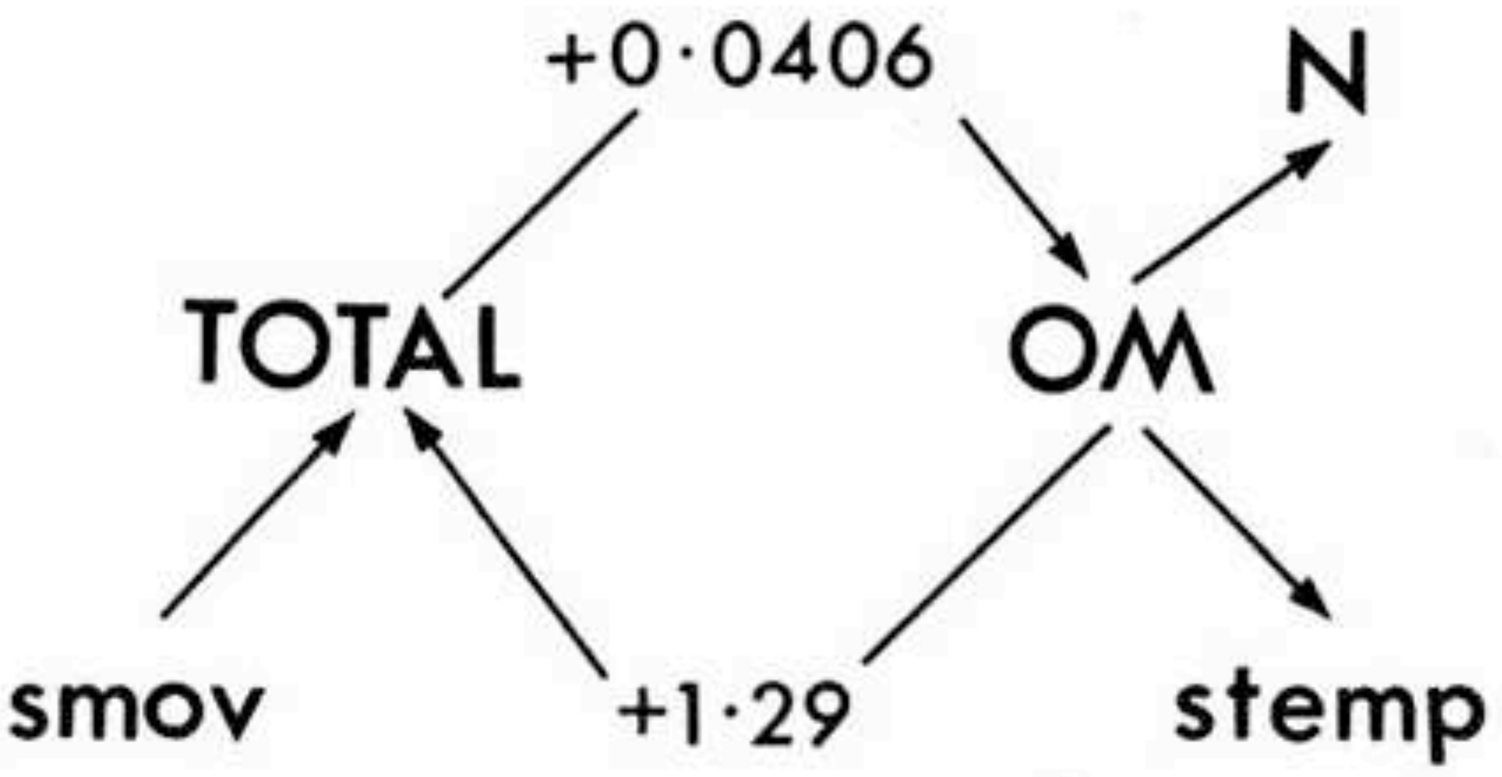
LTEMP = 0.88 + 1.08 CTEMP—3.43×10^{-3} CTEM2
+ 1.41×10^{-2} LIGHT—4.13×10^{-5} LIGH2
(1.5, 99) (ns54, **4, **24, **12, *44)
CO_2 = —21.8 + 2.37 LTEMP—0.0589 LTEM2
+ 0.162 LIGHT—1.21×10^{-3} LIGH2
(10.5, 80) (**19, **13, **9, **8, **11).

FEEDBACK

Feedback between variables in a problem would be recognised during the construction of the path diagram by arrows of opposite direction linking the same two variables either directly or through intermediary variables. Path analysis can attach quantitative coefficients to each path or link in such feedback relationships.

An example is taken from part of the previous complex example, in that total above-ground standing crop was given as one of the variables along with an index of micro-organism activity (soil temperature) in determining the organic matter content of the top soil. But organic matter content, through its effect on nutrient status, was one of the factors indirectly related to the standing crop of individual species and therefore total standing crop. Thus there was probably feedback between soil organic matter and total standing crop.

A simplified relationship is given below in which the only other variables influencing total standing crop was soil movement — the environmental variable with which it was most closely correlated.



The structural equations with path coefficients determined using this scheme were:

$$\begin{aligned} \text{OM} &= 16.4 + 0.0406\ \text{TOTAL} - 1.45\ \text{STEMP} \\ &\quad (7.3\ 31)\ \ ({**}33,\ {**}39,\ {**}33) \\ \text{TOTAL} &= 1.29\ \text{OM} + 19.2\ \text{S.MOV} \\ &\quad (27.5,\ 44)\ \ ({}^{\text{ns}}86,\ {**}10). \end{aligned}$$

The two main path coefficients are indicated in the diagram even though one did not reach statistical levels of significance. Positive feedback was indicated by the similarity in sign of both coefficients.

If coefficients were of opposite sign negative feedback occurs. Equilibrium is often obtained in those cases. An example is illustrated by Turner & Stevens (1959).

MIXED ESTIMATION

In the examples given to this stage all the quantitative estimates of path coefficients have been derived from the sample data. But there is an extension of the path analysis technique which allows *a priori* quantitative data to be incorporated into the solution.

In any investigation, the taking of measurements is usually prompted by the belief that there are, or may be, relationships between the variables concerned. Generally it is the qualitative aspects of these expected relationships which form the hypothesis which is tested against the sample data.

However, it is also reasonably common to reject, or at least question, the statistical inferences drawn from data, either by maintaining that a particular relationship exists even though it was statistically non-significant in the particular data, or by refusing to accept a "fortuitous" statistically significant relationship for which there seems to be no biological basis. The reason is that generally the investigator knows more about the relationships than he is able to incorporate into the initial hypothesis. For instance he may know whether it is a direct or inverse relationship between particular variables, or, indeed, may be able to specify the quantitative relationships to within certain limits. It is these *a priori* quantitative data which are difficult to include in most statistical techniques, but which are often the basis for criticism of the results subsequently obtained. Often it would be more desirable to include the *a priori* data in the initial statement of the problem along with the sample data and, if the two are not incompatible, to combine both sources of information.

In path analysis *a priori* data are incorporated into the computation as subsidiary equations stating the mean and variances of selected path coefficients or combinations of path coefficients in particular structural equations, e.g.

$$p_{12} = b \pm v$$

$$p_{12} + p_{32} - p_{42} = b + \text{appropriate variances and covariances.}$$

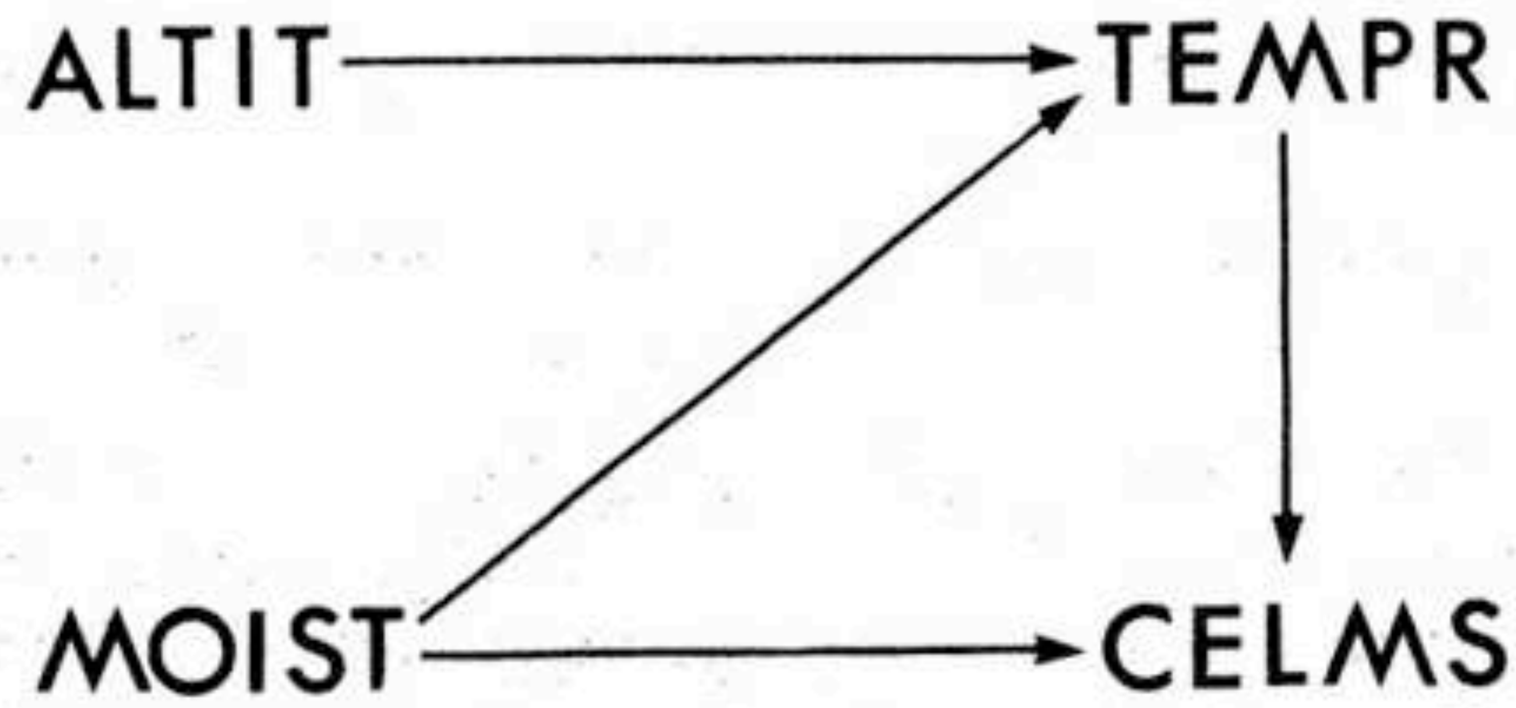
Several such restrictions can be applied simultaneously to each structural equation (presuming of course that they are not incompatible). The greater the accuracy of the *a priori* restrictions the more they will influence the final value of the path coefficients. Conversely, if the means or variances can be estimated only approximately then the *a priori* restrictions will have only a slight influence of the final values e.g. where only the sign of a path coefficient is known (say positive) and where this restriction has to be incorporated by specifying a positive number with a large variance. In all cases it is prudent to check that the variances of the *a priori* restrictions are not statistically different from the unrestricted estimates of the path coefficients before combining both estimates.

An example is a problem concerned with an alpine area and the relationship between altitude, soil temperature, soil moisture and the frequency of a particular species (Table 4). It was expected that soil temperature would be influenced by altitude and soil moisture while the frequency of

TABLE 4. *Description and Range of Variables in Mixed Estimation Example of Frequency of* Celmisia spectabilis.

Description of Variables	Code Name	Range
Altitude ((m above 1,000)/1000)	ALTIT	0.199 to 0.564
Soil moisture (% in gypsum blocks)	MOIST	84 to 116
Midsummer 30 cm soil temperature (°C)	TEMPR	8.3 to 12.2
Frequency of *Celmisia spectabilis* (%)	CELMS	0 to 44

the species would be influenced by soil temperature and soil moisture, and that altitude would exert an effect only indirectly through its effect on soil temperature, *viz.*

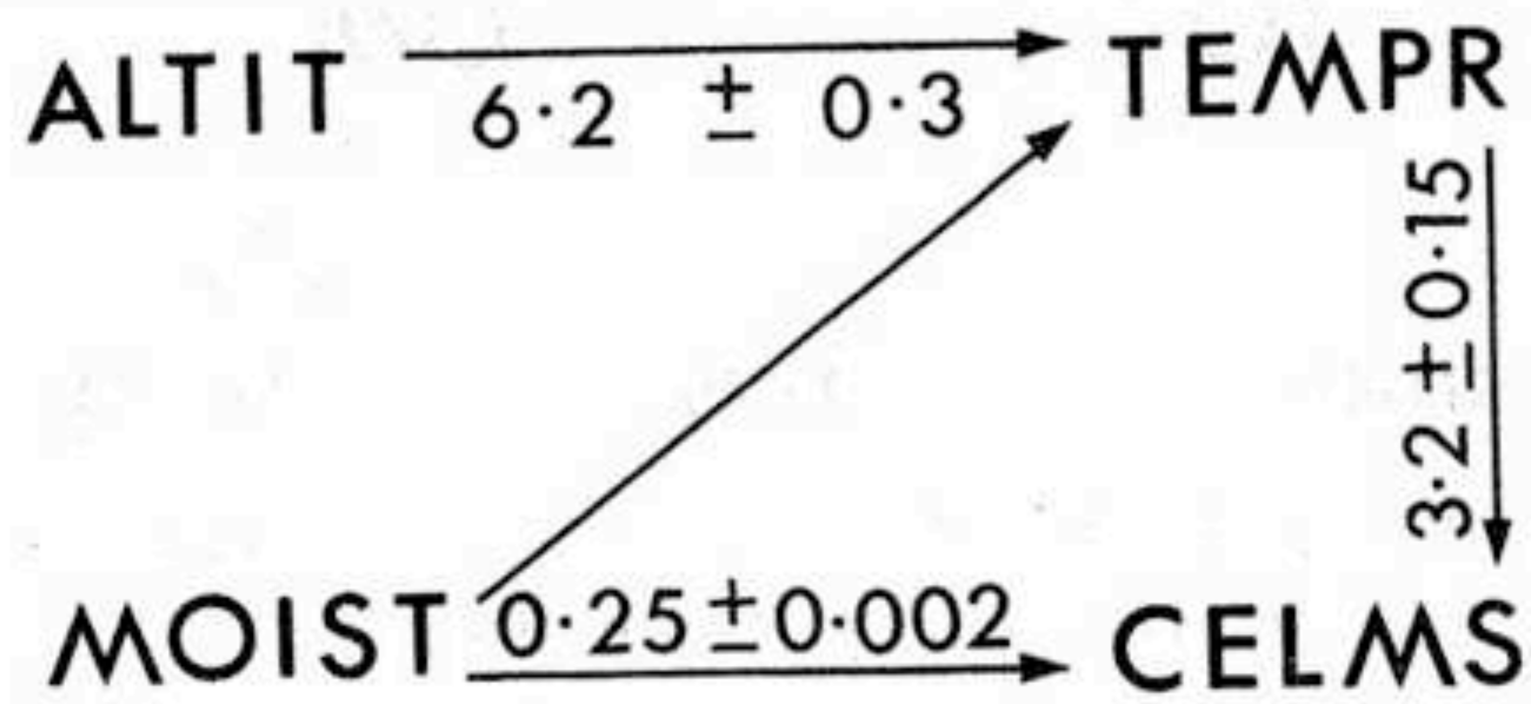


Path analysis of data from 20 sites gave the following structural equations and path coefficients (Mult. R^2. not given):

$$\text{TEMPR} = 13.6 - 5.55\ \text{ALTIT} - 0.0105\ \text{MOIST}$$
$$(0.63)\ (^{**}16,\ ^{**}25,\ ^{ns}171)$$
$$\text{CELMS} = -14.4 + 3.82\ \text{TEMPR} + 0.141\ \text{MOIST}$$
$$(13.7)\ (^{ns}127,\ ^{ns}140,\ ^{ns}265).$$

In these most of the path coefficients are not significant as indicated by their standard deviations, which are larger than the path coefficients.

But suppose there were other reasons for believing that such relationships existed — as, for example, that more extensive microclimate studies (hypothetical) had established that mean soil temperatures decreased 6.2°C per 1,000 m with a standard deviation of 0.3°, and that, in some manner, single factor experiments in growth cabinets had established that the frequency of *C. spectabilis* increased 3.2 ± 0.15 percent per 1°C in temperature and increased 0.25 ± 0.002 percent per 1 percent of soil moisture. Because of their variability the values based on the field measurements are not inconsistent with these:



An extension of path analysis combines both the *a priori* estimates of path coefficients and their variances, and those determined from the particular set of data. The details of the method are given by Theil & Goldberg (1961) and Hamilton (1968).

When the problem was solved using this additional information the following structural equations and path coefficients were obtained:

$$\text{TEMPR} = 14.4 - 5.81\ \text{ALTIT} - 0.0171\ \text{MOIST}$$
$$(0.62°)\ (^{**}7,\ ^{**}20,\ ^{ns}51)$$
$$\text{CELMS} = -45.7 + 3.20\ \text{TEMPR} + 0.248\ \text{MOIST}$$
$$(13.4)\ (^{**}34,\ ^{**}37,\ ^{**}18).$$

Comparison between the two pairs of equations show that the standard deviations of the path coefficients have been reduced, and that two of the coefficients for which there was no additional information had now reached statistical levels of significance. The coefficients in the last structural equation for the frequency of *C. spectabilis* were largely determined by the assumed accuracy of the additional information.

Comparison With Other Statistical Methods

The analysis of ecological data involves two aspects, (i) estimations of the magnitude of particular relationships, and (ii) decision on the reliability of these estimates in the light of chance or other errors. The first is curve fitting since its concern is with showing the graphical relationships within the data, whether this be by free-hand graphs or mathematical methods of fitting equations. The second is concerned with statistic, sampling theory and probability and will require some element of randomness. The distinction is warranted since the requirements of the data may be different for each. The two aspect are combined in most of the common statistical methods.

It has been shown in another paper (Scott 1969) that in terms of the user there can be some division of the common statistical methods based on: (i) subdivision into dependent and independent variables (ii) whether the variables are qualitative or quantitative and (iii) on the nature of random variation or error.

Path analysis is one of a group of techniques applicable where there are quantitative measurements of both the dependent and the independent variables. But along with Type 1 multiple regression of this group, it does not require dependent and independent variables to be random samples from a normal distribution of values. Thus there is a minimum of restrictions on sampling, and the various levels of the factor variables could be chosen by the investigator. This is a characteristic of only a few methods and is suited to the experimental approach where the response over a range of selected values is often required.

Path analysis is similar to multiple regression in giving explicitly a quantitative linear equation to the relationship between the variables, but differs in the requirements for the independent variables. In multiple regression the independent variables, as their name suggests, are assumed to be independent in their effect on the dependent variable. This is even so in Type 1 multiple regression, though in practice interaction, or curvilinear relationships, are often introduced. Path analysis withstands these criticisms in some circumstances. First the relationship between the independent variables in one relationship may be specified in associated relationships in which they appear as dependent variables. Second, the errors in some variables (intermediate effects) are taken into account when these are independent variables in other relationships. But, as in regression, the relationship between any introduced interaction or curvilinear terms are not taken into account in the mathematics of the method and they are treated as if they were further independent variables.

Where there is *a priori* quantitative data this may be used as a null hypothesis against which the sample data are tested in almost all the methods. But it appears that methods of combining this *a priori* data with the sample data have only been developed for multiple regression and path analysis.

The distinction between the methods is somewhat artificial in that they form a continuum of methods which, with one exception, are based on the least squares linear model, and where qualitative variables are treated as quantitative states. The differences are mainly in the assumption necessary to draw statistical inferences from them.

The statistical requirements of the data for the various methods are often more restrictive than commonly believed. The only justification in practice is that many of the methods have been shown to give good approximations even when the underlying assumptions are only partly valid.

Even where the statistical assumptions are not correct the mathematical computations may still give the best estimates of the required quantitative relations or equations for the data.

Synthesis of Ecological Data

1. *Path Analysis as a Conceptual Framework*

To this stage in the paper path analysis has been described primarily as a method of statistical analysis, and as such showed several advantages. The same features of the method also give it some potential as a means for synthesis of ecological data. The general problem of synthesis is first discussed and is followed by a discussion on the possible role of path analysis.

The aims of ecological research in a given situation are threefold:

(1) Isolate the variables which are important to the organisms concerned.

(2) Determine how these variables bring about their effect. This will depend on understanding the relationship between the various aspects of the environment and organisms, and may involve field, laboratory and controlled environment experimentation.

(3) In some manner synthesize the information from all sources so that quantitative predictions can be made about future behaviour from measurements of the appropriate variables.

Understanding of a particular problem will progress through several stages. Initially, one will be able only to speculate or hypothesize on the relevant variables and interactions. This will be followed by a stage where one can make qualitative statements about the relationships. Subsequently, as understanding and sophistication increase, it will be possible to make quantitative measurements and establish empirical quantitative relationships, which in turn may ultimately be replaced by exact theoretically established relationships which may or may not contain empirically determined components.

If synthesis is to be achieved at any of these stages then any method of data handling would have to satisfy both ecological and mathematical requirements.

The prime ecological requirement is that the variable must be treated in a manner which reflects biological, biochemical and physical understanding of the processes involved. Any problem or part of a problem may require working at various levels of complexity, e.g. molecular to global. While the ultimate understanding must be based on the exact functional relationships concerned, it may be sufficient in a particular problem to use empirical approximations. Another consideration is that a particular factor can influence an organism in a number of ways, e.g. through survival, rate of growth, or differentiation (Scott and Billings 1964, p. 267), and on a number of time scales (Billings 1952, Salisbury *et al.* 1968).

The prime mathematical requirement is that any method of synthesis make efficient exact quantitative manipulations on any of the relationships found whether they be qualitative, empirical quantitative or exact theoretical relationships. In the early stages of a problem there will be a requirement for a method with a statistical basis to determine whether a particular conclusion is justifiable on the basis of sample data, to construct confidence intervals and to discriminate between alternatives. However, as relationships are isolated and intensive work is directed at accurate quantifying, then there would be less demand for the statistical properties of a method and the problem would become one of curve fitting. Also, as understanding of a particular system is refined, there will be a tendency for a static description to become inadequate and for it to be replaced by a dynamic approach.

In an ecological situation where there are many variables, the number of interactions between them is likely to be finite. This is because the interactions arise from particular biological or physical processes linking particular variables (e.g. wind—by increasing the turbulent transfer of heat and water vapour away from a leaf). Therefore any variable is likely to be directly affected by only a few other variables and in turn have only a direct effect on some others. More complex situations, with their components, variables and interactions, are combinations of these simpler direct relationships.

That complex situations are combinations of simpler functional relationships between groups of variables seems the logical way to approach ecological problems. This is exemplified in the path analysis or systems approach. The path analysis diagram provides a convenient conceptual tool for showing the variables and relationships in a particular problem even if the mathematical aspects of the method are not used. Some comments on the use of the path analysis approach in this wider sense are relevant.

The use of the path analysis approach in a particular situation, although not making the problem any less complex, makes it manageable because attention can be directed at each link in the scheme as required. The analyst can take a group of variables and from all possible relationships that may exist determine which causal relationships are present and the basis of their functional relationships. Conversely the synthesist can use the concept to combine the known relationships between groups of variables to build up a scheme showing the relationships between many variables.

The path analysis approach is to be compared with some other methods of analysis and synthesis in ecology.

The holistic approach (e.g. Billings 1952), with its stress on many variables, and the implication of interaction between all variables, while acceptable in general terms, does not provide much help when faced with a particular problem where some relationships are known to be important, some unlikely and others impossible. Also it does not provide the necessary framework for handling quantitative information.

Another approach is to attempt to synthesize by reducing the environmental effect on organisms to a few factors, five in the case of Major (1951) and three each in the case of Loucks (1962) and Waring and Major (1964). But synthesis should not be regarded as the combination or merging of many factors into a single factor, but rather the ability to look at the effect of many interactions simultaneously.

A related empirical approach to synthesis is the use of multiple regression analysis (e.g. Coile 1952, Fritts 1958, Scott and Billings 1964). In this, the environmental variables considered relevant are used to determine which group of these most closely correlates with the variable of interest. The criticism of this approach is that it does not take into account any known relationship between the independent variables.

The weakness in all these approaches is in regarding ecological systems as only two stages with the organism or community on one hand and the environment on the other. A multistage process is a more realistic approximation of an ecosystem, where a factor "A" may influence a factor "B", which in turn affects "C", "D", etc. Hence the appeal of the path analysis approach.

Synthesis using the causal analysis approach is probably best exemplified by recent work on model building. Examples are the models of light interception by vegetation of de Wit (1965) and Duncan *et al.* (1967); or Watt's (1964) and Holling's (1965) models of the behaviour of insect populations; or Olsen's (1964) simulation of production of terrestrial vegetation. Such model building requires consideration of particular processes in all their complexity and then the combination of several such component processes to study larger schemes.

The interpretation of the path analysis approach has been in terms of cause and effect with the arrows on the diagram indicating that the interaction occurs in a particular direction. In a second sense the causal diagram is akin to a computer flow diagram in indicating the order in which variables are required or calculations made, while the coefficients indicate the magnitude of these interactions. This leads to a third interpretation on can *et al.* (1967); or Watt's (1964) and Holling's the arrows, namely the passage of time, in which the coefficients become rate coefficients. Intuitively this seems to be the most valuable approach in ultimately describing the dynamic behaviour of a system. Finally there will be a situation where the arrows would simply be weighting factors as when combining quantities of similar dimensions (transfer coefficients), or empirical conversion factors when combining quantities of different dimensions.

The comments in this section have given the path analysis approach a broader interpretation than is applicable to the particular statistical method and have been more fully developed in the IBP programme (Swartzemann *et al.* 1971).

2. *Path Analysis as a Mathematical Procedure for Synthesis*

From necessity, any synthesis of ecological data is usually concerned with the relationship between many variables. Within the causal multistage approach the path analysis technique can be used in many circumstances for estimating the required quantitative relationships.

When there is a single set of quantitative measurements of all of the variables relevant to a particular problem and when or if there is some understanding of the expected relationship between the variables, path analysis provides a means of determining the required quantitative coefficients, equations and associated tests of significance. The main requirements are that there is a complete set of data for all variables and that the relationship between groups of variables is suitably approximated by linear additive equations. The first examples given earlier in the paper demonstrate the use of the method in this context.

When there is additional quantitative information available the method of mixed estimation provides a method of combining information from different sources and of differing accuracies into a unified estimate. It is probably this property of unifying empirical estimates from different sources along with the potential stability of the path coefficients, which provides the method's greatest potential in synthesizing ecological data.

This can be further illustrated by another example from work given earlier in the paper. The example showing the use of path analysis in fitting curvilinear relationships was based on an experiment in which the CO_2 exchange of shoots of *Trifolium repens* was measured at a number of light intensities and temperatures. The relationship between leaf temperature, light intensity and chamber temperatures, and between CO_2 exchange, light intensity and chamber temperature, were the same as given previously, i.e.:

$$\begin{aligned} \text{LTEMP} &= 0.88 + 1.08\ \text{CTEMP} - 3.4\text{x}10^{-3}\ \text{CTEM2} \\ &\quad + 1.41\text{x}10^{-2}\text{LIGHT} - 4.13\text{x}10^{-5}\ \text{LIGH2} \\ &\quad (1.5, 99)\ ({}^{\text{ns}}54, {}^{**}4, {}^{**}24, {}^{**}12, {}^{*}44) \\ \text{CO2} &= -21.8 + 2.37\ \text{LTEMP} - 0.0589\ \text{LTEM2} \\ &\quad + 0.162\ \text{LIGHT} - 1.21\text{x}10^{-3}\ \text{LIGH2} \\ &\quad (10.5, 80)\ ({}^{**}19, {}^{**}13, {}^{**}9, {}^{**}8, {}^{**}11). \end{aligned}$$

A subsequent experiment (Scott 1970) with the same species was concerned with how the temperature at which the plant was grown influenced its CO_2 exchange at two light intensities and a similar range of temperatures. The difference between the two experiments was that, in the second, growing temperature was introduced as a further variable, and that measurements were made at only two light intensities. The description of variables and their range of values is given in Table 5 and the expected relationship between the variables in the following diagram:

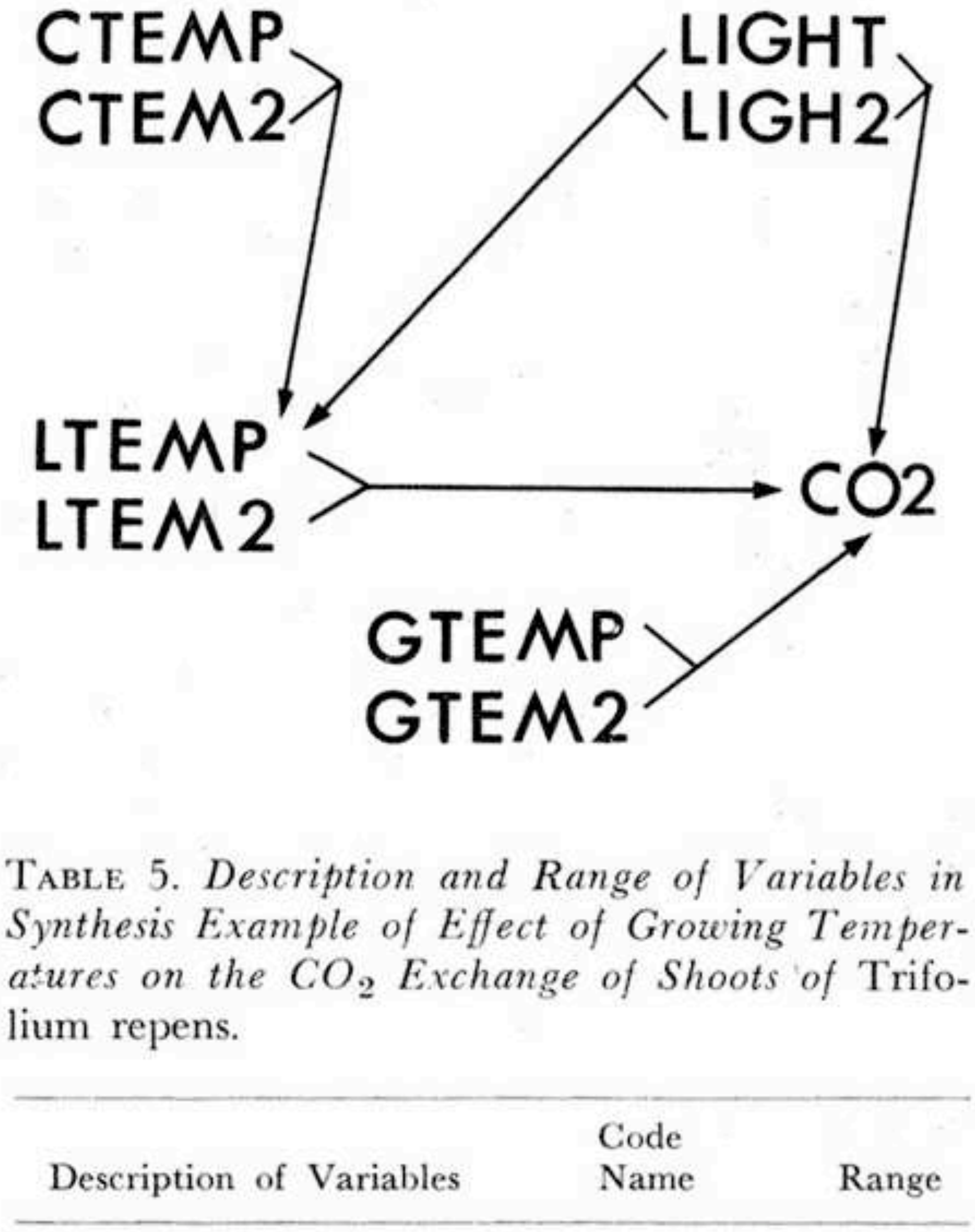


TABLE 5. *Description and Range of Variables in Synthesis Example of Effect of Growing Temperatures on the CO_2 Exchange of Shoots of* Trifolium repens.

Description of Variables	Code Name	Range
CO_2 exchange of shoots ($\text{mgm.g}^{-1}.\text{hr}^{-1}$)	CO2	—11 to 114
Temperature of fully illuminated leaves (°C)	LTEMP	2.3 to 45.
LTEMP^2	LTEM2	
Temperature of sampling chamber wall (°C)	CTEMP	0.6 to 40
CTEMP^2	CTEM2	
Light intensity in sampling chamber (watt.m^{-2} total)	LIGHT	0 or 1,00
LIGHT^2	LIGH2	
Temperature at which plant was grown (°C)	GTEMP	10°, 20°, o 30°
GTEMP^2	GTEM2	

The structural equations and path coefficient calculated from the data were as follows:

$$
\begin{aligned}
\mathrm{TEMP} = {} & 2.26 + 0.975\ \mathrm{CTEMP} - 5.46\times10^{-3}\ \mathrm{CTEM2} \\
& + 1.40\times10^{-2}\ \mathrm{LIGHT} - 3.10\times10^{-5}\ \mathrm{LIGH2} \\
& (1.5, 99)\ ({**}14, {**}3, {**}16, {**}22, {}^{\mathrm{ns}}100) \\
\mathrm{CTEM2} = {} & -56.2 + 1.00\ \mathrm{LTEMP} - 0.0300\ \mathrm{LTEM2} \\
& + 0.144\ \mathrm{LIGHT} - 8.62\times10^{-4}\ \mathrm{LIGH2} \\
& + 5.11\ \mathrm{GTEMP} - 0.121\ \mathrm{GTEM2} \\
& (15.8, 73)\ ({**}21, {*}48, {**}31, {**}22, \\
& {**}38, {**}24, {**}24).
\end{aligned}
$$

The equation for leaf temperature was very similar to that of the earlier experiment, except for the difference in the constant term. For CO_2 exchange the path coefficients were similar to those of the previous experiment except for the coefficient for LTEMP. The latter differed markedly from that in the first experiment but also had a larger standard deviation.

The structural equations and path coefficients of the second experiment were recalculated using the path coefficients for light, chamber and leaf temperature as *a priori* additional information and the following were obtained:

$$
\begin{aligned}
\mathrm{LTEMP} = {} & 2.26 + 0.975\ \mathrm{CTEMP} - 5.42\times10^{-3}\ \mathrm{CTEM2} \\
& + 1.44\times10^{-2}\ \mathrm{LIGHT} - 3.56\times10^{-5}\ \mathrm{LIGH2} \\
& (1.2, 99)\ ({**}8, {**}1, {**}3, {**}2, {**}7) \\
\mathrm{CO2} = {} & -68.1 + 2.46\ \mathrm{LTEMP} - 0.0567\ \mathrm{LTEM2} \\
& + 0.144\ \mathrm{LIGHT} - 1.05\times10^{-3}\ \mathrm{LIGH2} \\
& + 5.06\ \mathrm{GTEMP} - 0.120\ \mathrm{GTEM2} \\
& (14.5, 77)\ ({**}15, {**}3, {**}2, {**}2, \\
& {**}2, {**}22, {**}23).
\end{aligned}
$$

The result of using the additional information was to reduce the standard deviations of estimates of leaf temperature and CO_2 exchange by 21 percent and 7 percent respectively. It has also greatly reduced the standard deviations of the path coefficients for which there was *a priori* information and slightly reduced the standard deviations of the other variables.

As the example shows, the method is best suited to synthesizing relationships in problems where the number of variables being considered is expanding and where the results of simpler experiments can be used as the *a priori* restrictions in the more complex situations. The same example also showed the use of a previous experiment to determine the shape of a particular response curve (CO_2 exchange versus light intensity) and using this in a second more complex scheme where there were measurements at only two points on that particular response curve. If the coefficients determined in one experiment are to be used as restrictions in another then it would be desirable that other conditions were comparable and constant. This restriction may be desirable, but is not absolute because, if the particular relationship between the variables has been correctly identified, then the values obtained should be similar in whatever context they were obtained, irrespective of how other relationships or variables may differ between the two situations.

There are limitations to path analysis as a mathematical procedure for synthesis, but these will become a factor only when a particular problem has reached a degree of sophistication. Initially, the statistical characteristics of the method will be of importance in testing particular hypotheses and determining confidence intervals. As understanding of a particular problem increases and measurements are refined there will be less need for these statistical characteristics. Even in such circumstances path analysis may remain the best curve-fitting procedure for estimating unknown coefficients.

The computational method used in this paper is a single estimate from a two stage least squares solution. It is probably only when the path coefficients have reached a certain degree of stability that it would be justifiable in using further computational refinements such as the iterative methods using the two stage least squares, or in the three stage methods which determine the best estimate for all coefficients considered simultaneously (Zellner and Theil 1962).

The method as outlined probably does not make the most efficient use of data where there is a large amount of *a priori* data, for example where some path coefficients are known exactly or where there are estimates of particular path coefficients from several sources. But this should be a minor problem as efficient estimates should be available either as elaborations of the theory of path analysis with mixed estimation or from general principles of adjustment of data (Deming 1946). Path analysis is not ideally suited for relationships other than linear though, as the curvilinear example showed, it may be possible to get a good approximation using transformations.

One of the utilitarian purposes in studying any relationship is the hope that some processes can be optimised. In this path analysis forms a good precursor for the application of linear programming. Linear programming is another technique which was primarily developed by economists but which has recently been used in ecological problems (e.g. van Dyne 1966). The method determines the maximum possible values of a linear combination of a group of variables given that other linear combinations of these variables are subject to certain constraints. For example, given the relationship between individual species and environmental factors, the relative palatability of the species, the range of values and the relationship between the environmental variables, the technique could determine the most favourable site in terms of plant growth for grazing. Thus the structural equations from path analysis together with other restrictions could be used directly in linear programming and would overcome some of the difficulties in using multiple regression relationships in such contexts.

ACKNOWLEDGMENTS

Thanks are extended to Margaret S. Hamilton, G. A. Vignaux and the late A. C. Glenday, all of the Applied Mathematics Division, D.S.I.R., Wellington, for clarification of the method and the development of the computer programme (DATAPREP and SIMEQN) on which the work in this paper is based.

REFERENCES

AITKEN, A. C. 1944. *Statistical mathematics.* Oliver & Boyd (Lond) 153 pp.

BILLINGS, W. D. 1952. The environmental complex in relation to plant growth and distribution. *Quarterly Review of Biology* 27: 251-265.

CAMPBELL, E. D.; TURNER, M. E.; WRIGHT, M. F. 1960. *A handbook of path regression analysis.* Dept. of Biophysics and Biometry, Medical College of Virginia, Richmond.

COILE, T. S. 1952. Soil and growth of forests. *Advances in Agronomy* 4: 329-298.

DAUBENMIRE, R. F. 1953. Vegetational zonation in the Rocky Mountains. *Botanical Review* 9: 325-393.

DEMING, W. E. 1946. *Statistical adjustment of data.* Wiley (N.Y.) 261 pp.

DUNCAN, W. G.; LOOMIS, R. S.; WILLIAMS, W. A. HANAU, R. 1967. A model for simulating photosynthesis of plant communities. *Hilgardia* 38: 181-205.

FERRARI, T. J. 1963. Causal soil-plant relationships and path coefficients. *Plant and Soil* 19: 81-96.

FERRARI, T. J. 1964. Auswertung biologischer kettenprozesse mit hilfe von pfadkoeffizienten. *Biometrische Zeitschrift* 6: 89-102.

FRITTS, H. C. 1958. An analysis of radial growth of beech in a central Ohio forest during 1954-1955. *Ecology* 39: 705-720.

HAMILTON, M. S. 1968. A programme for the estimation of simultaneous linear equation models. Applied Mathematics Division, DSIR Report No. 4.

HOLLING, G. S. 1965. The functional response of predators to prey density and its role in mimicry and population regulation. *Memoirs Entomology Society Canada* 45: 5-60.

LI, C. C. 1955. *Population genetics.* 2nd ed. Chicago U.P. 366 pp.

LOUCKS, O. L. 1962. Ordinating forest communities by means of environmental scalers and phytosociological indices. *Ecological Monographs* 32: 137-166.

MAJOR, J. 1951. A functional, factorial approach to plant ecology. *Ecology* 32: 392-412.

OLSON, J. S. 1964. Gross and net productivity of terrestrial vegetation. *Journal of Ecology* 52 (Suppl.): 99-118.

SALISBURY, F. B.; SPOMER, G. G.; SOBRAL, M.; WARD, R. T. 1968. Analysis of an alpine environment. *Botanical Gazette* 129: 16-32.

SANDERSON, F. H. 1954. *Methods of crop forecasting.* Harvard U.P. 259 pp.

SCOTT, D. 1966. Interpretation of ecological data by path analysis. *Proceeding New Zealand Ecological Society* 13: 1-4.

SCOTT, D. 1969. Relationship between some statistical methods. *Proceedings of the New Zealand Ecological Society* 16: 58-64.

SCOTT, D. 1970. CO_2 exchange of plants 3. Temperature acclimatisation of three species. *New Zealand Journal of Botany* 8: 369-79.

SCOTT, D.; BILLINGS, W. D. 1964. Effect of environmental factors on the standing crop and productivity of an alpine tundra. *Ecological Monograph* 34: 243-270.

SCOTT, D.; MENALDA, P. H. 1970. CO_2 exchange of plants 2. Response of six species to temperature and light intensity. *New Zealand Journal of Botany* 8: 361-8.

SWARTZMAN, G. (co-ordinator) 1971. Some concepts of modelling. Tech. Report No. 32 Natural Resource Ecology Lab., Colorado State University.

THEIL, H. 1958. *Economic forecasts and policy.* North-Holland Publ. Co. (Amst). 567 pp.

THEIL, H.; GOLDBERGER, A. S. 1961. On pure and mixed statistical estimation in economics. *International Economic Review* 2: 65-78.

TINTER, G. 1952. *Econometrics.* Wiley (N.Y.)

TUKEY, J. W. 1954. Causation, regression and path analysis p. 35-66. *In* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. T. Lush (eds.) *Statistics and mathematics in biology.* Iowa State College Press (Ames) 632 pp.

TURNER, M. E.; STEVENS, C. D. 1959. The regression analysis of causal paths. *Biometrics* 15: 236-258.

VAN DYNE, G. M. 1966. Application and integration of multiple linear regression and linear programming in renewable resource analysis. *Journal of Range Management,* 19: 356-362.

WATT, K. E. F. 1962. Dynamic programming, "look ahead" "programming", and strategy of insect pest control. *General Systems* 9: 181-188 (from *Canadian Entomologist* 95).

WARING, R. H.; MAJOR, J. 1964. Some vegetation of the Californian Coastal Redwood region in relation to gradients of moisture, nutrients, light and temperature. *Ecological Monographs* 34: 167-215.

WIT DE, C. T. 1965. Photosynthesis of leaf canopies. *Verslagen landbouwkundig onderzoek Nederland* 2. 663: 1-57.

WRIGHT, S. 1918. On the nature of size factors. *Genetics* 3: 367.

WRIGHT, S. 1934. The method of path coefficients. *Annals Mathematical Statistics* 5: 161-215.

WRIGHT, S. 1954. The interpretation of multivariate systems. p11-33. In O. Kempthorne, T.A., Bancroft, J. W., Gowan, and J. L., Lush (eds.) *Statistics and mathematics in biology.* Iowa State College Press (Ames) 632 pp.

ZELLNER, A.; THEIL, H. 1962. Three stage least squares: Simultaneous estimation of simultaneous equations. *Econometrics* 30: 54-78.